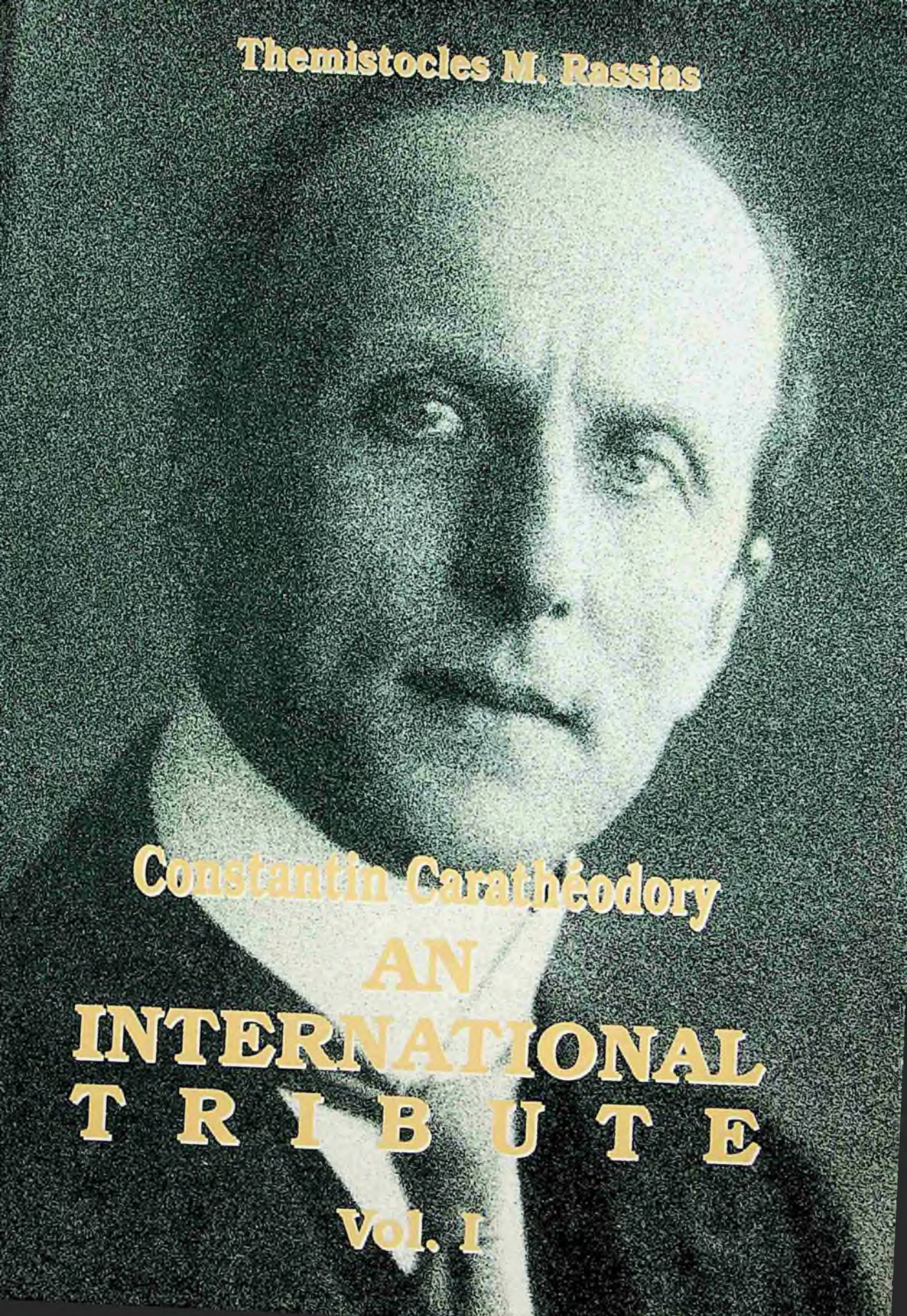
Themistocles M. Rassias



Constantin Carathéodory
AN INTERNATIONAL TRIBUTE
Vol. I

CONSTANTIN CARATHÉODORY: AN INTERNATIONAL TRIBUTE

VOL. I

Constantin Carathéodory (1873–1950)

Constantin Carathéodory

AN INTERNATIONAL TRIBUTE

Vol. I

Editor
Themistocles M. Rassias

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: 687 Hartwell Street, Teaneck, NJ 07666

UK office: 73 Lynton Mead, Totteridge, London N20 8DH

CONSTANTIN CARATHÉODORY: AN INTERNATIONAL TRIBUTE VOL. I

ISBN 981-02-0229-6
981-02-0544-9 (set)

Printed in Singapore by Utopia Press.

FOREWORD

This collection of articles by mathematicians from many countries is a tribute to the memory of Constantin Carathéodory. Among the leaders who in the first half of this century created a foundation for the future development of mathematics Carathéodory was one of the most original, and also one of the first to give new life to parts of classical mathematics that threatened to become stagnant.

His understanding of Riemann's mapping theorem was far ahead of his contemporaries', and his feat of promoting an obscure remark of H. A. Schwarz to become the famous Schwarz' Lemma was as important as it was generous. In the same area, I can personally never forget how impressed I was, and still am, by his invention of prime ends.

Carathéodory's contributions to the calculus of variations were probably the most important part of his work, but I wonder if he was not equally proud of what he did to real variables and measure theory. This was miles away from complex variables, but very much in the same spirit.

The initiative for this tribute was taken by Dr. Themistocles M. Rassias, who has also been the editor and the link with the contributors and the Publisher. As a fellow countryman of Carathéodory, he was the right person for the task and deserves everybody's gratitude.

Lars V. Ahlfors
Harvard University

PREFACE

Constantin Carathéodory (1873–1950) was born in Berlin on 13 September 1873 of Greek parents, and he died in Munich on 2 February 1950. His work covered several subjects of mathematics, including the calculus of variations, function theory, measure and integration, as well as applied mathematics. Carathéodory has also done very fundamental work in mechanics, thermodynamics, geometrical optics and relativity theory.

The first important contribution of Carathéodory to the calculus of variations was his proposal of a theory of discontinuous curves. This was his doctoral thesis "Über die diskontinuierlichen Lösungen in der Variationsrechnung" written while he was a student at the University of Göttingen and published in 1904. This monumental work was the starting point for several fundamental contributions to the calculus of variations by himself and other mathematicians, as well as the inspiration for certain early results in optimal control. In function theory, one of Carathéodory's significant achievements was a simplification of the proof of one of the most essential theorems of conformal representation. He succeeded in extending earlier results of Picard and Schwarz. In 1912, Carathéodory proved his celebrated kernel theorem on sequences of univalent functions. Later, in 1923 Charles Loewner introduced a new approach in function theory representing slit mappings in terms of a differential equation. Loewner, using the Carathéodory convergence theorem, proved the Bieberbach conjecture for the third coefficient. L. de Branges recently remarked that it was precisely this method which finally gave a full solution to the Bieberbach conjecture. In measure theory, as is well known, Carathéodory is the inventor of outer measures, and all that this implies in mathematics. In mechanics, thermodynamics, optics and relativity, besides his important scientific contributions, Carathéodory has significantly influenced the better understanding and the rigorous presentation of these fields.

An example of Carathéodory's wide-ranging influence in the interna-

tional mathematical community was seen during the first Fields Medals awards at the International Congress of Mathematicians, Oslo, 1936. The selection committee consisted of G. D. Birkhoff, Elie Cartan, C. Carathéodory, F. Severi, and T. Takagi. Two medals were awarded, one to L. V. Ahlfors and one to Jesse Douglas. It was C. Carathéodory who presented both their works during the opening of the International Congress.

These two volumes contain a series of scientific articles dedicated to the memory of Constantin Carathéodory. These articles deepen our understanding of some of the current research problems and theories in modern topics of calculus of variations, complex analysis, real analysis, differential equations, geometry and their applications, which are related to the work of Carathéodory. This presentation of concepts and methods makes this tribute an invaluable reference for teachers and other professionals in mathematics who are interested in pure and applied research, philosophy of mathematics, and mathematics education.

It is my pleasure to express my warmest thanks to all of the scientists who contributed to these two volumes, and, I would particularly like to extend my special appreciation to Professor L. V. Ahlfors for writing the Foreword to these volumes. I would also like to acknowledge the superb assistance in editing and composition that the staff of World Scientific Publishing Co. has provided in the preparation of this publication.

Athens, Greece
November, 1990

Themistocles M. Rassias

CONTENTS

Volume 1

Volume 2

CONSTANTIN CARATHÉODORY: AN INTERNATIONAL TRIBUTE

VOL. I

Constantin Carathéodory: An International Tribute (pp. 1–15)
edited by Th. M. Rassias

The Binomial Theorem in the Algebra $\mathbf{A}^+$

Lars V. Ahlfors[1]

(In memory of Constantin Carathéodory)

INTRODUCTION.

The standard Clifford algebra $\mathbf{A}_N$ (or $\mathbf{A}$) is the associative algebra over the reals with unit 1 and generators $e_1, \cdots, e_N$ subject to the relations $e_i^2 = -1$ and $e_i e_j = -e_j e_i$ for $i \neq j$. It is a vector space of dimension 2^N with a basis formed by all products $e_{i_1} e_{i_2} \cdots e_{i_k}$, in natural order, $0 \leqq k \leqq N$. For $k = 0$ the product equals 1.

$\mathbf{A}_N$ contains a smaller vector space $\mathbf{QC}^N$ (or $\mathbf{QC}$ when N has been fixed) formed by all elements that may be written in the form $z = x + y$ with $x \in \mathbf{R}$ and $y = y_1 e_1 + \cdots + y_N e_N$ with real $y_1, \cdots, y_N$. As the notation suggests, we shall regard z as a *generalized complex number* with $\operatorname{Re} z = x$ and $\operatorname{Im} z = y \in \mathbf{R}^N$.

We shall treat $\mathbf{QC}$ as a euclidean space with square norm $|z|^2 = x^2 + |y|^2$, $|y|^2 = y_1^2 + \cdots + y_N^2$. When dealing with two elements of $\mathbf{QC}$ we shall frequently denote them by $z = x + y$ and $w = u + v$. Their inner product is defined by $\langle z, w \rangle = xu + \langle y, v \rangle$, $\langle y, v \rangle = y_1 v_1 + \cdots + y_N v_N$.

The Clifford product of z and w, written as zw, is rarely in $\mathbf{QC}$ and therefore of little

[1]Research supported by NSF and Forschungsinstitut für Mathematik, ETH, Zürich

use. The algebraists have an easy way of avoiding this difficulty. They have invented a new algebra on $\mathbf{QC}$, called $\mathbf{A}^+$, in which the product of z and w is denoted by $z \cdot w$ and defined as $\frac{1}{2}(zw + wz)$. It is easy to see that $z \cdot w$ is indeed in $\mathbf{QC}$, and it is obvious that $z \cdot w = w \cdot z$, so that multiplication in $\mathbf{A}^+$ is commmutative.

The commutativity has been achieved at the price of revoking the associative law. Indeed, it is clear that $e_i \cdot e_i = -1$ while $e_i \cdot e_j = 0$ if $i \neq j$. It follows that $(e_i \cdot e_j) \cdot e_j = 0$ and $e_i \cdot (e_j \cdot e_j) = -e_i$, which shows that $\mathbf{A}^+$ is non-associative as soon as $N > 1$.

In the absence of associativity the commutative law plays only a limited role. For instance, it is essential to distinguish between $(a \cdot b) \cdot c$ and $a \cdot (b \cdot c)$. For products of more than three factors a more elaborate system of parentheses is needed.

Because $a \cdot a$ is the same as aa the notation a^2 may be used for both. More generally, a^n may be interpreted both as a power in the Clifford algebra and as a product of equal factors in $\mathbf{A}^+$. In the latter case it may itself occur as a factor in a product, but it should then be thought of as enclosed in a parenthesis. For instance, $a^2 \cdot b$ should be read as $(a \cdot a) \cdot b$ and distinguished from $a \cdot (a \cdot b)$.

A basic formula in $\mathbf{A}^+$ states that

$$(a \cdot b) \cdot a^2 = a \cdot (b \cdot a^2). \tag{1}$$

In terms of the Clifford algebra this is equivalent to

$$(ab + ba)a^2 + a^2(ab + ba) = a(ba^2 + a^2b) + (ba^2 + a^2b)a,$$

which is true by the associative law. In contrast, the simpler formula $a \cdot (a \cdot b) = a^2 \cdot b$ is false.

Any algebra which satisfies $a \cdot b = b \cdot a$ as well as (1) is known as a *Jordan algebra* after Paul Jordan, a mathematical physicist who in 1933 introduced this notion in connection with the early development of quantum mechanics. The algebraic idea caught the interest of professional algebraists, and the theory has become an important topic in nonassociative algebra.

The algebra A^+ is one of several special Jordan algebras, and as such much closer to complex numbers than the general notion. It is easier to use because there are many identities in A^+ which are not true in an arbitrary Jordan algebra.

Because $z \in \mathbf{QC}$ implies $z^n \in \mathbf{QC}$ for all n it is possible to consider power series of the form $\sum a_n z^n$ with suitable coefficients, preferably in QC. The writer believes that such series are a worthy subject of serious research, even though the likelihood of a close analogy with the complex case may be remote.

The theory of analytic continuation of complex power series makes important use of the binomial theorem, without which there would hardly be a starting point. The theorem turns out to be false in $\mathbf{A}^+$, but there is reason enough to look for a substitute. In this paper, which is of a preliminary nature, it will be shown that the binomial theorem can be rescued by adding a remainder term given by an explicit formula. The question of analytic continuation remains open.

Remark. The paper is essentially elementary and could be boring for true experts on nonassociative algebra. It is addressed mainly to readers who come from the side of classical complex analysis.

I. INDIVIDUAL POWERS IN $\mathbf{A}^+$.

1. Members of the set **QC** will be called *vectors*. In connections where only two vectors are involved we shall usually denote them by suggestive letters such as z and w, but in the case of more than two vectors the scarcity of suitable letters soon makes itself felt. It is therefore more practical to permit an arbitrary choice from one or more alphabets. For instance, the commutative law in $\mathbf{A}^+$ is best expressed by $a \cdot b = b \cdot a$. Similarly, the distributive law, which by convention is true in every algebra, takes the form $a \cdot (b + c) = a \cdot b + a \cdot c$.

The most important feature of the algebra $\mathbf{A}^+$ is its lack of associativity, which requires extreme caution. Otherwise, the multiplication is fully determined by the special rules $e_i^2 = -1$ and $e_i \cdot e_j = 0$ for $i \neq j$, together with the commutative and distributive laws. For purely imaginary vectors the special rules $y^2 = -|y|^2$ and $y \cdot v = -\langle y, v \rangle$ are in force.

The general multiplication formula in $\mathbf{A}^+$ reads

$$z \cdot w = (x + y) \cdot (u + v) = (xu - \langle y, v \rangle) + (xv + uy), \tag{1.1}$$

where the parentheses on the right serve only to isolate the real and imaginary parts. An important special case is

$$z^2 = (x^2 - |y|^2) + 2xy, \tag{1.2}$$

which is very similar to the corresponding formula in the complex case.

2. It is an important and remarkable property of **QC** that any power z^n of $z \in \mathbf{QC}$ is again in **QC**. This is easy to prove by induction, but there is a much more instructive way. If z is real there is nothing to prove, and we may therefore assume that $y \neq 0$. When this is so, $z = x + y$ can be rewritten as

$$z = x + |y| \frac{y}{|y|}. \tag{2.1}$$

Because $\left(\frac{y}{|y|}\right)^2 = -1$ it becomes obvious that the subalgebra generated by z is isomorphic to the one generated by $x + |y|i$. In other words, in order to find z^n it suffices to develop $(x + |y|i)^n$ and replace i by $\frac{y}{|y|}$.

This can be done quite explicitly by use of the binomial theorem. Even without carrying out the calculation one recognizes that the result will be of the form

$$z^n = \alpha_n(x, |y|) + \beta_n(x, |y|)\frac{y}{|y|}, \tag{2.2}$$

where α_n and β_n are homogeneous polynomials of total degree n in x and $|y|$. Moreover, α_n is even and β_n is odd in $|y|$. The denominator in $\frac{y}{|y|}$ cancels against a factor in β_n.

For later use we display the actual developments:

$$\begin{aligned} \alpha_n(x, |y|) &= \sum_k (-1)^k \binom{n}{2k} x^{n-2k}|y|^{2k} \\ \beta_n(x, |y|) &= \sum_k (-1)^k \binom{n}{2k+1} x^{n-2k-1}|y|^{2k+1}. \end{aligned} \tag{2.3}$$

Here and later it will be understood that the index k runs through all the integers for which the coefficients $\binom{n}{2k}$ and $\binom{n}{2k+1}$ are defined.

It should be noted that $\alpha_0 = 1, \beta_0 = 0$ and $\alpha_1 = x$, $\beta_1 = |y|$. The step from n to $n+1$ is given by

$$\alpha_{n+1} = x\alpha_n - |y|\beta_n, \quad \beta_{n+1} = |y|\alpha_n + x\beta_n \tag{2.4}$$

or in matrix form

$$\begin{pmatrix} \alpha_{n+1} \\ \beta_{n+1} \end{pmatrix} = \begin{pmatrix} x & -|y| \\ |y| & x \end{pmatrix} \begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix}. \tag{2.5}$$

The passage from $z = x + y \in \mathbf{QC}$ to the complex number $x + |y|i$ and *vice versa* will play an essential role in what follows. In particular, the argument φ of $x + |y|i$ is defined by $\cos\ \varphi = \frac{x}{|z|}$, $\sin\ \varphi = \frac{|y|}{|z|}$, $0 \leqq \varphi \leqq \pi$. Because $(\cos\ \varphi + i \sin\ \varphi)^n = \cos\ n\varphi + i \sin\ n\varphi$ it follows that

$$z^n = |z|^n(\cos\ n\varphi + \sin\ n\varphi \frac{y}{|y|}) \tag{2.6}$$

and on comparison with (2.1)

$$\alpha_n(x, |y|) = |z|^n \cos\ n\varphi, \quad \beta_n(x, |y|) = |z|^n \sin\ n\varphi. \tag{2.7}$$

We shall refer to $x + |y|i$ as the *complex image* of $x + y$. Thus $x + y$ and $x - y$ have the same complex image, and we agree that the argument of $x + |y|i$ shall also be considered the argument of $x \pm y$.

It is seen from (2.7) that $\alpha_n(x,|y|)$ and $\beta_n(x,|y|)$ are conjugate harmonic functions of the variables x and $|y|$. As such they satisfy the Cauchy-Riemann equations

$$\frac{\partial \alpha_n}{\partial x} = \frac{\partial \beta_n}{\partial |y|}, \quad \frac{\partial \beta_n}{\partial x} = -\frac{\partial \alpha_n}{\partial |y|} \tag{2.8}$$

as well as $\Delta \alpha_n = \Delta \beta_n = 0$.

This is an indication that the subject matter we are pursuing is not far removed from the classical theory of holomorphic functions.

3. From the preceding it is clear that the mapping of QC on itself which takes z to z^n is analytic. It is easy to compute all partial derivatives, but we shall limit ourselves to finding an explicit formula for the Jacobian of the mapping $z \to z^n$.

As a temporary notation we shall write $(x+y)^n = u + v$. The Jacobian is

$$D_n = \frac{\partial(u, v_1, \ldots, v_N)}{\partial(x, y_1, \ldots, y_N)} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y_j} \\ \frac{\partial v_i}{\partial x} & \frac{\partial v_i}{\partial y_j} \end{pmatrix}$$

in easily understandable notation. From $u = \alpha_n$ and $v_i = \beta_n \dfrac{y_i}{|y|}$ one obtains

$$\frac{\partial u}{\partial x} = n\alpha_{n-1}, \quad \frac{\partial v_i}{\partial x} = n\beta_{n-1}\frac{y_i}{|y|}.$$

Moreover,

$$\frac{\partial u}{\partial y_j} = \frac{\partial \alpha_n}{\partial |y|}\,\frac{y_j}{|y|} = -n\beta_{n-1}\frac{y_j}{|y|}$$

$$\frac{\partial v_i}{\partial y_j} = n\alpha_{n-1}\frac{y_i y_j}{|y|^2} + \frac{\beta_n}{|y|}\left(\delta_{ij} - \frac{y_i y_j}{|y|^2}\right).$$

For a more compact notation we shall henceforth write y as the column matrix $\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ and its transpose y^T as $(y_1 \cdots y_N)$. In terms of matrix multiplication this means that $y^T y = |y|^2$, a real number, while yy^T is the square matrix $\|y_i y_j\|$, $i,j = 1,\ldots,N$. It will also be expedient to write $n\alpha_{n-1} = a_n$, $n\beta_{n-1} = b_n$, $\beta_n/|y| = c_n$.

With these changes we obtain

$$D_n = \begin{pmatrix} a_n & -b_n \frac{y^T}{|y|} \\ b_n \frac{y}{|y|} & a_n \frac{yy^T}{|y|^2} + c_n(I_N - \frac{yy^T}{|y|^2}) \end{pmatrix}, \tag{3.1}$$

where I_N is the unit matrix in N dimensions.

4. The local quasiconformal nature of the mapping from z to z^n is determined by the matrix $D_n^T D_n$, and does not change when z is subjected to a conformal mapping. An easy computation based on (3.1) shows that

$$D_n^T D_n = \begin{pmatrix} a_n^2 + b_n^2 & 0 \\ 0 & (a_n^2 + b_n^2 - c_n^2)\frac{yy^T}{|y|^2} + c_n^2 I_N \end{pmatrix}. \tag{4.1}$$

Conjugate by $\begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix}$ where $K \in O(n)$ maps $\frac{y}{|y|}$ on e_1. Clearly, $\frac{yy^T}{|y|^2}$ is replaced by $e_1 e_1^T$, which is a diagonal matrix consisting of 1 followed by $N-1$ zeros. We conclude from this that $D_n^T D_n$ is conjugate to

$$\begin{pmatrix} a_n^2 + b_n^2 & 0 & 0 \\ 0 & a_n^2 + b_n^2 & 0 \\ 0 & 0 & c_n^2 I_{N-1} \end{pmatrix}. \tag{4.2}$$

Provided that $c_n \neq 0$ we may set $\lambda = (a_n^2 + b_n^2)/c_n^2$. The matrix (4.2) is then a multiple of

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & I_{N-1} \end{pmatrix}.$$

This has two eigenvalues, λ of multiplicity 2 and 1 of multiplicity $N-1$.

In trigonometric terms $\lambda = \left(\frac{n \sin \varphi}{\sin n\varphi)}\right)^2$, which implies $\lambda > 1$, except for $\lambda = 1$ if $n = 1$ or $\varphi = 0$. As for quasi-conformality, the mapping $z \to z^n$ is *quasiregular* with dilation λ and maps an infinitesimal sphere on a spheroid with two major and $N-1$ minor axes. It is regular except at points where $n\varphi$ is a multiple of π.

II. BASIC OPERATIONS ON THE ALGEBRA $\mathbf{A}^+$.

1. In the preceding section we were mainly concerned with powers, and for that purpose the specific properties of $\mathbf{A}^+$ were not actively involved. We shall now shift the attention to the ways of dealing with the intricate nature of non-associativity.

The author's experience is that the lack of the associative law involves many pitfalls, which can easily lead to serious mistakes. For this reason, when dealing with the algebra $\mathbf{A}^+$ it is very helpful to make more extensive use of matrix multiplication, which by itself is always associative.

The writer prefers to use matrices which act from the left on vectors written as column matrices, as already in Sec. I. We continue the practice of denoting a purely imaginary vector by a single letter y, z as $\begin{pmatrix} x \\ y \end{pmatrix}$, and its transpose as (x, y^T). The same applies to $w = \begin{pmatrix} u \\ v \end{pmatrix}$. The identity matrix is again denoted by I_N or I_{N+1}, as the case may be, and the subscript may be omitted when the dimensionality is taken for granted.

With the vector $z = x + y$ we associate the matrix

$$L(z) = \begin{pmatrix} x & -y^T \\ y & xI_N \end{pmatrix},$$

interpreted as a block matrix, in the obvious manner. The letter L is a reminder that the matrix acts from the left.

One verifies at once that

$$L(z)w = z \cdot w,$$

and this is the purpose of the notation. It follows that $L(z)w = L(w)z$ and $L(z)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = z$.

The following basic lemma serves to compare the products $L(z)L(w)$ and $L(z \cdot w)$.

LEMMA 1.

$$L(z)L(w) - L(z \cdot w) = \begin{pmatrix} 0 & 0 \\ 0 & (y^T v)I_N - yv^T \end{pmatrix}. \tag{1.1}$$

We recall that $y^T v$ is the same as the inner product (y, v), and that yv^T is the $N \times N$ matrix $\|y_i v_j\|$. The proof of the lemma is an easy verification, left to the reader. The power of the lemma is due to the three zeroes on the right.

The lemma implies

$$L(z)L(w)\begin{pmatrix}1\\0\end{pmatrix} = L(z\cdot w)\begin{pmatrix}1\\0\end{pmatrix} = L(w)L(z)\begin{pmatrix}1\\0\end{pmatrix}, \tag{1.2}$$

but it is not true that $L(z)L(w) = L(w)L(z)$. In fact, (1.1) shows that

$$L(z)L(w) - L(w)L(z) = \begin{pmatrix}0 & 0\\ 0 & vy^T - yv^T\end{pmatrix}. \tag{1.3}$$

The special case $w = z$ of the lemma yields

$$L(z)^2 - L(z^2) = \begin{pmatrix}0 & 0\\ 0 & |y|^2 I - yy^T\end{pmatrix}. \tag{1.4}$$

We shall use this to evaluate $z\cdot(z\cdot w) - z^2\cdot w$. This is precisely $(L(z)^2 - L(z^2))w$. Because of the zeros the product of the matrix on the right with $w = \begin{pmatrix}u\\v\end{pmatrix}$ is the same as with $\begin{pmatrix}0\\v\end{pmatrix}$, and hence equal to $|y|^2 v - \langle y, v\rangle y$, where we have dropped the superfluous zeros.

For many similar applications it is convenient to introduce the notation

$$\Lambda(y) = I_N - \frac{yy^T}{|y|^2}. \tag{1.5}$$

We remark that $\Lambda(y)$ is idempotent, i.e., $\Lambda(y)^2 = \Lambda(y)$. As for (1.4) it may be rewritten as

$$L(z)^2 - L(z^2) = |y|^2\begin{pmatrix}0 & 0\\ 0 & \Lambda(y)\end{pmatrix}.$$

2. As another application of Lemma 1 we shall derive an expression for $L(z^m)L(z^n) - L(z^{m+n})$. In (1.1), z and w shall be replaced by z^m and z^n, respectively, while on the right y and v become Im z^m and Im z^n. If arg $z = \varphi$ we know that the imaginary parts are $|z|^m \sin\, m\varphi\frac{y}{|y|}$ and $|z|^n \sin\, n\varphi\frac{y}{|y|}$, which proves

LEMMA 2.

$$L(z^m)L(z^n) - L(z^{m+n}) = |z|^{m+n}\sin\, m\varphi \sin\, n\varphi\begin{pmatrix}0 & 0\\ 0 & \Lambda(y)\end{pmatrix}.$$

This lemma will find important use in the next section.

III. THE BINOMIAL THEOREM IN $\mathbf{A}^+$.

1. It is a familiar fact that the theory of analytic continuation in the complex domain depends heavily on Newton's binomial theorem. To recall the details, the question of convergence is trivial, and all that is needed is to rearrange a power series in z so that it becomes one in $z - z_0$. For this purpose one appeals to the binomial theorem to obtain

$$z^n = \sum_{k=0}^{n} \binom{n}{k} z_0^k (z - z_0)^{n-k}.$$

A series $\sum_0^\infty a_n z^n$ becomes a double series

$$\sum_{n,k} \binom{n}{k} a_n z_0^k (z - z_0)^{n-k}$$

where k and n are restricted by $0 \leqq k \leqq n$. However, if we introduce $m = n - k$ it becomes

$$\sum_{m,k} \binom{m+k}{k} a_{m+k} z_0^k (z - z_0)^m,$$

where k and m run independently from 0 to ∞. Now, in terms of new coefficients

$$b_m = \sum_k \binom{m+k}{k} a_{m+k} z_0^k$$

we are led to the identity

$$\sum_0^\infty a_n z^n = \sum_0^\infty b_m (z - z_0)^m,$$

as desired.

More symmetrically, we could have replaced z_0 by z and $z - z_0$ by w, resulting in

$$\sum_n a_n (z + w)^n = \sum_{m,k} \binom{m+k}{k} a_{m+k} z^k w^m,$$

where the double series may be regarded as a power series either in z or in w. Observe that the symmetry in z and w is obvious from $\binom{m+k}{k} = \binom{m+k}{m}$.

2. If the binomial theorem were valid in $\mathbf{A}^+$ it would have to be written as

$$(z + w)^n = \sum_{k=0}^{n} \binom{n}{k} z^k \cdot w^{n-k}.$$

This is obvious for $n = 1$ and easy to verify for $n = 2$. We shall see that it fails for $n = 3$.

By commutativity, but without open or hidden use of the associative law, one has

$$(z+w)^3 = (z+w)\cdot(z^2+2z\cdot w+w^2)$$
$$= z^3+2z\cdot(z\cdot w)+z\cdot w^2+w\cdot z^2+2w\cdot(w\cdot z)+w^3.$$

By the definition of $L(z)$ (see Sec. II) $z\cdot(z\cdot w)=L(z)^2w$ and $w\cdot z^2=z^2\cdot w=L(z^2)w$. By use of II.(1.4-5) it follows that $2z\cdot(z\cdot w)+z^2\cdot w=3z^2\cdot w+2|y|^2\Lambda(y)v$. Similarly, $w^2\cdot z+2w\cdot(u\cdot z)=3z\cdot w^2\Lambda(v)y$, so that

$$(z+w)^3=(z^3+3z^2\cdot w+3z\cdot w^2+w^3)+2|y||v|\left(|y|\Lambda(y)\frac{v}{|v|}+|v|\Lambda(v)\frac{y}{|y|}\right). \tag{2.1}$$

This result supports the expectation that the binomial theorem is true except for an additive correction, and it seems like a good guess that the vector

$$\sigma(y,v)=|y|\Lambda(y)\frac{v}{|v|}+|v|\Lambda(v)\frac{y}{|y|} \tag{2.2}$$

will play a dominant role. The matrix $\Lambda(y)$ was defined by II.(1.2).

The formula (2.2) breaks down when either $y=0$ or $v=0$. In that case z and w commute, and the ordinary binomial theorem is in force. It is therefore no restriction to assume that y and v are both $\neq 0$. When this is so they enclose an angle ω with $\cos\,\omega=\langle y,v\rangle/|y||v|$. One verifies that $\sigma(y,v)$ can also be expressed by

$$\sigma(y,v)=|y|\left(\frac{v}{|v|}-\cos\,\omega\frac{y}{|y|}\right)+|v|\left(\frac{y}{|y|}-\cos\,\omega\frac{v}{|v|}\right). \tag{2.3}$$

3. For arbitrary n we shall denote the remainder in the binomial formula by $\rho_n(z,w)$ or simply ρ_n. It is thus defined by

$$(z+w)^n=\sum_k\binom{n}{k}z^k\cdot w^{n-k}+\rho_n(z,w). \tag{3.1}$$

Note that we have again refrained from spelling out the range of k.

We pass to the proof of a crucial lemma. In addition to $\varphi=\arg z$ we shall also need $\psi=\arg w$.

LEMMA 3.

$$\rho_{n+1} - (z+w)\cdot\rho_n = \left(\sum_k \binom{n}{k}|z|^k|w|^{n-k}\sin k\varphi\ \sin(n-k)\psi\right)\sigma(y,v).$$

PROOF: By (3.1), $\rho_{n+1} - (z+w)\rho_n$ is the same as

$$(z+w)\cdot\sum_k \binom{n}{k}z^k\cdot w^{n-k} - \sum_k \binom{n+1}{k}z^k\cdot w^{n+1-k}. \tag{3.2}$$

We concentrate on showing that the expression (3.2) can be identified with the right hand side in the lemma.

Because of the symmetry of the binomial coefficients z and w are interchangeable. Therefore,

$$(z+w)\cdot\sum_k \binom{n}{k}z^k\cdot w^{n-k} =$$

$$\sum_k \binom{n}{k}z\cdot(z^k\cdot w^{n-k}) + \sum_k \binom{n}{k}w\cdot(w^k\cdot z^{n-k}). \tag{3.3}$$

On invoking Lemma 2 (Sec. II)

$$z\cdot(z^k\cdot w^{n-k}) = L(z^{k+1})w^{n-k} + |z|^{k+1}\sin\varphi\sin k\varphi\begin{pmatrix}0 & 0\\ 0 & \Lambda(y)\end{pmatrix}\mathrm{Im}\ w^{n-k}.$$

Substitute $|y|$ for $|z|\sin\varphi$ and $|w|^{n-k}\sin(n-k)\psi\dfrac{v}{|v|}$ for $\mathrm{Im}\ w^{n-k}$. We obtain

$$\begin{aligned}\sum_k \binom{n}{k}z\cdot(z^k\cdot w^{n-k}) &= \sum_k \binom{n}{k}z^{k+1}\cdot w^{n-k} +\\ \sum_k \binom{n}{k}|z|^k|w|^{n-k}\sin\ k\varphi\ \sin(n-k)\psi\ |y|\Lambda(y)\frac{v}{|v|}&.\end{aligned} \tag{3.4}$$

The same is true when z and w are interchanged. At the same time we are free to let k and $n-k$ change places in any one of the sums in (3.4). Therefore, we have also

$$\begin{aligned}\sum_k \binom{n}{k}w\cdot(w^k\cdot z^{n-k}) &= \sum_k \binom{n}{k}z^k\cdot w^{n+1-k} +\\ \sum_k \binom{n}{k}|z|^k|w|^{n-k}\sin k\varphi\ \sin(n-k)\psi\ |v|\Lambda(v)\frac{y}{|y|}&.\end{aligned} \tag{3.5}$$

Since k may be allowed to run through all integers, nothing changes when k is replaced by $k-1$. This implies

$$\sum_k \binom{n}{k} z^{k+1} \cdot w^{n-k} = \sum_k \binom{n}{k-1} z^k \cdot w^{n+1-k}. \tag{3.6}$$

Also, $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$ for all k, even when the binomial coefficients are artificially defined.

Add (3.4) and (3.5). Because of (3.3) the left hand sides add up to the first term of (3.2). By (3.6) the first terms on the right cancel against the negative term in (3.2). Finally, by (2.1), the second terms on the right combine to form what is to the right of the equality sign in the lemma. This completes the proof. □

5. For convenience we shall reformulate Lemma 3 as the recursive formula

$$\rho_{n+1} = (z+w) \cdot \rho_n + R_n \sigma(y,v) \tag{4.1}$$

with

$$R_n = \sum_k \binom{n}{k} |z|^k |w|^{n-k} \sin k\varphi \ \sin(n-k)\psi. \tag{4.2}$$

Note that the R_n are *real numbers*. One sees at once that $R_1 = 0$, compatible with $\rho_1 = \rho_2 = 0$, and $R_2 = 2|y||v|$, which agrees with (2.1).

The system (4.1) is easy to solve for ρ_n by iteration, but only after rewriting it as

$$\rho_{n+1} = L(z+w)\rho_n + R_n \sigma(y,v). \tag{4.3}$$

In fact, one obtains the explicit formula

$$\rho_n = \sum_{k=1}^{n-2} R_{n-k} L(z+w)^{k-1} \sigma(y,v). \tag{4.4}$$

The nature of the numbers R_n as given by (4.2) is not immediately clear, but the use of the complex images $x + |y|i$ and $u + |v|i$ will be helpful. Because $(x + |y|i)^k = |z|^k(\cos k\varphi + i \sin k\varphi)$ and $(u + |v|i)^{n-k} = |w|^{n-k}(\cos(n-k)\psi + i \sin(n-k)\psi)$ one obtains

$$R_n = \sum_k \binom{n}{k} \mathrm{Im}(x + |y|i)^k \mathrm{Im}(u + |v|i)^{n-k}. \tag{4.5}$$

At first sight it could seem questionable to use the same i twice since in one factor it is supposed to be exchangeable by $\frac{y}{|y|}$ and in the other by $\frac{v}{|v|}$, but actually the imaginary parts are real numbers and we have used i only in its original sense.

It is also easy to see that (4.5) may be rewritten as

$$R_n = -\frac{1}{2}\left[\mathrm{Re}\sum_k \binom{n}{k}(x + |y|i)^k(u + |v|i)^{n-k} - \mathrm{Re}\sum_k \binom{n}{k}(x + |y|i)^k(u - |v|i)^{n-k}\right].$$

In this form the complex version of the binomial theorem is available, and we arrive at the relatively simple formula

$$R_n = -\frac{1}{2}\mathrm{Re}\{[x + u + (|y| + |v|)i]^n - [x + u + (|y| - |v|)i]^n\}. \tag{4.6}$$

Together (4.4) and (4.6) yield an acceptable answer to the problem at hand. When written out the result takes the form

$$\rho_n = -\frac{1}{2}\sum_{k=1}^{n-2} \mathrm{Re}\{[x+u+(|y|+|v|)i]^{n-k} - (x+u+(|y|-|v|)i]^{n-k}\} L(z+w)^{k-1}\sigma(y,v). \tag{4.7}$$

This rather pedestrian formula can almost certainly be simplified, but the writer's present research along this line is not ready for publication. It is his hope that younger brains will be attracted to the many open problems in this area.

References

1. Albert, A.A., "On Jordan algebras of linear transformations", Trans. Am. Math. Soc. 59, 1946.
2. Albert, A.A., edit. Studies in Modern Algebra, MAA, 1963.
3. Braun, H. and Koecher, M., Jordan-Algebren, Springer, 1966.
4. Jacobson, N., "Structure and representation of Jordan algebras", Am. Math. Soc. 1968.

Lars V Ahlfors
Department of Mathematics
Harvard University
Cambridge, MA 02138
USA

Constantin Carathéodory: An International Tribute (pp. 16–46)
edited by Th. M. Rassias

ON SOLUTIONS OF SOME CLASSES OF DIFFERENTIAL EQUATIONS OF RIEMANN–PAPPERITZ TYPE AND THE EXTENSION OF RIEMANN P–FUNCTION

M.A. Al–Bassam

ABSTRACT. In a previous paper ([9], p. 7) the author has generalized the Riemann–Papperitz second order differential equation, of which the Gauss hypergeometric second order equation is a particular case. This generalization is represented by an integro–differential equation of order (m–n, n) (the highest integral term of order m–n and the highest derivative of order n) with m singular points. In this article the nth order differential equation with (n+1) singular points has been studied, analysed and solved. In addition, a particular case of a class of fifth order differential equation with six singular points has been solved and its solutions have been obtained in terms of the hypergeometric functions $F_D^{(4)}$. It has been shown that the nth order differential equation with (n+1) singular points has $2n^2(n+1)$ branch solutions, where n is the order of the equation, (n+1) the number of singular points including (∞) and 2n is the number of transformations which leave the integrals unaltered. These transformations have been obtained and developed here by the author. In general if m is the number of singular points of the equation then the number of branch solutions is 2mn(m+1). Also, in this article the extended Riemann P–Function (M–Functions) for these equations have been studied, discussed and obtained. The M–Functions associated with equations of different orders and nth order equations have been found. It must be mentioned that the study of solutions have been carried out by Generalized Analysis (Fractional Calculus) and the use of properties of the integro–differential operator of generalized (fractional) order.

1. Introduction

In previous articles, ([9], [10], [12], [13], [14], [15], [22], [23]), [27] the author has studied and developed new operational methods for obtaining solutions of differential, integro–differential and integral equations. This may be achieved through representation of these equations by their equivalent operator (transform) equations. When equivalent operator equations are obtained, then their solutions are those of the corresponding differential or integro–differential or integral equations. Methods for solving such operator equations are based upon the use of generalized calculus (fractional calculus) and the operational properties of the integro–differential operator $({}_{a}^{x}I^{\alpha})$ of generalized order [8].

In an article [9] the author has studied equivalence properties of Gauss's hypergeometric equations and their corresponding operator equations. Then he has established the existence of Kummer's twenty four solutions as listed in ([5], pp. 87–88).

In this article equivalence properties between a class of general integro–differential equations of Gauss–Papperitz–Riemann type and their operator equations will be discussed. In dealing with the general case, it would be sufficient to study first in detail a particular case as the treatment and approach in both cases are similar and later in the work the generalized Riemann–Papperitz linear differential equations will be discussed. So, if (i,j) denotes the order of the integro–differential equation where i represents the integral order and j indicates the differential order, then the total order of the equation will be (i+j). Thus our study will include the particular case of the integro–differential equations of fifth total order. These my be represented by equations of order (0,5) or (1,4) or (2,3) or (3,2) or (4,1) as they are all equivalent. It may be interesting to point out that the fifth order equations or the equivalent forms (i,j) (i = 0,1,2,3,4), (j = 1,2,3,4,5) are satisfied by hypergeometric functions of the type $F_D^{(4)}(A,B_1,B_2,B_3,B_4;\ C;\ x,y,z,w)$, where y = y(x), z = z(x) and w = w(x).

In addition to Gauss's hypergeometric equations, other ordinary and

partial differential equations may have many branch solutions as it was indicated by various authors. Baily ([7], p. 78) has mentioned that the hypergeometric function F satisfies a certain type of hypergeometric partial differential equations with "at least six solutions" of this type. Also it has been indicated by Appell and Kampé de Fériet that there are sixty solutions of these equations, ([6], [7]). On similar subjects some work have appeared in [25]. Recently, in an article [27] the author has studied a class of third order differential equations and has shown that such a class possesses seventy two branch solutions. Also, it has been shown that it is associated with the M–Function (the Riemann extended P–Function).

In our work, in dealing with ordinary differential equations, it will be revealed that the number of branch solutions depends upon the order of the equation, the number of singular points of the equation and the number of transformations leaving the principal integrals unaltered. Thus the number of solutions of the general case of the integro–differential equation, presented in this work, with total order n and m is the number of singular points of the equation inlcuding $(-\infty)$ or $(+\infty)$ equals $2nm(m-1)$, where $2(m-1)$ is the number of transformations leaving the integrals unaltered. In dealing with fifth order equations of six singular points including $(-\infty)$, then the total of branch solution would be three hundreds as it will be shown later in this work.

2. Preliminaries, Definitions and Some Hypergeometric Identities

Some definitions and properties of integro–differential operators of generalized order, transformations and hypergeometric functions and identities, together with their references, will be given here as they may be needed in this work.

Definition 1.

If f(x) is a real–valued function of class $C^{(n)}$ on $a \leq x \leq b$ and Re $\alpha + n > 0$, then

$$ {}_{a}^{x}I^{\alpha} f = \frac{1}{\Gamma(\alpha+n)} D_x^n \int_a^x (x-t)^{\alpha+n-1} f(t), \quad (n=0,1,2,\ldots) \tag{2.1} $$

When $a \to -\infty$

$$\underset{-\infty}{\overset{x}{I}}{}^{\alpha} f = \lim_{x \to -\infty} \overset{x}{\underset{a}{I}}{}^{\alpha} f = D_x^n \underset{-\infty}{\overset{x}{I}}{}^{\alpha+n} f \ , \tag{2.2}$$

where $D_x^n = \frac{d^n}{dx^n}$ and Γ is the gamma function.

Details and properties of this operator may be found in [8] and [9].

Definition 2.

If $\underset{(a;-\infty)}{\overset{x}{I}}{}^{-\alpha} f = 0$, Re $\alpha - n \geq 0$, $(n = 0,1,2,...)$ and

(1) $f \in C^{(n)}$ on $[a,b]$, then $\overset{x}{\underset{a}{I}}{}^{\alpha} 0 = f(x) = \sum_{k=1}^{n} \frac{C_k (x-a)^{\alpha-k}}{\Gamma(\alpha-k+1)}$, (2.3)

(2) $f \in C^{(n)}$ on $(-\infty,b]$, then $\underset{-\infty}{\overset{x}{I}}{}^{\alpha} 0 = f(x) = \sum_{i=1}^{n} \frac{C_i x^{i-1}}{\Gamma(i)}$, (2.4)

where C_i are arbitrary constants ([9], pp. 5–6).

Remarks: When $\overset{x}{\underset{a}{I}}{}^{-\alpha} F = 0$, then $D_x^3 \overset{x}{\underset{a}{I}}{}^{3-\alpha} F = 0$ which implies that $\overset{x}{\underset{a}{I}}{}^{3-\alpha} F = \overset{x}{\underset{a}{I}}{}^{3} 0 = C_1(x-a)^2 + C_2(x-a) + C_3$ and hence

$$F = C_1 \frac{(x-a)^{\alpha-1}}{\Gamma\alpha)} + C_2 \frac{(x-a)^{\alpha-2}}{\Gamma(\alpha-1)} + C_3 \frac{(x-a)^{\alpha-3}}{\Gamma(\alpha-2)} \tag{2.5}$$

C_1, C_2 and C_3 are arbitrary constants. These results may also be obtained from (2.3) if C_k (the arbitrary constants) are chosen such that $C_k = 0$ for $(k = 4,...,n)$.

Definition 3:

The hypergeometric functions in the variables x_i, $(i = 1,2,...,n)$ may be given by

$$F_D^{(n)}(A,B_1,B_2,...,B_n;\ C;\ x_1,...,x_n) = \frac{\Gamma(C)}{\Gamma(A)\Gamma(C-A)} \times$$

$$\int_0^1 u^{A-1}(1-u)^{C-A-1} \prod_{i=1}^{n} (1-ux_i)^{-B_i} \, du \tag{2.6}$$

where A,B_i,C $(i = 1,2,...,n)$ are numbers and Re $C >$ Re $A > 0$; $|\mathrm{Arg}(1-x_i)| < \pi$. Also,

$F_D^{(n)}(A,B_1,...,B_n;\ C;\ x_1,x_2,...,x_n) =$

$$\sum_{k_1,\ldots,k_n=0}^{\infty} \frac{(A)_{k_1+\ldots+k_n}(B_1)_{k_1}\ldots(B_n)_{k_n}}{(C)_{k_1+\ldots+k_n}} \frac{x_1^{k_1}x_2^{k_2}\ldots x_n^{k_n}}{\Gamma(k_1+1)\Gamma(k_2+1)\ldots\Gamma(k_n+1)}, \tag{2.7}$$

$(|x_i| < 1,\ i = 1,2,\ldots,n)$, where

$$(B)_r = B(B+1)\ldots(B+r-1), r \geq 1,\ (B_0) = 1,\ B \neq 0\ .$$

From the above we may conclude the following:–

(a) The integrals (2.6) remain unaltered under the 2(n+1) transformations

(i) $u = v$, (ii) $u = 1-v$

(iii) $u = \frac{v}{1-x_k+vx_k}$, (iv) $u = \frac{1-v}{1-vx_k}$,

$(k = 1,2,\ldots,n)$. (2.8)

(b) (i) is clearly the identity transformation.

(c) In applying (ii) to (2.6) we find that

$$F_D^{(n)}(A,B_1,\ldots,B_n;\ C;\ x_1,\ldots,x_n) = \prod_{i=1}^{n}(1-x_i)^{-B_i}$$

$$F_D^{(n)}\left[C-A,B_1,\ldots,B_n;\ C;\ \frac{x_1}{x_1-1},\ \frac{x_2}{x_2-1},\ldots,\ \frac{x_n}{x_n-1}\right] \tag{2.9}$$

(d) In applying (iii) $\left(u = \frac{v}{1-x_i+vx_i}\ ,\ i = 1,2,\ldots,n\right)$ to the integrals (2.6) we have

$$F_D^{(n)}(A,B_1,\ldots,B_n;\ C;\ x_1,\ldots,x_n) = (1-x_i)^{-A}$$

$$F_D^{(n)}\left[A,B_1,\ldots,B_{i-1},\ C-\sum_{i=1}^{n}B_i,B_{i+1},\ldots,B_n;\ C;\ \frac{x_1-x_i}{1-x_i},\ldots,\right.$$

$$\left.\frac{x_{i-1}-x_i}{1-x_i},\ \frac{x_i}{x_i-1},\ \frac{x_{i+1}-x_i}{1-x_i},\ldots,\ \frac{x_n-x_i}{1-x_i}\right] \tag{2.10}$$

(e) If we let $u = \frac{1-v}{1-vx_i}$, $(i = 1,2,\ldots,n)$ in (2.6) we find that

$$F_D^{(n)}(A,B_1,\ldots,B_n;\ C;\ x_1,\ldots,x_n) = (1-x_i)^{C-A-B_i}$$

$$\prod_{\substack{r=1\\ r\neq i}}^{n}(1-x_r)^{-B_r}\ F_D^{(n)}\left[C-A,B_1,\ldots,B_{i-1},\ C-\sum_{k=1}^{n}B_k,B_{i+1},\ldots,\right.$$

$$B_n;\ C;\ \frac{x_i-x_1}{1-x_1},\ \frac{x_i-x_2}{1-x_2},\ldots,\frac{x_i-x_{i-1}}{1-x_{i-1}},x_i,\frac{x_i-x_{i+1}}{1-x_{i+1}},\ldots,\frac{x_i-x_n}{1-x_n}\Big], \tag{2.11}$$

3. Equivalence Properties

It has been shown [9] that the integro–differential equation of Riemann–Papperitz type

$$y^{(n)} + \sum_{p=-1}^{m-2} \varphi_p \sum_{1\le i_1<i_2<\ldots<i_{p+2}\le m} \frac{W - \sum_{k=1}^{p+2} \alpha_{i_k} + 1}{\prod_{k=1}^{p+2} (x+a_{i_k})} \, {}_a I^{p-n+2}_x \, y(x) = f(x) \tag{3.1}$$

where $y^{(n)} = \frac{d^n y}{dx^n}$, $\varphi_{-1} = 1$, $\varphi_r = w(w-1)\ldots(w-r)$, $(r = 0,1,\ldots,m-2)$, α_i, a_i are numbers, $y \in C^{(n)}$, $f \in C$ on $[a,b]$, m,n are positive integers and $\mathrm{Re}(\lambda-w) > 0$, $(\lambda = 1,2,\ldots)$, is equivalent to the operator equation

$${}_a I^{-w}_x \prod_{i=1}^{m} (x+a_i)^{\alpha_i} \, {}_a I^{-1}_x \prod_{i=1}^{m} (x+a_i)^{1-\alpha_i} \, {}_a I^{w-n+1}_x \, y(x) = f(x) . \tag{3.2}$$

If in (3.2) we put $m = 5$, $n-1 = s$ and $f(x) = 0$, then the operator equation

$${}_a I^{-w}_x \prod_{i=1}^{5} (x+a_i)^{\alpha_i} \, {}_a I^{-1}_x \prod_{i=1}^{5} (x+a_i)^{1-\alpha_i} \, {}_a I^{w-s}_x \, y = 0 \tag{3.3}$$

represents an integro–differential equation of order (0,5) when $s = 4$, of order (1,4) if $s = 3$, of order (2,3) if $s = 2$, of order (3,2) if $s = 1$ and of order (4,1) when $s = 0$. For example the form of order (2,3) may be given by the integro–differential equation

$$y''' + \sum_{i=1}^{5} \frac{w-\alpha_i+1}{(x+a_i)} y'' + w \sum_{1\le i<j\le 5} \frac{w-\alpha_i-\alpha_j+1}{(x+a_i)(x+a_j)} y' +$$

$$+ w(w-1) \sum_{1\le i<j<k\le 5} \frac{w-\alpha_i-\alpha_j-\alpha_k+1}{(x+a_i)(x+a_j)(x+a_k)} y + \prod_{p=0}^{2} (w-p)$$

$$\sum_{1\le i<j<k<\ell\le 5} \frac{w-\alpha_i-\alpha_j-\alpha_k-\alpha_\ell+1}{(x+a_i)(x+a_j)(x+a_k)(x+a_\ell)} \, {}_a I^{x}_x \, y +$$

$$\prod_{p=0}^{3}(w-p)\,\frac{w-\sum_{i=1}^{5}\alpha_i+1}{\prod_{i=1}^{5}(x+a_i)}\,I_a^{x_2}\,y = 0 \quad .$$

This equation can be transformed to an equation of any one of the order indicated above, and so they are equivalent regarding their respective solutions.

It may be more convenient to deal with and study the form of order (0,5) represented by the fifth order ordinary differential equation

$$y^{v} + \sum_{i=1}^{5}\frac{w-\alpha_i+1}{x+a_i}\,y^{iv} + w\sum_{1\le i<j\le 5}\frac{w-\alpha_i-\alpha_j+1}{(x+a_i)(x+a_j)}\,y''' +$$

$$w(w-1)\sum_{1\le i<j<k\le 5}\frac{w-\alpha_i-\alpha_j-\alpha_k+1}{(x+a_i)(x+a_j)(x+a_k)}\,y'' + \prod_{p=0}^{2}(w-p)$$

$$\sum_{1\le i<j<k<\ell\le 5}\frac{w-\alpha_i-\alpha_j-\alpha_k-\alpha_\ell+1}{(x+a_i)(x+a_j)(x+a_k)(x+a_\ell)}\,y' + \prod_{p=0}^{3}(w-p)$$

$$\frac{w-\sum_{i=1}^{5}\alpha_i+1}{\prod_{i=1}^{5}(x+a_i)}\,y = 0 \quad , \tag{3.4}$$

which is equivalent to the operator equation when s = 4:

$$I_a^{x-w}\prod_{i=1}^{5}(x+a_i)^{\alpha_i}\,I_a^{x-1}\prod_{i=1}^{5}(x+a_i)^{1-\alpha_i}\,I_a^{x_{w-4}}\,y = 0 \tag{3.5}$$

Representation Of Fifth Order Equations By Operator Equations

It can be easily shown that any fifth order differential equation of the form (3.4) can be represented by its equivalent oeprator equation if w and α_i (i = 1,2,3,4,5) are determined. The equation:

$$\prod_{i=1}^{5}(x+a_i)Y^{v} + (Ax^4+Bx^3+Cx^2+Dx+E)Y^{iv} + (Fx^3+Gx^2+Hx+R)Y'''$$

$$+ (Jx^2+Kx+L)Y'' + (Mx+N)Y' + SY = 0 \tag{3.6}$$

is equivalent to the operator equation whenever:

$$H = 3w(w+1) \sum_{1\le i<j\le 5} a_i a_j - 2w\Big[\alpha_1 \sum_{2\le i<j\le 5} a_i a_j + \alpha_2 \sum_{\substack{1\le i<j\le 5\\(i,j\neq 2)}} a_i a_j + \alpha_3 \sum_{\substack{1\le i<j\le 5\\(i,j\neq 3)}} a_i a_j + \alpha_4 \sum_{\substack{1\le i<j\le 5\\(i,j\neq 4)}} a_i a_j + \alpha_5 \sum_{1\le i<j\le 4} a_i a_j\Big]$$

$$R = w(w+1) \sum_{1\le i<j<k\le 5} a_i a_j a_k - w\Big[\alpha_1 \sum_{2\le i<j<k\le 5} a_i a_j a_k + \alpha_2 \sum_{\substack{1\le i<j<k\le 5\\(i,j,k\neq 2)}} a_i a_j a_k + \alpha_3 \sum_{\substack{1\le i<j<k\le 5\\(i,j,k\neq 3)}} a_i a_j a_k + \alpha_4 \sum_{\substack{1\le i<j<k\le 5\\(i,j,k\neq 4)}} a_i a_j a_k + \alpha_5 \sum_{1\le i<j<k\le 4} a_i a_j a_k\Big]$$

$$J = 10w(w^2-1) - 6w(w-1) \sum_{i=1}^{5} \alpha_i$$

$$K = 4w(w^2-1) \sum_{i=1}^{5} a_i - 3w(w-1)\Big[\alpha_1 \sum_{i=2}^{5} a_i + \alpha_2 \sum_{\substack{i=1\\i\neq 2}}^{5} a_i + \alpha_3 \sum_{\substack{i=1\\i\neq 3}}^{5} a_i + \alpha_4 \sum_{\substack{i=1\\i\neq 4}}^{5} a_i + \alpha_5 \sum_{i=1}^{4} a_i\Big] \qquad (3.6.1)$$

$$L = w(w^2-1) \sum_{1\le i<j\le 5} a_i a_j - w(w-1)\Big[\alpha_1 \sum_{2\le i<j\le 5} a_i a_j + \alpha_2 \sum_{\substack{1\le i<j\le 5\\(i,j\neq 2)}} a_i a_j + \alpha_3 \sum_{\substack{1\le i<j\le 5\\(i,j\neq 3)}} a_i a_j + \alpha_4 \sum_{\substack{1\le i<j\le 5\\(i,j\neq 4)}} a_i a_j + \alpha_5 \sum_{1\le i<j\le 4} a_i a_j\Big]$$

$$M = 5w(w^2-1)(w-2) - 4w(w-1)(w-2) \sum_{i=1}^{5} \alpha_i$$

$$N = w(w^2-1)(w-2)\sum_{i=1}^{5} a_i - w(w-1)(w-2)\Big[\alpha_1 \sum_{i=2}^{5} a_i + \alpha_2 \sum_{\substack{i=1 \\ i\neq 2}}^{5} a_i + \alpha_3 \sum_{\substack{i=1 \\ i\neq 3}}^{5} a_i + \alpha_4 \sum_{\substack{i=1 \\ i\neq 4}}^{5} a_i + \alpha_5 \sum_{i=1}^{4} a_i\Big]$$

$$S = w(w-1)(w-2)(w-3)(w - \sum_{i=1}^{5} \alpha_i + 1) .$$

The values of w and α_i (i = 1,2,...,5) may be easily obtained from equations (3.6.1) above. Also, if a differential equation is given in the form (3.4), then it would not be difficult to find the values of α_i (i = 1,...,5) and w and consequently it can be expressed by an equivalent operator equation.

4. Solutions of Equations

Solutions of (3.4) may be obtained by finding solutions of its equivalent operator equation (3.5).

By applying property (2.5), by using the equality $D_x^4 \, {}_a^x I^{4-w} = {}_a^x I^{-w}$ in (3.5) and performing the inverse operations of the integro–differential operator of generalized order we find that

$$y(x;a) = {}_a^x I^{4-w} \prod_{i=1}^{5} (x+a_i)^{\alpha_i - 1}\Big[K + {}_a^x I \prod_{i=1}^{5} (x+a_i)^{-\alpha_i}\Big\{C_1 \frac{(x-a)^{w-4}}{\Gamma(w-3)} + C_2 \frac{(x-a)^{w-3}}{\Gamma(w-2)} + C_3 \frac{(x-a)^{w-2}}{\Gamma(w-1)} + C_4 \frac{(x-a)^{w-1}}{\Gamma(w)}\Big\}, \qquad (S)$$

where K and C_i (i = 1,2,3,4) are arbitrary constants.

If we let $C_1 = \Gamma(w-3)$, $C_2 = \Gamma(w-2)$, $C_3 = \Gamma(w-1)$, $C_4 = \Gamma(w)$, then the fundamental solutions may be written as

$$y_1(x;a) = K \, {}_a^x I^{4-w} \prod_{i=1}^{5} (x+a_i)^{\alpha_i - 1} \qquad (S_1)$$

$$y_2(x;a) = {}_a^x I^{4-w} \prod_{i=1}^{5} (x+a_i)^{\alpha_i - 1} \; {}_a^x I \prod_{i=1}^{5} (x+a_i)^{-\alpha_i} (x-a)^{w-4} \qquad (S_2)$$

$$y_3(x;a) = {}_a^x I^{4-w} \prod_{i=1}^{5} (x+a_i)^{\alpha_i - 1} \; {}_a^x I \prod_{i=1}^{5} (x+a_i)^{-\alpha_i} (x-a)^{w-3} \qquad (S_3)$$

$$y_4(x;a) = {}_a^x I^{4-w} \prod_{i=1}^{5} (x+a_i)^{\alpha_i - 1} \; {}_a^x I \prod_{i=1}^{5} (x+a_i)^{-\alpha_i} (x-a)^{w-2} \qquad (S_4)$$

$$y_5(x;a) = {}_a^x I^{4-w} \prod_{i=1}^{5} (x+a_i)^{\alpha_i - 1} \; {}_a^x I \prod_{i=1}^{5} (x+a_i)^{-\alpha_i} (x-a)^{w-1} \qquad (S_5)$$

The fundamental solutions (S_i), $(i = 1,...,5)$ may be determined according to the singular points of the differential equation (3.4) or (3.6). These singular points are: $-a_i$, $-\infty$ or $+\infty$ $(i = 1,...,5)$. Thus, the lower limits (a) of the integrals may be represented by these singular points. It can be easily seen that these solutions are linearly independent.

Now, for each lower limit we may obtain five fundamental solutions. Consequently thirty principal solutions are obtained by using the values of singular points as lower limits of the integrals. According to the transformations (2.8) each fundamental solution may be expressed in ten forms and the total number of branch solutions would be three hundreds.

In general the number of solutions of an equation of nth order with m singular points including $(-\infty)$ may be estimated as follows: The equation has n fundamental solutions, a number of nm principal solutions and $2nm(m-1)$ branch solutions since there are $2(m-1)$ transformations which may leave the integrals unaltered as indicated by (2.8).

Verification of Solutions:

To show that S_i $(i = 1,2,...,5)$ are solutions of (3.5) we may show that they satisfy the equation. It would be sufficient to show that any one of the solutions satisfy the equation. If $y_2(x;a)$ is substituted in (3.5) and by using the operational properties of the operator of generalized order, with the fact that ${}_a^x I^{\alpha}\, {}_a^x I^{\beta} = {}_a^x I^{\alpha+\beta}$ (as shown in [8]), then the left hand side of (3.5) takes the form:

$$\prod_{a}^{x-w}\prod_{i=1}^{5}(x+a_i)^{\alpha_i}\left[\prod_{i=1}^{5}(x+a_i)^{-\alpha_i}(x-a)^{w-4}\right] = D_x^4\, I_a^{4-w}(x-a)^{w-4}$$
$$= D_x^4\,\Gamma(w-3) = 0$$

4. A Study of A Particular Case

A particular case of (3.5) where $a_1 = 0$, $a_2 = 1$, $a_3 = -1$, $a_4 = -r$, $a_5 = -p$ will be studied. In this case the operator equation takes the form

$$I_a^{x-w}\, x^{\alpha_1}(1+x)^{\alpha_2}(1-x)^{\alpha_3}(r-x)^{\alpha_4}(p-x)^{\alpha_5}\, I_a^{x-1}\, x^{1-\alpha_1}(1+x)^{1-\alpha_2}(1-x)^{1-\alpha_3}$$
$$(r-x)^{1-\alpha_4}(p-x)^{1-\alpha_5}\, I_a^{x-w-4}\, y = 0 \qquad (4.1)$$

which is equivalent to (3.4) when a_i $(i = 1,...,5)$ are replaced by its values given above. This is also equivalent to the fifth order differential equation

$$x(1+x)(1-x)(r-x)(p-x)y^{v} - (A_1x^4+B_1x^3+C_1x^2+D_1x+E_1)y^{iv} -$$
$$(F_1x^2+H_1x+R_1)y''' - (J_1x^2+K_1x+L_1)y'' - (M_1x+N_1)y' - S_1y = 0 \qquad (4.2)$$

where

$$A_1 = 5(w-1) - \sum_{i=1}^{5}\alpha_i$$

$$B_1 = -4(w+1)(r+p) - \left[r\sum_{\substack{i=1\\ i\neq 4}}^{5}\alpha_i + p\sum_{i=1}^{4}\alpha_i + \alpha_2 - \alpha_3\right]$$

$$C_1 = 3(w+1)(rp-1) - rp(\alpha_1-\alpha_2-\alpha_3) + (r+p)(\alpha_3-\alpha_2) + \alpha_1 - \alpha_4 - \alpha_5$$

$$D_1 = 2(w+1)(r+p-rp) - r(\alpha_1+\alpha_5) + rp(\alpha_2-\alpha_3) - \alpha_4 p$$

$$E_1 = -\,-4(w+1)rp - \alpha_1 rp$$

$$F_1 = 10w(w+1) - 4w\sum_{i=1}^{5}\alpha_i$$

$$G_1 = -6w(w+1)(r+p) + 3w\left[(r+p)\sum_{i=1}^{3}\alpha_i + p\alpha_4 + r\alpha_5 + \alpha_2 - \alpha_3\right]$$

$$H_1 = 3w(w+1)(rp-1) - w\left[rp\sum_{i=1}^{3}\alpha_i + (r+p)(\alpha_2-\alpha_3) - \alpha_4 - \alpha_5\right]$$

$$R_1 = w(w+1)(r+p-rp) - w\left[rp(\alpha_3-\alpha_2) + r(\alpha_1+\alpha_5) + p\alpha_4\right]$$

$$J_1 = 10w(w^2-1) - 6w(w-1)\sum_{i=1}^{5}\alpha_i$$

$$K_1 = -4(w^2-1)(r+p) + 3w(w-1)\left[(r+p)\sum_{i=1}^{3}\alpha_i + p\alpha_4 + r\alpha_5 + \alpha_2 - \alpha_3\right]$$

$$L_1 = w(w^2-1)(rp-1) - w(w-1)\left[rp\sum_{i=1}^{3}\alpha_i + (r+p)(\alpha_2-\alpha_3) - \alpha_4 - \alpha_5\right]$$

$$M_1 = 5w(w^2-1)(w-2) - 4w(w-1)(w-2)\sum_{i=1}^{5}\alpha_i$$

$$N_1 = -w(w^2-1)(w-2)(r+p) - w(w-1)(w-2)\left[(r+p)\sum_{i=1}^{3}\alpha_i + p\alpha_4 + r\alpha_5 + \alpha_2 - \alpha_2\right]$$

$$S_1 = \prod_{p=0}^{3}(w-p)\left(w - \sum_{i=1}^{5}\alpha_i + 1\right) \qquad (4.3)$$

Solutions of (4.1)

Solutions of (4.1) satisfy the equivalent differential equation (4.2). These fundamental solutions as in the general case indicated by (S_i) may be given by:

$$y_1(x;a) = K\, I_a^{4-w}\, x^{\alpha_1-1}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-1}$$

$$y_2(x;a) = I_a^{4-w}\, x^{\alpha_1-1}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-1}\, I_a^{x}\, x^{-\alpha_1}$$
$$(1+x)^{-\alpha_2}(1-x)^{-\alpha_3}(r-x)^{-\alpha_4}(p-x)^{-\alpha_5}(x-a)^{w-4}$$

$$y_3(x;a) = I_a^{4-w}\, x^{\alpha_1-1}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-1}\, I_a^{x}\, x^{-\alpha_1}$$
$$(1+x)^{-\alpha_2}(1-x)^{-\alpha_3}(r-x)^{-\alpha_4}(p-x)^{-\alpha_5}(x-a)^{w-3}$$

$$y_4(x;a) = I_a^{4-w}\, x^{\alpha_1-1}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-1}\, I_a^{x}\, x^{-\alpha_1}$$
$$(1+x)^{-\alpha_2}(1-x)^{-\alpha_3}(r-x)^{-\alpha_4}(p-x)^{-\alpha_5}(x-a)^{w-2}$$

$$y_5(x;a) = I_a^{x_{4-w}}\, x^{\alpha_1-1}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-1}\, I_a^{x}\, x^{-\alpha_1}$$

$$(1+x)^{-\alpha_2}(1-x)^{-\alpha_3}(r-x)^{-\alpha_4}(p-x)^{-\alpha_5}(x-a)^{w-2}\ .$$

Solutions $y_i(x;a)$, $(i = 1,\ldots,5)$ are determined according to the singular points of the differential equation (4.2). These are: 0,1,−1,r,p and $(-\infty)$. Therefore the lower limit (a) of the integrals above may take these values of the singular points. For each singular point as a lower limit we find six basic solutions. Hence a number of thirty principal solutions are obtained. But each one of these can be expressed in ten forms by applying transformations (2.8), with $(k = 1,2,3,4)$, to solutions represented by integrals of the forms (2.6), which may yield hypergeometric functions of the forms $F_D^{(4)}$. Therefore a total of three hundred branch solutions may be obtained. Thus this number of solution forms are found when ten transformations are applied to the thirty solutions $y_i(x;a)$, $(i = 1,2,\ldots,5)$, where a may be any one of the six singular points of the equation.

It would be sufficient to discuss some few cases of these solutions such as $y_1(x;a)$, $y_2(x;1)$ and $y_3(x,-\infty)$, as the other solutions can be similarly studied.

$\underline{y_1(x;0)}$:

By expressing $y_1(x;a)$ in the integral form, with lower limit $a = 0$, and by applying the transformation $t = xu$ we find that

$$y_1(x;0) = \frac{K}{\Gamma(4-w)}\int_0^x (x-t)^{3-w}\, t^{\alpha_1-1}(1+t)^{\alpha_2-1}(1-t)^{\alpha_3-1}(r-t)^{\alpha_4-1}$$

$$(p-t)^{\alpha_5-1}\,dt$$

$$= \frac{K\,\Gamma(\alpha_1)}{\Gamma(\alpha_1-w+4)}\, x^{\alpha_1-w+3}\, r^{\alpha_4-1}\, p^{\alpha_5-1}\, F_D^{(4)}\Big[\alpha_1, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4,$$

$$1-\alpha_5;\ \alpha_1-w+4;\ -x, x,\ \frac{x}{r},\ \frac{x}{p}\Big]\ . \qquad (4.4)$$

In applying (2.8) for $(k = 1,2,3,4)$ to (4.4), i.e. the tranformations:

$$u = v,\ u = 1-v,\ u = \frac{v}{1-x_i+vx_i},\ u = \frac{1-v}{1-vx_i} \qquad (4.5)$$

$(i = 1,2,3,4)$, with $x_1 = -x$, $x_2 = x$, $x_3 = \frac{x}{r}$ and $x_4 = \frac{x}{p}$, as shown by (2.9), (2.10) and (2.11), we would get the following ten branches:

If $M_1 = \frac{K\,\Gamma(\alpha_1)}{\Gamma(\alpha_1-w+4)}$, then

$$y_{1,1}(x;0) = M_1 r^{\alpha_4-1} p^{\alpha_5-1} x^{\alpha_1-w+3} F_D^{(4)}\Big[\alpha_1, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4, 1-\alpha_5; \alpha_1-w+4; -x, x, \frac{x}{r}, \frac{x}{p}\Big]$$

$$y_{1,2}(x;0) = M_1 x^{\alpha_1-w+3} (1+x)^{\alpha_2-1} (1-x)^{\alpha_3-1} (r-x)^{\alpha_4-1} (p-x)^{\alpha_5-1} F_D^{(4)}\Big[4-w, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4, 1-\alpha_5; \alpha_1-w+4; \frac{x}{x+1}, \frac{x}{x-1}, \frac{x}{x-r}, \frac{x}{x-p}\Big]$$

$$y_{1,3}(x;0) = M_1 r^{\alpha_4-1} p^{\alpha_5-1} x^{\alpha_1-w+3} F_D^{(4)}\Big[\alpha_1, \sum_{i=1}^{5} \alpha_i - w, 1-\alpha_3, 1-\alpha_4, 1-\alpha_5; \alpha_1-w+4; \frac{x}{x+1}, \frac{2x}{x-1}, \frac{x(1+r)}{(1+x)r}, \frac{x(1+p)}{(1+x)p}\Big]$$

$$y_{1,4}(x;0) = M_1 r^{\alpha_4-1} p^{\alpha_5-1} x^{\alpha_1-w+3} (1-x)^{-\alpha_1} F_D^{(4)}\Big[\alpha_1, 1-\alpha_2, \sum_{i=1}^{5} \alpha_i - w, 1-\alpha_4, 1-\alpha_5; \alpha_1-w+4; \frac{2x}{1-x}, \frac{x}{x-1}, \frac{x(1-r)}{(1-x)r}, \frac{x(1-p)}{(1-x)p}\Big]$$

$$y_{1,5}(x;0) = M_1 r^{\alpha_1+\alpha_4-1} p^{\alpha_5-1} x^{\alpha_1-w+3} (r-x)^{-\alpha_1} F_D^{(4)}\Big[\alpha_1, 1-\alpha_2, 1-\alpha_3, \sum_{i=1}^{5} \alpha_i - w, 1-\alpha_5; \alpha_1-w+4; \frac{(r+1)x}{x-r}, \frac{(r-1)x}{r-x}, \frac{x}{x-r}, \frac{(r-p)x}{p(r-x)}\Big]$$

$$y_{1,6}(x;0) = M_1 r^{\alpha_4-1} p^{\alpha_1+\alpha_5-1} x^{\alpha_1-w+3} (p-x)^{-\alpha_1} F_D^{(4)}\Big[\alpha_1, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4, \sum_{i=1}^{5} \alpha_i - w; \alpha_1-w+4; \frac{(p+1)x}{x-p}, \frac{(p-1)x}{p-x}, \frac{(p-r)x}{r(p-x)}, \frac{x}{x-p}\Big]$$

$$y_{1,7}(x;0) = M_1 x^{\alpha_1-w+3} (1+x)^{\alpha_2-w+3} (1-x)^{\alpha_3-1} (r-x)^{\alpha_4-1} (p-x)^{\alpha_5-1} \times$$

$$\times F_D^{(4)}\left[4-w, \sum_{i=1}^{5}\alpha_i-w;\ 1-\alpha_3, 1-\alpha_4, 1-\alpha_5;\ \alpha_1-w+4;\ -x,\ \frac{2x}{x-1},\right.$$
$$\left.\frac{(r+1)x}{x-r},\ \frac{(p+1)x}{x-p}\right]$$

$$y_{1,8}(x;0) = M_1 x^{\alpha_1-w+3}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-w+3}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-1}$$
$$F_D^{(4)}\left[4-w, 1-\alpha_2, \sum_{i=1}^{5}\alpha_i-w,\ 1-\alpha_4, 1-\alpha_5;\ \alpha_1-w+4;\right.$$
$$\left.\frac{2x}{x+1},\ x,\ \frac{(1-r)x}{r-x},\ \frac{(1-p)x}{p-x}\right]$$

$$y_{1,9}(x;0) = M_1 r^{w-4} x^{\alpha_1-w+3}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-w+3}(p-x)^{\alpha_5-1}$$
$$F_D^{(4)}\left[4-w, 1-\alpha_2, 1-\alpha_3, \sum_{i=1}^{5}\alpha_i-w, 1-\alpha_5;\ \alpha_1-w+4;\right.$$
$$\left.\frac{(1+r)x}{r(1+x)},\ \frac{(1-r)x}{r(1-x)},\ \frac{x}{r},\ \frac{(p-r)x}{r(p-x)}\right]$$

$$y_{1,10}(x;0) = M_1 p^{w-4} x^{\alpha_1-w+3}(1+x)^{\alpha_2-1}(1-x)^{\alpha_3-1}(r-x)^{\alpha_4-1}(p-x)^{\alpha_5-w+3}$$
$$F_D^{(4)}\left[4-w, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4, \sum_{i=1}^{5}\alpha_i-w;\ \alpha_1-w+4;\right.$$
$$\left.\frac{(1+p)x}{p(1+x)},\ \frac{(1-p)x}{p(1-x)},\ \frac{(r-p)x}{p(r-x)},\ \frac{x}{p}\right]$$

$\underline{y_2(x;1):}$

This solution may be written in the integral form (after replacing a by 1):

$$y_2(x;1) = \frac{(-1)^{w-4}}{\Gamma(4-w)}\int_1^x (x-t)^{3-w}\, t^{\alpha_1-1}(1+t)^{\alpha_2-1}(1-t)^{\alpha_3-1}(r-t)^{\alpha_4-1}$$
$$(p-t)^{\alpha_5-1}\left[\int_1^t z^{-\alpha_1}(1+z)^{-\alpha_2}(1-z)^{w-\alpha_3-4}(r-z)^{-\alpha_4}\right.$$
$$\left.(p-z)^{-\alpha_5}\, dz\right]dt$$

If $u = \frac{1-z}{1-t}$ in $I = \int_1^t z^{-\alpha_1}(1+z)^{-\alpha_2}(1-z)^{w-\alpha_3-4}(r-z)^{-\alpha_4}(p-z)^{-\alpha_5}\, dz$

Then,

$$I = -z^{-\alpha_2}(r-1)^{-\alpha_4}(p-1)^{-\alpha_5}(1-t)^{w-\alpha_3-3} F_D^{(4)}\left[w-\alpha_3-3, \alpha_1, \alpha_2, \alpha_4, \alpha_5; w-\alpha_3-2; 1-t, \frac{1-t}{2}, \frac{1-t}{1-r}, \frac{1-t}{1-p}\right],$$

and

$$y_{2,1}(x;1) = \frac{(-1)^{w-1}(r-1)^{-\alpha_4}(p-1)^{-\alpha_5}}{w-\alpha_3-2} \int_1^x 4-w\ x^{\alpha_1-1}(1+x)^{\alpha_2-1}(1-x)^{w-4}$$
$$(r-x)^{\alpha_4-1}(p-x)^{-\alpha_5} F_D^{(4)}\left[w-\alpha_3-3, \alpha_1, \alpha_2, \alpha_4, \alpha_5; w-\alpha_3-2; 1-x, \frac{1-x}{2}, \frac{1-x}{1-r}, \frac{1-x}{1-p}\right] .$$

The other branch solutions $y_{2,j}(x;1)$, $(j = 2,3,\ldots,10)$ may be obtained by applying the same transformations (4.5) to $F_D^{(4)}$ in $y_{2,1}(x;1)$.

$y_3(x;-\infty)$:

In this case where $a \to -\infty$ in y_2, y_3, y_4 and y_5, it is convenient, for the evaluation of the integrals, to replace $(x-a)$ in y_2 by unity, in y_3 by x, in y_4 by x^2 and y_5 by x^3. This is justified by the choice of arbitrary constants (2.4), where in the case of y_2, $(C_1 = 1, C_i = 0$ for $i \neq 1)$, for the case of y_3, $(C_i = 0$ for all $i \neq 2$ and $C_2 = 1)$, for y_4, $(C_i = 0$ for all $i \neq 3$, $C_3 = 1)$ and for the case of y_5, $(C_i = 0$ for all $i \neq 4$ and $C_4 = 1)$. We have

$$y_3(x;-\infty) = \frac{1}{\Gamma(4-w)} \int_{-\infty}^{x} (x-t)^{3-w}\ t^{\alpha_1-1}(1+t)^{\alpha_2-1}(1-t)^{\alpha_3-1}(r-t)^{\alpha_4-1}$$
$$(p-t)^{\alpha_5-1}\left[\int_{-\infty}^{t} z^{1-\alpha_1}(1+z)^{-\alpha_2}(1-z)^{-\alpha_3}(r-z)^{-\alpha_4}\right.$$
$$\left.(p-z)^{-\alpha_5}\, dz\right] dt .$$

If $z = tu^{-1}$ in $I = \int_{-\infty}^{t} z^{1-\alpha_1}(1+z)^{-\alpha_2}(1-z)^{-\alpha_3}(r-z)^{-\alpha_4}(p-z)^{-\alpha_5}\, dz$,

then,

$$I = \frac{(-1)^{1-\sum_{i=3}^{5}\alpha_i}\, t^{2-\sum_{i=1}^{5}\alpha_i}}{\sum_{i=1}^{5}\alpha_i - 1} F_D^{(4)}\left[\sum_{i=1}^{5}\alpha_i-2, \alpha_2, \alpha_3, \alpha_4, \alpha_5;\ \sum_{i=1}^{5}\alpha_i-1;\right.$$

$$\left. -\frac{1}{t}, \frac{1}{t}, \frac{r}{t}, \frac{p}{t}\right]$$

and consequently,

$$y_{3,1}(x;-\infty) = \frac{(-1)^{1-\sum_{i=2}^{5}\alpha_i}}{\sum_{i=1}^{5}\alpha_1 - 1} \int_{-\infty}^{x} t^{4-w}\, x^{1-\sum_{i=2}^{5}\alpha_i} (1+x)^{\alpha_2-1} (1-x)^{\alpha_3-1}$$

$$(r-x)^{\alpha_4-1} (p-x)^{\alpha_5-1} F_D^{(4)}\left[\sum_{i=1}^{5}\alpha_i-2, \alpha_2, \alpha_3, \alpha_4, \alpha_5;\right.$$

$$\left.\sum_{i=1}^{5}\alpha_i-1;\ -\frac{1}{x}, \frac{1}{x}, \frac{r}{x}, \frac{p}{x}\right] .$$

Also $y_{3,j}(x;-\infty)$, (j = 2,3,...,10) may be found by using the transformations (4.5) in the hypergeometric functions $F_D^{(4)}$ in the integral of $y_{3,1}(x;-\infty)$.

By a similar approach other branch solutions may be found. These solutions are obtained from the five fundamental solutions $y_i(x;a)$, (i = 1,2,...,5) which take anyone of the six values of the singular points yielding thirty principal solutions. The three hundred branch solutions are found by applying to each principal solution the ten transformations (4.5).

5. Generalized Riemann–Papperitz Equations

The general case of (3.1) where m = n and f(x) ≡ 0.

In this case we find that the nth order linear differential equation with n singular points:

$$y^{(n)} + \sum_{i=1}^{n} \frac{w-\alpha_i+1}{x+a_i} y^{(n-1)} + w \sum_{1\le i<j\le n} \frac{w-\alpha_i-\alpha_j+1}{(x+a_i)(x+a_j)} y^{(n-2)}$$

$$+ w(w-1) \sum_{1\le i<j<k\le n} \frac{w-\alpha_i-\alpha_j-\alpha_k+1}{(x+a_i)(x+a_j)(x+a_k)} y^{(n-3)}$$

$$+ w(w-1)(w-2) \sum_{1\le i<j<k<\ell\le n} \frac{w-\alpha_i-\alpha_j-\alpha_k-\alpha_\ell+1}{(x+a_i)(x+a_j)(x+a_k)(x+a_\ell)} y^{(n-4)}$$

$$+\ldots + \prod_{i=0}^{n-2} (w-i) \frac{w - \sum_{i=1}^{n} \alpha_i + 1}{\prod_{i=1}^{n} (x+a_i)} y = 0 , \qquad (5.1)$$

where $y^{(n)} = \frac{d^n y}{dx^n}$ is equivalent to the operator equation

$$\prod_{a}^{x-w} \prod_{i=1}^{n} (x+a_i)^{\alpha_i} \prod_{a}^{x-1} \prod_{i=1}^{n} (x+a_i)^{1-\alpha_i} \prod_{a}^{x^{w-n+1}} y = 0 . \qquad (5.2)$$

Representation by operator Equations:

The equations in the form (5.1) can be put in the form similar to (3.6), that is in the form:

$$\prod_{i=1}^{n} (x+a_i) Y^{(n)} + \left[\sum_{i=1}^{n} A_i x^{n-i} \right] Y^{(n-1)} + \left[\sum_{i=1}^{n-1} B_i x^{n-i-1} \right] Y^{(n-2)}$$

$$+\ldots+ (M_1 x + M_2) Y' + SY = 0 . \qquad (5.3)$$

Any equation of the form (5.3) can be represented by an operator equation of the form (5.2) by determining w and α_i (i = 1,2,...,n) from equations similar to (3.6.1), through the establishment of equivalence properties such as in the case of the fifth order differential equations, which give the relations between the coefficients A_i, B_j,...,M_1, M_2 and S, (i = 1,2,...,n), (j = 1,2,...,n–1) and that of the parameters w and α_i (i = 1,...,n). These are all have become possible through the use of (fractional calculus) Generalized Calculus and the integro–differential operator of generalized order.

It may be pointed out that this writer has generalized the second order Riemann's equations and Papperitz equations ([26], p. 206 & p. 283) to nth order integro–differential equations as given by (3.1) ([9], p. 7) and has found its equivalent operator equations.

A Study of Solutions:

Solutions of (5.1) may be obtained by finding solutions of its equivalent operator equations (5.2). By using properties of the integro–differential operator of generalized order as applied to (5.2). Since $D_x^{n-1} \prod_{a}^{x(n-1)-w} = \prod_{a}^{x-w}$, then

$$y(x;a) = \overset{x}{\underset{a}{I}}^{(n-1)-w} \prod_{i=1}^{n} (x+a_i)^{\alpha_i-1}\left[K + \overset{x}{\underset{a}{I}} \prod_{i=1}^{n} (x+a_i)^{-\alpha_i}\right]\left\{\sum_{i=1}^{n-1} C_i \frac{(x-a)^{w-i}}{\Gamma(w-i+1)}\right\} \tag{S}$$

where C_i, K are arbitrary constants.

Let $C_i = \Gamma(w-i+1)$, then we have the n–fundamental solutions:–

$$y_i(x;a) = K \overset{x}{\underset{a}{I}}^{(n-1)-w} \prod_{i=1}^{n} (x+a_i)^{\alpha_i-1} \tag{S_i}$$

$$y_j(x;a) = \overset{x}{\underset{a}{I}}^{(n-1)-w} \prod_{i=1}^{n} (x+a_i)^{\alpha_i-1} \overset{x}{\underset{a}{I}} \prod_{i=1}^{n} (x+a_i)^{-\alpha_i} (x-a)^{w-j+1} \tag{S_j}$$

(j = 2,3,...,n). These fundamental solutions (S_i) (i = 1,2,...,n) may be determined according to the singular points of the differential equation (5.1). These singular points are given the set $\{\bar{a}_1, \bar{a}_2, ..., \bar{a}_n, -\infty \text{ or } \infty\}$, where ($\bar{a}_i = -a_i$). Thus the lower limits (a) of the integrals in $\{S_i\}$ may be determined by these singular points in evaluating the n solutions. For each singular point we obtain n principal solutions. Thus we have n(n+1) principal solutions since the number of singular points is (n+1) including ($-\infty$). But each principal solution amy be expressed in 2n equivalent integrals as indicated by (2.8). Therefore equation (5.1) has $2n^2(n+1)$ branch solutions. These solutions are expressed in the hypergeometric functions $F_D^{(n-1)}$. The principal solutions are linearly independent and they can be easily verified that they satisfy the equation as shown in section (4) of this article for the case n = 5. Evaluation and other details of these solution can be dealt with in the same approach as the one used here in dealing with the fifth order differential equation. The number (a) which appear in (S_i), (i = 1,2,...,n) may be replaced by the singular points in evaluating the integrals. When $a = -\infty$ the replacement and evaluation of solutions follow the same way as that dealt with in section 4 of the article.

6. The Extended Riemann P–Function

1. Riemann P–Function:

Riemann P–Function is the scheme

$$P\left\{\begin{array}{cccc} A_1 & A_2 & A_3 & \\ \alpha & \beta & \gamma & x \\ \alpha' & \beta' & \gamma' & \end{array}\right\}, \tag{6.1}$$

with the exponent α, β, γ, α', β', γ' and the singular points A_i, i = 1,2,3 is associated with the Gauss' hypergeometric equation

$$x(1-x)y'' + \{c - (a+b+1)x\} - aby = 0 \tag{6.2}$$

where the singular points are $\{1, -\infty, 0\}$, ([5], pp. 82–104) and ([16], pp. 206–209, 281–296). Kummer's twenty four solutions of this equation are associated with six principal branch sets each of which contains four branch solutions, i.e. $\{P^{(\alpha)}, P^{(\alpha')}\}$, $\{P^{(\beta)}, P^{(\beta')}\}$ and $\{P^{(\gamma)}, P^{(\gamma')}\}$. It has been shown that these functions have interesting properties ([5], pp. 88–92). In a previous work the author has shown that these sets are associated with the equations singular points. In fact, these solutions have been obtained through integrals with singular points as lower limits as shown for the fifth order equation in section 4. If in equation (3.1), m = 3 and n = 2, then the equation takes the form

$$y'' + \sum_{i=1}^{3} \frac{w-\alpha_i+1}{x+a_i} y' + w \sum_{1\le i<j\le 3} \frac{w-\alpha_i-\alpha_j+1}{(x+a_i)(x+a_j)} y$$

$$+ w(w-1) \frac{w - \sum_{i=1}^{3} \alpha_i + 1}{\prod_{i=1}^{3} (x+a_i)} I_a^x y = 0 . \tag{6.3}$$

This is reduced to second order differential equation of Riemann type if

$$w - \sum_{i=1}^{3} \alpha_i + 1 = 0 \tag{6.4}$$

This condtion is satisfied by the values of w and α_i which are given by matrix equation

$$\begin{bmatrix} w \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = U \begin{bmatrix} \alpha+\beta+\gamma \\ \alpha'+\beta+\gamma \\ \alpha+\beta'+\gamma \\ \alpha+\beta+\gamma' \end{bmatrix} \tag{6.5}$$

where U is the (4×4) unit matrix, α, β, γ, α', β', γ' are the indices of Riemann P–Function where $\Sigma(\alpha+\alpha') = \alpha+\beta+\gamma+\alpha'+\beta'+\gamma' = 1$. By applying condition (6.4) to equation (6.3), choosing the singular points 1, 0, ∞ and if $a_2 \to \infty$, then (6.3) is reduced to Gauss' equation of the form (6.2)

([9], pp. 8–9).

2. The M–Functions (the extended Riemann's P–Functions) for the third and fourth order equations

If m = n = 3 in (3.1) or (3.2) we notice that the differential equation

$$y''' + \sum_{i=1}^{3} \frac{w-\alpha_i+1}{a_i+x}\, y'' + w \sum_{1\le i<j\le 3} \frac{w-\alpha_i-\alpha_j+1}{(a_i+x)(a_j+x)}\, y' + w(w-1)\, \frac{w-\sum_{i=1}^{3}\alpha_i+1}{\prod_{i=1}^{3}(a_i+x)}\, y = 0 \qquad (6.6)$$

is equivalent to the operator equation

$$I_a^{x-w} \prod_{i=1}^{3} (a_i+x)^{\alpha_i} I_a^{x-1} \prod_{i=1}^{3} (a_i+x)^{1-\alpha_i} I_a^{x w-2} = 0 \qquad (6.7)$$

This operator equation is equivalent to the third order linear differential equation

$$\prod_{i=1}^{3} (a_i+x)\, y''' + (Ax^2+Bx+C)y'' + (Dx+E)y' + Sy = 0 \qquad (6.6.1)$$

if:

$$\begin{aligned}
A &= 2(w+1) - \sum_{i=1}^{3} \alpha_i \\
B &= 2(w+1) \sum_{i=1}^{3} a_i - [\alpha_1(a_2+a_3) + \alpha_2(a_1+a_3) + \alpha_3(a_1+a_2)] \\
C &= (w+1) \sum_{1\le i<j\le 3} a_i a_j - (\alpha_1 a_2 a_3 + \alpha_2 a_1 a_3 + \alpha_3 a_1 a_2) \\
D &= w[3(w+1) - 2 \sum_{i=1}^{3} \alpha_i] \qquad (6.8) \\
E &= w(w+1) \sum_{i=1}^{3} a_i - w\,[\alpha_1(a_2+a_3) + \alpha_2(a_1+a_3) + \alpha_3(a_1+a_2)] \\
S &= w(w-1)(w - \sum_{i=1}^{3} a_i + 1)\ .
\end{aligned}$$

It is clear that (6.3) is the same equation as (6.6.1). These third order equations would have seventy two branch solutions, as it has been indicated

in this article, for the equation has four singularities including $(-\infty)$ and there are six transformations as mentioned in (2.8) with $(k = 1,2)$. The number of these transformations is determined by twice the number of singularities excluding $(-\infty)$.

If in (3.1), $m = 4$ and $n = 3$ we would have an operator equation similar to (6.7) with $m = 4$ and it would be equivalent to the integro–differential equation

$$y''' + \sum_{i=1}^{4} \frac{w-\alpha_i+1}{a_i+x}\, y'' + w \sum_{1\le i<j\le 4} \frac{w-\alpha_i-\alpha_j+1}{(a_i+x)(a_j+x)}\, y' + w(w-1) \sum_{1\le i<j<k\le 4} \frac{w-\alpha_i-\alpha_j-\alpha_k+1}{(a_i+x)(a_j+x)(a_k+x)}\, y + w(w-1)(w-2) \times \frac{w-\sum_{i=1}^{4} \alpha_i+1}{\prod_{i=1}^{4} (a_i+x)} \int_a^x y = 0 \qquad (6.9)$$

This equation is reduced to (6.6) if

$$w - \sum_{i=1}^{4} \alpha_i + 1 = 0 \qquad (6.10)$$

and $a_4 \to \pm \infty$.

Condition (6.10) is satisfied by the values of w and α_i given by matrix equation

$$\begin{bmatrix} w \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = U \begin{bmatrix} \alpha+\beta+\gamma+\delta \\ \alpha+\beta+\gamma+\delta \\ \alpha'+\beta'+\gamma'+\delta' \\ \alpha''+\beta''+\gamma''+\delta'' \\ \alpha+\beta+\gamma+\delta \end{bmatrix} \qquad (6.11)$$

where U is the (5×5) unit matrix, α, β, γ, δ, α', β', γ', δ', α'', β'', γ'', δ'' are the indices of the M–Function (The extended Riemann P–Function) where $\Sigma(\alpha+\alpha'+\alpha'') = 1$, which may be given by the scheme

$$M \left\{ \begin{matrix} A_1 & A_2 & A_3 & A_4 & \\ \alpha & \beta & \gamma & \delta & x,\ y(x) \\ \alpha' & \beta' & \gamma' & \delta' & \\ \alpha'' & \beta'' & \gamma'' & \delta'' & \end{matrix} \right\} \qquad (6.12)$$

where $A_i = -a_i$ $(i = 1,\ldots,4)$ are the singular points of equation (6.6) or (6.6.1). These equations possess seventy two branch solutions, twelve principal solutions and three fundamental solutions. The branch sets

corresponding to the singular points are $\{M^{(\alpha)}, M^{(\alpha')}, M^{(\alpha'')}\}$, $\{M^{(\beta)}, M^{(\beta')}, M^{(\beta'')}\}$, $\{M^{(\gamma)}, M^{(\gamma')}, M^{(\gamma'')}\}$, $\{M^{(\delta)}, M^{(\delta')}, M^{(\delta'')}\}$. Each one of these branch sets $M^{(\alpha)}$,..., etc. represents six solutions. The study of these sets may result in interesting properties. The branch solutions are of the hypergeometric forsm F_1. As to the fourth order differential equation the M–Function may be given by the scheme:

$$M\left\{\begin{array}{ccccc|c} A_1 & A_2 & A_2 & A_4 & A_5 & \\ \alpha & \beta & \gamma & \delta & e & \\ \alpha' & \beta' & \gamma' & \delta' & e' & x,\ y(x),\ z(x) \\ \alpha'' & \beta'' & \gamma'' & \delta'' & e'' & \\ \alpha''' & \beta''' & \gamma''' & \delta''' & e''' & \end{array}\right\}$$

with the condition $w - \sum_{i=1}^{5} \alpha_i + 1 = 0$, where α_i and w are given by the matrix equation

$$\begin{bmatrix} w \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} = U \begin{bmatrix} \alpha+\beta+\gamma+\delta+e \\ \alpha+\beta+\gamma+\delta+e \\ \alpha'+\beta'+\gamma'+\delta'+e' \\ \alpha''+\beta''+\gamma''+\delta''+e'' \\ \alpha'''+\beta'''+\gamma'''+\delta'''+e''' \\ \alpha+\beta+\gamma+\delta+e \end{bmatrix}$$

where U is (6×6) unit matrix and $\Sigma\alpha + \Sigma\alpha' + \Sigma\alpha'' + \alpha''' = 1$. The principal sets of branch solutions which correspond to the singular points may be given by: $\{M^{(\alpha)}, M^{(\alpha')}, M^{(\alpha'')}, M^{(\alpha''')}\}$, $\{M^{(\beta)}, M^{(\beta')}, M^{(\beta'')}, M^{(\beta''')}\}$, $\{M^{(\gamma)}, M^{(\gamma')}, M^{(\gamma'')}, M^{(\gamma''')}\}$, $\{M^{(\delta)}, M^{(\delta')}, M^{(\delta'')}, M^{(\delta''')}\}$ and $\{M^{(e)}, M^{(e')}, M^{(e'')}, M^{(e''')}\}$. Each set represents eight branch solutions and so the total of branch solution for the fourth order differential equations of this type is one hundred sixty branch solutions in the forms of hypergeometric function $F_D^{(3)}$.

7. M–Function for the Fifth and nth Order Equations

If in (3.2) we put m = 6, n = 5, f(x) = 0 we have the operator equation equivalent to the integro–differential equation of order (1,5):

$$y^{v} + \sum_{i=1}^{6} \frac{w-\alpha_i+1}{x+a_i}\, y^{iv} + w \sum_{1\le i<j\le 6} \frac{w-\alpha_i-\alpha_j+1}{(x+a_i)(x+a_j)}\, y'''$$

$$+ w(w-1) \sum_{1\le i<j<k\le 6} \frac{w-\alpha_i-\alpha_j-\alpha_k+1}{(x+a_i)(x+a_j)(x+a_k)}\, y'' +$$

$$+ \prod_{r=0}^{2} (w-r) \sum_{1\le i<j<k<\ell\le 6} \frac{w-\alpha_i-\alpha_j-\alpha_k-\alpha_\ell+1}{(x+a_i)(x+a_j)(x+a_k)(x+a_\ell)}\, y' +$$

$$+ \prod_{p=0}^{3} (w-p) \sum_{1\le i<j<k<\ell<m\le 6} \frac{w-\alpha_i-\alpha_j-\alpha_k-\alpha_\ell-\alpha_m+1}{(x+a_i)(x+a_j)(x+a_k)(x+a_\ell)(x+a_m)}\, y +$$

$$+ \prod_{q=0}^{4} (w-q) \frac{w-\sum_{i=1}^{6} \alpha_i+1}{\prod_{i=1}^{6} (a_i+x)}\, \mathop{I}_{a}^{x}\, y = 0 \qquad (7)$$

If in (7):

$$w - \sum_{i=1}^{6} \alpha_i + 1 = 0 \qquad (7.1)$$

and $a_6 \to \pm \infty$, then the equaiton is reduced to equation (3.4) and its equivalent operator equation (3.5). In addition w and α_i (i = 1,2,...,6) are given by the matrix equation

$$\begin{bmatrix} w \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} = U \begin{bmatrix} \alpha & +\beta & +\gamma & +\delta & +e & +\nu \\ \alpha & +\beta & +\gamma & +\delta & +e & +\nu \\ \alpha' & +\beta' & +\gamma' & +\delta' & +e' & +\nu' \\ \alpha'' & +\beta'' & +\gamma'' & +\delta'' & +e'' & +\nu'' \\ \alpha''' & +\beta''' & +\gamma''' & +\delta''' & +e''' & +\nu''' \\ \alpha^{iv} & +\beta^{iv} & +\gamma^{iv} & +\delta^{iv} & +e^{iv} & +\nu^{iv} \\ \alpha & +\beta & +\gamma & +\delta & +e & +\nu \end{bmatrix} \qquad (7.2)$$

where U is (7×7) unit matrix. Condition (7.1) and equation (7.2) implies that

$$\Sigma\alpha + \Sigma\alpha' + \Sigma\alpha'' + \Sigma\alpha''' + \Sigma\alpha^{iv} = 1 \qquad (7.3)$$

The M–Function for the fifth order equations of this type may be given by the scheme:

$$M \left\{ \begin{matrix} A_1 & A_2 & A_2 & A_4 & A_5 & A_6 \\ \alpha & \beta & \gamma & \delta & e & \nu \\ \alpha' & \beta' & \gamma' & \delta' & e' & \nu' \\ \alpha'' & \beta'' & \gamma'' & \delta'' & e'' & \nu'' \\ \alpha''' & \beta''' & \gamma''' & \delta''' & e''' & \nu''' \\ \alpha^{iv} & \beta^{iv} & \gamma^{iv} & \delta^{iv} & e^{iv} & \nu^{iv} \end{matrix} \quad x,\ y(x),\ z(x),\ w(x) \right\}$$

where $A_i = -a_i$ (i = 1,2,...,6) are the singular points. The principal sets

of branch solutions which correspond to these singular points may be given by the following:

The sets corresponding to the singular point A_1 are $\{M^{(\alpha)}, M^{(\alpha')}, M^{(\alpha'')}, M^{(\alpha''')}, M^{(\alpha^{iv})}\}$, to the singular point A_2 are $\{M^{(\beta)}, M^{(\beta')}, M^{(\beta'')}, M^{(\beta''')}, M^{(\beta^{iv})}\}$, to the point A_3 are $\{M^{(\gamma)}, M^{(\gamma')}, M^{(\gamma'')}, M^{(\gamma''')}, M^{(\gamma^{iv})}\}$, to the point A_4 are $\{M^{(\delta)}, M^{(\delta')}, M^{(\delta'')}, M^{(\delta''')}, M^{(\delta^{iv})}\}$, to the point A_5 are $\{M^{(e)}, M^{(e')}, M^{(e'')}, M^{(e''')}, M^{(e^{iv})}\}$, and the sets corresponding to the singular point A_6 are $\{M^{(\nu)}, M^{(\nu')}, M^{(\nu'')}, M^{(\nu''')}, M^{(\nu^{iv})}\}$.

Each set contains ten branch solutions and thus the total branch solutions of the fifth order differential equations of Riemann–Papperitz type is three hundreds solutions. These solutions are in the forms of the hypergeometric functions $F_D^{(4)}$ as shown in section 4.

The Differential Equation of nth order:

If in equations (3.1) and (3.2) we let $m = n+1$, then the resulting equation would be an integro–differential equation of order (1,n) which is reduced to nth order differential equations of Riemann–Papperitz type (5.1) if the conditions

$$w - \sum_{i=1}^{n+1} \alpha_i + 1 = 0 \tag{7.4}$$

and $a_{n+1} \to \pm \infty$. This condition is satisfied by the values of w and α_i given by the matrix equation

$$\begin{bmatrix} w \\ \alpha_1 \\ \alpha_3 \\ \vdots \\ \alpha_n \\ \alpha_{n+1} \end{bmatrix} = U \begin{bmatrix} \sum_{i=1}^{n+1} \alpha_i^{(0)} \\ \sum_{i=1}^{n+1} \alpha_i^{(0)} \\ \vdots \\ \sum_{i=1}^{n+1} \alpha_i^{(n-1)} \\ \sum_{i=1}^{n+1} \alpha_i^{(0)} \end{bmatrix} \tag{7.5}$$

where U is (n+2×n+2) unit matrix. Condition (7.4) and equation (7.5) imply that $\sum_{i=1}^{n+1} \sum_{j=0}^{n-1} \alpha_i^{(j)} = 1$.

The M–Function for the nth order linear differential equation of this type may be given by the scheme:–

$$M\left\{\begin{array}{cccc|c} A_1 & A_2 & \dots & A_n & A_{n+1} & \\ \alpha_1^{(0)} & \alpha_2^{(0)} & \dots & \alpha_n^{(0)} & \alpha_{n+1}^{(0)} & \\ \alpha_1^{(1)} & \alpha_2^{(1)} & \dots & \alpha_n^{(1)} & \alpha_{n+1}^{(1)} & x,\ y_1(x),\dots,\ y_{n-2}(x) \\ \vdots & & & & & \\ \alpha_1^{(n-1)} & \alpha_2^{(n-1)} & & \alpha_n^{(n-1)} & \alpha_{n+1}^{(n-1)} & \end{array}\right\}$$

where $A_i = -a_i$, (i = 1,2,...,n+1).

The principal sets of branch solutions which correspond to the singular points may be given by: $\{M^{(\alpha_i^{(j)})}\}$, (i = 1,2,...,n+1), (j = 0,1,2,...,n–1). For fixed i and j each set $\{M^{(\alpha_i^{(j)})}\}$ contains 2n branch solutons. Thus the total number of branch solutions of the nth order differential equation (5.1) is $2n^2(n+1)$. For fixed i $M^{\alpha_i^{(j)}}$, (j = 0,1,...,n–1) are the sets corresponding to the singular point A_i.

It would be interesting and of importance if a study and analysis are done on such M–Functions, similar to the study made on Riemann's P–Function ([5], pp. 83–92).

REFERENCES

[1] J.L. Liouville, "Mémoire sur le changement de la variabledans le calcul des differentielle à indices quelconques", J. Ecole Polytech., 15, Section 24 (1835).

[2] B. Riemann, "Versuch einer allgemeinen Auffassung der Integration und Differentiation", Gesammelte Mathematische Werke, Leipzig, (1876).

[3] H.J. Holmgren, "Om differentialkalkylen med indices of hvilken natur som helst", Kungl. Svenska Vetnskaps–Akademins Handlingar, Bd. 5, n. 11, Stockholm (1965–66)), pp. 1–83.

[4] M. Riesz, "L'integral de Rieman–Liouville et le problem de Cauchy", Acta Mathematica, Vol. 81 (1949), pp. 1–223.

[5] E.G.C. Poole, "Introduction to the theory of Linear Differential Equations", Dover Publications, New York, 1960.

[6] Appell and Kampé de Fériet, "Fonctions Hypergeometrique et Hyperspheriques, polynome d'Hermite" Gauthier–Villars, Paris, (1926).

[7] W.N. Bailey, "Generalized Hypergeometric Series", Cambridge University Press, 1935.

[8] M.A. Al–Bassam, "Some properties of Holmgren–Reisz transform", Annali della Scoula Noremale Superiore di Pisa, Scienze Fisiche e Mathematiche, Serie III, Vol. VX, Fasc. 1–11, (1961), pp. 1–24.

[9] M.A. Al–Bassam, "Concerning Holmgren–Reisz transform equations of Gauss–Riemann type", Rendiconti del Circolo Matematico di Palermo, Serie II, Vol. XI, (1962), pp. 1–20.

[10] M.A. Al–Bassam, "On certain types of H–R transform equations and their equivalent differential equations", Journal fur die Reine und Angewandte Mathematik, Band 26, (1964), pp. 91–100.

[11] M.A. Al–Bassam, "Some existence theorems on differential equations of generalized order", Journal fur die Reine und Angewandte Mathematik, Band 218, (1965), pp. 70–78.

[12] M.A. Al–Bassam, "On an integro–differential equation of Legengre–Velterra type", Portugaliae Math., Vol. 25, Fasc. 1, (1966), pp. 53–61.

[13] M.A. Al–Bassam, "H–R transform equations of Laguerre type", Bulletin, College of Science, University of Baghdad, Vol. 9, (1966), pp. 181–184.

[14] M.A. Al–Bassam, "On Laplace's second order linear order differential equations and their equivalent H–R transform equations", Journal fur die Reine und Angewandte Mathematik, Band 225, (1967), pp. 76–84.

[15] M.A. Al–Bassam, "On some differential and integro–differential equations associated with Jacobi's differential equations", Journal fur die Reine und Angewandte Mathematik, Band 288, (1976), pp. 211–217.

[16] M.A. Al–Bassam, "H–R transform in two dimensions and some of its applications", "Fractional Calculus and its Applications, Proceedings of the Internaitonal Conference, Univ. of New Haven, June 1974, Edited by B. Ross, Lecture Notes in Mathematics, 9457), Springer–Verlag, New York, pp.91–105.

[17] M.A. Al–Bassam, "On fractional calculus and its applications to the theory of ordinary differential equations of generalized order", Nonlinear Analysis and Applications, Vol. 80, Marcel Dekker Inc., 1982, New York, pp. 305–331.

[18] K.B. Oldham and J. Spanier, "The fractional calculus", Academic Press, New York, (1974).

[19] M.A. Al–Bassam, "On fractional analysis and its applications", Modern Analysis and its Applications, Edit. H.L. Manocha, Prentice Hall of India Ltd., New Delhi, 1986,pp. 269–307.

[20 M.A. Al–Bassam, "Some applications of fractional calculus to differential equations", Research Notes in Mathematics, Pitman Advanced Publishing Program, London 1985.

[21] M.A. Al–Bassam, "On generalized power series and generalized operational calculus and its applications", Nonlinear Analysis, World Scientific Publishing Co., Pte Ltd., New Jersey, 1987.

[22] M.A. Al–Bassam, "Application of fractional calculus to differential equations of Hermite's type", Indian Journal of Pure and Applied Mathematics, 16(9), pp. 1009–1016.

[23] M.A. Al–Bassam, "Some applications of generalized calculus to differential and integro–differential equations", Mathematical Analysis and Applications, International Conference, Kuwait University, Pergamon Press, pp. 61–76.

[24] K. Nishimoto, "Applications of fractional calculus to a fourth order linear ordinary differential equation of Fuchs type", Journal of the College of Engineering, Nihon University, Series B, Vol. 29, March 1988.

[25] Harold exton, "Multiple Hypergeometric Functions and Applications", Ellis Harwood Limited, Chichister, (1976).

[26] E.T. Whittaker and G.N. Watson, "A course of Modern Analysis", Cambridge University Press, 1962.

[27] M.A. Al–Bassam, "On fractional operator equations and solutions of a class of third order equations", Bolletino U.M.I. (7) 3–B (1989).

[28] M.A. Al–Bassam, "Application of Fractional Calculus to a class of integor–differential equations of Riemann–Papperitz–Gauss type", Proceedings of the third International Conference on "Fractional Calculus", Nihon University, Tokyo, Japan, May 29–June 2, 1989. (To appear).

[29] M.A. Al–Bassam, "Applications of fractional calculus to a class of third order differential equations", Proceedings of the Fourth Conference on Differential Equations, Rousse, Bulgaria, August 12–19, 1989, (To appear).

[30] M.A. Al–Bassam, "A unified class of differential equations", Math Japonica 34, No. 4(1989), 513–532.

M.A. Al–Bassam
Department of Mathematics
Kuwait University
P.O. Box 5969
13060 Safat, Kuwait

Constantin Carathéodory: An International Tribute (pp. 47–52)
edited by Th. M. Rassias

ON THE FUNCTIONAL EQUATION $|T(x) \cdot T(y)| = |x \cdot y|$

C. Alsina and J.L. Garcia-Roig

We characterize all functions T from a real Hilbert space $(E, \cdot)$ of dimension ≥ 2 into itself satisfying the functional equation $|T(x) \cdot T(y)| = |x \cdot y|$. We study also this equation in a restricted domain.

Let $(E, \cdot)$ be a real Hilbert space (always assumed to be of dimension $n \geq 2$) and consider, for a map T from E into itself, the functional equation

$$|T(x) \cdot T(y)| = |x \cdot y|, \quad \text{for all } x, y \text{ in } E. \tag{1}$$

It is known the solution of (1) in the case T bijective (see [3] Theorem 3) and T continuous and dim $E < \infty$ (see [2]). Following a suggestion of W. Benz, here we will consider the general case.

To begin with, we observe that any solution T of (1) satisfies the following properties, for all x, y in E and λ in R:

(a) $\|T(x)\| = \|x\|$;

(b) $T(x) = 0$ if and only if $x = 0$;

(c) $T(x) \cdot T(y) = 0$ if and only if $x \cdot y = 0$;

(d) $|\cos A(x, y)| = |\cos A(T(x), T(y))|$, where $A(u, v)$ denotes the angle between u and v;

(e) $T(\lambda x) = \pm \lambda T(x)$;

(f) Area $(T(x), T(y))$ = Area (x, y).

It is also important to establish, for any such T, the following

Lemma 1. If T satisfies (1) then T transforms planes into planes.

Proof. We can take for a plane π an orthonormal basis e_1, e_2 and extend it to a maximal complete orthonormal system $\{e_i\}_{i\in I}$ in E. Then, because of (1), $\{T(e_i)\}_{i\in I}$ is an orthonormal set of E which can be extended with a certain (possibly empty) set $\{u_j\}_{j\in J}$ to a complete orthonormal set. Therefore, for any v in π, $v = a_1e_1 + a_2e_2$, we have by (1) that $T(v) = \sum_{i\in I} b_iT(e_i) + \sum_{j\in J} c_ju_j$, for some c_j and b_i with $|b_i| = |a_i|$, $i = 1, 2$. Using (1) and (a) we have

$$\sum_{i=1}^{2} |a_i|^2 = \|v\|^2 = \|T(v)\|^2 = \sum_{i\in I} b_i^2 + \sum_{j\in J} c_j^2,$$

whence $c_j = 0$, for all $j \in J$ and $b_i = 0$ whenever $i \neq 1, 2$. Thus $T(v)$ belongs necessarily to the plane spanned by $T(e_1)$ and $T(e_2)$.

Remark. As a consequence of the proof of Lemma 1, we can further assert that if $v = \lambda e_1 + \mu e_2$ then $Tv = \pm\lambda Te_1 \pm \mu Te_2$. We shall write $w = \lambda Te_1 + \mu Te_2$ and $\overline{w} = \lambda Te_1 - \mu Te_2$ (and similarly, $\overline{v} = \lambda e_1 - \mu e_2$) and use this notation since it reminds us of complex conjugation. So we have

$$Tv \in \{\pm w, \pm\overline{w}\}. \tag{2}$$

The study of T is, after Lemma 1, essentially reduced to the case of R^2.

Lemma 2. Any map $T : R^2 \to R^2$ satisfying the functional equation (1) is of the form

$$T(x) = \epsilon(x) \cdot S(x) \tag{3}$$

where S is an orthogonal transformation of R^2 and ϵ is an arbitrary map from R^2 into the set $\{\pm 1\}$.

Proof. Obviously any function T of the form (3) satisfies (1). Conversely, assume that T satisfies (1). By applying a suitable rotation we can further suppose that $e_1 = (1,0)$ is fixed under T. Then, with our previous notation, $Tv \in \{\pm v, \pm\overline{v}\}$.

Now we distinguish two cases:

Case 1. All points in R^2 are eigenvectors for T. Their eigenvalues are obviously ± 1 and we are done.

Case 2. There exists a noneigenvector u of T, which by (2) has to satisfy $Tu = \pm\overline{u}$, with $\overline{u} \notin < u >$, or in other words, if $u = (a,b)$, with $a \cdot b \neq 0$. We claim that in this case we must have for all v in R^2

$$T(v) = \pm\overline{v}. \tag{4}$$

Obviously (4) is true for the lines generated by $e_1, e_2 = (0,1)$ and u. Now if there exists v not in these lines and not satisfying (4), i.e., such that $Tv = \pm v$, we would obtain

$$|\cos A(\overline{u}, v)| = \frac{|\overline{u} \cdot v|}{\|\overline{u}\| \cdot \|v\|} = \frac{|T(u) \cdot T(v)|}{\|T(u)\| \cdot \|T(v)\|}$$

$$= |\cos A(T(u), T(v))| = |\cos A(u, v)|$$

which yields the contradiction that v should be on one of the lines generated by e_1 or e_2. From this the lemma follows.

These results immediately entail the following

Theorem 1. Let E be a real Hilbert space (of dimension ≥ 2) and let $T : E \to E$ satisfy (1). Then $T = \epsilon \cdot S$, where ϵ maps E into $\{\pm 1\}$ and S is a linear isometry of E onto a closed subspace of E.

Remark. Obviously the image of S is the whole of E if T is assumed to send a complete orthonormal set of E onto another such one and only in this case.

We now turn our attention to the restricted functional equation

$$|T(x) \cdot T(y)| = |x \cdot y|, \text{ for all } x, y \text{ in } E \text{ with } \|x\| = 1. \tag{5}$$

We prove first the following

Lemma 3. If E is an inner product space of dimension at least 2 (not necessarily Hilbert) and $T : E \to E$ satisfies (5) then $\|T(x)\| \geq \|x\|$, for all $x \in E$.

Proof. For any x in E, $x \neq 0$, from (5) we have that $\left\|T\left(\frac{x}{\|x\|}\right)\right\| = 1$ and $\left|T\left(\frac{x}{\|x\|}\right) \cdot T(x)\right| = \|x\|$, i.e., $T\left(\frac{x}{\|x\|}\right) \cdot T(x) = \pm\|x\|$. Then we have

$$0 \leq \left\|T\left(\frac{x}{\|x\|}\right) - \frac{1}{\|x\|}T(x)\right\|^2$$

$$= \left\|T\left(\frac{x}{\|x\|}\right)\right\|^2 + \frac{\|T(x)\|^2}{\|x\|^2} - \frac{2}{\|x\|}T\left(\frac{x}{\|x\|}\right) \cdot T(x). \tag{6}$$

If $T\left(\frac{x}{\|x\|}\right) \cdot T(x) = +\|x\|$ then (6) yields $\|T(x)\| \geq \|x\|$.

If $T\left(\frac{x}{\|x\|}\right) \cdot T(x) = -\|x\|$ then from (6) and the triangle inequality of the norm we obtain:

$$\frac{\|T(x)\|^2}{\|x\|^2} + 3 = \left\|T\left(\frac{x}{\|x\|}\right) - \frac{1}{\|x\|}T(x)\right\|^2 \leq \left(\left\|T\left(\frac{x}{\|x\|}\right)\right\| + \frac{\|T(x)\|}{\|x\|}\right)^2$$

$$= 1 + \frac{\|T(x)\|^2}{\|x\|^2} + 2\frac{\|T(x)\|}{\|x\|},$$

so that, again, $\|T(x)\| \geq \|x\|$.

In order to see that, in general, (5) is not equivalent to (1) we will exhibit the following

Example 1. Let $E = l_2$ and take for T Bernoulli's shift: $T(e_n) = e_{n+1}$, $n = 1, 2, \ldots$, where $\{e_i\}_{i \in N}$ is the usual complete orthonormal set in l_2, and extend it by linearity and continuity except for $v = 2e_1$ which can be sent to $e_1 + 2e_2$. Then T satisfies (5) but is not a solution of (1) because, e.g., $\|T(v)\| > \|v\|$.

However we can show the following

Theorem 2. If $T : E \to E$ satisfies (5) and sends a maximal orthonormal system of E into another maximal such system (this is obviously the case if E is finite-dimensional) then (1) holds, i.e., in this case the restricted condition (5) is equivalent to (1).

Remark. The condition of T on maximal orthonormal systems can be replaced by that of preservation of norms ($\|T(v)\| = \|v\|$, for all v) as the following proof makes clear.

Proof. Under the hypotheses assumed on T, if (5) is satisfied then we proceed, as in the proof of Lemma 1, to consider any $v = \sum_{i \in I} a_i e_i$ and $T(v) = \sum_{i \in I} b_i T(e_i)$, where $\{e_i\}_{i \in I}$ is a maximal orthonormal system of E (and so is $\{T(e_i)\}_{i \in I}$). Thus by (5) we immediately have that $|a_i| = |b_i|$ for all $i \in I$ and therefore $\|T(v)\| = \|v\|$, for all v in E. Then we have for any $v \neq 0$ in E:

$$\left| T(v) \cdot T\left(\frac{v}{\|v\|}\right) \right| = \left| v \cdot \frac{v}{\|v\|} \right| = \|v\| = \|T(v)\| = \|T(v)\| \cdot \left\| T\left(\frac{v}{\|v\|}\right) \right\|,$$

i.e., we have equality in the Cauchy-Schwarz inequality and consequently $T(v) = \pm \|v\| T\left(\frac{v}{\|v\|}\right)$. From this the theorem follows at once.

ACKNOWLEDGEMENT

We thank Prof. W. Benz (Hamburg) for his interesting remarks concerning the problem treated in this paper.

REFERENCES

1. J. Aczél, *Functional Equations and their Applications*, Academic Press, New York (1966).
2. C. Alsina, J.L. Garcia Roig, *"On continuous preservation of norms and areas"*. *Aeq. Math.* 38 (1989), 211-215.
3. J.S. Lomont, P. Mendelson, *"The Wigner unitary-antiunitary theorem"*. *Ann. of Math.* 78 (1963), 548-559.

C. Alsina and J.L. Garcia-Roig
Sec. Matemàtiques i Informàtica (ETSAB)
Universitat Politècnica Catalunya
Diagonal 649
E08028 Barcelona
SPAIN

Constantin Carathéodory: An International Tribute (pp. 53–64)
edited by Th. M. Rassias

MULTICRITERIA OPTIMIZATION

Alexis Bacopoulos

In this presentation we define a convenient framework and give some results in multicriteria optimization. Theorems of existence, characterizations uniqueness and computation of best approximations are given here in three mutually related contexts of optimality.

The publication of these two volumes in commemoration of the mathematician C. Carathéodory gives me the apportunity to mention here how ch I have personnally been influenced in my study of Mathematics by his amental works in Real Analysis, Complex Analysis (Funktionentheo- Partial Differential Equations and Calculus of Variations [1–4]. As is own, these masterpieces are *only part* of his scientific contribution, arathéodory has also contributed fundamentally, both in form and in various other fields of Mathematics and Physics. (For a com- of scientific works and a historical review see also other a these two volumes as well as his Collected Mathe

Approxi

Problem I. $\min_{p \in \Pi_n} F(p)$, where $F(p) \equiv \|w_1(f-p)\|_\infty + \|w_2(f-p)\|_\infty$. The (weight) functions w_1 and w_2 are strictly positive on $[a,b]$. If $w_1(x) \equiv 1$ and $w_2(x) \equiv \frac{1}{f(x)}$ we may think of simultaneous absolute and relative approximation in the sup norm.

Problem II. $\min_{p \in \Pi_n} F(p)$, where $F(p) \equiv \|f-p\|_\infty + \|f-p\|_2$.

In case of, say, power transmission considerations, one may think of F as total cost, comprised of the sum of initial cost (insulation) plus operating cost.

Problem III. $\min_{q \in K} F(q)$, where $F_{q \in K}(q) = \|f-q\|_\alpha + \|f-q\|_\beta$ and K is a convex and closed proper subset of S, S is a linear space, $f \in S \sim K$, $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ two general norms on S.

We remark that Problem III is much more general than I and II. For example, q here may be a multidimentional polynomial. K may be infinite dimensional. As expected therefore, any characterizations that one gets will be accordingly general.

Related to the above three typical sum norm problems, we may also consider the corresponding *max* norm optimization Problems I′, II′ and III′, defined by the composite norm $\max\{\|f-p\|_\alpha, \|f-p\|_\beta\}$. There is yet another framework for considering approximation with two (or more) criteria of proximity, which we have called *vectorial* or *vec* approximation [6, 7]. Essentially synonymous terminology to vec approximation is *multicriteria optimization* (in Operations Research) and *Pareto optimality* (in Economics) [8, 9]. As we shall see this is a "natural" setting for imbedding problems of sum and max approximation, in the sense that vectorial approximation preserves the structure of the component norms useful for theorem solving and efficient computation [7, 10, 11, 12].

Vectorial Approximation

Let $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ be two norms defined on a linear space S and let $f \in S \sim K$ be a given function to be approximated by approximations $p \in K \subset S$. K is assumed to be a closed, convex, proper subset of S. Let

$G(p) = (\|f-p\|_\alpha, \|f-p\|_\beta)$ and define the partial ordering $\leqq$ on $G(K)$ by

$$G(p) \leqq G(q) \Leftrightarrow \begin{cases} \|f-p\|_\alpha & \leq \|f-q\|_\beta \text{ and} \\ \|f-p\|_\alpha & \leq \|f-q\|_\beta \end{cases}$$

we shall write $G(p) < G(q)$ iff $G(p) \leqq G(q)$ and $G(p) \neq G(q)$.

Definition. We say that p is a *best vec* approximation if there does not exist a $q \in K$ such that $G(p) < G(q)$.

Definition. The minimal set M is given by

$$M = \{G(p) : p \in K \text{ is a best vec approximation}\}.$$

Notation. In Fig. 1 Λ is the 45° bisector of the $\|\cdot\|_\alpha, \|\cdot\|_\beta$ orthogonal axes.

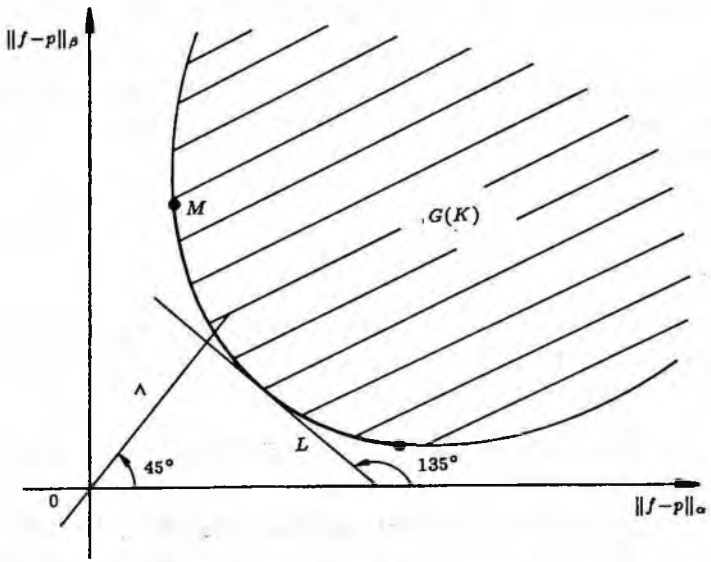


Fig. 1.

L is the supporting line to $G(K)$ which makes 135° angle with the $\|\cdot\|_\alpha$ axis.

Theorem 1. M is a convex decreasing arc.

Proof. The decreasing part follows from the partial ordering and the minimality of $M \subset G(K)$. The convexity is a consequence of the convexity of the norms.

Assuming existence and uniqueness of the ordinary one-norm best approximations (denoted by p_α and p_β relative to the norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$, respectively), we remark that

Corollary. M is a point iff $G(p_\alpha) = G(p_\beta)$.

Theorem 2. Let p_s be a best sum approximation. Then $G(p_s) \subseteq M \cap L$. In case p_s is unique, then $G(p_s) = M \cap L$. Similarly, if p_m denotes a best max approximation, then $G(p_m) = M \cap \Lambda$ (assuming $M \cap \Lambda \neq \phi$).

The proof of Theorem 2 is an easy consequence of the definitions, convexity and, in the case of $G(p_m) = M \cap \Lambda$, the continuity of the best approximation operator.

Main Results

In what follows we make the same assumptions as in Problem I and introduce a notation which is reminiscent of Chebychev's domain of ideas.

$$\underline{\overline{X}}_{+1} = \{x : w_1(x)(f(x) - p(x)) = +\|w_1(f-p)\|\}$$
$$\underline{\overline{X}}_{+2} = \{x : w_2(x)(f(x) - p(x)) = +\|w_2(f-p)\|\}$$
$$\underline{\overline{X}}_{-1} = \{x : w_1(x)(f(x) - p(x)) = -\|w_1(f-p)\|\}$$
$$\underline{\overline{X}}_{-2} = \{x : w_2(x)(f(x) - p(x)) = -\|w_2(f-p)\|\}$$
$$\underline{\overline{X}}_{p} = \underline{\overline{X}}_{+1} \cup \underline{\overline{X}}_{+2} \cup \underline{\overline{X}}_{-1} \cup \underline{\overline{X}}_{-2}$$

The sign function $\sigma(x)$ on $\overline{\underline{X}}_p$ is defined by

$$\sigma(x) = -1 \quad \text{when } x \in \overline{\underline{X}}_{-1} \cup \overline{\underline{X}}_{-2} \qquad \text{and}$$
$$\sigma(x) = +1 \quad \text{when } x \in \overline{\underline{X}}_{+1} \cup \overline{\underline{X}}_{+2}\,.$$

Theorem 3. Consider the vectorial analogue to Problem I. Then p is a best vec approximation to f if and only if there exist $n+2$ points $x_1 < x_2 < \ldots < x_{n+2}$ in $\overline{\underline{X}}_p \subset [a,b]$ satisfying

$$\sigma(x_i) = (-1)^{i+1}\sigma(x_1)\,.$$

Theorem 4. Each best vec approximation is unique; i.e., given $\mu \in M$ there is only one $p \in \prod_n$ such that $G(p) = \mu$.

Note that this uniqueness does not contradict the fact that the minimal set M has, in general, an infinite number of points, all of which, by definition, correspond to (unique) best vectorial approximations. Likewise, the easily shown existence of M proves the existence of best solutions.

The proof of Theorem 3 is technical and shall be omitted in this presentation. For a complete proof see [7]. We remark however that from what we know now it would have been wrong to expect a generalisation of alternation alone to provide a characterization of best sum approximations. This suggests that the vectorial context is better suited for these type of problems. For the equivalence among vectorial convex minima and weighted sum convex minima see [10, 11]. Now combining Theorems 2, 3 and 4 we obtain the following

Corollary. Each best sum approximation p_s (solution to Problem I) is characterized by the generalized alternation property of Theorem 3 with the added constraint that $G(p_s) \subseteq M \cap L$.

This Corollary will be used for the efficient computation of best sum approximations featuring *quadratic convergence*. In case p_s is unique, we may think of its characterization as the generalized oscillation of Theorem 3 with the constraint

$$(\|w_1(f-p_s)\|\,, \|w_2(f-p)\|) = M \cap L\,.$$

In the Remes-type computation that follows for the best sum approximation, we denote by p^s the polynomial approximation at the s-th cycle of the Algorithm. Note that this superscript notation should not be confused with the notation p_s for the best sum approximations where s here is a subscript. The symbol $v_i^s = 1$ or 2 will signify the "activity" indices, which assume the values 1 or 2 depending on whether the supremum of the generalied alternation at the s-th cycle is attained by the 1st or the 2nd norm. It is helpful in understanding the Algorithm to think of (D^s, A^s) as approximations at the s-th cycle of the 2-dimensional vector $M \cap L$.

We further define the following functions and recursive relations:

$$E(p,x) = f(x) - p(x)\,,\ N_1(p) = \|E_1(p,x)\|\,,$$
$$E_1(p,x) = w_1(x)(f(x) - p(x))\,,\ N_2(p) = \|E_2(p,x)\|\,,$$
$$E_2(p,x) = w_2(x)(f(x) - p(x))\,,$$
$$\lambda_i^{s+1} = \prod_{\substack{j=1 \\ j\neq i}}^{n+2} \frac{1}{x_i^{s+1} - x_j^{s+1}}\,,$$
$$A^{s+1} = \frac{\sum_{j=1}^{n+2} |\lambda_j^{s+1}|[|E(p^s, x_j^{s+1})| - D^s(2 - v_j^{s+1})/(w_1(x_j^{s+1}))]}{\sum_{j=1}^{n+2} |\lambda_j^{s+1}|(v_j^{s+1} - 1)/(w_2(x_j^{s+1}))}$$
$$p^{s+1}(x) = \sum_{j=1}^{n+2} S_j^{s+1} \prod_{\substack{i=1 \\ i\neq j}}^{n+2} (x - x_i^{s+1})$$

where

$$S_j^{s+1} = f(x_j^{s+1})\lambda_j^{s+1} - \operatorname{sgn} E(p^s, x_{n+2}^{s+1})|\lambda_j^{s+1}|B_j^{s+1}$$

and

$$B_j^{s+1} = \frac{K(2 - v_j^{s+1})}{w_1(x_j^{s+1})} + \frac{A^{s+1}(v_j^{s+1} - 1)}{w_2(x_j^{s+1})}\,.$$

Algorithm. Assume that the sth cycle of the algorithm has been completed and we have a number A^s, a polynomial $p^s(x)$ of degree less than or equal to n and a sequence of ordered pairs $(x_i^s, v_i^s), i = 1, 2, \ldots, n+2$, such that

$$a \leqq x_1^s < x_2^s < \ldots < x_{n+2}^s \leqq b\,,$$

v_i^s is either 1 or 2, for $i = 1, 2, \ldots, n+2$ and

$$\text{if } v_i^s = 1 \text{ , then } |E_1(p^s, x_i^s)| = D^s$$
$$\text{if } v_i^s = 2 \text{ , then } |E_2(p^s, x_i^s)| = A^s .$$

Assume furthermore that $E(p^s, x_i^s)$ is alternately positive and negative as i varies from 1 to $n+2$.

Following an exchange-type argument it may be shown inductively that after the $(s+1)$-cycle of this Algorithm the alternation in sign of $E(p^{s+1}, x_i^{s+1})$ is maintained and that a contraction type argument yields indeed quadratic convergence to the solution p_s.

The details of convergence as well as the initialization arguments may be found in [7].

A numerical example for the case of simulataneous absolute and relative approximation is given here. For let $f(x) = e^x$ on [0,1], $w_1(x) \equiv 1, w_2(x) = e^{-x}, \prod_n = \prod_4, p_i$ be the best approximation corresponding to $w_i(x), i = 1, 2$.

Computational Example

Polynomials	p_1	p_2	p_s
Coefficient of x^4	0.21259	0.16995	0.10692
Coefficient of x^3	−0.41751	−0.18094	0.30215
Coefficient of x^2	1.22217	0.82266	−0.33051
Coefficient of x^1	0.69981	0.90786	1.39083
Coefficient of x^0	1.01975	1.00366	1.11867
Errors			
Absolute error	0.01975	0.07346	0.02346
Relative error	0.01975	0.00366	0.00908
Sum error	0.03951	0.07711	0.03254

We now turn our attention to Problem II, which is to minimize in $\prod_n$ (efficiently) the expression $\|f-p\|_\infty + \|f-p\|_2$. As with Problem I, we shall deal with Problem II by imbedding it in a vectorial framework. We remark that the uniqueness of each best vectorial approximation here is an immediate consequence of the strict convexity of the L_2-norm.

It follows from Theorem 1 that the arc M may be described by a convex function of one variable with domain $[\|f-p_1\|_\infty, \|f-p_2\|\infty]$ and range $[\|f-p_2\|_2, \|f-p_1\|_2]$. Our problem now may be stated as follows: Given a couple of real numbers d, d' satisfying $(d,d') \in M$, find the $p \in \prod_n$ which is the solution of the equation $G(p) = (d,d')$. We shall denote this (unique) solution by p_d.

It may be easily seen that both end points of the minimal set M may be obtained numerically using the standard algorithms for the max-norm and the L_2-norm, respectively. We now reformulate the problem: Find the best vectorial approximation p_d whose error in the Chebychev norm equals a prescribed value $d, \|f-p_1\|_\infty \le d \le \|f-p_2\|_\infty$. It is now clear that the desired polynomial p_d is the unique solution to the problem

$$\begin{aligned} &\min \ \|f-p\|_2 \\ &\text{subject to } \|f-p\|_\infty \le d. \end{aligned}$$

Since the number of constraints here is infinite, we proceed by solving a sequence of quadratic programming problems, each with a finite number of constraints. The sequence of solutions $\{p_k\}$ is shown to converge to the theoretical solution p_d.

Algorithm. At the kth step we have from the preceeding steps a finite set of points $X^k \subset [a,b]$. We solve the quadratic program

$$\begin{aligned} &\min \ \|f-p\|_2 \\ &\text{subject to } |f(x)-p(x)| \le d, \ \ x \in X^k. \end{aligned}$$

Denoting by p_k the solution of this problem, we calculate a point $x_k \in [a,b]$ such that

$$|f(x_k)-p_k(x_k)| = \|f-p_k\|_\infty .$$

We form $X^{k+1} = X^k \cup \{x_k\}$ and proceed to the next cycle. At the beginning

X^1 may be an arbitrary set, containing a maximum of $|f(x) - p_L(x)|$.

Feasibility and Convergence of the Algorithm

We denote by $c = (c_0, \dots, c_n)$ the coefficient vector of a polynomial $p = \sum_{i=0}^{n} c_i g_i$, where $\{g_i\}$ is the orthonormal basis in $\prod_n$ of Legendre polynomials shifted to the interval $[a, b]$. This representation of p is used in order to express the objective function in the quadratic form

$$\|f - p\|_2^2 = \text{constant} + \sum_{i=0}^{n} (c_i - F_i)^2 ,$$

where

$$\text{constant} = \int_a^b f^2 - \sum_{i=0}^{n} F_i^2 \text{ and } F_i = \int_a^b f g_i .$$

Using standard Quadratic Programming notation and removing the absolute values from the constraints, the program at each cycle becomes

$$\begin{aligned} \min \ & -2c^T F_\cdot + c^T c \\ \text{subject to} \ & -c^T g_\cdot(x) \le d - f(x) , \ x \in X^k \\ & c^T g_\cdot(x) \le d + f(x) , \ x \in X^k , \end{aligned}$$

where $F_\cdot = (F_0, F_1, \dots, F_n)$ and $g_\cdot(x) = (g_0(x), g_1(x), \dots, g_n(x))$.

We now make the following definitions:

$$\begin{aligned} r(c\,; x) &= f(x) - \sum c_i g_i(x) \\ \Delta^k(c) &= \max_{x \in X^k} |r(c; x)| \\ \Delta(c) &= \|r(c; x)\| \\ |c| &= \sum |c_i| \\ c^k &= \text{the coefficient vector of the polynomial } p_k , \\ &\quad \text{which solves the Q. P. problem at step } k . \end{aligned}$$

To show now that the sequence $\{p_k\}$ converges to the best vectorial approximation p_d, first note that unless $p_d = p_2$, at least one constraint is

active (equality), i.e., there exists $x \in X^k$ such that $|r(c^k; x)| = d$. Otherwise, since the L_2-norm is convex, this would imply that $\|f - p_k\|_2$ is a global minimum, i.e., $p_k = p_2$. But this is impossible since, by hypothesis there exists $x \in X^1 \subset X^k$ such that $|f(x) - p_2(x)| = \|f - p_2\|_\infty$. Next observe that the sequence $\{\|f - p_k\|_2\}$ is bounded from above by $\|f - p_d\|_2 = d'$. The sequence $\{c^k\}$ is thus located in the ball $\{c \in R^{n+1} : \|f - \sum_k c_i g_i\|_2 \leq d'\}$.

such that $|b - c^k| < \delta$ for all $k \geq N$. Now, for any $c, c' \in R^{n+1}$ we have

$$|r(c'; x) - r(c; x)| \leq B|c' - c|\,,$$

where $B = \max_i \max_x |g_i(x)|$. This implies the following inequalites:

$$|r(c'; x) \leq |r(c; x)| + B|c' - c|$$

and

$$\Delta(c') \leq \Delta(c) + B|c' - c|\,.$$

Using the last inequalities it follows that $|c^k - c^I| \leq 2\delta$ and that

$$\begin{aligned} d \leq \Delta(b) &\leq \Delta(c^I) + B\delta \\ &= |r(c^I; x_I)| + B\delta \\ &\leq |r(c^k; x_I)| + 3B\delta \\ &\leq d + 3B\delta\,. \end{aligned}$$

This shows that for every cluster point b of $\{c^k\}$, $\Delta(b) = d$. Thus the sequence $\{\|f - p_k\|_\infty\}$ converges to d. Since also $\|f - p_k\|_2 \leq \|f - p_d\|_2$ and since the best vectorial approximation p_d is unique, it follows from $\leq$-minimality that the sequence $\{p_k\}$ defined by the algorithm converges to p_d.

In the next Corollary the set D is defined as follows: Consider the orthogonal system of reference with axes $\|f - p\|_\infty, \|f - p\|_2$ (as in Fig. 1) D is the projection of $M \cap L$ on the $\|f - p\|_\infty$-axis.

Corollary. Each best sum approximation p_s (solution to Problem II) is the solution of the quadratic program of the Algorithm corresponding to each $d \in D$. p_s is unique iff D is a singleton.

An approximate d may be found by approximating M by a parabola. C. A. Botsaris and the author are currently collaborating in finding an accurate d efficiently as well as devising acceleration techniques for the above

Algorithm. In addition we investigate certain methods general enough to apply to Problem III.

For the sake of simplicity we have presented up to now only polynomial approximations. Yet, some of the above theorems are also valid for rational function approximations, so-called varisolvent functions, spline functions with variable knots etc. Some of these techniques have been used jointly with I. Chryssoverghi on non-convex optimal control problems [13].

We conclude with a vectorial version of Problem III. This follows from general results on convexity which were obtained jointly with Ivan Singer [14]. We shall omit the proof.

In this final theorem, S will be assumed to be a linear space, K a proper convex closed subset of S, $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ any two norms on S. Furthermore, we define the set D by

$$D = \{d : \inf_{q \in K} \|f - q\|_\alpha \leq d \leq \inf_{q \in B} \|f - q\|_\alpha\}$$

where

$$B = \{r \in K : \|f - r\|_\beta = \inf_{q \in K} \|f - q\|_\beta\}\,.$$

Theorem 5. An element $p \in K$ is a best vectorial approximation iff there exists a $d \in D$ and $\Phi \in S^*$ satisfying

$$\|\Phi\|_\beta = 1$$
$$\Phi(f - p) = \|f - p\|_\beta$$

and

$$\operatorname{Re} \Phi(p - q) \leq 0 \quad \text{for all } q \in K \text{ satisfying} \quad \|f - q\|_\alpha \leq d\,.$$

References

1. C. Carathéodory, *Vorlesungen über Reelle Funktionen*, Leipzig-Berlin, 1927.
2. C. Carathéodory, *Theory of Functions of a Complex Variable*, Vol. 1, 2nd English edition, Chelsea Publishing Company, New York, N. Y., 1983.
3. C. Carathéodory, *Theory of Functions of a Complex Variable*, Vol. 2, 2nd English edition, Chelsea Publishing Company, New York, N. Y., 1983.
4. C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the First Order*, Chelsea Publishing Company, New York, N. Y., 1990.

5. C. Carathéodory, *Gesammelte Mathematische Schriften*, Vols. I–V, C. H. Beck'sche Verlag, München, 1954–57.
6. A. Bacopoulos, *Topology of a general approximation system and applications*, Journal of Approximation Theory 4 (1971), 147–158.
7. A. Bacopoulos and B. Gaff, *On the reduction of a problem of minimization of n+1 variables to a problem of one variable*, SIAM Journal of Numerical Analysis 8 (1971), 97–103.
8. V. Pareto, *Cours d' Économie Politique*, Lausanne, Rouge, 1876.
9. L. Cesari and M. B. Suryanarayana, *Existence theorems for Pareto problems of optimization*, in: Calculus of Variations and Control Theory, pp. 139–154, Academic Press, Inc., New York, 1976.
10. S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics*, Pergamon Press, London-Paris, 1959.
11. K. J. Arrow, E. W. Barankin and D. Blackwell, *Admissible points of convex sets*, in: Contributions to the Theory of Games (eds. H. W. Kuhn and A. W. Tucker), pp. 87–91, Princeton Univ. Press, Princeton, 1953.
12. L. Cesari and M. B. Suryanarayana, *Existence theorems for Pareto optimization in Banach spaces*, Bull. Amer. Math. Soc. 82 (1976), 306–308.
13. I. Chryssoverghi and A. Bacopoulos, *Discrete approximation of related optimal control problems*, Journal of Optimization Theory and Applications 65 (1990), 395–407.
14. A. Bacopoulos and I. Singer, *On convex vectorial optimization in linear spaces*, Journal of Optimization Theory and Applications 21 (1977), 175–188.

Alexis Bacopoulos
Department of Mathematics
National Technical University
Zographou Campus
(15773) Athens
Greece

Constantin Carathéodory: An International Tribute (pp. 65–75)
edited by Th. M. Rassias

CARATHÉODORY AND HARVARD

Garrett Birkhoff

Elsewhere in this volume, my Harvard colleague Lars Ahlfors describes Constantin Carathéodory's beneficial influence on his career, mentioning the fact that Carathéodory served with Élie Cartan and my father, G. D. Birkhoff, on the Fields Committee awarding him one of its first two medals. I shall describe how Carathéodory also influenced crucially (if fortuitously) the careers of two other Harvard mathematicians: Marshall Stone and myself. Carathéodory's work on (Lebesgue) measure and integration, reprinted in [3, pp. 249–494], was an important factor in this influence, and so I shall first recall some aspects of it.

Carathéodory and Measure Theory

Carathéodory's approach to measure theory is summarized in pp. 338–41 of his *Vorlesungen über reelle Funktionentheorie*. As is explained at the end of [3, pp. 249–77], it stems from a beautiful 1914 paper entitled "Über das linears Mass der Punktmengen", reprinted in [3, pp. 249–75], followed by a two-page historical note. In 1919, he applied Lebesgue measure to give the first rigorous interpretation of Poincaré's 1890 Wiederkehrsatz (recurrence theorem) [3, 296–300].

"Recurrent point-groups" were also the final theme of a major paper by G. D. Birkhoff printed in 1920 by the Acta Mathematica (43, 1–119), and several of his later papers (see [1, pp. 111–394]).

The introduction to Carathéodory's 1919 paper observes that Poincaré recognized many basic properties of recurrent motions arising in volume-

preserving flows, but could not formulate them precisely because the theory of the Lebesgue integral did not yet exist. Twelve years later, G. D. Birkhoff would make the same observation in connection with ergodic theory (see below).

Harvard in 1928

To understand Carathéodory's influence on Marshall Stone (and indirectly on ergodic theory), one needs to know something about mathematical activities at Harvard in the years preceding 1928. In 1923, G. D. Birkhoff had inaugurated a course (Math. 16) on "Space, Time, and Relativity" intended to explain to students familiar with second-year calculus Einstein's revolutionary ideas. Its announcement read:

> At the basis of the theory of relativity there lies a concept of space and time which is a radical modification of the concept usually entertained. It is the primary aim of the course to present the "space-time" of relativity in a manner devoid of unnecessary mathematical complication, and to indicate a few simple physical applications. In accomplishing this purpose attention will be confined principally to the case of a single spatial and a single temporal dimension.

The text for the course was his *Relativity and Modern Physics* (Harvard Univ. Press, 1923, 1927).

By 1927, G. D. Birkhoff was also speculating about the new quantum mechanics, Schrödinger's equations are not invariant under the Lorentz group; in his retiring presidential address to the American Mathematical Society, G. D. Birkhoff proposed a "perfect fluid" model for the hydrogen atom that *is* invariant.‡

Von Neumann had also become interested in the Schrödinger equation, which was the final topic (in §10) of a paper which he coauthored with Hilbert and L. Nordheim [6, 104–33]. In a sequell [6, 151–207], von Neumann explicitly reinterpreted Schrödinger's *unbounded* linear differential operators in the context of an axiomatically defined complex Hilbert space, with special attention to its spectral theory. These could not be treated by the earlier methods of Hilbert, Schmidt, and F. Riesz, which

‡[1, pp. 737–63]

were designed for the inverse *bounded* linear integral operators.

In 1925–26, he had supervised the doctoral theses of Marshall Stone and Bernard Koopman. Good friends, the two men had spent the year 1924–25 in Paris as Sheldon Fellows, after graduating from Harvard College. Stone's thesis, like that of G. D. Birkhoff in 1907, was concerned with expansions of solutions of linear ordinary differential equations in infinite series of eigenfunctions.† Koopman's thesis was concerned with the three-body problem of celestial mechanics.

Carathéodory and Stone

Carathéodory was Visiting Lecturer at Harvard during the second half of the academic year 1927–28. G. D. Birkhoff was travelling around the world that term, analyzing oriental music and art as background for a primarily philosophical book on *Aesthetic Measure* (Harvard Univ. Press, 1933). He was also preparing a lecture on that subject entitled "Quelques éléments mathématiques de l'art", to be delivered in Florence's Palazzio Vecchio at the International Mathematical Congress that September.‡

At Harvard, Carathéodory gave an advanced half-course (Math. 32) on 'The Theorem of Picard and its Generalizations", which led through modular and triangular functions and "theorems of Picard, Landau, Schottky, Julia" to hypergeometric functions and the computation of the constants, according to that year's departmental pamphlet. He also gave Math. 16, which had been given by G. D. Birkhoff in 1925–26 and 1926–27.

In the previous fall, Stone had taught a standard basic graduate half-course (Math. 10b) founded by Byerly at least 30 years earlier, concerned with solutions by infinite series of linear boundary value problems arising in mathematical physics. This provided a preparation for the specialized course on the topic of his thesis (Math. 35), which he taught in the spring term. To all outward appearances, he was destined for a career in classical analysis; however, things worked out very differently!

---

†The history of this theory of "Sturm-Liouville series" to 1912 was reviewed by Bôcher (*International Congress of Mathematics.*, Cambridge Univ. Press, vol. 1, 1913, 163–95).

‡Atti del Congresso Int. dei Matematici, tomo I, 315–33.

Two Unpublished Papers[†]

Carathéodory was an editor of the *Mathematische Zeitschrift*, founded by Julius Springer around 1920. When he left Harvard, he gave Stone proofsheets of articles about to appear in that journal. Among these was an article by von Neumann, deriving the main spectral theorem for symmetric linear operators on Hilbert space. Stone quickly recognized from his thesis the relevance of the classical theory of *self-adjoint* unbounded linear differential operators, and soon was reformulating his methods in a Hilbert space context, publishing announcements in a series of notes to the Proceedings of our National Academy of Sciences.[‡] He also submitted a long paper to the *Transactions* of the American Mathematical Society, in which he presented a full and systematic treatment of his results.

Before Stone's paper was typeset, von Neumann published a now famous *second* derivation of the spectral theorem for unbounded self-adjoint operators in the *Math. Annalen* **102** (1929), 49–131 and 370–477, and withdrew the paper typeset by the *Zeitschrift*. Naturally irritated, Springer gave von Neumann the option of paying a large sum to cover the cost of typesetting, or of writing a *book* which would treat in a readable way the flood of new ideas that he had been publishing.* Von Neumann chose the second option.

When Stone read von Neumann's new derivation, he withdrew *his* paper from the *Transactions* in turn. Fortunately, the editors of this journal (Dunham Jackson and J. D. Tamarkin) encouraged Stone to present his independent methods and results in extended book form. This Stone did, in his celebrated *Linear Transformations in Hilbert Space* (A.M.S., 1932). In the same year, Springer published von Neumann's *Mathematische Grundlagen der Quantenmechanik*, and Banach's *Theorie des Opérations Linéaires* came out. The establishment of functional analysis as a major branch of mathematics may be said to date from that year [1].

[†]For Stone's own description of the events involved, see his letter to me published in [1, p. 309].

[‡]Vol. 15 (1929), 198–200 and 423–25, and vol. 16, pp. 172–72.

*In a dozen papers before 1930: see ##6, 8, 9, 10, 17, 18, 19, 23, 24, 25, 28 and 29, in addition to his withdrawn paper!

Carathéodory and Ergodic Theory

In the same exciting years, ergodic theory was also founded, essentially by von Neumann (already at Princeton), together with Harvard's Stone, Koopman, and G. D. Birkhoff. As Carathéodory had observed in his 1919 paper on Poincaré's Wiederkehrsatz, one can apply Lebesgue measure theory to the phase space Ω of any autonomous Hamiltonian dynamical system. Koopman (who had gone to Columbia university) noted that the measure-preserving "flow" in phase space associated with passing time therefore induces a group of *unitary* transformations on the Hilbert space $L^2(\Omega)$ given by the Riesz-Fischer theorem. In his third note to the National Academy *Proceedings*, Stone had provided a spectral resolution theorem for such unitary groups. Using Koopman's observation and Stone's theorem, von Neumann soon proved his Mean Ergodic Theorem.[§]

Koopman, Stone, and von Neumann were still under 30, and discussed their ideas freely among each other and with G. D. Birkhoff, himself not yet 50, who had taught two of them. Within weeks of hearing about von Neumann's Mean Ergodic Theorem, G. D. Birkhoff invented a new method to prove a stronger Pointwise Ergodic Theorem, more closely related to the work that he and Carathéodory had done earlier. In a historical note written with Koopman [1, pp. 462–5], Carathéodory's role is made clear.

My Undergraduate Thesis

I now come to Carathéodory's influence on my own career, viewed retrospectively. This took place in 1931 and 1932, the years in which ergodic theory was born and functional analysis was established. It began with my undergraduate thesis; to understand the following personal digression, one should know something about the tutorial system used by the Harvard Mathematics Dept. from around 1928 to World War II.

In pursuance of the principle that "all real education is self-education", in which Harvard's President Lowell believed strongly, honors candidates were encouraged to take reading courses under faculty guidance. It was my privelege to have Marston Morse as my tutor, and he encouraged me to read Chapter VII of the third (1914) edition of de la Vallée-Poussin's *Cours d'Analyse Infinitésimale*. This chapter had been inserted (without

[§]See George Mackey's article, *Von Neumann and the early days of ergodic theory*, Proc. Symp. Pure Maths. 50 (1990) 25–38. This does not mention Carathéodory's role.

exercises) as a "completely new foundation for the theory of the Lebesgue integral", into what had before been a calculus text for French-speaking university students of mathematics.

I understood the principles involved, and then read Hausdorff's beautiful *Grundzüge der Mengenlehre*, from which I learned many things: axioms for "teilweise geordneten Mengen" (now called posets), rings and σ-rings of sets, transfinite numbers, Hausdorff spaces, and why (p. 469) there exists no measure in $\mathbb{R}^3$ that is invariant under all rigid rotations. I also read Fréchet's famous 1906 thesis, which introduced me to functional analysis and gave me another perspective into general topology. Morse saw me for a half-hour every week or two, and checked up on my understanding of all this material.

In my senior year, I then read Carathéodory's 1914 paper on linear measure [3, pp. 249–75], and admired its three axioms for outer measure. Its last section sketches very clearly the concept of p-dimensional outer measure μ_p^* in q-dimensional space, developed by Hausdorff four years later. I read Hausdorff's paper (*Math. Annalen* 79, 157–79); p need not be an integer, and one can define the Hausdorff *dimension* $d[S]$ of any Borel set $S \subset \mathbb{R}^q$ as that number $d \leq q$ such that

$$\mu_p^*(S) = \begin{cases} \infty & \text{if } p < d\,, \\ 0 & \text{if } p > d\,. \end{cases} \tag{1}$$

My 80-page undergraduate thesis built on some of the ideas I had read about. I was unable to prove the following plausible (to me at that time) conjecture: if $x(t)$ is a rectifiable curve in $\mathbb{R}^n$, and S_k denotes the Borel set of all points $\underline{a}$ such that $\underline{x}(t) = \underline{a}$ has k solutions, then

$$\sum \mu(S_k) = \int |\underline{x}'(t)|dt\,. \tag{2}$$

This seemed plausible, because every function of bounded variation is differentiable almost everywhere. Of course it is false: there exist nonconstant functions of bounded variation whose derivative vanishes almost everywhere.† It was good that I did not give a "proof" of my conjecture!

†De la Vallée-Poussin's book defined absolute continuity, but did not give a graphic *example* of a function that is continuous without being absolutely continuous.

Carathéodory and Lattice Theory

My mathematical immaturity in 1932 should be obvious from the preceding discussion. My career plan was to become a mathematical physicist; I had taken a graduate course in quantum mechanics, and Cambridge University (which had no official Ph. D. program) accepted me as a research student, with R. H. Fowler as my advisor. But besides mathematical physics, topology, and measure theory, I also had a fourth interest: browsing in the Harvard departmental library, I had begun reading the book *Finite Groups* by Miller, Blichfeldt, and Dickson.

After learning about the two groups of orders 4 and 6, and the five groups of order 8, I became fascinated with the problem of determining all finite groups of given order n. Working by myself in Munich that summer, very naively, I did rediscover Kronecker's fundamental theorem (of around 1870?), which states that every finite *commutative* group is the product $C_1 \times \ldots \times C_r$ of cyclic groups of orders n_i, where each n_i is a divisor of n_{i-1}. But I was basically floundering as regards the area of research to pursue.

At that juncture, I fortunately asked Carathéodory to give me an interview. He graciously invited me to come to his home for tea, where I met his son and daughter. After tea, he showed me his library, a room perhaps $25' \times 15'$ in size, containing many multishelved steel stacks lined with books. I had never seen a personal library like that before!

Dazzled, I shyly told him of my interest in finite groups. He advised me that, to become well-informed on that subject, I should read Speiser's *Gruppentheorie*. He added that, to "learn about algebra in general", I should read van der Waerden's *Moderne Algebra*, which was "creating quite a stir in Germany."

I bought both books, and studied them like bibles during the following year at Cambridge University. I abandoned mathematical physics, which I had previously intended to make my major field. While greatly enjoying Hardy's brilliant lectures (he taught number theory without mentioning ideals!), I conversed more seriously about unsolved problems in group theory with Philip Hall.

Gradually, I concentrated on the *structure* of finite groups, and by spring I had rediscovered the basic axioms characterizing lattices, modular lattices, and distributive lattices, whose importance was first recognized by Dedekind around 1900. My first paper on lattices was published in the

Proceedings of the Cambridge Philosophical Society that October, which also published two years later a second paper concerned with the definition and basic properties of "algebras" in general. These papers proved to be very timely; lattice theory had a major renaissance in the 1930s.† Both owed much to the good advice given to me by Carathéodory in Munich in 1932.

Two Representation Theorems

Many of the definitions and theorems of my 1933 paper on "lattices" had already been stated around 1900 in two papers by Dedekind, in connection with algebraic number theory. What are now called distributive lattices were called "Dualgruppen von Idealtypus" by Dedekind. Hausdorff's 1914 book had defined *rings* and *fields* of sets in its §7, and it is obvious that every ring of sets is a distributive lattice.‡ Theorem 25.2 of my 1933 paper was (in different language) the converse representation theorem: every distributive lattice is isomorphic with a ring of sets. The proof was by transfinite induction, for which I referred to van der Waerden's book.

Shortly after I returned to Harvard that fall, Marshall Stone gave a colloquium lecture in which he proved (among other things) a closely related representation theorem: that every Boolean algebra is isomorphic with a field of sets.* We had a friendly chat after his lecture, and agreed not to make priority an issue!

In the next few years, I would prove a sharper theorem relating *finite* distributive lattices to T_0-spaces (it was already well-known that every finite Boolean algebra is isomorphic with the field of *all* subsets of some finite set.), while Stone would prove much deeper representation theorems about infinite Boolean algebras, in two long and famous papers.

---

†See the book *Die Entstehung der Verbandstheorie* by H. Mehrtens for a scholarly account of this renaissance; pp. 1–11 of my article in *Trends in Lattice Theory* (J. C. Abbott, ed.), van Nostrand, 1970, gives a more personal account.

‡The word "ring" seems to have been first used in its usual modern sets by A. Fraenkel around 1908; Hausdorff's meaning is of course different. I called distributive lattices "C-lattices" in my 1933 paper.

*Stone had announced his result in Abstract 39-5-86 of the Bull. A.M.S. See also Proc. Nat. Acad. Sci. 20 (1934) 197–202.

Stone and Carathéodory, II

In the first of these papers, Stone explains that his interest in the subject "arose in connection with the spectral theory of symmetric transformations in Hilbert space", the theory in which he had become involved seven years earlier by the proofsheets of von Neumann's unpublished paper given to him by Carathéodory. His second paper proved the now famous Stone-Cech compactification theorem, which asserts that every completely regular Hausdorff space X can be extended to a *compact* Hausdorff space $\beta(X)$, which contains X as a dense subset.

The compactification $\beta(X)$ is a maximal compactification; in a casual conversation with me, Stone once emphasized how many others there are. For example, the open unit disk Δ can be compactified to the Riemann sphere by adding a point at infinity, or to the projective plane by adding a projective "line at infinity", as well as by $\beta(\Delta)$. I have often wondered whether, in this case, the imaginary "points" added to form $\beta(\Delta)$ are not the same as the "prime ends" used by Carathéodory in his well-known work on conformal representation, and whether he may not have mentioned them in lectures attended by Stone in 1928.

Algebraic Measure Theory

In two papers published in 1938 and reprinted in [3, pp. 302–51], Carathéodory began to develop new algebraic foundations for the theories of measure and integration. Because *sets* ("Mengen") are treated as *elements* in Boolean algebra, he referred to these elements as *Soma* (Greek for the German word "Körper"), referring to what are usually called "fields of sets" in English as "Körper von Soma". These papers were clearly stimulated by the 1935 papers of Stone and Ore.† Since Ore's work was stimulated by my 1933 paper on lattices (which he called "structures"), Carathéodory's papers constitute in some sense further interaction with Harvard.

A 1942 paper [3, pp. 443–73] actually dealt with Boolean *lattices*, emphasizing their definition in terms of the disjointness relation $a \wedge b = 0$. This defines a mapping $\alpha \to a^{\perp}$ (or $A^{\perp}$), the *set* of all b disjoint from a, and one can *define* $a \leq b$ to mean that $A^{\perp} \supset B^{\perp}$. This has many interesting interpretations and applications, to an earlier theorem of Crlivenko and

†Cf. Stone, Am. J. Math. 37 (1935) 703–28; O. Ore, Annals of Math. 36 (1935) 406–37.

later developments in the theory of orthomodular lattices.

In his last years, Carathéodory gave these and later papers a "systematic and unified exposition". Carefully edited by P. Finsler, A. Rosenthal, and R. Steuerwald, this exposition is available in both German and English editions [4].

There will never be a last word on measure theory. For example, the fractional-dimensional measure (2) initiated by Carathéodory and Hausdorff is one of the two main ideas underlying the theory of *fractals* developed by Benoit Mandelbrot [5], the other being that of a semigroup of self-similar transformations. The importance of fractals for interpreting many natural and mathematical phenomena is now fully recognized; since Mandelbrot was a Visiting Lecturer at Harvard for some years before going permanently to Yale, this represents still another connection, between Carathéodory and Harvard.

My own changing attitudes toward the algebraization of measure theory are expressed in three editions of my book *Lattice Theory*, all of which pay tribute to Carathéodory. The 1940 edition (p. 99) begins by stating that measure theory "has been perfected by Carathéodory". The 1948 edition states on p. 181 that its "formulation is due to the genius of C. Carathéodory". This deeper account of *Borel algebras* in the 1967 edition (pp. 254–66) attributes to Carathéodory's 1914 paper the fundamental concepts of outer and regular measures (p. 363). However, all of these editions were written without studying Carathéodory's highly original theory of soma. Some historically minded expert on measure theory should surely correlate this theory with other treatments!

References

1. G. D. Birkhoff, *Collected Scientific Papers*, vol. II, American Mathematical Society, 1950.
2. G. Birkhoff and E. Kreyszig, *The establishment of functional analysis*, Historia Math. 11 (1984) 258–321.
3. C. Carathéodory, *Gesammelte Mathematische Schriften*, Bd. IV, Becksche Verlag, Munich, 1954.
4. C. Carathéodory, *Mass und Integral und ihre Algebraisierung*, Birkhäuser, 1956; *Measure and Integration*, Chelsea, 1963.
5. B. B. Mandelbrot, *Fractals: Form, Chance, and Dimension*, Freeman, 1977.
6. J. von Neumann, *Collected Works*, vol. II, Pergamon Press, 1961.

Garrett Birkhoff
Harvard University
Department of Mathematics
Cambridge, MA 02138
USA

Constantin Carathéodory: An International Tribute (pp. 76–90)
edited by Th. M. Rassias

CARATHEODORY EXTENSION PROCESS AND APPLICATIONS TO WEIERSTRASS-TYPE INTEGRALS

Primo Brandi *Anna Salvadori*

We present the Caratheodory-type process which allows to extend the Burkill-Cesari integral to a Borel measure. Moreover we present some applications to the Weierstrass functionals of the calculus of variations.

1. Introduction

As it is well-known, the Weiertsrass functionals of the calculus of variations are defined by means of a Burkill-Cesari integration process over an appropriate set function (see e.g. [5b] for a survey). This approach presents the remarkable advantage of a direct and constructive definition, as a limit process over a finite summation, and it preserves a precise geometrical meaning connected to the underlying variety. On the other hand, some topics - as semicontinuity - appear harder to solve in this framework than in other ones, as Lebesgue-Stieltjes or Serrin contexts. Thus a great interest holds a connection among these different approaches.

In particular we refer here to the comparison between Weierstrass and Lebesgue-Stieltjes functionals of the calculus of variations. As we will show, the key result on this direction is the possibility of supporting Burkill-Cesari integrable set functions with a suited measu

re. This idea, which Cesari primarily developed in [4bc], makes an essential use of the classical Caratheodory extension process[3]. The consequent applications to Weierstrass-type integrals allows to get important results. In particular, the Weiertrass functionals admit a representation in terms of a suitable Lebesgue-Stieltjes integral, both in the parametric and non-parametric setting. Furthermore, in the non-parametric setting, the Weierstrass integral is greater than the corresponding Lebesgue one,in general,and the two functionals coincide if and only if the underlying variety is absolutely continuous. In other words, the classical Tonelli-type theorem (given for the length of a curve primarily) is still valid for the general Weierstrass functional. Anyway these are only some of the "capacities" developed in force of the associated measure.

In this note we first illustrate the Caratheodory process adopted for the Burkill-Cesari integral, in the general case of a set function with values on a Banach space; then we present the more recent and advanced applications to Weierstrass-type functionals.

2. Extension of Burkill-Cesari Integral to a Measure

Let A be a metric space, we denote by $\mathcal{M}$, $\mathcal{G}$ and $\mathcal{B}$ the family of all the subsets of A, that of the open sets and the Borel σ-algebra, respectively. Let $\{I\}\subset\mathcal{M}$ be a fixed sub-family of compact sets, that we shall call "intervals". For finite system D we meant a finite collection of non-overlaping intervals, i.e. $D=\{I_1,\dots,I_N\}$ with $I_i^\circ \neq \emptyset$ and $I_i^\circ \cap I_j = \emptyset$, $i\neq j$, $i,j=1,\dots,N$. Let $(D_t)_{t\in T}$ be a given net of finite systems such that $\inf_T \max\{\operatorname{diam}(I), I\in D_t\} = 0$.
Let $s:\mathcal{M}\times\mathcal{M}\to\{0,1\}$ be the function defined by $s(H,K) = 1$ when $H\subset K$, $s(H,K) = 0$ otherwise. Let ϕ: $\{I\}\to E$ be a given interval function, with E real Banach space.

The function ϕ is said to be *integrable in the sense of Burkill-Cesari* (BC-integrable) over $M \in \mathcal{M}$ ([4b]) if the limit below exists

$$\lim_{T} \sum_{I \in D_t} s(I,M)\, \phi(I) = \int_M \phi \, .$$

In order to provide existence and hereditarity for this algorithm, Cesari introduced the following definition ([4b]).

The function ϕ is *quasi-additive* (q.a.) if:

(q.a.) given $\varepsilon>0$ there exists t_1 such that, for every $t_0 >> t_1$ there exists t_2 with the property that for every $t>>t_2$, put $D_{t_0} = [I]$ and $D_t = [J]$, we have

$$\text{i)} \quad \sum_I \Big| \sum_J s(J,I)\phi(J) - \phi(I) \Big| < \varepsilon$$

$$\text{ii)} \quad \sum_J \Big[1 - \sum_I s(I,J)\, |\phi(J)| \Big] < \varepsilon .$$

Moreover the function ϕ is said to be *(o)-quasi-additive* ((o)-q.a.) if (q.a.) holds with $s(J,I)$ substituted by $s(J,I^\circ)$. Of course ((o)-q.a.) is stronger than (q.a.).

The function ϕ is said to be of *bounded variation* (BV) if:

$$\overline{\lim_{T}} \sum_{I \in D_t} |\phi(I)| < +\infty .$$

The following proposition is a key result on the theory of BC-integral (see Cesari [4bc], Breckenridge [2], Brandi-Salvadori [1ab]).

Theorem 2.1. *If ϕ is q.a. and BV, the interval functions ϕ, $|\phi|$, $\langle z,\phi\rangle^+$, $\langle z,\phi\rangle^-$, $z \in E'$, are BC-integrable on every set $M \in \mathcal{M}$.*

Thus, under the assumptions of Theorem 2.1, we can consider the functions $|\nu|$, ν_z^+, $\nu_z^- : \mathcal{M} \to R_o^+$, $z \in E'$, defined by

$$|\nu|(M) = \inf \int_G |\phi| \quad , \quad \nu_z^+(M) = \inf \int_G \langle z,\phi\rangle^+ \quad \nu_z^-(M) = \inf \int_G \langle z,\phi\rangle^-$$

where the infimum is taken with respect to all the sets $G\in\mathcal{G}$ with $G\supset M$. Furthermore, let $\nu_z:\mathcal{M}\to R$, $z\in E'$, and $\nu:\mathcal{M}\to E''$ be respectively defined by

$$\nu_z(M) = \nu_z^+(M) - \nu_z^-(M) \quad \text{and} \quad \langle\nu(M),z\rangle = \nu_z(M)\ , \ z\in E'.$$

Then it can be shown that $|\nu|$, ν_z^+ and ν_z^-, $z\in E'$, are *Caratheodory outer measures*, and in force of Caratheodory extension process, the following result can be proved (see Cesari [4bc], Breckenridge[2], Brandi-Salvadori[1b]).

Theorem 2.2. *Suppose that ϕ is (o)-q.a. and BV, then the functions $|\nu|$ and ν are measures over $\mathcal{B}$. Moreover $|\nu|$ coincides with the total variation of ν.*

Note that, for every $G\in\mathcal{G}$, we have $\nu(G) = \int_G \phi$ and $|\nu|(G) = \int_G |\phi|$; in other words, the measures ν and $|\nu|$ extend (from $\mathcal{G}$ to $\mathcal{B}$) the BC-integral of ϕ and $|\phi|$ respectively.

2.1. Properties of the extension measure. In order to point out some useful properties of the measures just defined, let us strengthen our setting by assuming that a subnet $(D_{t_n})_{n\in N}$ exists such that:

- for every $n\in N$ and $I\in D_{t_n}$ we have that $I = \bigcup_{J\subset I} J$ where $J\in D_{t_{n+1}}$;
- the intervals $\{I\}^* = \{I\in D_{t_n}, n\in N\}$ are connected;
- $\lim_{n\to\infty} \max\{\mathrm{diam}(I), I\in D_{t_n}\} = 0$.

Theorem 2.3. *Suppose that ϕ is (o)-q.a. and BV, then the measures ν and $|\nu|$ satisfy the conditions*

i) $\lim_{n\to\infty} |\nu|(G - \bigcup_{I\in D_{t_n}} I^\circ) = 0$, $G\in\mathcal{G}$;

ii) for every $I \in \{I\}^*$ *we have* $\int_{I^\circ} \phi = \nu(I^\circ) = \nu(I) = \int_I \phi$.

Now let $\lambda : \{I\} \to R^+$ be an interval function (o)-q.a. and BV and denote by μ the measure which extends its BC-integral. We say that ϕ is *AC^* with respect to* λ if ϕ is absolutely continuous with respect to λ on $\{I\}^*$. In [1e] the following result is proved.

Theorem 2.4. *Suppose that* ϕ *and* λ *are (o)-q.a. and BV. Then* ϕ *is AC^* with respect to* λ *iff the measure* ν *is AC with respect to* μ.

Let us consider now the sequence of step functions $(\eta_n)_{n\in N}$ with $\eta_n : A \to E$ defined by

$$\eta_n(a) = \begin{cases} \dfrac{\nu(I)}{\mu(I)} , & a \in I^\circ, \ \mu(I) \neq 0, \ I \in D_{t_n} \\ 0 , & \text{otherwise} \end{cases}$$

About its convergence, the following result holds (see [1ec]).

Theorem 2.5. *Suppose that* E *is a reflexive Banach space. A function* $\frac{\delta\nu}{\delta\mu} : A \to E$ *exists such that* $\eta_n \to \frac{\delta\nu}{\delta\mu}$ μ*-a.e. on* A. *Moreover, if* ϕ *is AC^* with respect to* λ *and the* σ*-algebra generated by* $\{I\}^*$ *coincides with* $\mathcal{B}$, *we have that* $\frac{\delta\nu}{\delta\mu} = \frac{d\nu}{d\mu}$, *where* $\frac{d\nu}{d\mu}$ *denotes the Radon-Nikodym derivative.*

We wish to point out that the above convergence result is proved by a suitable connection between BC-integration process the theory of martingales.

Remark 2.6. Let us consider now a remarkable particular case of our setting. Let $A = [a_o, b_o] \times [c_o, d_o]$ be a closed rectangle and let $\{I\}$ be a dense family of closed subrectangles. Consider the class of all the finite partitions D of A into rectangles of $\{I\}$, directed by the mesh function $\delta(D) = \max\{\mathrm{diam}(I), I \in D\}$. Finally let $\phi : \{I\} \to R$ be an (o)-q.a. and BV rectangle function and let $\lambda(I) = \mathrm{meas}(I)$, $I \in \{I\}$.

In this context the following representation holds (see[1e]).

If the derivative $D\Phi$ *of the rectangle function* Φ *exists a.e., then*

$$\frac{\delta\nu}{\delta\mu} = D\Phi \quad a.e. \ .$$

2.2. The generalized area of a BV surface and its extension measure.

Let $z: R_o=[a_o,b_o]\times[c_o,d_o] \to R$ be a BV surface. As it is well-known ([1d]), two sequences $(x_n)_n$ in $]a_o,b_o[$ and $(y_n)_n$ in $]c_o,d_o[$ respectively exist such that

$$\int_{c_o}^{d_o} |z(x_n-0,y)-z(x_n+0,y)|dy \neq 0 \quad \text{and} \quad \int_{a_o}^{b_o} |z(x,y_n-0)-z_n(x,y_n+0)|dx \neq 0$$

where $z(x_n\pm0,y)$, $z(x,y_n\pm0)$ denote the essential limits.

Let $S=\{(x,y)\in R_o: x=x_n \text{ or } y=y_n,\ n\in N\}$, let $\{R\}$ be the family of all the subrectangles of R_o and let $\{R\}_S$ be that of the subrectangles whose sides intersect S in a set of null linear measure. Let $\mathcal{D}_S$ be the collection of all the finite systems $D=[R]$ of rectangles $R\in\{R\}_S$ such that $\bigcup_{R\in D} R = \hat{R}\in\{R\}_S$. Finally let $\delta: \mathcal{D}_S \to R^+$ be the mesh function defined by $\delta(D) = \max\{\text{meas}(R_o-\hat{R}), \text{diam}(R), R\in D\}$.

Let us consider the rectangle function $\Phi:\{R\}\to R$ defined by

$$\Phi(R) = \left[(\Phi_1(R))^2+(\Phi_2(R))^2+(\text{meas}(R))^2\right]^{\frac{1}{2}},$$

with

$$\Phi_1(R) = \int_c^d V_y(z,[a,b])\,dy \ , \quad \Phi_2(R) = \int_a^b V_x(z,[c,d])\,dx \ , \quad R=[a,b]\times[c,d]$$

where V_y and V_x denote the generalized variation.

In [1d] we have proved that Φ *is q.a. and BV with respect to* $\mathcal{D}_S,\delta$ *and* $\int_{R_o}\Phi$ *coincides with the generalized area* α *of the surface* z ([4a]). Moreover Φ is (o)-q.a. with respect to a dense subfamily of $\mathcal{D}_S$, thus *its BC-integral can be extended to a measure* $\nu:\mathcal{B}\to R$

Among the others, in [1d] we have pointed out the following properti

es of the measure ν , which can be deduced by the general results illustrated in section 2.1

i) $\nu(\{x,y\}) = 0$ *for every* $(x,y)\in R_o$;

ii) $\nu(\partial R_o) = 0$;

iii) $\nu(\{x\}\times[c,d]) = \int_c^d |z(x-0,y)-z(x+0,y)|\,dy$ *and*

$$\nu([a,b]\times\{y\}) = \int_a^b |z(x,y-0)-z(x,y+0)|\,dx;$$

iv) $\nu(R^\circ) = \alpha(z,R)$ *for every* $R\in\{R\}$, *and* $\nu(R) = \alpha(z,R)$ *if* $R\in\{R\}_S$;

v) $\nu(B) = \nu_a(B) + \nu_s(B) = \int_B \left(\left[z_x(x,y)\right]^2 + \left[z_y(x,y)\right]^2 + 1 \right)^{\frac{1}{2}} dxdy + \nu_s(B)$

is the Lebesgue decomposition.

3. Applications to Weierstrass-type Functionals of the Calculus of Variations

The results summarized in section 2 have got interesting applications on the Weierstrass approach to problems of Calculus of Variations, as we shall show in what follows.

3.1. The parametric case. Let (K,d) be a metric space, E be a uniformly convex Banach space and B be a Banach space. Consider the functions $p:\{I\} \to K$, $\phi:\{I\} \to E$ and $F: K\times E \to B$.

Let us denote by $\Phi:\{I\} \to B$ the interval function

$$\Phi(I) = F\big(p(I),\phi(I)\big) ;$$

according to Cesari ([4b]), the BC-integral of the function Φ (when it exists) is called the *parametric Weierstrass-integral of the Calculus of Variations* and denoted by $\int_A F(p,\phi)$.

Let us start by recalling the more advanced result on the existence of this integral (see [1h]). On this purpose, we state first the key assumption.

The couple (p,ϕ) is said *Γ-quasi additive* (Γ-q.a.) ([1h]) if (Γ-q.a.) given $\varepsilon>0$ there exist $0<\sigma<\varepsilon$ and t_1 such that, for every $t_0>>t_1$ there exists t_2 with the property that for every $t>>t_2$ we have

i) $$\sum_I \Big| \sum_{J\in\Gamma_I} \phi(J) - \phi(I)\Big| < \varepsilon$$

ii) $$\sum_I \Big| \sum_{J\notin\Gamma_I} s(J,I)\phi(J)\Big| < \varepsilon$$

iii) $$\sum_J \Big[1 - \sum_I s(J,I)\ |\phi(J)|\Big] < \varepsilon$$

where we put $D_{t_0} = [I]$, $D_t = [J]$ and Γ_I denotes a subfamily (even empty) of the set $\{J\subset I : d(p(J),p(I))<\sigma\}$.

Note that : *if the couple (p,ϕ) is Γ-q.a. then ϕ is q.a..*

On the integrand F we assume the following classical condition

(F) F is bounded and uniformly continuous on $K\times S_1$, where $S_1=\{x\in E:|x|=1\}$, and $F(k,\cdot)$ is positively homogeneous of degree one over E, $k\in K$.

Theorem 3.1. (Existence) *Suppose that*

- *the integrand F satisfies assumption (F);*
- *the couple (p,ϕ) is Γ-q.a. and ϕ is BV;*

then the interval function Φ is q.a. and BV. Thus $\int_M F(p,\phi)$ exists, for every $M\in\mathcal{M}$.

Now, in force of the above mentioned Caratheodory extension process, we can represent the Weierstrass functional in terms of a suitable Lebesgue-Stieltjes integral. On this purpose, let $\nu:\mathcal{B}\to E$ denote the vectormeasure associated to the interval function ϕ, as in section 2. Moreover assume that our setting satisfies the assumptions of section 2.1.

Let us consider the sequences of step functions $(p_n)_{n\in N}$ and $(\eta_n)_{n\in N}$, with $p_n: A \to K$ and $\eta_n: A \to E$ defined by

$$p_n(a)=\begin{cases} p(I) & a\in I^\circ,\ I\in D_{t_n} \\ k^* & \text{otherwise} \end{cases}, \quad \eta_n(a)=\begin{cases} \dfrac{\nu(I)}{|\nu(I)|} & a\in I^\circ,\ I\in D_{t_n} \\ 0 & \text{otherwise} \end{cases}, \ n\in N$$

where $k^*\in K$ is fixed and $\nu(I)/|\nu(I)|= 0$ if $\nu(I) = 0$.
In force of Theorem 2.5 we have that $\eta_n \to d\nu/d|\nu|$ ν-a.e. .
Moreover we shall suppose that the following condition is satisfied

(c) there exists a ν-measurable function $\P: A \to K$ such that

$$p_n \to \P \quad \nu\text{-almost everywhere.}$$

Thus we have that $\sum_{I\in D_{t_n}} F(p(I),\nu(I)) = \int_{A_n} F(p_n,\eta_n)\, d|\nu|$, where $A_n = \bigcup_{I\in D_{t_n}} I$.

Moreover it can be proved that $\int_A F(p,\phi) = \lim_{n\to\infty} \int_{A_n} F(p_n,\eta_n)$.

As a consequence, the following representation result can be deduced (see [Ihbc] for the details).

Theorem 3.2. (Representation) *Assume that all the assumptions of Theorem 3.1 are satisfied and moreover suppose that*

- *K is compact;*
- *the function ϕ is (o)-q.a. and the sequence $(p_n)_n$ satisfies (c);*

then for every $G\in\mathcal{G}$ we have

$$\int_G F(p,\phi) = \int_G F\left(\P, \frac{d\nu}{d|\nu|}\right) d|\nu|.$$

3.2. The non-parametric case. Let (C,d) be a compact metric space and consider the functions $q: \{I\} \to C$, $\phi: \{I\} \to R^n$, $\lambda: \{I\} \to R^+$ and $f: C\times R^n \to R^m$.

Let us denote by $\Psi:\{I\} \to R^m$ the interval function

$$\Psi(I) = \lambda(I)\, f\left(q(I), \frac{\phi(I)}{\lambda(I)}\right) \quad ;$$

according to Vinti ([5a]), the BC-integral of the function Ψ (when it exists) is called the *non-parametric Weierstrass-integral of the Calculus of Variations* and denoted by $\int_A f(q, \frac{\phi}{\lambda})$.

As it is well-known ([5a]), the non-parametric Weierstrass functional can be handled as a parametric one, by associating to f a suitable parametric integrand F ; i.e. $F: C \times R^{n+1} \to R^m$ defined by

$$F(u;t,v) = \begin{cases} |t| f(u, \frac{v}{|t|}) & , \text{ if } t \neq 0 \\ \lim_{\tau \to 0} F(u;\tau,v) & , \text{ if } t = 0 \end{cases}$$

Moreover note that, *if the function* f *satisfies the assumption*

(f) $f(u,\cdot)$ is convex, for every $u \in C$;

the function $f(u,v)/1+|v|$ is bounded and uniformly continuous;

then the associated integrand F *satisfies condition* (F).

Thus, as an application of Theorems 3.1 and 3.2 the following results can be proved (see [1ic] for the details).

Theorem 3.3. (Existence) *Suppose that*

- *the integrand* f *satisfies assumption* (f);
- *the couple* $(q;(\lambda,\phi))$ *is* Γ*-q.a. and* (λ,ϕ) *is BV;*

then the interval function Ψ *is q.a. and BV. Thus* $\int_M f(q, \frac{\phi}{\lambda})$ *exists, for every* $M \in \mathcal{M}$.

Now assume that our setting satisfies the conditions of section 2.1. Under the assumption that (λ,ϕ) is (o)-q.a. and BV, let μ and ν denote the measures which extend the BC-integral of λ and ϕ respectively.

Theorem 3.4. (Representation) *Assume that all the assumptions of Theorem 3.3 are satisfied and moreover suppose that*

- (λ,ϕ) *is (o)-q.a. and condition* (c) *holds (i.e.* $q_n \to \xi$ *, (μ,ν)-a.e.);*

then for every $G \in \mathcal{G}$ *we have*

$$\int_G f(q,\frac{\phi}{\lambda}) = \int_G F(\xi;\ \frac{d\mu}{d|(\mu,\nu)|},\frac{d\nu}{d|(\mu,\nu)|})\ d|(\mu,\nu)|.$$

Furthermore, in non-parametric setting, we can prove more; indeed the following result holds,which allows to compare the Weierstrass functional with the corresponding Lebesgue one, according to the classical theorem that Tonelli gave for the length of a curve (see [1i]).

Theorem 3.5. (Comparison) *Under the assumptions of Theorem 3.4, assume that* f *is non-negative, then for every* $G \in \mathcal{G}$ *we have*

$$(*) \qquad \int_G f(q,\frac{\phi}{\lambda}) \geq \int_G f\left(\xi;\frac{\delta\nu}{\delta\mu}\right) d\mu\ .$$

Moreover suppose that

$$f(u,v) \geq -1 + M|v|, \quad (u,v) \in C\times R^n;$$

then the equality sign holds in $(*)$ *iff* ϕ *is* AC^* *with respect to* λ. *And in this case* $\frac{\delta\nu}{\delta\mu} = \frac{d\nu}{d\mu}$.

In order to illustrate the results of sections 3.1 and 3.2, we take into consideration the following particular cases (see [1hilmno] for the details).

3.3. The Weierstrass-integrals over a BV curve. Let $x:[a,b] \to R^n$ be a BV curve and denote by $Z = \{c \in]a,b[: x(c) = x(c+0) = x(c-0)\}$, then $[a,b] - Z$ is a null set. Let $\{I\}$ be the family of all the closed intervals in $[a,b]$ whose end-points belong to Z and let $\mathcal{D}$ be the collection of the finite subdivisions of the type $D = \left[I_1,\ldots,I_N\right]$ with

$I_i=[\alpha_i,\alpha_{i+1}]\in\{I\}$, $i=1,\dots,N$. We consider the mesh function $\delta:\mathcal{D}\to R^+$ defined by $\delta(D) = \max\{(\alpha_1-a),(b-\alpha_{N+1}), \text{meas}(I)\ ,\ I\in D\}$. Let $p_o:\{I\}\to R$, $p_x:\{I\}\to R^n$, $\lambda:\{I\}\to R^+$ and $\Delta x:\{I\}\to R^n$ be the interval functions defined by

$p_o(I)\in I$, arbitrarily choosen;

$p_x=(p_1,\dots,p_n)$ and $p_i(I)=\gamma_i\ \text{infess}(x_i,I) + (1-\gamma_i)\ \text{supess}(x_i,I)$, where $0\le\gamma_i\le 1$ are given constants, $i=1,\dots,n$;

$\lambda(I) = \text{meas}\ (I)$;

$\Delta x(I) = x(\beta) - x(\alpha)$, if $I=[\alpha,\beta]$.

Finally, let $F:K\times R^n\to R^m$ and $f:C\times R^n\to R$ be given with $K\supset x([a,b])$ and $C\supset \text{graph } x$.

In the present particular case, the Weierstrass integrals $\int_{[a,b]} F(p_x,\Delta x)$ and $\int_{[a,b]} f(p_o,p_x;\frac{\Delta x}{\lambda})$ (when they exist) are called *the parametric and non-parametric Weierstrass functionals over the curve* x, respectively and shortly denoted by $W_F(x)$ and $W_f(x)$.

It can be proved that Δx is (o)-q.a. and BV with respect to $\mathcal{D},\delta$; thus let ν_x denote the associated measure (i.e. the Stieltjes measure associated to x). Moreover condition (c) is satisfied by the function $\P_x:[a,b]\to R^n$ defined by

$$\P_i(t) = \gamma_i \min(x_i(t+0),x_i(t-0)) + (1-\gamma_i)\max(x_i(t+0),x_i(t-0)),\quad 1\le i\le n.$$

Hence, in force of Theorems 3.1-3.5 the following results hold.

Theorem 3.6. *Suppose that* K *is compact and the integrand* F *satisfies condition* (F); *then* $W_F(x)$ *exists and the representation holds*

$$W_F(x) = \int_a^b F(\P_x,\ d\nu_x/d|\nu_x|)d|\nu_x|;$$

in particular, if x *is AC , then*

$$W_F(x) = \int_a^b F(x(t),x'(t))\ dt\ .$$

Theorem 3.7. *Suppose that* C *is compact and the integrand* f *satisfies all the assumptions of Theorem 3.5; then* $W_f(x)$ *exists and the relation holds*

$$W_f(x) \geq \int_a^b f(t,x(t),x'(t))\,dt$$

and the equality sign holds iff x *is AC.*

3.4. The multiple Weierstrass-integral over a BV surface. Let $z:R_o \to R$ be a BV surface and let $\mathcal{D}_S$ and δ be defined as in section 2.2. Let us consider the rectangle function $\phi=(\phi_1,\phi_2) :\{R\} \to R^2$ defined by

$$\phi_1(R) = \int_c^d \Big[z(b-0,y)-z(a+0,y)\Big]dy \quad , \quad \phi_2(R) = \int_a^b \Big[z(x,d-0)-z(x,c+0)\Big]dx \ ,$$

where $R=[a,b]\times[c,d]$. Moreover let $q:\{R\} \to R^2$ and $\lambda:\{R\} \to R^+$ be defined by $q(R)\in R$ arbitrarily choosen, $\lambda(R) = \text{meas}(R)$.

In the present particular case, the Weierstrass integral $\int_{R_o}\lambda f(q,\frac{\phi}{\lambda})$ (when it exists) is called *the multiple Weierstrass functional over the surface* z, and shortly denoted by W(z).

It can be proved that ϕ is (o)-q.a. and BV with respect to $\mathcal{D}_S,\delta$; thus let $\nu=(\nu_1,\nu_2)$ denote the measure which extends the BC-integral of ϕ.

As an application of Theorems 3.3 and 3.5 the following result can be proved.

Theorem 3.8. *Suppose that* C *is compact and the integrand* f *satisfies all the assumptions of Theorem 3.5; then* W(z) *exists and the relation holds*

$$W(z) \geq \int_{R_o} f(x,y,z_x(x,y),z_y(x,y))\,dxdy$$

where the equality sign holds iff $z\in W^{1,1}$.

REFERENCES

1. P.Brandi - A.Salvadori, (a) *Sull'estensione dell'integrale debole alla Burkill-Cesari ad una misura,* Rend. Circ. Mat. Palermo 30 (1981), 207-234;

(b) *Un teorema di rappresentazione per l'integrale parametrico del Calcolo delle Variazioni alla Weierstrass,* Ann. Mat. Pura Appl. 124 (1980), 39-58;

(c) *Martingale ed integrale alla Burkill-Cesari,* 67 (1979), 197-203;

(d) *Sull'area generalizzata,* Atti Sem. Mat. Fis. Univ. Modena 28 (1979), 33-62;

(e) *The non-parametric integral of the Calculus of Variations as a Weierstrass integral. I.- Existence and representation,* J. Math. Anal. Appl. 107 (1985), 67-95;

(f) *The non-parametric integral of the Calculus of Variations as a Weierstrass integral. II.- Some applications,* J. Math. Anal. Appl. 112 (1985), 290-313;

(g) *L'integrale del Calcolo delle Variazioni alla Weierstrass lungo curve BV e confronto con i funzionali integrali di Lebesgue e Serrin,* Atti Sem. Mat. Fis. Univ. Modena 35 (1987), 319-325;

(h) *A quasi-additivity type condition and the integral over a BV variety,* to appear in Pacific. J. Math.;

(i) *On the non-parametric integral over a BV surface,* J. Nonlinear Anal. 13 (1989), 1127-1137;

(l) *On the lower semicontinuity of certain integrals of the Calculus of Variations,* J. Math. Anal. Appl. 144 (1989), 183-205;

(m) *On Weierstrass-type integrals over BV varieties,* to appear on Rend. Accad. Naz. Lincei;

(n) *L'integrale multiplo del Calcolo delle Variazioni per superfici discontinue*, to appear on Atti Sem. Mat. Fis. Univ. Modena;
(o) *On the definition and properties of a variational integral over a BV curve*, Dip. Mat. Univ. Perugia - Rapporto tecnico n.12 (1988).

2. J.C.Breckenridge, *Burkill-Cesari integrals of quasi additive interval functions*, Pacific J. Math. 37 (1971), 635-654.

3. C.Caratheodory, *Vorlesungen über Reelle Funktionen*, Leipzig, Berlin (1927).

4. L.Cesari, (a) *Sulle funzioni a variazione limitata*, Ann. Scuola Norm. Sup. Pisa 5 (1936), 299-312;
(b) *Quasi additive set functions and the concept of integral over a variety*, Trans. Amer. Math. Soc. 102 (1962), 94-113;
(c) *Extension problem for quasi additive set functions and Radon-Nikodym derivatives*, Trans. Amer. Math. Soc. 102 (1962), 114-145.

5. C.Vinti,(a)*L'integrale di Weierstrass e l'integrale del Calcolo delle Variazioni in forma ordinaria*, Atti Accad. Sci. Lett. Arti Patermo 19 (1958), 51-82;
(b) *Non-linear integration and Weierstrass integral over a manifold, connection with theorems on martingales*, J. Optim. Theor. Applic. 41 (1983), 213- 237.

Primo Brandi - Anna Salvadori
Dipartimento di Matematica
Università degli Studi
06100 Perugia, Italy

Constantin Carathéodory: An International Tribute (pp. 91–96)
edited by Th. M. Rassias

A PROPERTY OF GENERALIZED CONVEX FUNCTIONS

Dobiesław Brydak

We consider a linear two-parameter family F of functions defined and twice differentiable on an interval I. Assuming the hypotheses (i)–(v) below we prove that a twice differentiable function ψ: $I \to \mathbb{R}$ is either strictly convex or strictly concave with respect to F iff for every two points $x_1, x_2 \in I, x_1 < x_2$, there exists a unique $c \in (x_1, x_2)$ such that $\psi'(c) = \varphi'(c)$, where $\varphi \in F$ is the unique function satisfying the equalities: $\varphi(x_1) = \psi(x_1)$, $\varphi(x_2) = \psi(x_2)$ (this theorem characterizes the strictly convex or strictly concave functions in the usual sense, where F is the family of all straight lines).

The generalized convex functions with respect to a two-parameter family of functions were defined by E. F. Beckenbach [1]. In this paper we prove that the generalized convex functions with respect to a linear family of functions have a property similar to a property of convex functions. We shall apply the obtained result to linear differential inequalities of second order.

Let F be a linear two-parameter family of real finite functions defined in an interval I having their graphs in a region D and satisfying the following hypotheses

(i) each $\varphi \in F$ is a twice differentiable function;

(ii) for every two points $(x_1, y_1), (x_2, y_2) \in D, x_1 \neq x_2$, there is a unique member of F such that $\varphi(x_1) = y_1$ and $\varphi(x_2) = y_2$;

(iii) for every point $(x_0, y_0) \in D$ and every real number y_0', there is a unique member φ of the family F such that $\varphi(x_0) = y_0$ and $\varphi'(x_0) = y_0'$;

(iv) there exist two linearly independent functions $u, v \in F$ such that

$u'(x) \neq 0$ in I and the wronskian

$$W(x) = \begin{vmatrix} u(x), & v(x) \\ u'(x), & v'(x) \end{vmatrix} \neq 0 \text{ for } x \in I;$$

(v) the function $\lambda(x) := \frac{v'(x)}{u'(x)}$ for $x \in I$ is an increasing function.

Remark. Because of (iv), we have in fact

$$F = \{\varphi : I \to \mathbb{R} : = au + bv, (a,b) \in \mathbb{R}\}.$$

Following E. F. Beckenbach [1] we define the strictly convex (resp. concave) functions as follows:

Definition. The function ψ will be called a strictly convex (resp. concave) function with respect to the family F, provided hypotheses (i) and (ii) are fulfilled and for every $x_1, x_2, x \in I$, with $x_1 < x < x_2$, we have $\psi(x) < \varphi(x)$ (resp. $\psi(x) > \varphi(x)$), where

$$\varphi(x_1) = \psi(x_1) \text{ and } \varphi(x_2) = \psi(x_2). \tag{1}$$

The following lemma will be useful in the sequel

Lemma. Let hypotheses (i)–(iii) be fulfilled. A function $\psi \in C^1(I)$ is a strictly convex (resp. concave) function with respect to the family F iff for every $x_0 \in I$ we have $\psi(x) > \varphi(x)$ (resp. $\psi(x) < \varphi(x)$) for $x \in I \setminus \{x_0\}$, where φ satisfies

$$\varphi(x_0) = \psi(x_0), \varphi'(x_0) = \psi'(x_0), \varphi \in F. \tag{2}$$

This lemma has been proved in [3].

Now we are going to generalize the well known theorem saying that a differentiable function is either strictly convex or strictly concave iff the Lagrange's mean value theorem is fulfilled by this function at a unique point. Namely we are going to prove the following

Theorem 1. Let hypotheses (i)–(v) be fulfilled. A twice differentiable function ψ is either strictly convex or strictly concave with respect to the

family F if and only if for every two points $x_1, x_2 \in I, x_1 < x_2$, there eixsts a unique point $x_0 \in I$ such that $x_1 < x_0 < x_2$ and

$$\psi'(x_0) = \varphi'(x_0)\,, \tag{3}$$

where φ is the unique member of F satisfying (1).

Proof. Let ψ be a twice differentiable function strictly convex with respect to F. If ψ is strictly concave, the proof is similar. Let, further, $\varphi \in F$ satisfy condition (1), where $x_1 < x_2$ are arbitrarily fixed points of I. Thus, by virtue of the Lemma, φ satisfies (2) neither for $x_0 = x_1$ nor for $x_0 = x_2$, whence $\psi'(x_1) \neq \varphi'(x_1)$ and $\psi'(x_2) \neq \varphi'(x_2)$. It follows from the convexity of ψ that

$$\psi'(x_1) < \varphi'(x_1) \text{ and } \psi'(x_2) > \varphi'(x_2)\,. \tag{4}$$

Hence, in view of the Darboux property for derivatives, there exists a point $x_0 \in (x_1, x_2)$ such that equality (3) holds. We are going to prove that such a point is unique.

It is obvious that ψ is strictly convex iff $\psi - \varphi$ is strictly convex. Therefore we may confine ourselves to the case where $\psi(x_1) = \psi(x_2) = 0, \psi'(x_0) = 0$ and $\varphi(x) = 0$ for $x \in I$. It also follows from hypothesis (iii) that for every $x \in I$ there exists a unique $\varphi \in F$ such that $\varphi(x) = \psi(x)$ and $\varphi'(x) = \psi'(x)$ and there are functions $\alpha(x)$ and $\beta(x)$ such that

$$\psi(x) = \alpha(x)u(x) + \beta(x)v(x), \psi'(x) = \alpha(x)u'(x) + \beta(x)v'(x)\,. \tag{5}$$

Hence

$$\beta(x) = \frac{u(x)\psi'(x) - u'(x)\psi(x)}{W(x)}\,, \alpha(x) = \frac{v(x)\psi'(x) - v'(x)\psi(x)}{W(x)}\,. \tag{6}$$

Let us observe that our family F consists of the solutions of the differential equation

$$L[y](x) := W(x)y''(x) - W'(x)y'(x) + V(x)y(x) = 0\,, \tag{7}$$

where V is the wronskian of the system of functions u' and v'. It follows from (5), (6) and (7) that α and β are differentiable and

$$\left.\begin{array}{c} \alpha'(x) = -v(x)\frac{L[\psi](x)}{W(x)}\,, \beta'(x) = u(x)\frac{L[\psi](x)}{W(x)}\,, x \in I \\ \alpha'(x)u(x) + \beta'(x)v(x) = 0\,, \alpha'(x)u'(x) + \beta'(x)v'(x) = L[\psi](x), x \in I \end{array}\right\} \tag{8}$$

Denote by A the wronskian of the functions α and β. One can calculate from (5)–(8) that

$$AW = \psi L[\psi], x \in I. \tag{9}$$

Since $W(x)$ is continuous, we may assume, in views of (iv), that $W(x) > 0$ in I. It has been proved in [3] that ψ is strictly convex with respect to F if and only if it satisfies the differential inequality

$$L[\psi](x) > 0, x \in I. \tag{10}$$

It follows from the definition of generalized convexity that $\psi(x) < 0, x \in I$, therefore we obtain from (9) and (10) that $A(x) < 0, x \in I$. Let us put

$$g(x) := \frac{\alpha(x)}{\beta(x)} - \lambda(x), x \in [x_1, x_2]. \tag{11}$$

Since λ is differentiable in I, in view of (i) and (v), we have

$$\begin{aligned} g'(x) &= \frac{-\alpha'(x)\beta(x) + \alpha(x)\beta'(x)}{[\beta(x)]} - \lambda'(x) \\ &= \frac{A(x)}{[\beta(x)]^2} - \lambda(x) \end{aligned} \tag{12}$$

$x \in I$.

Hence $g'(x) < 0$ for $x \in [x_1, x_2]$, by virtue of negativity of A and hypothesis (v). Therefore the function g is strictly decreasing in $[x_1, x_2]$, whence there may exist at most one point $x_0 \in [x_1, x_2]$ such that $g(x_0) = 0$, i.e., such that $\psi'(x_0) = 0$. And such a point actually exists what has already been proved.

Now let ψ be a twice differentiable function on I and let for every two points $x_1, x_2 \in I, x_1 < x_2$, there exists exactly one point $x_0 \in [x_1, x_2]$ such that inequality (3) holds, for $\varphi \in F$ satisfying (1). Further let us assume that ψ is neither strictly convex nor strictly concave with respect to F. Thus there exist $x_1, x_2 \in I, x_1 < x_2$ and $t_1, t_2 \in (x_1, x_2), t_1 < t_2$, such that

$$\psi(t_1) < \varphi(t_1), \psi(t_2) > \varphi(t_2), \tag{13}$$

where $\varphi \in F$ satisfies (1). It follows from the continuity of ψ and φ that there is a $t_0 \in (t_1, t_2)$ such that $\psi(t_0) = \varphi(t_0)$. Denote

$$\begin{aligned} t_3 &:= \inf \{t : t_1 < t \leq t_0, \psi(t) = \varphi(t)\}, \\ t_4 &:= \sup \{t : x_1 \leq t < t_1, \psi(t) = \varphi(t)\}. \end{aligned}$$

It follows from (1) and (13) that $t_1 < t_3 \leq t_0$ and $x_1 \leq t_4 < t_1$, whence $t_4 < t_3 \leq t_0$. Since $\varphi, \psi \in C^1(I)$, we have $\psi'(t_3) \geq \varphi'(t_3), \psi'(t_4) \leq \varphi'(t_4)$, thus there exists an $x_3 \in (x_1, t_0)$ such that $\psi'(x_3) = \varphi'(x_3)$, by virtue of the Darboux property for ψ' and φ'. Similarly we can prove that there exists an $x_4 \in (t_0, x_2)$ such that $\psi'(x_4) = \varphi'(x_4)$. Since $x_3 \neq x_4$, it contradicts the assumption that there exists a unique point such that (3) holds. This contradiction ends that proof of the theorem.

Let us consider a second order homogeneous equation

$$L[y] := y'' + p(x)y' + q(x)y = 0\,, x \in I \tag{14}$$

and the inequalities

$$L[\psi] > 0\,, x \in I \tag{15}$$

$$L[\psi] < 0\,, x \in I\,, \tag{16}$$

where ψ is a twice differentiable function on I. E. F. Bonsall [2] proved that ψ is a solution of (15) (resp. (16)) iff it is strictly convex (resp. concave function) with respect to the family F of all solutions of (14). As a simple application of Theorem 1 and that of Bonsall we have the following

Theorem 2. Let p and $q \geq 0$ be continuous functions on I and let F be the family of all solutions of (14). Moreover, let hypotheses (i)–(iv) be fulfilled.

A function ψ, twice differentiable in I, is a solution of either (15) or (16) if and only if for every $x_1, x_2 \in I, x_1 < x_2$, there exists a unique point $x_0 \in (x_1, x_2)$ such that (3) is fulfilled, where φ is a solution of (14) satisfying (1).

Proof. Let ψ be a twice differentiable function on I. We are going to prove that hypothesis (v) is also fulfilled. Indeed, it is easy to calculate the derivative of the function λ:

$$\lambda'(x) = q(x)\frac{W(x)}{[u(x)]^2} \text{ for } x \in I\,.$$

Therefore λ is an increasing function in I and Theorem 2 is a simple consequence of Theorem 1.

Let us observe that if a function ψ is a solution of either

$$L[\psi] + r(x) > 0 \text{ for } x \in I \tag{17}$$

or

$$L[\psi] + r(x) < 0 \text{ for } x \in I, \tag{18}$$

where r is a given function, then taking an arbitrary fixed solution y_0 of equation

$$L[y] + r(x) = 0 \text{ for } x \in I, \tag{19}$$

we have $\psi(x) = \psi_0(x) + y_0(x), x \in I$, where ψ_0 is an arbitrary solution of (15), when ψ satisfies (17), and ψ_0 is an arbitrary solution of (16), when ψ satisfies (18). Therefore, as an easy consequence of Theorem 2, we have the following

Corollary. Let $p, q \geq 0$ and r be continuous functions in I and let the family F of all solutions of Eq. (14) satisfy hypotheses (i)–(iv).

A function ψ, twice differentiable in I, is a solution of either (13) or (14) if and only if for every $x_1, x_2 \in I, x_1 < x_2$, there exists a unique point $x_0 \in (x_1, x_2)$ such that equality (3) holds, where φ is a solution of (15) satisfying (1).

References

1. E. F. Beckenbach, *Generalized convex functions*, Bull. Amer. Math. Soc. **43** (1937), 363–371.
2. E. F. Bonsall, *The characterization of generalized convex functions*, Quart. J. Math. 1 (1950), 100–111.
3. D. Brydak, *Applications of generalized convex functions to second order differential inequalities*, General Inequalities 4, ISNM 71, Birkhäuser Verlag (1984), 297–305.

Dobiesław Brydak
Pigonia 4/4
31-228 Kraków
Poland

Constantin Carathéodory: An International Tribute (pp. 97–118)
edited by Th. M. Rassias

ON THE EIGENVALUE PROBLEM FOR QUASILINEAR ELLIPTIC OPERATORS

Raffaele Chiappinelli

Abstract. We study the existence of eigenfunctions of arbitrary fixed L^2 norm for quasilinear elliptic operators with odd coefficients, on a bounded domain of $\mathbb{R}^N$, in absence of global coercivity of the corresponding functional.

1. Introduction.

Let Ω be a bounded open subset of $\mathbb{R}^N$ with smooth boundary $\partial\Omega$. We consider the quasilinear eigenvalue problem

$$(1.1)\qquad \begin{cases} -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} a_i(x, u, \nabla u) + a_0(x, u, \nabla u) = \mu u & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

where each $a_i = a_i(x, t, p)$ $(i = 0, 1, \ldots, N)$ is a Caratheodory function on $\Omega \times \mathbb{R} \times \mathbb{R}^N$ (ie. measurable in x for each $(t, p) \in \mathbb{R} \times \mathbb{R}^N$ and continuous in (t, p) for a.a. $x \in \Omega$). It is our aim to establish in this note, by means of the critical point theory of Liusternik and Schnirelmann, the existence of infinitely many distinct eigenfunctions for (1.1) under mild coercivity assumptions on the coefficients a_i. More precisely, we shall give conditions (which for linear operators reduce to boundedness and uniform ellipticity of the coefficients) ensuring that for each $r > 0$ there exist eigenfunction-eigenvalue pairs $(u_n(r), \mu_n(r))$, $n = 1, 2, \ldots,$

solving (1.1) in the generalised sense, with $\int_\Omega u_n^2(r) = r^2$ for each n while

$$\int_\Omega |\nabla u_n(r)|^2 \to \infty, \qquad \mu_n(r) \to +\infty \qquad (n \to \infty).$$

Throughout this paper, the a_i are assumed to satisfy the following conditions for a.a. $x \in \Omega$ and all $(t,p) \in \mathbb{R} \times \mathbb{R}^N$:

H1) *(Growth)*

$$|a_i(x,t,p)| \le c(|t|^r + |p|) + d \qquad (i = 1, \dots, N)$$
$$|a_0(x,t,p)| \le c(|t|^s + |p|^q) + d$$

where (assuming for simplicity $N \ge 3$) $0 < r < \frac{N}{N-2}$, $0 < s < \frac{N+2}{N-2}$, $0 < q < \frac{N+2}{N}$. Here and elsewhere c, d denote unspecified positive constants.

H2) *(Monotonicity)*

$$\sum_{i=1}^N |a_i(x,t,p) - a_i(x,t,p')||p_i - p_i'| > 0 \qquad (p \ne p')$$

H3) *(Ellipticity)*

$$\sum_{i=1}^N a_i(x,t,p)p_i \ge \nu|p|^2 - d \qquad (\nu > 0).$$

Let $H_0^1(\Omega)$ denote the closure in $H^1(\Omega)$ (the usual first Sobolev space over Ω) of $C_0^\infty(\Omega)$, the family of all smooth functions having compact support in Ω. We equip $H_0^1(\Omega)$ with the scalar product and norm

$$(u,v) = \int_\Omega \nabla u \cdot \nabla v \quad , \quad \|u\|^2 = \int_\Omega |\nabla u|^2.$$

An eigenfunction-eigenvalue pair (an eigenpair for short) of (1.1) is a pair (u,μ) with $u \in H_0^1(\Omega)$, $u \ne 0$ and $\mu \in \mathbb{R}$ which solves (1.1) in the variational sense, ie.

$$a(u,v) := \sum_{i=1}^{N} \int a_i(x,u,\nabla u)\frac{\partial v}{\partial x_i} + \int a_0(x,u,\nabla u)v = \mu \int uv \tag{1.2}$$

for all $v \in H_0^1(\Omega)$ (integrals are over Ω, unless otherwise stated). $a(u,v)$ is the generalized Dirichlet form associated with the quasilinear operator in (1.1); under the above assumption H1) it is well defined for all $u, v \in H_0^1(\Omega)$ and satisfies an inequality of the form

$$|a(u,v)| \le c(\|u\|^\gamma + 1)\|v\|, \qquad \gamma = \max\{r,s,q\} \tag{1.3}$$

with $c > 0$. This is a straightforward consequence of H1) via Hölder's inequality and the Sobolev embedding theorem $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$, $p \le \frac{2N}{N-2}$; see e.g. ([2], Lemma 3).

We further assume that the operator in (1.1) is an Euler-Lagrange operator of the calculus of variations, ie. there exists a map $F = F(x,t,p)$ defined over $\Omega \times \mathbb{R} \times \mathbb{R}^N$ and of class C^1 in (t,p) in the Caratheodory sense, such that

$$a_0 = \frac{\partial F}{\partial t}, \qquad a_i = \frac{\partial F}{\partial p_i} \qquad (i = 1, \dots, N).$$

Assuming F obeys growth restrictions similar to those in H1), it is then well-known (e.g. [2], p. 35) that setting

$$f(u) := \int F(x,u,\nabla u)$$

f is of class C^1 on $X := H_0^1(\Omega)$ and

$$f'(u)v = a(u,v) \qquad (u,v \in X). \tag{1.4}$$

Here $f'(u)$ denotes the derivative of f at the point u, which is then a bounded linear form on X, and $f'(u)v$ is the value of $f'(u)$ at the point $v \in X$. On the other hand, if $g(u) := \frac{1}{2}\int u^2$, then $g'(u)v = \int uv$ so that (1.2) can be rewritten $f'(u)v = \mu g'(u)v$ $(v \in X)$, ie.

$$f'(u) = \mu g'(u). \tag{1.5}$$

Therefore, if we add the normalization condition $g(u) = const.$, our original problem consists (in its weak form) in finding the constrained critical points of f over the manifold $N = \{u \in X : g(u) = c\}$ in X, the eigenvalues appearing as Lagrange multipliers: see e.g. the discussion in section 2. Equivalently, adding the side condition $f(u) = const$, one would look for critical points of g over $M = \{u \in X : f(u) = c\}$.

If we further assume that, for $(x, t, p) \in \Omega \times \mathbb{R} \times \mathbb{R}^N$,

$$a_i(x, -t, -p) = -a_i(x, t, p) \qquad (i = 0, 1, \ldots, N)$$

then f is an even functional (like evidently g) and this in turn allows, under further technical assumptions to be specified below, to make use of Liusternik-Schnirelmann's (LS) theory and thereby prove the existence of infinitely many critical points.

In doing this, one has evidently in mind the corresponding classical results concerning the linear uniformly elliptic eigenvalue problem

$$\text{(1.6)} \qquad \begin{cases} -\sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) + a_0(x)u = \mu u & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

having L^∞ coefficients $a_{ij} = a_{ji}$ $(i, j = 1, \ldots, N)$ and a_0, whose eigenvalues (forming an infinite sequence tending to $+\infty$) and corresponding eigenfunctions are related, by the Courant-Weyl principle, to minimax (or maximin) of the quadratic functionals associated with (1.6) over appropriate subsets of X.

The above program has been thoroughly developed by Browder in [2], in the very general context of eigenvalue problems of the type

$$\sum_{|\alpha| \le m} (-1)^{|\alpha|} D^\alpha A_\alpha(x, u, \ldots, D^m u) = \mu \sum_{|\beta| \le k} (-1)^{|\beta|} D^\beta B_\beta(x, u, \ldots, D^k u)$$

involving quasilinear operators of arbitrary even order $2k < 2m$ in generalized divergence form, considered together with the boundary conditions given implicity by a closed subspace V of the Sobolev space $H^m(\Omega)$, with $H_0^m(\Omega) \subset V \subset H^m(\Omega)$.

We concentrate for simplicity on the specific eigenvalue problem (1.1), for which Browder's result ([2], Thm. 23) reads as follows:

THEOREM 1. *Assume that, for a.a.* $x \in \Omega$ *and all* $(t, p) \in \mathbb{R} \times \mathbb{R}^N$,

H4) $$a_0(x,t,p)t \geq -c_0 t^2 - e|p|^2 - d$$

with $0 \leq e < \nu$ *and* $0 \leq c_0 < \lambda_1(\nu - e)$, *where* ν *is as in H3) and* λ_1 *is the first eigenvalue of* $-\Delta u = \lambda u$ *in* Ω, $u = 0$ *on* $\partial\Omega$.

Then, for sufficiently large c, *there exists an infinite sequence* $(u_n(c), \mu_n(c))$ *of eigenpairs of (1.1), normalized by* $f(u_n) = c$ *for all* n. *These are characterized by* $g(u_n) = c_n$, *where*

(1.7) $$c_n = \sup_{H_n} \inf_{H} g(u).$$

In (1.7), H_n is the family of all compact, symmetric subsets H of $M_c = \{u \in X : f(u) = c\}$ with $\gamma(H) \geq n$, where $\gamma(H)$ denotes the genus of H (Actually, in [2] the equivalent notion of category of H is used: see [7] for a proof of the equivalence).

In [2], the LS principle is used as follows. Assumption H4) implies that

$$\int a_0(x,u,\nabla u)u \geq -c_0 \int u^2 - \varepsilon \int |\nabla u|^2 - d' \geq -\left(\frac{c_0}{\lambda_1} + \varepsilon\right) \int |\nabla u|^2 - d'$$

where we have used the Poincarè inequality $\int |\nabla u|^2 \geq \lambda_1 \int u^2$. Therefore, the Dirichlet form $a(u,v)$ in (1.2) satisfies, taking into account H3),

$$a(u) := a(u,u) = \sum_{i=1}^{N} \int a_i(x,u,\nabla u)\frac{\partial u}{\partial x_i} + \int a_0(x,u,\nabla u)u$$

$$\geq \left(\nu - \frac{c_0}{\lambda_1} - \varepsilon\right) \int |\nabla u|^2 - d''.$$

In view of H4), we have in conclusion for all $u \in X$

(1.8) $$a(u) \geq c\|u\|^2 - d, \qquad c > 0$$

so that $a(u) \to +\infty$ if $\|u\| \to \infty$; in other words, a is coercive on X. This implies ([2], p. 33) that $f(u) = \int F(x,u,\nabla u)$ is coercive on X, too. Then for each c the manifold $M_c = \{u \in X : f(u) = c\}$ is bounded; this last fact is then crucially employed, together with (1.8) above, when showing that g^{-1} satisfies (for c large) the compactness condition of

Palais-Smale on M_c. One can then conclude that the numbers c_n defined in (1.7) are critical levels of g^{-1} on M_c, which in fact implies that there exists (u_n, μ_n) with $f'(u_n) = \mu_n g'(u_n)$, $u_n \in M_c$ and $g(u_n) = c_n$.

It should be noted that the coercivity of a is used again in order to show ([2], p. 49-50) that for c large M_c is starlike (ie. diffeormorphic to the unit sphere in X) and therefore contains subsets of arbitrary genus (ie. $H_n \neq \phi$ in the above notation for all n).

We have to mention that the coercivity assumption in [2] has been here specialized to the present context of $H_0^1(\Omega)$, corresponding to zero Dirichlet boundary conditions in (1.1), on separating for convenience in H3) and H4) the conditions on the top order coefficients from those on a_0. The original assumption for a general subspace V of $H^1(\Omega)$ would read (see e.g. eq. (18) in [2])

$$\sum_{i=1}^{N} a_i(x,t,p)p_i + a_0(x,t,p)t \geq \nu(|p|^2 + t^2) - d.$$

As remarked above, a result similar to Theorem 1 holds for a much wider class of problems, and even in the context of second order operators acting in $H_0^1(\Omega)$, it allows to consider more general second members $\mu b_0(x,u)$ rather than merely μu. However, when considered for the specific equation (1.1), the above statement shows some inconvenients which we now describe. These become more apparent when looking at the linear and semilinear counterpart of (1.1), ie. (1.6) and

$$\text{(1.9)} \qquad \begin{cases} -\sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left(a_{i,j}(x)\frac{\partial}{\partial x_i}\right) + a_0(x,u) = \mu u & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

We first remark that no conclusion of the kind concerning (1.6) about the asymptotic behaviour of u_n and μ_n as $n \to \infty$ is explicity stated.

Let us next concentrate on the coercivity assumption H4). When referred to (1.6), this becomes

$$a_0(x) \geq -c_0 \quad , \quad c_0 < \lambda, \nu$$

and is for instance satisfied on imposing a smallness condition on $\|a_0\|_{L^\infty(\Omega)}$. But the existence of infinitely many eigenpairs (u_n^0, μ_n^0) is here granted whatever $a_0 \in L^\infty(\Omega)$, as classical spectral theory shows.

Likerwise, any simple semilinear problem such as

$$\begin{cases} -\Delta u - |u|^{\alpha} u = \mu u & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{1.10}$$

is not covered by the above theory if $\alpha > 0$. Indeed, the corresponding functional

$$a(u) = \int |\nabla u|^2 - \int |u|^{\alpha+2}$$

is not coercive on $H_0^1(\Omega)$, for if u_0 is any fixed vector in $H_0^1(\Omega)$,

$$a(tu_0) = t^2 \int |\nabla u_0|^2 - t^{\alpha+2} \int |u_0|^{\alpha+2} \qquad (t > 0)$$

so that $a(tu_0) \to -\infty$ as $t \to +\infty$ if $u_0 \neq 0$. Note that here

$$a_0(x, t, p)t = a_0(x, t)t = -|t|^{\alpha+2}$$

which violates condition H4). Nevertheless, (1.10) does possess countably many eigenvalues (for restricted α) as shown for instance in [4].

Finally, the restriction "c large" for the existence of eigenfunctions on the level set $f(u) = c$ appears to be rather severe, both intrinsically (as referred to linear theory again) and from a different viewpoint. In fact, if in (1.9) the nonlinearity $a_0(x, u)$ is small enough (in a sense to be made precise) and $a_0(x, 0) = 0$, one expects bifurcation to take place from the eigenvalues of (1.6): see e.g. Theorem 4.3 of [8] for a typical result of this kind, though in an abstract context. However, if $f(u) \leq c\|u\|^{\delta} + d$ $(\delta > 0)$, as will be the case under the assumptions above, then studying eigenfunctions lying on $f(u) = c$ with c large implies considering only solutions of large norm, which evidently rules out any bifurcation analysis of (1.9). We refer again to [4] for an investigation of this kind.

To overcome these difficulties, we propose a simple alternative approach to the LS analysis of (1.1). Namely, rather than considering "maximin" of g over $M = \{u : f(u) = const\}$, we study "minimax" of f over $N = \{u : g(u) = const\}$. Note that N is an unbounded submanifold of X; nevertheless, it is plainly diffeomorphic to the unit sphere in X for each $c > 0$, and this permits to use LS theory for f over N once the corresponding Palais-Smale condition is established. But this can be done on relaxing significantly the coercivity condition H4). The outcome about (1.1) is indeed as follows:

THEOREM 2. *Assume that*

H5) $$a_0(x,t,p)t \geq -c|t|^{\sigma+1} - \varepsilon|p|^2 - d$$

where $0 \leq \varepsilon < \nu$, $0 \leq c$ *is arbitrary, and* $\sigma < 1 + \frac{4}{N}$. *Then for each* $r > 0$, *(1.1) has countably many eigenpairs* $(u_n(r), \mu_n(r))$ *such that* $\int u_n^2(r) = r^2$ $(n = 1, 2, \ldots)$; *moreover,* $\mu_n(r) \to^+ \infty$ *and* $\|u_n(r)\| \to \infty$ *as* $n \to \infty$, *for each* $r > 0$.

Here the eigenfunctions $u_n(r)$ are characterized as critical points of f on

$$N_r = \left\{ u \in X : \int u^2 = r^2 \right\} = \left\{ u \in X : g(u) = \frac{r^2}{2} \right\}$$

associated with the critical values

(1.11) $$c_n(r) = \inf_{K_n(r)} \sup_K f(u)$$

where

$$K_n(r) = \{K \subset N_r : K \text{ compact, symmetric, } \gamma(K) \geq n\}.$$

Note that the semilinear problem (1.10) is covered by H5) as long as $\alpha < \frac{4}{N}$. On the other hand, in the linear problem (1.6) the above condition is satisfied with $\sigma = 1$ and $c = \|a_0\|_{L^\infty(\Omega)}$. Furthermore, using homogeneity it is immediate to check that if H3) holds with some d then it holds with $d = 0$; H2) is then an obvious consequence of this. In conclusion, our hypotheses reduce to the classical boundedness and uniform ellipticity conditions for the coefficients of (formally selfadjoint) second order operators.

We shall see in Proposition 4.4 that under H5) the functional a is coercive on N_r (for each $r > 0$) in the sense that

(1.12) $$a(u) \to +\infty \qquad as \qquad \|u\| \to \infty, \quad u \in N_r$$

a similar property being then enjoyed by f. It is easily seen that the requirement that a be coercive on N_r is equivalent to the property:

(1.13) $$N_r^c := \{u \in N_r : a(u) \leq c\} \quad \textit{is bounded for each } c.$$

The above result will be proved as a consequence of a more general theorem which can be seen as an abstract version of the present way to apply the LS principle. Note that (1.5) can be written

(1.14) $$Au = \mu Bu$$

where $A, B : X \to X$ are defined by duality as follows:

$$(Au, v) = a(u, v) \qquad (Bu, v) = \int uv \qquad (u, v \in X).$$

It is shown in [2] (Lemmas 3 and 7) that A is countinuous and bounded on bounded subsets while B is strongly countinuous (see below). We then have from (1.4)

$$(Au, v) = f'(u)v \qquad (Bu, v) = g'(u)v \qquad (u, v \in X)$$

with $f(u) = \int F(x, u, \nabla u)$ and $g(u) = \frac{1}{2} \int u^2$. This is expressed by saying that A and B are gradient mappings in X with potentials f, g respectively.

Let us now consider in general a real Hilbert space and two mappings A, B in X for which we recall the following definitions:

- A is of type $(S)_1$ [1] if $x_n \rightharpoonup x$ and $Ax_n \to y$ imply $x_n \to x$ ($\rightharpoonup$ and $\to$ denote weak and strong convergence in X respectively);

- B is strongly (sequentially) continuous ("completely continuous" in [2]) if $x_n \rightharpoonup x$ implies $Bx_n \to Bx$.

We shall assume without further mention that all operators appearing in the sequel are continuous and bounded on bounded subsets of X. Note however that both these conditions are implied by strong continuity: in fact in this case the operator is evidently continuous and is also compact, ie. maps bounded subsets of X onto relatively compact subsets of X, as is easily checked.

We also recall for further reference that, for a strongly continuous gradient mapping B, the corresponding potential g is weakly (sequentially) continuous, ie. $g(u_n) \to g(u)$ whenever $u_n \rightharpoonup u$ in X; see e.g. [1], p. 61.

Finally, adding a constant if necessary, we may and shall assume that the potential f of any gradient operator is normalized by $f(0) = 0$.

THEOREM 3. *Let X be a real, infinitedimensional, separable Hilbert space and let $A, B : X \to X$ be odd gradient mappings with potentials f, g respectively. Assume $(Bu, u) > 0$ for $u \neq 0$. Then for each $r > 0$ the set $N_r = \{u \in X : g(u) = r\}$ is a C^1 submanifold of X. Assume further:*

i) A is of type $(S)_1$; f is bounded below and coercive on N_r.

ii) B is strongly continuous; for each $u \neq 0$, $g(tu) \to +\infty$ as $t \to +\infty$.

Then the eigenvalue problem (1.14) *has countably many distinct eigenpairs* (u_n, μ_n) *with* $u_n \in N_r$, *ie.* $g(u_n) = r$ *for all* n. *If moreover*

iii) $f(u) \to +\infty$ *implies* $\|u\| \to \infty$;

iv) $a(u) := (Au, u)$ *is coercive on* N_r;

v) (Bu, u) *is bounded above on* N_r,

then $\mu_n \to +\infty$ *and* $\|u_n\| \to \infty$ *as* $n \to \infty$.

The proof of Theorem 3 is postponed to Section 3, after recalling a suitable version of the LS principle. In Section 2 we shall discuss for further reference some useful facts about the Palais-Smale condition. Finally in Section 4 we concentrate again on the concrete quasilinear eigenvalue problem (1.1), on showing that under the given assumptions on the coefficients a_i the above general results applies, thereby proving Theorem 2.

2. Some remarks on the Palais-Smale condition.

Let f be a C^1 functional on X. Given a C^1 submanifold M in X with tangent space $T_x(M)$ at $x \in M$, the derivative of $f_M := f_{|M}$ at $x \in M$, denoted $f'_M(x)$, is just the restriction of $f'(x)$ to $T_x(M)$; and a critical point of f_M is a point where $f'_M(x) = f'(x)_{|T_x(M)} = 0$. The search for critical points of a functional on a manifold rests heavily on the following compactness property:

DEFINITION 2.1. *f is said to satisfy the Palais-Smale (PS) condition on M if any sequence $x_n \subset M$ along which $f(x_n)$ is bounded and $f'_M(x_n) \to 0$ contains a convergent subsequence.*

Note that $f'_M(x_n) \to 0$ means $\|f'_M(x_n)\|_{Z_n} \to 0$, $Z_n = (T_{x_n}(M))^*$. We shall now derive in detail some more viable criterion to verify the PS condition: compare e.g. Section 6 of [3].

LEMMA 2.2. *Let X be a Hilbert space, M a closed linear subspace of X, and let P be any liner bounded projection onto M (ie. $P^2 = P$, $P(X) = M$). Given $F \in X^*$, let a be the unique vector in X such that $F(x) = (x, a)$, $x \in X$, and let $f = F_{|M}$ be the restriction of F to M. Then if Q denotes the adjoint to P, we have*

$$f(v) = (v, Qa) \qquad (v \in M) \tag{2.1}$$

and

$$\|P\|^{-1}\|Qa\| \leq \|f\| \leq \|Qa\| \tag{2.2}$$

where $\|f\| = \sup\{|f(v)|/\|v\|, v \in M, v \neq 0\}$ *is the norm of* f *in* M^* *while* $\|P\|$ *is the operator norm of* P. *Note in particular that* $f = 0$ *iff* $Qa = 0$ *and* $\|f\| = \|P_0 a\|$ *if* P_0 *is the orthogonal projection onto* M *(recall* $Q_0 = P_0$ *and* $\|P_0\| = 1$ *in this case).*

PROOF.

i) For each P, $FP := F \circ P \in X^*$ and so there exists a unique $a_p \in X$ such that $F(Px) = (x, a_p), x \in X$; moreover, $\|FP\| = \|a_p\|$. But

$$F(P_x) = (Px, a) = (x, Qa) \qquad (x \in X)$$

which gives (2.1) (recall $x = Px$ iff $x \in M$) and shows that $a_p = Qa$, whence $\|FP\| = \|Qa\|$.

ii) For $v \in M$, $|f(v)| = |F(Pv)| \leq \|FP\|\|v\|$, so we get at once $\|f\| \leq \|FP\|$.

As for the first inequality in (2.2), we note it is trivial if $f = 0$; indeed, this is equivalent to $FP = 0$ and so $\|a_p\| = \|Qa\| = 0$. Assume then $FP \neq 0$; given $\varepsilon > 0$, let $x_0 \neq 0$ be such that $\|F(Px_0)\| > (\|FP\| - \varepsilon)\|x_0\|$. Note this implies $PX_0 \neq 0$.

As $\|Px\| \leq \|P\|\|x\|$ for all $x \in X$, we then have $\|F(Px_0)\| > c_\varepsilon \|Px_0\|$ with $c_\varepsilon = (\|FP\| - \varepsilon)\|P\|^{-1}$; therefore, $|f(v_0)| > c_\varepsilon \|v_0\|$ with $v_0 = Px_0 \neq 0$, whence $\|f\| > c_\varepsilon$, which gives the result on letting $\varepsilon \to 0$.

COROLLARY 2.3. *Let* f *be a* C^1 *functional on* X *and let* Ax *denote the gradient of* f *at* x, *ie.* $(Ax, v) = f'(x)v$, $v \in X$. *Let* M *be a* C^1 *manifold in* X *with tangent space* $T_x(M)$ *at* $x \in M$. *Given any bounded linear projection* P_x *of* X *onto* $T_x(M)$, *we have for the derivative* f'_M *of* f *on* M

$$f'_M(x)v = (v, Q_x A_X) \qquad (v \in T_x(M)) \tag{2.3}$$

where Q_x *is the adjoint to* P_x; *moreover,*

$$\|P_x\|^{-1}\|Q_x A_X\| \leq \|P_x^0 Ax\| = \|f'_M(x)\| \leq \|Q_x Ax\| \tag{2.4}$$

where $\|f'_M(x)\|$ *is the norm of* $f'_M(x)$ *in* $(T_x(M))^*$ *and* P_x^0 *is the orthogonal projection onto* $T_x(M)$.

In particular, $x \in M$ *is a critical point of* f_M *iff* $Q_x Ax = 0$.

Assume now that M is the level set of another C^1 functional g on X: $M = M_c(g) = \{x \in X : g(x) = c\}$, c a real constant. If c is not a critical value of g, ie. $g'(x) \neq 0$ for $x \in M_c(g)$, then $M_c(g)$ is indeed a C^1 submanifold of X of codimension 1, whose tangent space at a point x is $T_x(M) = \{v \in X : g'(x)v = 0\}$, ie.

$$T_x(M) = \{v \in X : (Bx, v) = 0\}$$

with B denoting the gradient of g. It is then useful to recall the following:

LEMMA 2.4. *Let a, b be fixed vectors in X with $(a,b) \neq 0$. Then any $u \in X$ has a unique decomposition $u = u_1 + u_2$ with $(u_1, a) = 0$ and $u_2 = c\, b$, $c \in \mathbb{R}$.*

PROOF. Assume $u = u_1 + c\, b$. Then $u_1 = u - c\, b$ and $(u_1, a) = (u - c\, b, a) = 0$, whence $c = \frac{(u,a)}{(a,b)}$. Therefore,

$$u = u - \frac{(u,a)}{(a,b)} b + \frac{(u,a)}{(u,b)} b.$$

If we set

$$P\, u = u - \frac{(u,a)}{(a,b)} b \qquad Q\, u = u - \frac{(u,b)}{(a,b)} a \qquad (u \in X) \tag{2.5}$$

it is then easy to check that P, Q are linear bounded projections onto $a^\perp$, $b^\perp$ respectively and $(Pu, v) = (u, Qv)$ for all $u, v \in X$, so that Q is the adjoint of P. Evidently, if $a = b$ then $P = Q =$ orthogonal projection onto $a^\perp$.

Therefore, if $M = \{x \in X : g(x) = c\}$ and $C : X \to X$ is any map satisfying $(Bx, Cx) \neq 0$ for $x \neq 0$, then setting

$$P_x^C(u) = u - \frac{(u, Bx)}{(Bx, Cx)} Cx \qquad (u \in X) \tag{2.6}$$

we obtain for each $x \neq 0$ a projection onto $(Bx)^\perp = \{v \in X : (Bx, v) = 0\} = T_x(M)$, and the orthogonal projection P_x^0 is obtained taking $C = B$ (we assume here and henceforth $Bx \neq 0$ for all $x \neq 0$: this implies in particular that $M_c(g)$ is a regular submanifold for each $c \neq 0$ if $g = (0)$.

In conjunction with Corollary 2.3, we then have the explicit representations for $f'_M(x)$:

$$Q_x^C(Ax) = Ax - \frac{(Ax, Cx)}{(Bx, Cx)} Bx =: A_C(x) \tag{2.7}$$

and in particular

$$P_x^0(Ax) = Ax - \frac{(Ax, Bx)}{\|Bx\|^2} Bx =: A_0(x) \tag{2.8}$$

in the sense that (see 2.3)

$$f'_M(x)u = (A_C(x), u) \qquad (u \in T_x(M)).$$

For this reason, A_C may be called the gradient of f_M (along C). Critical points of f_M are then characterized as zeros of A_C, ie. points where

$$Ax = \frac{(Ax, Cx)}{(Bx, Cx)} Bx$$

Note this is the same as to say $Ax = \mu Bx$ (ie. $f'(x) = \mu g'(x)$ in X^*) for some real μ, because then taking inner products we see that μ is necessarily equal to $\frac{(Ax,Cx)}{(Bx,Cx)}$.

The relevant point in connection with the (PS) condition defined in (2.1) is the estimate

$$\|P_x^C\|^{-1}\|A_C(x)\| \le \|A_0(x)\| = \|f'_M(x)\| \le \|A_C(x)\| \tag{2.9}$$

which follows from (2.4) via the above definitions of A_0, A_C. Indeed, the first shows that "$f'_M(x_n) \to 0$" in the statement of (PS) can be replaced by "$A_0(x_n) \to 0$". However, in most instances it is useful to translate the (PS) condition in terms of A_C rather than A_0; to this purpose, we note that from the definition (2.6) of P_x^C,

$$\|P_x^C(u)\| \le \left(1 + \frac{\|Bx\|\|Cx\|}{|(Bx, Cx)|}\right)\|u\|, \qquad u \in X.$$

Therefore, if $|(Bx, Cx)| \ge d > 0$ on bounded subsets of M then (assuming as always B, C bounded on bounded sets) it follows that for each bounded subset $K \subset M$, there exists $c > 0$ so that $\|P_x^C\| \le c$ $(x \in K)$.

We can now conclude these remarks with the following criterion:

PROPOSITION 2.5. *Assume f coercive on M and (Bx, Cx) bounded away from 0 on each bounded subset of M. Suppose that any sequence $(x_n) \subset M$ such that $f(x_n)$ is bounded and $A_C(x_n) \to 0$ contains a convergent subsequence. Then f satisfies (PS) on M.*

PROOF. Assume $f(x_n)$ bounded and $A_0(x_n) \to 0$. Then (x_n) is bounded by the coercivity assumption on f. By the above remark, $\|P_{x_n}^C\| \le c$ and so $A_C(x_n) \to 0$, because $\|A_C(x_n)\| \le c\,\|A_0(x_n)\|$ by (2.9) above. The conclusion now follows.

3. The Liusternik-Schnirelmann principle. Proof of Theorem 3.

Let us recall that given a closed, symmetric (with respect to the origin) subset A of X with $0 \notin A$, the genus of A, written $\gamma(A)$, is defined as

$$\gamma(A) = \min\{n \in \mathbb{N} : \textit{there exists a continuous odd map } h : A \to \mathbb{R}^n \setminus \{0\}\}.$$

If there is no such number, we put $\gamma(A) = \infty$. Furthermore, we set $\gamma(\phi) = 0$. We refer to e.g. [7] or [8] for the properties of γ and bound ourselves to recall the following:

(3.1) $\gamma(A) = \gamma(B)$ if there exists an odd homeorphism of A onto B;

(3.2) If $\gamma(A) > k$, V is a k-dimensional subspace of X and $V^{\perp}$ is a topological supplement to V, then $A \cap V^{\perp} \neq \phi$.

We are now in a position to state the LS principle in a form suited to our context. Due to the cited equivalence [7] between LS category and the genus defined above, this is just a special version of Theorem 20 in [2].

THEOREM 4. *Let X be a real infinitedimensional Hilbert space, let f, g be two even C^1 functionals on X, and let $M = M_c(g)$ be the level set $\{u \in X : g(u) = c\}$ with $0 \notin M$. Assume:*

i) $g'(u)u > 0$ for $u \in M$, so that in particular M is a C^1 submanifold of X;

ii) M is starlike, ie. each ray through the origin hits M in exactly one point;

iii) f is bounded below on M and satisfies (PS) on M.

For $n = 1, 2, \ldots$ set

$$c_n = \inf_{K_n} \sup_{K} f(u) \tag{3.3}$$

where $K_n = \{K \subset M$: K compact, symmetric, $\gamma(K) \geq n\}$. Then for each n there exist $u_n \in M$ and $\mu_n \in \mathbb{R}$ such that $f(u_n) = c_n$ and $f'(u_n) = \mu_n g'(u_n)$.

REMARK 3.1. We note that condition ii) is essential in two ways. First, it is shown in Browder ([2], Theorem 19) that in a Banach space having C^1 norm on its unit sphere S, a starlike manifold M is C^1 diffeomorphic to S via the radial projection $p(u) = \frac{u}{\|u\|}(u \neq 0)$;

this allows to "transplant" (retaining the PS condition) the original problem from M to S, ie. to a smooth manifold if X has smooth norm, as in the present Hilbert space case. Recall the original requirement for the LS principle to hold is that M be of class C^2; see [2] for the complete discussion.

Furthermore, if X_n is any n-dimensional subspace of X, then $\gamma(S \cap X_n) = n$, a consequence of the Borsuk-Ulam theorem (e.g. [7], p. 180); as p is evidently an odd homeomorphism of M onto S and $p(M \cap X_n) = S \cap X_n$, it follows from (3.1) above that $K_n \neq \phi$ for all $n \in \mathbb{N}$, so that the definition (3.3) of c_n is meaningful for each n.

PROOF OF THEOREM 3.

a) We begin by showing that under the assumptions of this Theorem, the functionals f, g related to A, B satisfy the requirements of the LS principle, Theorem 4 above. First, f and g are even since A, B are odd. Next, since $g(0) = 0$ and $(Bu, u) > 0$ for $u \neq 0$, the assumptions $0 \notin N_r = \{u : g(u) = r\}$ and $g'(u)u > 0$ for $u \in N_r$ are satisfied for each $r > 0$.

Likewise, let us prove that N_r is starlike for each $r > 0$. Let $u \in X$, $u \neq 0$; we have to show that there exists a unique $t > 0$ such that $tu \neq N_r$, ie. $g(tu) = r$. To do this, we werely consider the map $\varphi(t) := g(tu)$ and observe that, since $(Bu, u) > 0$ for $u \neq 0$ by assumption, then

$$\varphi'(t) = g'(tu)u = (B(tu), u) = t^{-1}(B(tu), tu) > 0 \qquad (t > 0)$$

so that φ is (continuous and) strictly increasing on $[0, \infty[$; moreover, $\varphi(0) = 0$ while $\varphi(t) = g(tu) \to +\infty$ as $t \to +\infty$ by assumption. The result now follows.

b) Let us now show that f satisfies (PS) on N_r. We first note that N_r is weakly (sequentially) closed: indeed, if $u_n \rightharpoonup u_0$ for some sequence $(u_n) \subset N_r$, then since g is weakly continuous we have $r = g(u_n) \to g(u_0)$ and so $u_0 \in N_r$. Next we claim that (Bu, u) is bounded away from 0 on each bounded subset of N_r; for if not, there would exist a bounded sequence $(u_n) \subset N_r$ with $(Bu_n, u_n) \to 0$. We can assume that (u_n) converges weakly to some $u_0 \in N_r$; by the strong continuity of B, we then have $Bu_n \to Bu_0$ and so $(Bu_0, u_0) = 0$, hence $u_0 = 0$ by the definiteness assumption on B, contradiction since $0 \notin N_r$.

Therefore, in view of Proposition 2.5, with the choice $C_u = u$, it will be enough to prove that a sequence $(u_n) \subset N_r$ contains a convergent subsequence whenever $f(u_n)$ is bounded and (see (2.7))

$$\overline{A}u_n := Au_n - \frac{(Au_n, u_n)}{(Bu_n, u_n)} Bu_n \to 0. \tag{3.4}$$

This can be deduced as follows. Since $f(u_n)$ is bounded and f is coercive on N_r, (u_n) is bounded in X and therefore (passing to a subsequence if necessary) we can assume $u_n \rightharpoonup u_0$.

By the strong continuity of B, we have $Bu_n \to Bu_0$. Moreover, as A is bounded on bounded sets while $(Bu_n, u_n) \geq c > 0$ by the above remark, the sequence $\frac{(Au_n, u_n)}{(Bu_n, u_n)}$ of real numbers is bounded and so we can assume if converges too. By (3.4), we then have that Au_n converges strongly in X; and as A satisfies $(S)_1$ by assumption, we conclude that $u_n \to u_0$.

The requirements of Theorem 4 are therefore all satisfied, and this proves the first statement in Theorem 3. However, at this stage we do not know that there are infinitely many distinct eigenfunctions of the pair (A, B) on N_r. Clearly, since $f(u_n(r)) = c_n(r)$, this will be accomplished if we prove:

$$(3.5) \qquad c_n(r) = \inf_{K_n(r)} \sup_{K} f(u) \to +\infty \qquad \text{as} \quad n \to \infty.$$

To do this, we adapt to our context an argument appearing e.g. in [6], p. 365. Let us first prove a result of more general interest.

LEMMA 3.2. *Let X be an infinitedimensional, separable Hilbert space, and let (v_n) be a complete orthonorhal system in X. Let $N \subset X$ be weakly closed with $0 \notin N$. Then given any bounded subset N' of N, there exists an integer n_0 (depending on N') such that $N' \cap X_{n_0}^{\perp} = \phi$, where $X_n := span\{v_1, \ldots, v_n\}$ $(n \in N)$ and $X_n^{\perp}$ denotes the orthogonal supplement to X_n.*

PROOF. For each n. let P_n denote the orthogonal projection of X onto X_n; then $u \in X_n^{\perp}$ if and only if $P_n u = 0$. Furthermore, $u_n \rightharpoonup u$ in X implies $P_n u_n \to u$; indeed, for all $v \in X$,

$$(P_n u_n, v) = (u_n, P_n v) \to (u, v)$$

by the selfadjointness of P_n and the assumed completeness of (v_n), which implies $P_n v \to v$.

Given $N' \subset N$, N' bounded, assume by contradiction that $N' \cap X_n^{\perp} \neq \phi$ for all n; then there exists a sequence $(u_n) \subset N'$ with $P_n u_n = 0$ for all n. As N' is bounded, we can assume that (u_n) converges weakly to some $u_0 \in N$ (N is weakly closed by assumption). Therefore, $P_n u_n \to u_0$ by the above mentioned property and so $u_0 = 0$. But $0 \notin N$, contradiction.

c) PROOF OF THE CLAIM (3.5). Fix $r > 0$; recall N_r is weakly closed by the strong continuity of B, and $0 \notin N_r$. Also note the sequence $c_n(r)$ is nondecreasing since $K_{n+1}(r) \subset K_n(r)$. Assume thus by contradiction that, for some $d \in \mathbb{R}$, $c_n(r) < d$ for all n; then by the definition of $c_n(r)$, there would exist a sequence (A_n) of compact symmetric subsets of N_r, with $\gamma(A_n) \geq n$ for each n, such that

$$f(u) \leq d \qquad \text{for} \quad u \in \bigcup_{n=1}^{\infty} A_n.$$

Let $N_r^d = \{u \in N_r : f(u) \leq d\}$. By the coercivity of f on N_r, N_r^d is bounded (see (1.13)) and so by the above Lemma, there exists $n_0 \in \mathbb{N}$ such that $N_r^d \cap X_{n_0}^{\perp} = \phi$. Then $A_n \cap X_{n_0}^{\perp} = \phi$ for all n, which contradicts the property (3.2) of the genus as soon as $n > n_0$.

d) CONCLUSION OF THE PROOF OF THEOREM 3. Let $(u_n(r), \mu_n(r))$ be the eigenpair corresponding to $c_n(r)$, ie. $f(u_n(r)) = c_n(r)$ and

$$Au_n(r) = \mu_n(r)Bu_n(r) \tag{3.6}$$

with $u_n(r) \in N_r$. Since by assumption $\|u\| \to \infty$ whenever $f(u) \to +\infty$, it follows that $\|u_n(r)\| \to \infty$ by the fact $c_n(r) \to +\infty$ just proved above.

Moreover, taking the inner product with $u_n(r)$ in (3.6) we have

$$a(u_n(r)) = (Au_n(r), u_n(r)) = \mu_n(r)(Bu_n(r), u_n(r)).$$

Since $(Bu, u) \leq D_r$ on N_r by hypothesis, then

$$\mu_n(r) = (Bu_n(r), u_n(r))^{-1} a(u_n(r)) \geq D_r^{-1} a(u_n(r))$$

and the final assertion now follows from $\|u_n(r)\| \to \infty$ and the assumed coercivity of a on N_r.

4. Proof of Theorem 2.

Let us first state a technical result, which is a direct consequence (via Hölder's inequality) of the Sobolev embedding $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$, $\bar{p} = \frac{2N}{N-2}$ (see e.g. [5]):

LEMMA 4.1. *Let $p : 1 \leq p \leq \frac{N+2}{N-2}$ (so that $2 \leq p+1 \leq \bar{p}$) and let $\beta = \beta(p) = \frac{N}{\bar{p}}(\bar{p}-(p+1)) = (p+1) - \frac{N}{2}(p-1)$. Then there exists $c > 0$ such that*

$$\|u\|_{p+1}^{p+1} \leq c\|\nabla u\|_2^{p+1-\beta}\|u\|_2^{\beta} \tag{4.1}$$

for all $u \in H_0^1(\Omega)$. (Here and henceforth, $\|u\|_p$ denotes the norm of u in $L^p(\Omega)$).

Let us now go back to (1.1). We recall for the reader's convenience that the operators and functionals of interest are here as follows:

$$a(u, v) = \sum_{i=1}^{N} \int a_i(x, u, \nabla u)\frac{\partial v}{\partial x_i} + \int a_0(x, v, \nabla u)v$$

$$f(u) = \int F(x, u, \nabla u) \qquad g(u) = \frac{1}{2}\int u^2 \tag{4.2}$$

$$(Au, v) = a(u, v) = f'(u)v \qquad (Bu, v) = g'(u)v.$$

LEMMA 4.2. *Under the growth assumption H1), f satisfies:*

$$|f(u)| \le c\|u\|^{\delta} + d \qquad (u \in X) \tag{4.3}$$

with $c, d > 0$ and $\delta = \gamma + 1$ $(\gamma = \max\{r, s, q\})$.

PROOF. We have

$$f(u) - f(0) = \int_0^1 \frac{d}{ds} f(su)ds = \int_0^1 f'(su)u\,ds = \tag{4.4}$$

$$= \int_0^1 a(su, u)ds.$$

By (1.3),

$$|a(u, u)| \le c(\|u\|^{\gamma+1} + \|u\|)$$

whence, for $s > 0$,

$$|a(su, u)| = s^{-1}|a(su, su)| \le s^{-1}c(s^{\gamma+1}\|u\|^{\gamma+1} + s\|u\|) =$$

$$= c(s^{\gamma}\|u\|^{\gamma+1} + \|u\|).$$

Therefore, from (4.4),

$$|f(u) - f(0)| \le \frac{c}{\gamma + 1}\|u\|^{\gamma+1} + c\|u\|$$

which gives the result.

LEMMA 4.3. *Under the assumption H3) and H5), there exist constants $0 < \alpha < 1$ and $\beta > 0$ such that*

$$a(u) \ge c_1\|u\|^2 - c_2\|u\|^{2\alpha}\|u\|_2^{\beta} - d \qquad (u \in X) \tag{4.5}$$

where $c_1 > 0$, $c_2, d \ge 0$. A similar inequality is satisfied by f.

PROOF.

i) By H5),

$$\int a_0(x,u,\nabla u)u \geq -c\int |u|^{\sigma+1} - \varepsilon\int |\nabla u|^2 - d.$$

Using this together with the ellipticity condition H3), we get

$$a(u) = a(u,u) \geq (\nu-\varepsilon)\int |\nabla u|^2 - c\int |u|^{\sigma+1} - d =$$

$$= c_1\|u\|^2 - c_2\|u\|_{\sigma+1}^{\sigma+1} - d. \tag{4.6}$$

ii) Let us show that a similar inequality holds for f. Let $u \in X$, $\|u\| \geq 1$; write $u = rv$, $\|v\| = 1$. We have

$$f(u) - f(v) = f(rv) - f(v) = \int_1^r f'(sv)v\,ds = \tag{4.7}$$

$$= \int_1^r a(sv,v)\,ds.$$

By (4.6), as $\|v\| = 1$,

$$a(sv,sv) \geq c_1 s^2 - c_2 s^{\sigma+1}\|v\|_{\sigma+1}^{\sigma+1} - d$$

whence, for $s > 0$,

$$a(sv,v) \geq c_1 s - c_2 s^{\sigma}\|v\|_{\sigma+1}^{\sigma+1} - ds^{-1}.$$

Therefore, from (4.7),

$$f(u) - f(v) \geq c_1' r^2 - c_2' r^{\sigma+1}\|v\|_{\sigma+1}^{\sigma+1} - d\log r - d' =$$

$$= c_1'\|u\|^2 - c_2\|u\|_{\sigma+1}^{\sigma+1} - d\log\|u\| - d' =$$

$$\geq c_1''\|u\|^2 - c_2'\|u\|_{\sigma+1}^{\sigma+1} - d''.$$

As, by Lemma 4.2, $|f(v)| \leq c$ for all $v \in X$ with $\|v\| = 1$, we conclude that (4.6) is satisfied by f too (with different constants c_1, c_2, d).

iii) By Lemma 4.1,

$$\|u\|_{\sigma+1}^{\sigma+1} \leq c\|u\|^{2\alpha}\|u\|_2^{\beta}$$

where $2\alpha = \sigma + 1 - \beta = (\sigma - 1)\frac{N}{2}$. Note that the assumption $\sigma < 1 + \frac{4}{N}$ is equivalent to $\alpha < 1$. Using this in (4.6), we get the desired inequality.

With the aid of the above Lemmas, we are now able to prove Theorem 2. This will be accomplished by the following:

PROPOSITION 4.4. *Assume that the coefficients a_i of the quasilinear differential operator in (1.1) satisfy the conditions stated in the Introduction, with* H4) *replaced by* H5). *Then the operators and functionals A, B, f, g related to* (1.1) *by* (4.2) *satisfy all the assumptions of Theorem 3.*

a) Proof of the assumptions concerning B and g.
The relations $(Bu, u) > 0$ $(u \neq 0)$, $g(tu) \to \infty$ if $u \neq 0$ and $t \to \infty$, $(Bu, u) \leq D_r$ on $N_r = \{u : \int u^2 = r^2\}$ follow trivially from the explicit expression

$$(Bu, u) = 2g(u) = \int u^2.$$

Furthermore, B is strongly continuous by the compact embedding $X = H_0^1(\Omega) \to L^2(\Omega)$.

b) Proof of the assumptions concerning A and f.
i) As the proof of Browder ([2], p. 27-30) shows, the monotonicity and ellipticity assumptions H2), H3) ensure that A is of type (S), ie.

$$u_n \to u \quad \text{and} \quad (Au_{n_X} - Au, u_n - u) \to 0 \quad \text{imply} \quad u_n \to u.$$

But as pointed out by Amann ([1], p. 57) and as can be easily checked, (S) is a stronger property shan $(S)_1$.

ii) From Lemma 4.3, we have that if $\|u\|_2 = r$ then

$$a(u) \geq c_1\|u\|^2 - c_2\|u\|^{2\alpha}r^{\beta} - d \qquad (c_1 > 0)$$

and a similar inequality holds for f, too. Since $\alpha < 1$ (by the assumption $\sigma < 1 + \frac{4}{N}$), it follows that both a and f are bounded below and coercive on N_r for all $r > 0$.

iii) Finally from Lemma 4.2 we have

$$f(u) \leq c\|u\|^{\delta} + d \qquad (u \in X)$$

which evidently implies that $\|u\| \to \infty$ whenever $f(u) \to +\infty$.

REMARK 4.5. Condition H5) is for instance satisfied if

$$|a_0(x,t,p)| \leq c(|t|^\sigma + |p|) + d \qquad \left(\sigma < 1 + \frac{4}{N}\right)$$

ie. $s = \sigma$ and $q = 1$ in the growth assumption H1) on a_0. Indeed, the above inequality implies

$$\begin{aligned} |a_0(x,t,p)t| &\leq c(|t|^{\sigma+1} + |p||t|) + d|t| \\ &\leq c'(|t|^{\sigma+1} + |p||t|) + d'. \end{aligned}$$

For each $\varepsilon > 0$, $2|p||t| \leq \frac{t^2}{\varepsilon} + \varepsilon|p|^2$ and so (assuming w.l.o.g. $\sigma \geq 1$)

$$|a_0(x,t,p)t| \leq c''|t|^{\sigma+1} + \varepsilon|p|^2 + d''$$

which implies H5) as soon as $\varepsilon < \nu$.

REFERENCES

[1] Amann, H., *Liusternik-Schnirelmann theory and nonlinear eigenvalue problems,* Math. Ann. 199 (1972), 55-72.

[2] Browder, F.E., *Existence theorems for nonlinear partial differential equations,* Proc. Symp. Pure Math. 16 (1970), 1-60.

[3] Browder, F.E., *Nonlinear eigenvalue problems and group invariance,* in Functional Analysis and Related Fields (F.E. Browder, Editor) Springer (1970), 1-58.

[4] Chiappinelli, R., *Remarks on bifurcation for elliptic operators with odd nonlinearity,* Israel J. Math. 65 (1989), 285-292.

[5] Chiappinelli R., *On spectral asymptotics and bifurcation for elliptic operators with odd superlinear term,* Nonlinear Anal. TMA 13 (1989), 871-878.

[6] Krasnoselskii, M.A., Topological methods in the theory of nonlinear integral equations. Pergamon Press, Oxford, 1964.

[7] Rabinowitz, P.H., *Some aspects of nonlinear eigenvalue problems,* Rocky Mt. J. Math. 3 (1973), 161-202.

[8] Rabinowitz, P.H., *Variational methods for nonlinear eigenvalue problems,* in Eigenvalues of Nonlinear Problems (G. Prodi, Editor) Cremonese (1974), 141-195.

Raffaele Chiappinelli
Dipartimento di Matematica
Universitá della Calabria
87036 Rende (CS), Italy.

Constantin Carathéodory: An International Tribute (pp. 119–141)
edited by Th. M. Rassias

DYNAMICAL SYSTEMS CREATED FROM SEMIDYNAMICAL SYSTEMS

Krzysztof Ciesielski

The paper presents the method of construction of dynamical systems from the given semidynamical systems. There are also given some theorems concerning this construction and characterizing the obtained systems and phase spaces.

0. Introduction

In the semidynamical system the movement is defined only for positive values of time t. However, we may ask about "the past" of a given point x in a phase space for a particular value of t. There may be many different points which reach x after time t; on the other hand there may be no such point. When the system is a dynamical system, such point is always unique for every value of t. Thus the structure of semisystems may be quite different from the structure of systems, the latter being much better.

The natural question is: Can we reconstruct semidynamical systems to get dynamical systems? In this paper a construction is presented which allows us to do so. Roughly speaking, we get the better structure of systems but we can destroy a good structure of topological space. After "glueing" the points in a suitable way we get the semi-systems with negative unicity. The first part of the paper is devoted to the description of a new structure and the characterization of conditions to get dynamical systems. In the next chapter some necessary conditions for the systems to get "not too bad" topological space after construction are given. Finally, some remarks on planar systems are presented.

The basic properties of semidynamical systems may be found in [1],

[10], [14], [18] and [20]. Many investigations on semidynamical systems were presented in many papers not cited here, in particular written by P. Bajaj, N. P. Bhatia, K. M. Das, S. Elaydi, S. K. Kaul, S. S. Lakshmi, M. Nishihama, A. Pelczar and S. H. Saperstone.

Chapter I. Preliminaries

By a semidynamical system we mean a triplet $(X, \mathbb{R}_+, \pi)$, where π is a continuous map from $\mathbb{R}_+ \times X \to X$ such that $\pi(0,x) = x$ for every $x \in X$ and $\pi(t, \pi(s,x)) = \pi(t+s,x)$ for every $x \in X$ and $t, s \in \mathbb{R}_+$. By $\pi^+(x)$ we denote the set $\pi(\mathbb{R}_+ \times \{x\})$ and call it a positive trajectory through x. We put $F(t,x) = \{y \in X : \pi(t,y) = x\}$ (and call it a cut of a funnel through x) and $F([s,t],x) = \cup\{F(u,x) : s \leq u \leq t\}$ (and call it a section of a funnel through x). By a funnel we mean the set $\cup\{F(t,x) : t \geq 0\}$. We say that a point x is a point of negative unicity if $F(t,x)$ contains at most one element for each $t \geq 0$. The system is said to be the system with negative unicity if every point of X is a point of negative unicity. The point x is called a start point if $F(t,x) = \emptyset$ for any $t > 0$. It is known ([1]) that when X is a manifold (without boundary) then a semidynamical system has no start points.

A function $\sigma : (\alpha, 0] \to X$ is called a solution through x if $\sigma(0) = x$, $\pi(t, \sigma(s)) = \sigma(t+s)$ whenever $t, t+s \in (\alpha, 0]$ and $t \geq 0$. When this function is maximal (with respect to inclusion of images) the solution is called a left maximal solution. By a negative trajectory through x we mean an image of a left maximal solution through x. By a trajectory we mean an union of the negative and positive trajectories through a given point. By $L^+(x)$ we denote the set $\{y \in X : \pi(t_n, x) \to y$ for some sequence $\{t_n\} \to \infty\}$. By $L^-_\sigma(x)$, where σ is a solution through x, we denote the set $\{y \in X : \sigma(t_n) \to y$ for some sequence $\{t_n\}$ such that $t_n \to -\infty\}$. A set M is said to be stable if for each neighbourhood U of M and $x \in M$ there is a neighbourhood V of x with $\pi(t, V) \subset U$ for every $t \geq 0$.

Replacing in the definition of semidynamical system $\mathbb{R}_+$ by $\mathbb{R}$ we get the definition of dynamical system.

With respect to a given semidynamical system we may classify each point $x \in X$ into one of the three sets. A point x is said to be a stationary point if $\pi(t,x) = x$ for each $t \geq 0$. A point $x \in X$ is said to be a periodic point if there exists a $T > 0$ such that $\pi(T,x) = x$ and x is not a stationary

point. The smallest T with the above properties is called the period of x. A point $x \in X$ is said to be a regular point if it is neither stationary nor periodic. We say that a point x merges to stationary (periodic) point, if there is a time t such that $\pi(t,x)$ is stationary (periodic).

The systems $(X, \mathbb{R}_+, \pi)$ and $(X, \mathbb{R}_+, \rho)$ are said to be isomorphic if there is a continuous mapping $\phi : \mathbb{R}_+ \times X \to \mathbb{R}_+$ such that $\phi(0,x) = 0$ and the mapping $\phi(\cdot, x) : \mathbb{R}_+ \to \mathbb{R}_+$ is a homeomorphism for each $x \in X$ and $\pi(t,x) = \rho(\phi(t,x),x)$ for each $(t,x) \in \mathbb{R}_+ \times X$. Note that the isomorphism does not change the trajectories, it changes only "the speed of moving along the trajectories". Also, it does not change the character of singular points and the negative unicity points.

In 1977 R. C. McCann introduced the definition of the negative escape time of a point $x \in X$ (which is, intuitively, the minimal time length of all negative trajectories through x) and proved some properties of isomorphisms of semidynamical systems ([13]). The definition presented below is a little simpler than the definition given by McCann, but when the system has no start points, then these definitions are equivalent.

1.1. Definition. By a negative escape time $N(x)$ of x we define $N(x) = \inf\{s \in (0,\infty] : (-s,0]$ is a left maximal solution through $x\}$.

Then McCann's results give us the following

1.2. Theorem. When x is a locally compact metric space and the system $(X, \mathbb{R}_+, \pi)$ has no start points, then the system is isomorphic to a system $(X, \mathbb{R}_+, \pi')$ which has infinite negative escape time for each $x \in X$.

We will also use B–H escape time, which is, intuitively, "the maximal time length of all negative trajectories through x" (see [1]):

1.3. Definition. By B–H escape time we mean $\sup\{s \in (0,\infty] : (-s,0]$ is a left maximal solution through $x\}$.

We will also use the important

1.4. Theorem. ([8], [13]). If the semidynamical system on a locally

compact space X has no start points and it has the infinite negative escape time $N(x)$ for each $x \in X$, then it can be extended to the semidynamical system π^* on X^*, where $X^* = X \cup \{\infty\}$ is the one point compactification of X and ∞ is a stationary point for the new system.

Chapter II

Assume that a semisystem $(X, \mathbb{R}_+, \pi)$ is given. Let us state

2.1. Definition. We define:
$x \simeq y \leftrightarrow$ there exists an s with $\pi(s,x) = \pi(s,y)$

It can be easily verified that this relation is an equivalence relation. Now let us put
$X(\pi) = X/\simeq$ (the quotient space) and

$$\phi(\pi)(t,[x]) = \begin{cases} [\pi(t,x)] & \text{for } t \geq 0 \\ [z] : z \in F(t,x) & \text{for } t < 0 \end{cases}.$$

We have to verify if the definition is correctly stated. Let us take x, y with $x \simeq y$. Then there is an s with $\pi(s,x) = \pi(s,y)$. Thus for $t \geq s$, of course $\pi(t,x) = \pi(t,y)$. Let us take $t \in [0,s]$. Then $\pi(s-t, \pi(t,x)) = \pi(s-t, \pi(t,y))$ as $\pi(t,x) = \pi(t,y)$, so $\pi(t,x) \simeq \pi(t,y)$. Now consider $t < 0$. Then $s - t > 0$. Take a $z \in F(t,x)$ and a $v \in F(t,y)$. We have $\pi(s-t,z) = \pi(s,x) = \pi(s,y) = \pi(s-t,v)$ and $z \simeq v$, which finishes the proof.

2.2. Remark. For every t and x all the points of $F(t,x)$ belong to the same equivalence class. By $[F(t,x)]$ we will denote the class $[y]$, where y is an element of $F(t,x)$.

2.3. Remark. We have: $[y] = \cup\{\pi_t^{-1}(\pi(t,y)) : t \geq 0\} = \cup\{F(t, \pi(t, y)) : t \geq 0\}$. Note also that for a stationary point x we have $[x] = F(x)$ (and thus also for any point which merges to a stationary point).

2.4. Theorem. The triplet $(X(\pi), \mathbb{R}_+, \phi(\pi))$ is a semidynamical system with negative unicity.

Proof. The unicity of system follows directly from the definition. We have to check only the continuity of ϕ for $t \geq 0$. Let us take a neighbourhood U^* of $[\pi(t,x)]$. We look for a neighbourhood V^* of $[x]$ and a δ with $\phi((t-\delta, t+\delta), V^*) \subset U^*$. There exists a neighbourhood U of $\pi(t,x)$ with $U^* = \{[y] : y \in U\}$. There are a $\delta > 0$ and a neighbourhood V of x with $\pi((t-\delta, t+\delta), V) \subset U$. Let us put $V^* = \{[y] : y \in V\}$. This set is a neighbourhood of $[x]$ in $X(\pi)$. Take a $[z] \in V^*$ and an $s \in (t-\delta, t+\delta)$. There is a $y \in V$ with $y \in [z]$. Then $\pi(s,y) \in U$ and $\phi(\pi)(s,[z]) = [\pi(s,z)] = [\pi(s,y)] \in U^*$. This finishes the proof.

2.5. Remark. Of course the obtained semisystem need not be a dynamical system for an obvious reason: the value of $\phi(t,x)$ need not be defined for every $t < 0$. It is obvious that it is defined only for t smaller than B–H escape time of x. However, in many cases this problem can be omitted after using Theorem 1.2.

2.6. Example. From the planar semidynamical system with trajectories shown in Fig. 1 we obtain the planar dynamical system (we glue together the points of $F(t,(0,0))$ for each t).

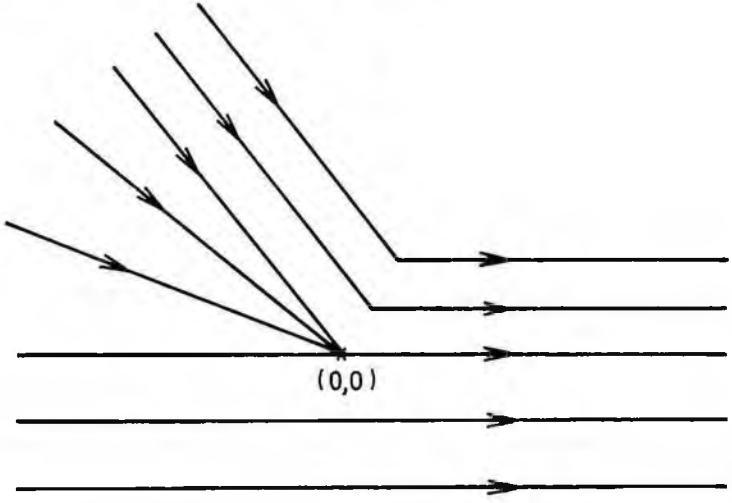


Fig. 1.

2.7. Example. From the semidynamical system with trajectories presented in Fig. 2 we get the system on the ball in the plane.

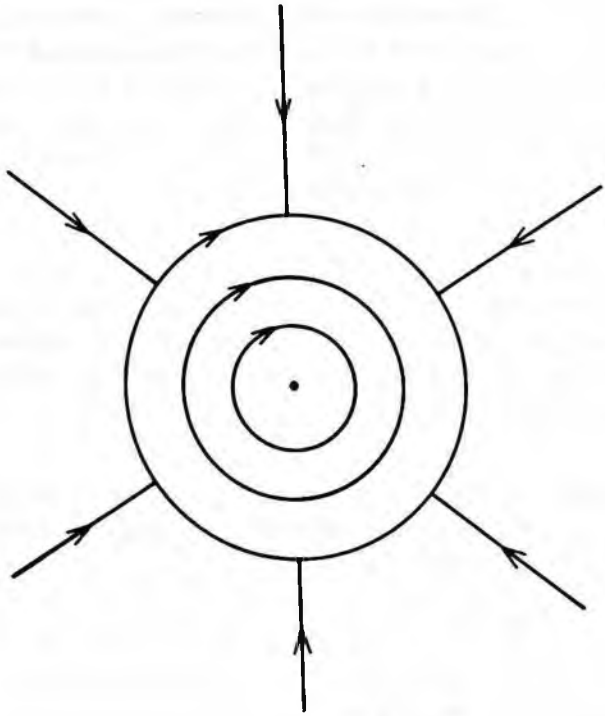


Fig. 2.

The interesting chapter of the theory of semidynamical systems has its source in the infinite dimensional dynamical systems (compare for instance [11], [16], [17] and [18]). We would not develop here this aspect, but only present the following

2.8. Example. Consider $X = \{f : [0,\infty) \to \mathbb{R}, f \text{ continuous}, f \text{ bounded}\}$. The function π where $\pi(t,f)(s) = f(s+t)$ for $s \geq 0$ defines a semidynamical system. The class $[f]$ equals $\{g :$ there exists an $s_0 \geq 0$ such that $f(s) = g(s)$ for $s \geq s_0\}$. Thus the phase space $X(\pi)$ is the space of germs.

Before coming to the main theorems of this chapter we present

2.9. Proposition. The function $\phi(\pi)$ is continuous in $(0, x)$ for each $x \in X$.

Proof. According to Theorem 2.4 we need only to show that for each neighbourhood U^* of $[x]$ there is a neighbourhood V^* of $[x]$ and a $\delta > 0$ with $\phi((t-\delta, 0], V^*) \subset U^*$. Using Theorem 4.4 in [1] we obtain that there is a δ and a neighbourhood V of x such that $F([0, \delta], V) \subset U$, where by U we mean the same neighbourhood of x as in the proof of Theorem 2.3. Let us take a $[z] \in V^*$ and an $s \in (0, \delta)$. There exists a $y \in V$ with $y \in [z]$. We have $F(s, y) \in U$ and $[F(s, y)] \in U^*$. But $[F(s, y)] = [F(s, z)] = \phi(\pi)(s, [z])$ which finishes the proof.

2.10. Theorem. The function: $t \to \phi(\pi)(t, [x])$ is continuous in its domain for each $[x] \in X(\pi)$.

Proof. Because of Theorem 2.4 and Proposition 2.9 we need only check a continuity for $t < 0$. Let us take $[F(-t, x)]$ and a neighbourhood U^* of $[F(-t, x)]$. The set $U = \cup\{y : [y] \in U^*\}$ is open in X and contains $F(-t, x)$. Let us take a left maximal solution σ through x such that $\sigma(t) = y$ and assume that domain σ is not equal to $[t, 0]$. The set U is a neighbourhood of y. Every solution is continuous ([1]), so there is a δ such that $(t-\delta, t+\delta)$ is contained in domain σ and $\sigma(t-\delta, t+\delta) \subset U$. Thus for every $s \in (t-\delta, t+\delta)$ we have $F(s, x) \cap U \neq 0$, so $[F(s, x)] \subset U^*$.

Now suppose that for each left maximal solution σ through y, domain σ is contained in $[t, 0]$. As above we find a δ with $[F(s, x)] \subset U^*$ for each $s \in [t, t+\delta)$. However, for every $s < t$ the set $F(s, x)$ is empty, so $\phi(\pi)(s, [x])$ is defined only for $s \in [t, \infty)$. We have shown that $\phi(\pi)(\cdot, x)$ is continuous in its domain.

2.11. Remark. The general dynamical systems with the function $t \to \pi(t, x)$ continuous were investigated for instance in [14] and [15].

The following theorem gives the sufficient condition for the system to be dynamical.

2.12. Theorem. Assume that $F(t, x) \neq \emptyset$ for each $x \in X$ and $t > 0$.

Assume also that the function $F : \mathbb{R}_+ \times X \to 2^X$ is upper semicontinuous, i.e., for every neighbourhood U of $F(t,x)$ there is a neighbourhood V of x and a δ such that $F((t-\delta,t+\delta),V)) \subset U$ (compare [12]). Then $(X(\pi),\mathbb{R},\phi(\pi))$ is a dynamical system.

Proof. Let us take a neighbourhood U^* of $[F(t,x)]$ (compare Remark 2.2). From the semicontinuity of a multivalued function F it follows that there is a V and a δ such that $F((t-\delta,t+\delta),V) \subset U$. The set $V^* = \{[y] : y \in V\}$ is a neighbourhood of $[x]$, as V is a neighbourhood of x. To finish the proof we need to verify if $\phi(\pi)((t-\delta,t+\delta),V^*) \subset U^*$. Take an $s \in (t-\delta,t+\delta)$ and a $[z] \in V^*$. There is a $y \in V$ with $y \in [z]$. We have $F(s,y) \subset U$ and $[F(s,y)] \subset U^*$. By definition we obtain that $[F(s,y)] = \phi(\pi)(s,[y]) = \phi(\pi)(s,[z])$, so the last point belongs to U^*.

Using the above theorem we can get some results stating when the constructed structure is a dynamical system.

2.13. Theorem. Let $(X,\mathbb{R}_+,\pi)$ be a semidynamical system without start points on a locally compact and first countable space. Assume that for every t and every x there is a neighbourhood W of x such that the closure of W : $\mathrm{Cl}W$ is compact and $F(t,\mathrm{Cl}W)$ is compact as well. Then $(X(\pi),\mathbb{R},\phi(\pi))$ is a dynamical system.

The theorem follows from Theorem 1.17 in [3] and Theorem 2.12.

2.14. Theorem. Let X be a locally compact, paracompact and first countable space. Assume that a semidynamical system $(X,\mathbb{R}_+,\pi)$ (without start points) has an infinite negative escape time for each $x \in X$. Then the system $(X(\pi),\mathbb{R},\phi(\pi))$ is a dynamical system.

This is a consequence of Theorem 2.12 and Proposition 2.9 in [3].

2.15. Theorem. Assume that X is a locally compact metric space and that the system $(X,\mathbb{R}_+,\pi)$ has no start points. Then this system is isomorphic to a system $(X,\mathbb{R}_+,\pi')$ such that the system $(X(\pi'),\mathbb{R},\phi(\pi'))$ is a dynamical system.

The theorem is a corollary from Proposition 2.8 in [3], Theorem 4.1 in

[13] and Theorem 2.12.

2.16. Remark. Theorem 2.20 in [3] and Theorem 2.4 in this paper suggest that the obtained system could be a local dynamical system. However, we cannot use the cited result, as the space $X(\pi)$ obtained in our construction need not be locally compact and metric.

2.17. Remark. In Theorem 2.15 we use at first the isomorphism of semi-systems and then the construction of "getting unicity". However, one may suggest changing the direction: firstly to get the unicity and then try to use isomorphism. The natural question is if we get the same result. Example 2.18 shows that the answer is negative. Moreover, we may get "very bad" topological space, which will not allow us to construct the isomorphism.

2.18. Example. Let us take $X = \mathbb{R}^2 \setminus (-\infty, -1] \times \{0\}$ and π with the trajectories presented in Fig. 3. The "speed of points" "along the first variable" is the same for all points. Thus $N((0,0)) = 1$. After the suitable isomorphism we get the system homeomorphic to the semidynamical system presented in Example 2.6, which after the identification gives the classical dynamical system $\tilde{\pi}(t,(x,y)) = (t+x,y)$ on $\mathbb{R}^2$. If we make the construction from Definition 2.1 at first, then we obtain the semidynamical system with negative unicity and with $N([x]) = \infty$ for each $[x]$. However, the obtained system is not dynamical. As one can easily verify, the function $\phi(\pi)$ is not continuous for $t < -1$ in the points $[F(-t,(0,0))]$ and the space $X(\pi)$ is not homeomorphic to the plane (consider also the points $[F(-t,(0,0))]$ for $t < -1$).

2.19. Remark. In [7] S. Elaydi introduced a completely different method of obtaining dynamical systems on the base of semisystems. This result gives a system in the space of negative maximal solutions of semisystems.

Chapter III

Throughout this section we assume that $N(x) = \infty$ for every $x \in X$ in

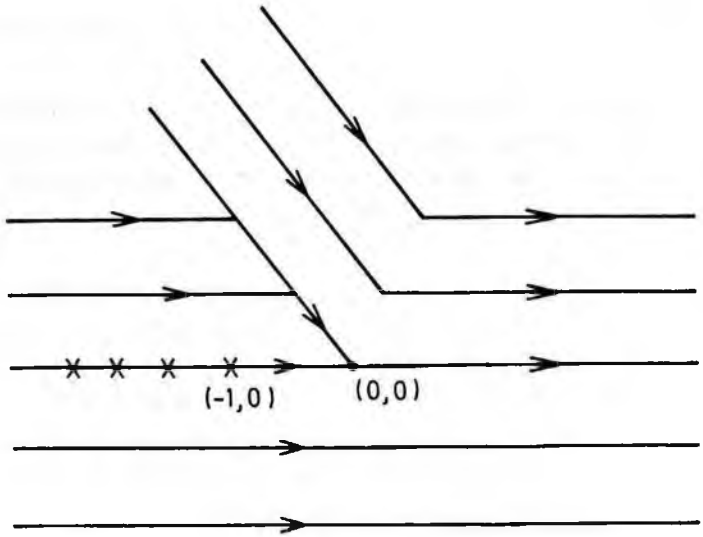


Fig. 3.

a given semidynamical system $(X, \mathbb{R}_+, \pi)$. This assumption is quite natural (compare [8], [13] and Theorems 1.2 and 2.15).

Using the construction introduced in Chapter II we get the better structure of system. However, we may destroy the structure of space. The obtained space $X(\pi)$ may behave very well (compare Examples 2.6, 2.7 and 2.8), but also very bad.

3.1. Example. Let us take the semidynamical system with the trajectories presented in Fig. 4. After the suitable construction we get the space $\mathbb{R}\times\mathbb{Z}$, but with non-euclidean topology. For an even integer k the basic neighbourhoods of (x,k) are the intervals $(x-\delta, x+\delta)\times\{k\}$, but for an odd integer k the basic neighbourhoods of (x,k) are the unions: $(x-\delta_1, x+\delta_1)\times\{k-1\}\cup(x-\delta_2, x+\delta_2)\times\{k\}\cup(x-\delta_3, x+\delta_3)\times\{k+1\}$. Note that the space $\mathbb{R}^2(\pi)$ is not Hausdorff and even not T_1.

Using the basic topological properties we get

3.2. Proposition. The space $X(\pi)$ is T_1 if and only if the set $\{y : y\in[x]\}$ is closed for each $x\in X$.

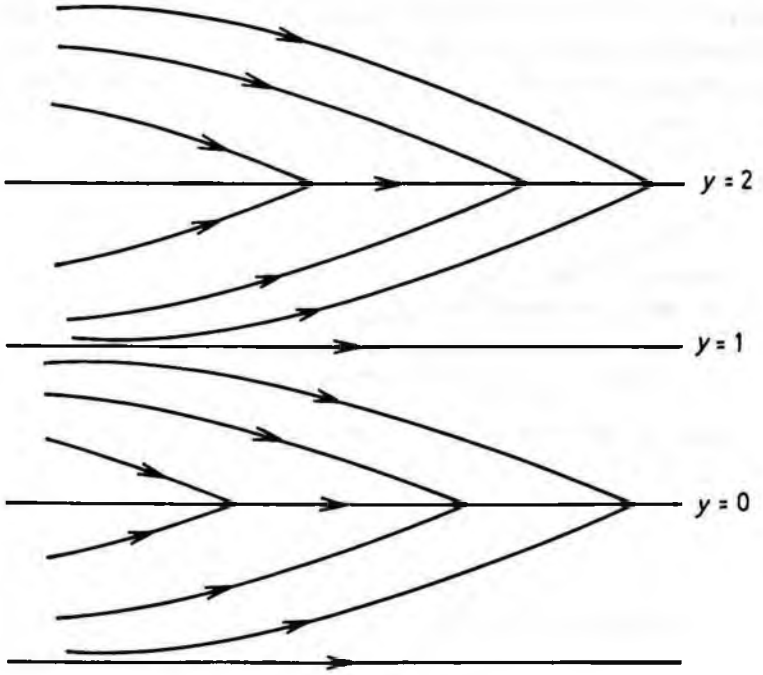


Fig. 4.

Generally, the set $\{y : y \in [x]\}$ need not be closed (however the set $F(t,x)$ is closed for each t — compare [13]), which can be seen on the previous example and also the following

3.3. Example. In the semidynamical system with trajectories shown in Fig. 5 the topological space $\mathbb{R}^2(\pi)$ is not T_1. The space $\mathbb{R}^2(\pi)$ is equal to $\mathbb{R}^2$, but with non-euclidean topology. For each $x \in \mathbb{R}$ and a neighbourhood U^* of $[(x,0)]$ the point $[(x,e^{-x})] \in U^*$. Note that $\{y : y \in [(x,e^{-x})]\} = \{x\} \times [e^{-x},0)$ is not closed.

The condition from Proposition 3.2 does not assure the good behaviour of space. Consider the following

3.4. Example. Let us take the semidynamical system with the trajectories shown in Fig. 6. The points $(0,y)$ for $y \in \mathbb{R}$ are stationary points, (1,0) is not a point of negative unicity. The points $(0,\alpha)(\alpha \in [-1,1])$ are

equal to $L_\sigma^-((1,0))$ for different solutions σ through (1,0). It is easy to verify (as in the proof of Theorem 3.6) that $(0,\alpha)$ and $(0,\beta)$ do not possess disjoint neighbourhoods for $-1 \leq \alpha < \beta \leq 1$, so $\mathbb{R}^2(\pi)$ is not Hausdorff.

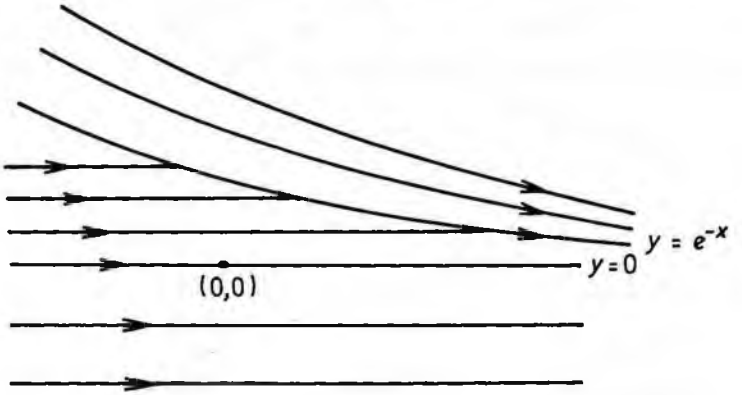


Fig. 5.

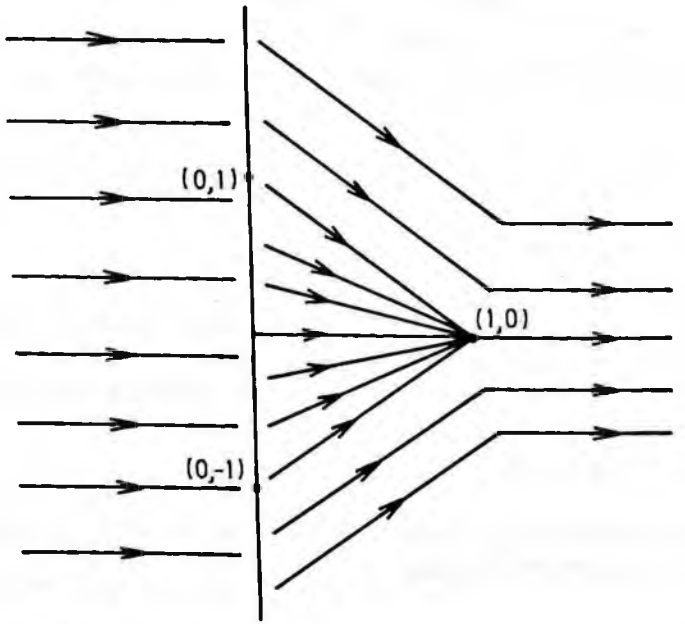
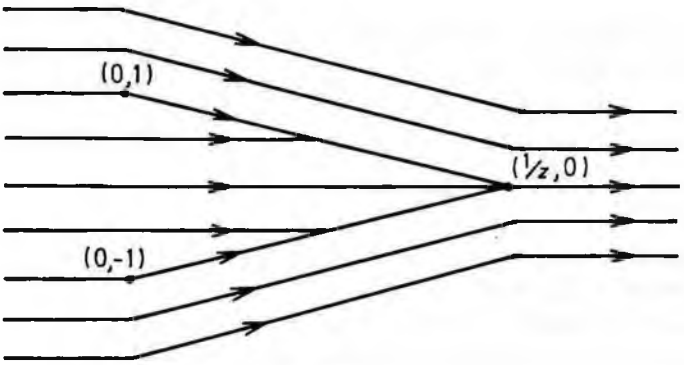


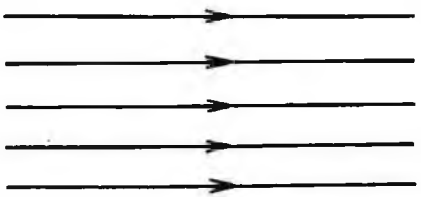
Fig. 6.

3.5. Example. Let $X = \mathbb{R}^3$. If $z > 0$ then the trajectories on the plane $\mathbb{R}^2 \times \{z\}$ are shown in Fig. 7(a), if $z \leq 0$, they are presented in Fig. 7(b). The system seems to behave very well, however, every neighbourhoods of $[(0,-1,0)]$ and $[(0,1,0)]$ in $\mathbb{R}^3(\pi)$ have nonempty intersections, as $(0,-1,z) \simeq (0,1,z)$ for each $z > 0$.



$z > 0$

(a)



$z \leqslant 0$

(b)

Fig. 7

The above examples show also another application of the constructed systems. We may use the presented construction to observe that some

orbits of semi-systems are "closed to each other" because of the behaviour of all orbits in semi-system. In particular, in Example 3.5 this is the case of trajectories through $(0,-1,0)$ and $(0,1,0)$.

Below we present some theorems giving necessary conditions for the space $X(\pi)$ to be Hausdorff.

3.6. Theorem. Assume that $X(\pi)$ is a Hausdorff space. Then for each point $x \in X$: if for two left maximal solutions σ_1 and σ_2 we have $\sigma_1(t) \to y, \sigma_2(t) \to z$ when $t \to -\infty$, and y, z are stationary points, then $y = z$.

Proof. Suppose to the contrary that σ_1, σ_2, y, z are as in the theorem and $y \neq z$. Take the arbitrary neighbourhoods U^* of $[y]$ and V^* of $[z]$. As in the proof of Theorem 2.4 we construct the neighbourhoods U of y and V of z. For a sufficiently large $-t$ we have $\sigma_1(t) \in U$ and $\sigma_2(t) \in V$, so $F(-t,x)$ is not disjoint from U as well as from V. Thus $[F(-t,x)] \in U^*$ and $[F(-t,x)] \in V^*$ which shows that these sets are not disjoint. The space $X(\pi)$ is not Hausdorff.

Without loss of generality instead of locally compact spaces we may consider compact spaces. This is because of Theorem 1.4.

For compact spaces we have

3.7. Theorem. If the space X is compact and $X(\pi)$ is Hausdorff, then for each $x \in X$ and $z \in X$ either the set $\{y : y \in [z]\}$ has the nonempty intersection with $L_\sigma^-(x)$ for each left maximal solution σ through x, or it has the empty intersection with all these sets.

Proof. From the elementary properties of $L_\sigma^-(x)$ we have that this set is closed. The space X is compact, so $L_\sigma^-(x)$ is compact and nonempty for each x and σ. Suppose to the contrary that there exist σ and σ_0 with $y \in L_\sigma^-(x)$ and $y \notin [z]$ for every $z \in L_{\sigma_0}^-(x)$. Take a sequence $t_n \to -\infty$ with $\sigma(-t_n) \to y$. From the compactness of X we may assume that $\sigma_0(-t_n) \to z \in L_{\sigma_0}^-(x)$ (taking a subsequence, if necessary). From the hypothesis we have the disjoint neighbourhoods U^* of $[y]$ and V^* of $[z]$. The sets U and V defined as in the previous proof are disjoint and

for a sufficiently large n we have $\sigma(-t_n) \in U, \sigma_0(-t_n) \in V$. This is a contradiction, as $[\sigma(-t_n)] = [\sigma_0(-t_n)]$, so $U^* \cap V^* \neq \emptyset$.

Even under the assumptions of Theorem 3.7 the sets $L_\sigma^-(x)$ need not be the same for all σ. This can be seen from the following

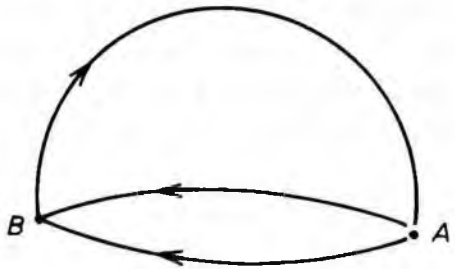
3.8. Example. Consider the semidynamical system on a compact subset of $\mathbb{R}^3$ described below. The point $A = (1,0,0)$ is a stationary point. For $B = (-1,0,0)$ we have two different negative solutions through B with trajectories on the circle $x^2 + y^2 = 1, z = 0$, which tend to A as $t \to -\infty$. The positive solution through B tends also to A, but along the trajectory on the semi-circle $x^2 + z^2 = 1, z \geq 0, y = 0$ (see Fig. 8(a)). Let us take a point $C = (0,0,\frac{1}{2})$. In the positive direction the solution through C tends to $(0,0,0)$ on a straight line. In the negative direction we have two negative solutions through C which eventually are on the surface $x^2 + y^2 + z^2 = 1$ and as the negative limit sets they have two Jordan curves joining A and B described above (one of negative trajectories through C with its limit set is presented in Fig. 8(b)). These two curves have different negative limit sets for different solutions through C. However, we can join all the points from these curves and construct the equivalence classes containing one (for $z > 0$) or two ($z = 0$) elements. The obtained dynamical system is presented (after the suitable homeomorphism) in Fig. 8(c). Using the same method it is possible to construct a similar example on a closed semi-ball in $\mathbb{R}^3$ and get the dynamical system on a semi-ball in $\mathbb{R}^2$.

Note that it may happen that $X(\pi)$ is Hausdorff, but $X^*(\pi^*)$, where X^* is the one point compactification of X, is not. This can be observed in the following

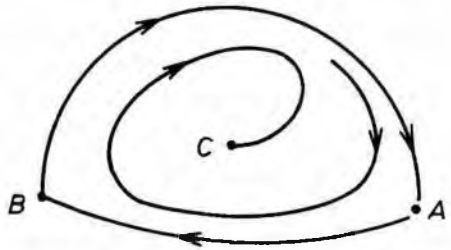
3.9. Example. Consider the semidynamical system on $\mathbb{R}\times(0,\infty)$ with trajectories shown in Fig. 9. The classes of equivalence equal to the segments $\{x\} \times (0, e^{-x}]$ are closed in $X(\pi)$, but not in $X^*(\pi^*)$, as they are not compact and possess the convergent subsequences.
However, the converse holds.

3.10. Proposition. Suppose that the space $X^*(\pi^*)$ is Hausdorff. Then the space $X(\pi)$ is Hausdorff as well.

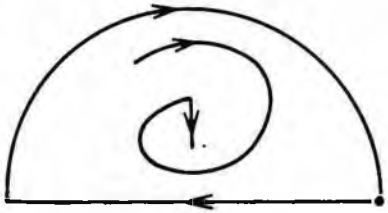
The proof is obvious, so it will be omitted here.



(a)



(b)



(c)

Fig. 8.

Fig. 9.

3.11. Remark. Considering the change of structure from semidynamical systems to dynamical systems, we may compare the investigated properties of the semi-systems and the systems created by them and find the similarities and differences. This seems to be an interesting subject of investigation which will also help to enlarge the knowledge of the structure of semidynamical systems.

Chapter IV

Throughout this section we assume that a semidynamical system on $\mathbb{R}^2$ with $N(x) = \infty$ for each $x \in \mathbb{R}^2$ is given. Semidynamical systems fulfilling these assumptions have many interesting properties (compare [4], [5] and [6]). In particular, the set $F(t, x)$ for a non-stationary point x can be fully characterized. We recall

4.1. Theorem ([4]). For a non-stationary point x there is a t_0 such that the set $F(t, x)$ is a point for $t \leq t_0$ and an arc, i.e., the set homeomorphic to the interval $[0, 1]$, for $t > t_0$.

4.2. Theorem ([6]). If the set $F(t, x)$ is an arc (for a non-stationary point x), then there is a $\delta > 0$ such that $F([t - \delta, t + \delta], x)$ is homeomorphic to the square. The boundary of this set is equal to the union of the arcs $F(t - \delta, x), F(t + \delta, x)$ and two segments of trajectories through the endpoints of $F(t + \delta, x)$. Each non-end point of the arc $F(t, x)$ is an interior point of this set.

Using this, we may state

4.3. Theorem. Assume that $\mathbb{R}^2(\pi)$ is a Hausdorff space. Then for a non-stationary point x the set $\cup\{y : y \in [x]\}$ is a one dimensional manifold (possible with boundary) or a point.

Proof. Denote the set defined above by E and assume that E is not a singleton set. Take a point $y \in E$. Then, according to the definition of the equivalence relation, there is an $s > 0$ and an x such that $y \in F(s, x)$ and $F(s, x)$ is an arc. If y is not an end point of $F(s, x)$, we may take a small δ and (using Theorem 4.2) find a neighbourhood of y, required in the

definition of manifold (as $\{y : y \in [F(\alpha, x)]\} \cap \{y : y \in [F(\beta, x)]\} = \emptyset$ for $|\alpha - \beta| < \delta$). From Theorem 4.1 we have that the set $F(t + s, \pi(t, x))$ is an arc for every $t \geq 0$. Thus (after repeating the above construction, if necessary) we need only consider the case when y is an end point of all arcs $F(t + s, \pi(t, x))$.

Using Schönflies Theorem about the planar homeomorphisms we may assume that $F([s - \delta, s + \delta], x)$ is equal to $[-1, 1]^2$ with $y = (1, 0)$ (see Fig. 10) and $\{1\} \times [-1, 1]$ being a segment of a trajectory through y. Define B_r as $\{(x, y) : (x - 1)^2 + y^2 < r^2, x > 1\}$.

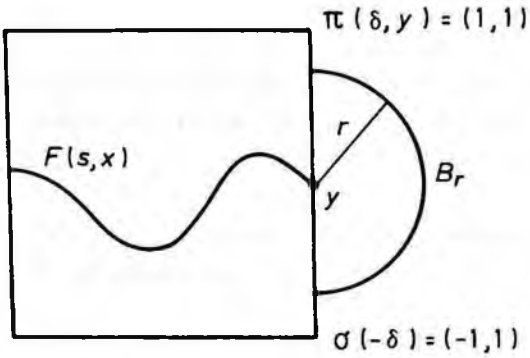


Fig. 10.

Note that for $t_1 < t_2$ we have $F(t_1 + s, \pi(t_1, x)) \subset F(t_2 + s, \pi(t_2, x))$. Using Remark 2.3 we can write the set E as $\phi([0, \infty))$ where ϕ is a continuous function with $\phi(0) = 0$. If there is no $u_n \to \infty$ with $\phi(u_n) \to y$ then $B_r \cap \phi([0, \infty)) = \emptyset$ for some r and the theorem is proved (y is a boundary point). Assume that y is an accumulation point of $\phi(t)$ as $t \to \infty$ (Fig. 11). If there is another accumulation point z of $\phi(t)$, then $z \in E$, because $\mathbb{R}^2(\pi)$ is Hausdorff and E is closed. Thus z is a non-end point of $F(t + s, \pi(t, x))$ for some t and using again Theorem 4.2 we show (as in the beginning of the proof) that there is a neighbourhood U of z with $U \cap E = \phi((\alpha, \beta))(\alpha < \beta < \infty)$. We have proved that z is the only one accumulation point of $\phi(t)$ $(t \to \infty)$, so $\phi(t) \to z$ as $t \to \infty$ and the set $\phi([0, \infty))$ is homeomorphic to the circle. This finishes the proof.

4.4. Remark. The one dimensional manifold may be only in one of four shapes. All of them may be obtained as a class of equivalence. We may

obtain a segment (see Example 2.6, $x = (-1, 0)$), a half-line (see Example 3.3, $x = (0, 0)$), a line (see Fig. 12, $x = (0, 0)$) and a circle (see Fig. 13, $x = (0, 1)$).

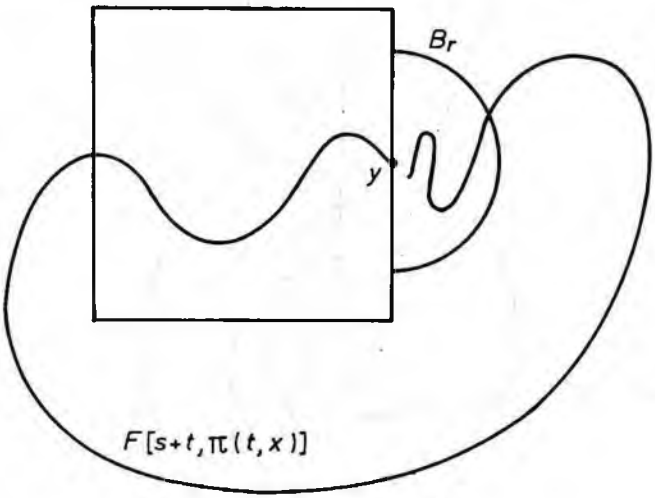


Fig. 11.

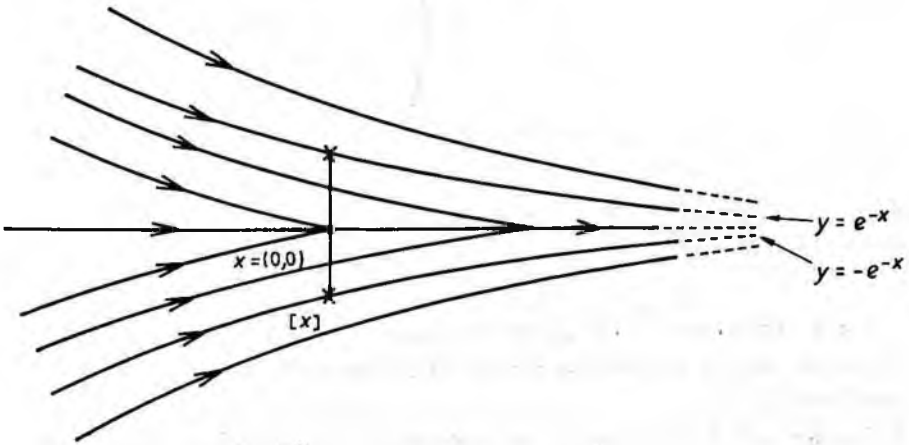


Fig. 12.

Consider again Examples 2.6 and 3.1. In the case of Example 2.6 we got a "good" space and in the system there were many trajectories with all

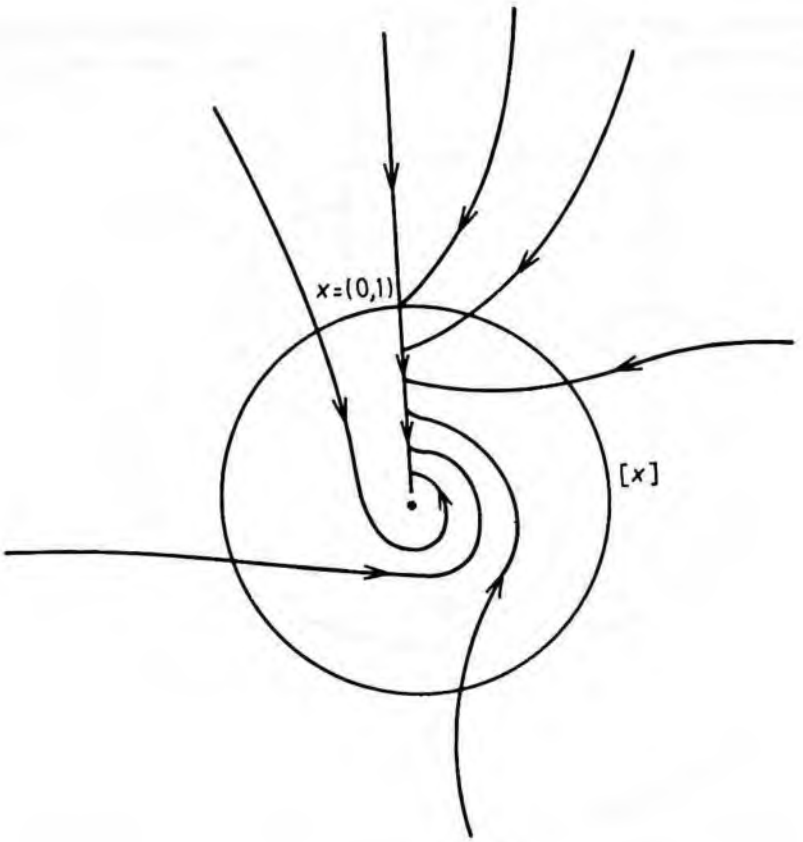


Fig. 13.

the points belonging to them being the points of negative unicity. This is not accidentally. We have

4.5. Theorem. In the planar semidynamical system there is at most countable number of pairwise disjoint trajectories that do not fulfill the condition (4.5.1):

(4.5.1) for each x belonging to the trajectory T we have $[x] = \{x\}$.

Proof. Take a trajectory T which does not fulfill (4.5.1). Thus there are an x and a t with $F(t,x) \cap T \neq \emptyset$ and $F(t,x)$ being an arc. Using

Theorem 4.2 we may find a nonempty subset U of $\mathbb{R}^2$, such that $U \subset F([0, t+\delta], x)$ for some δ. Since then U is contained in the union of elements given by the equivalence classes of the points of T.

When two trajectories are disjoint, then all their equivalence classes are different, which follows easily from Definition 2.1. This means that the open sets, constructed for different trajectories as above, are disjoint. There is at most countable number of such sets in the plane, so the number of trajectories fulfilling (4.5.1) is at most countable.

4.6. Remark. Note that the set of negative unicity points in planar semidynamical system is large, as the set of non-unicity points is of first Baire category (compare [4]).

The following theorem shows that from a particular property of the obtained system we may conclude about the good structure of a given semi-system on a big set.

4.7. Theorem. Assume that there is a nonempty open subset U of $\mathbb{R}^2(\pi)$ homeomorphic with an open subset V of $\mathbb{R}^2$. Then there is an invariant set K of the second Baire category such that for each $x \in K$ the point x is of negative unicity and the system $(K, \mathbb{R}, \pi_{|\mathbb{R}\times K})$ is a dynamical system (for a negative t we put as $\pi(t, y)$ the unique point of $F(t, y)$).

Proof. Define K as the set of all points from the trajectories not fulfilling (4.5.1). All the points of K are of negative unicity and thus $\pi(t, x)$ is defined for all $x \in K$ and $t \in \mathbb{R}$. Using Proposition 2.9 from [3] and considering the restriction of π to $\mathbb{R}_+ \times K$ we get that π is continuous on $\mathbb{R}\times K$ (with $\pi(t, x)$ defined for $t < 0$ as above) and so the system is dynamical. We need only to show that K is of second Baire category.

Take the function $\phi : \mathbb{R}^2 \ni x \to [x] \in \mathbb{R}^2(\pi)$, which is continuous. The set $W = \phi^{-1}(U)$ is open in $\mathbb{R}^2$. We prove that $\phi_{|K\cap W}$ is a homeomorphism onto its image. It is enough to prove that $\phi^*_{|(K\cup\{\infty\})\cap W}$ defined on the one point compactification space is a homeomorphism onto its image. Of course $\phi^*_{|(K\cup\{\infty\})\cap W}$ is continuous and bijective (according to the definition of K). Take $[x_n] \subset \phi^*(K \cup \{\infty\}) \cap U, [x_n] \to [x_0] \in \phi^*(K \cup \{\infty\}) \cap U$. We have to show that $x_n \to x_0$.

Suppose to the contrary that there is a subsequence $\{x_{n_k}\}$ of $\{x_n\}$

not tending to x_0. By the compactness of $\mathbb{R}^2 \cup \{\infty\}$ we may assume that $x_{n_k} \to y$. Then $[x_{n_k}] \to [y]$. But $[x_0] \in U$ and $[x_{n_k}] \in U$ for large k, the last set being a Hausdorff space, so $[y] = [x_0]$ and $y = x_0$.

Now consider $N = \mathbb{R}^2 \backslash K$ and $\phi(N) \cap U$. According to Theorem 4.5 the set $\phi(N)$ has at most countable number of trajectories. Let us say that they are given by $\{y_n\}$. The set $\phi(N) \cap U = U \cup \bigcup\{\pi([k, k+1], y_n) : k \in \mathbb{Z}, n \in \mathbb{N}\}$ is of first Baire category in U, so $\phi(K) \cap U$ is of second Baire category in U. We have shown that $\phi_{|K \cap W}$ is a homeomorphism, so if $K \cap W$ was of first category then also $\phi(K \cap W) = U \cap \phi(K)$ would be of first category, which is a contradiction. Thus $W \cap K$ is of second category in W. Since then the set K must be of second Baire category in $\mathbb{R}^2$, which finishes the proof.

Acknowledgment

The author wishes to thank Professor Themistocles M. Rassias for his kind invitation to participate in this memorial volume dedicated to Constantin Carathéodory.

References

1. N. P. Bhatia and O. Hajek, *Local Semidynamical Systems*, Lecture Notes in Mathematics **90**, Springer-Verlag, 1969.
2. N. P. Bhatia and G. P. Szegō, *Stability Theory of Dynamical Systems*, Springer-Verlag, 1970.
3. K. Ciesielski, *Continuity in semidynamical systems*, Annales Polonici Mathematici **46** (1985), 61–70.
4. K. Ciesielski, *Kneser type theorems and Baire properties for planar semidynamical systems*, in Differential Equations: Stability and Control, edited by S. Elaydi, to be published by Marcel Dekker in 1990.
5. K. Ciesielski, *The Jordan Curve Theorem for the funnel in a planar semidynamical system*, University of Warwick, preprint.
6. K. Ciesielski, *The topological characterization of the section of a funnel in a planar semidynamical system*, University of Warwick, preprint.
7. S. Elaydi, *Semidynamical systems with nonunique global backward extensions*, Funkcialaj Ekvacioj **26** (1983), 173–187.
8. S. Elaydi, *Semidynamical systems with nonunique global backward extensions II: the negative aspects*, Funkcialaj Ekvacioj **27** (1984), 85–100.

9. S. Elaydi and S. K. Kaul, *Semiflows with global extensions*, Nonlinear Analysis **10** (1986), 713–726.
10. O. Hajek, *Dynamical Systems in the Plane*, Academic Press, 1968.
11. D. Henry, *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Mathematics **840**, Springer-Verlag, 1981.
12. K. Kuratowski, *Topology*, vol. II, New York, 1968.
13. R. C. McCann, *Negative escape time in semidynamical systems*, Funkcialaj Ekvacioj **20** (1977), 39–47.
14. A. Pelczar, *General Dynamical Systems* (in Polish), Lecture Notes of the Jagiellonian University **293**, Kraków 1978.
15. A. Pelczar, *Remarks on removable instabilities of sets in pseudo-dynamical semisystems*, Ahmadu Bello University Report, Zaria, 1/1981.
16. T. M. Rassias, *Foundations of Global Nonlinear Analysis*, Teubner-Texte zur Mathematik, Band 86, Leipzig 1986.
17. K. P. Rybakowski, *The Homotopy Index and Partial Differential Equations*, Springer-Verlag, 1988.
18. S. H. Saperstone, *Semidynamical Systems in Infinite Dimensional Spaces*, Applied Mathematical Sciences **37**, Springer-Verlag, 1981.
19. R. Srzednicki, *On funnel sections of local semiflows*, Bulletin of the Polish Academy of Sciences, Mathematics **34** (1986), 203–209.
20. K. S. Sibirskij, A. S. Šube, *Poludinamičeskije sistemy*, Topologičeskaja teorija (in Russian), Štiinca 1987.

Krzysztof Ciesielski
Jagiellonian University
Mathematics Institute
Reymonta 4, 30–059 Kraków
Poland

Constantin Carathéodory: An International Tribute (pp. 142–180)
edited by Th. M. Rassias

THE PROBLEM OF THE LOCAL SOLVABILITY OF THE LINEAR PARTIAL DIFFERENTIAL EQUATIONS

Andrea Corli and Luigi Rodino

1. INTRODUCTION

We want to present a survey of the problem of the local solvability, reviewing shortly the main results obtained in these last thirty years; emphasis will be given on methods involving canonical transformations and Fourier integral operators.

Let us consider a linear partial differential operator of order m:

$$(1.1) \qquad P = \sum_{|\alpha|\le m} c_\alpha(x) D^\alpha,$$

with C^∞ coefficients $c_\alpha(x)$ defined in an open subset Ω of R^n; the notations are standard: $\alpha=(\alpha_1,\dots\alpha_n)\in Z^n_+$, $D^\alpha=D^{\alpha_1}_{x_1}\dots D^{\alpha_n}_{x_n}$ where $D_{x_j}=-i\partial_{x_j}$, and $|\alpha|=\alpha_1+\dots+\alpha_n$. Set the following definition.

Definition 1.1. *The operator* P *is said to be locally solvable at* $x_0\in\Omega$ *if there exists a neighborhood* U *of* x_0, *contained in* Ω, *such that for every* $f\in C^\infty(\Omega)$ *there is a Schwartz distribution* $u\in\mathcal{D}'(\Omega)$ *solution of the equation* Pu=f *in* U.

The problem of the local solvability plays an important rôle in the theory of the linear partial differential operators; precisely: one wants to find necessary and/or sufficient conditions on the symbol

$$(1.2) \qquad p(x,\xi) = \sum_{|\alpha|\leq m} c_\alpha(x)\xi^\alpha,$$

in order that the corresponding operator P is locally solvable at a given $x_0 \in \Omega$.

Let us first observe that all the linear partial differential operators with constant coefficients are locally solvable at any point $x_0 \in \mathbf{R}^n$, in view of the existence of a fundamental solution. On the other hand, if we assume the coefficients $c_\alpha(x)$ are analytic in Ω, and there exists a multi-index α, $|\alpha|=m$, such that $c_\alpha(x_0)\neq 0$, then we can deduce, as a consequence of the Cauchy-Kovalevsky theorem, that for every analytic function f, defined in a neighborhood of x_0, there exist another neighborhood V of x_0 and an analytic solution u in V of the equation Pu=f.

This solvability result is not valid when we require f to be only an indefinitely differentiable function, even maintaining the analyticity of the coefficients. The first example of a non-locally-solvable operator was given in 1957 by Lewy [1], who proved that the equation

$$(1.3) \qquad D_{x_1}u+iD_{x_2}u+i(x_1+ix_2)D_{x_3}u=f$$

has no C^1 solution u in a neighborhood of the origin in $\mathbf{R}^3$ (actually, no distribution-solution in any nonvoid open subset Ω of $\mathbf{R}^3$), for suitable functions $f \in C^\infty(\mathbf{R}^3)$. In fact, by means of elementary computations one may check that, if (1.3) is satisfied in a neighborhood of the origin by some $u \in C^1(\mathbf{R}^3)$ for a C^∞ real-valued function f depending on the variable x_3 only, then f must be analytic there.

The example of Lewy was the starting point for the study of large classes of linear partial differential equations, which do not admit local solutions; among the first works we quote Hörmander [3], Mizohata [1] and Nirenberg and Treves [1].

Hörmander [3] (1960) gave a necessary condition for local solvability, involving the principal symbol of the operator P:

$$(1.4) \qquad p_m(x,\xi) = \sum_{|\alpha|=m} c_\alpha(x)\xi^\alpha;$$

precisely, if we have

$$(1.5) \qquad p_m=0, \quad \{p_m, \overline{p_m}\}\neq 0$$

at some point $x=x_0$, $\xi=\xi_0$, then P is not locally solvable at x_0. Hörmander based the proof on the following remark: if P is locally solvable at x_0 then there exist a neighborhood V of x_0 and constants C, k and N such that for all $f,v\in C_0^\infty(V)$

$$(1.6)\qquad \left|\int fv\,dx\right| \le C\sum_{|\alpha|\le k}\sup|D^\alpha f|\cdot\sum_{|\beta|\le N}\sup|D^\beta({}^tPv)|,$$

where the formal adjoint (or transposed) ${}^tP$ is defined by the identity

$$\int_\Omega \psi\,{}^tP\phi\,dx = \int_\Omega \phi P\psi\,dx, \qquad \phi,\psi\in C_0^\infty(\Omega).$$

One is then reduced to prove that (1.6) is contradicted by suitable sequences of test functions, if (1.5) is valid.

Condition (1.5) is satisfied by the Lewy example (1.3). There exist unsolvable operators which do not satisfy (1.5); in fact, Mizohata [1] (1962) proved that the operator in $\mathbf{R}^2$

$$(1.7)\qquad M=D_{x_1}+ix_1^hD_{x_2}$$

is not locally solvable at the points of the x_2-axis if h is any odd integer, whereas (1.5) holds if and only if h=1.

The nonsolvability of M follows easily from Hörmander's estimates (1.6), and it can be also directly shown if we limit ourselves to consider C^1 solutions in neighborhoods V of the origin. In fact, for any fixed neighborhood V, let us take $f\in C_0^\infty(V)$, $f(x_1,x_2)\ge 0$ with support in the half plane $x_1>0$. Assuming the existence of a solution of Mu=f in V, we decompose u in its odd and even part with respect to the variable x_1. Letting M act on $u=u_{odd}+u_{even}$, with odd h, it is easy to conclude that $\int f(x_1,x_2)dx_1dx_2=0$, which contradicts the assumptions on f (cf. Grushin [2]).

Starting from the example of Mizohata, in 1970 Nirenberg and Treves [2], [3] gave a general necessary (and sufficient) condition for the local solvability of all the operators of principal type, i.e. the operators for which $d_\xi p_m(x,\xi)\neq 0$ where $p_m(x,\xi)=0$. Their result will be presented in Section 2; we shall also give there a cheap proof, using the Fourier integral operators of Hörmander [7] and the related theory of the canonical transformations of Carathéodory [1].

Let us now turn attention to sufficient conditions for local solvability. It will be convenient to recall first the definition of hypoellipticity.

Definition 1.2. *The operator* P *in* (1.1) *is said to be hypoelliptic in* Ω *if*

$$\text{singsupp } Pu = \text{singsupp } u, \qquad \textit{for all } u \in \mathcal{D}(\Omega);$$

i.e. all solutions u *of* Pu=f *are* C^∞ *where* f *is* C^∞.

The elliptic operators are hypoelliptic, in view of the well known theorem on the regularity of their solutions; other relevant examples of hypoelliptic operators are given by the heat operator and other parabolic-type operators. The Mizohata operator M is hypoelliptic if h in (1.7) is even.

Hypoellipticity and local solvability are closely related; in fact, we have that *the hypoellipticity of the formal adjoint* ${}^tP$ *in a neighborhood of* $x_0 \in \Omega$ *implies the local solvability of* P *at the same point* (Treves [1], Theorem 52.2, or Yoshikawa [1]). So, for example, the elliptic operators are locally solvable.

Observe that also the strictly hyperbolic operators are locally solvable, since we can always solve the Cauchy problem for the related nonhomogeneous equation.

Operators of principal type with real-valued principal symbols have been proved to be locally solvable by Hörmander [2] (1960); this class contains both the strictly hyperbolic operators and the elliptic operators with real coefficients. The result of Hörmander [2] will be stated and proved in Section 2, following the microlocal presentation of Duistermaat and Hörmander [1]. The techniques of the Fourier integral operators, canonical tranformations and pseudo-differential operators will be also shortly reviewed in Section 2. We shall not dwell too long upon operators of principal type, since there are already several survey papers about them, referring at different times on progresses made in the researches: Treves [2],[3], Egorov [4], Treves [6], Dieudonné [1]. Moreover there are the books of Egorov [5] and, above all, Hörmander [10].

On the other hand, there is no such a survey work concerning the field of operators with *multiple* characteristics, i.e., operators which are not of principal type. There have been many contributions in this area, during the last twenty years; some results have been proved independently at different times, and some noteworthy intersections deserve to be pointed out. So we think it is useful to collect local solvability results about this kind of operators; this will be the content of Section 3.

We shall address ourselves mainly to the results on *nonsolvability*. Sufficient

conditions for solvability can be deduced from theorems of hypoellipticity, as already observed, or directly from the existence of a right parametrix; results of this type for operators with multiple characteristics will be left outside for lack of space. For the same reason we shall not discuss solvability and nonsolvability of systems of equations nor operators on Lie groups; the study of the range of unsolvable operators will be also omitted.

Under these restrictions, our survey will be certainly not exhaustive of all the existing works; however several lines of research will be discussed in detail, especially for the case in which the characteristics are double. Under the same restrictions the bibliography contains almost all papers we know about unsolvability results.

Section 4 will be devoted to the study of the local solvability in the frame of the Gevrey classes; relevant references in this connection are the former works of the authors, Rodino [2], [3], Corli [1], [2], Rodino and Corli [1].

2. CANONICAL TRANSFORMATIONS AND OPERATORS OF PRINCIPAL TYPE.

Changes of variables $y=y(x):\Omega_x \to \Omega'_y$, with C^∞ inverse $x=x(y)$, are often used in the study of the local solvability. In fact, the operator P is solvable at x_0, if and only if the corresponding operator P' in the new variables y is solvable at $y_0=y(x_0)$; on the other hand the principal symbol $p'_m(y,\eta)$ of P' is given by

(2.1) $$p'_m(y,\eta) = p_m(x(y),[(\partial x/\partial y)^t]^{-1}\eta),$$

where $p_m(x,\xi)$ is the principal symbol of P, and a suitable choice of $y=y(x)$ may then lead to easy espressions of $p'_m(y,\eta)$. We want to show how we may reduce ourselves, more generally, to an operator P' with principal symbol

(2.2) $$p'_m(y,\eta) = p_m(\chi(y,\eta))$$

where now $(x,\xi)=\chi(y,\eta)$ is an arbitrary C^∞ transformation involving both x and ξ, which is assumed to be homogeneous with respect to the dual variables, and *canonical*, i.e. preserving the symplectic two-form:

$$\sum_{j=1}^{n} dy_j \wedge d\eta_j = \sum_{j=1}^{n} dx_j \wedge d\xi_j .$$

A general study of the canonical tranformations was given by Carathéodory [1]; in

particular, it is proved there that a canonical tranformation is always "generated" by a function $\omega(x,\eta)$; precisely, assume $\omega(x,\eta)$ is C^∞ homogeneous with respect to η of degree 1, and suppose also

$$\det \partial^2\omega(x,\eta)/\partial x\partial\eta \neq 0,$$

then, setting $\xi=\omega_x(x,\eta)$ and $y=\omega_\eta(x,\eta)$ and solving with respect to (x,ξ) or (y,η), we obtain a homogeneous canonical tranformation χ. In the opposite direction, given χ, we may find the generating function ω. Observe that a change of variables $y=y(x)$ corresponds to the canonical transformation in (2.1) generated by $\omega(x,\eta)=y(x)\eta$.

We are now able to give a precise meaning to P', transformed of P under the canonical transformation χ. As in Hörmander [7], Egorov [2], we begin by defining the Fourier integral operator

$$(2.3) \qquad Ef(x) = (2\pi)^{-n} \int e^{i\omega(x,\eta)} b(x,\eta) Ff(\eta) d\eta$$

where the phase $\omega(x,\eta)$ is the generating function of χ and the amplitude $b(x,\eta)$ is assumed for the moment $=1$ in R^n; Ff denotes the Fourier transformation of f. The operator E has an inverse E^{-1} which can be expressed again as Fourier integral operator. Setting then

$$(2.4) \qquad P' = E^{-1}PE$$

we obtain

$$P'f(y) = (2\pi)^{-n} \int e^{iy\eta} p'(y,\eta) Ff(\eta) d\eta$$

where the principal symbol $p'_m(y,\eta)$ of $p'(y,\eta)$ is given by the formula (2.2). Observe that $p'(y,\eta)$ is not any more polynomial with respect to η, in general, but it can be expressed by means of an asymptotic expansion

$$(2.5) \qquad p'(y,\eta) \approx \sum_{j=0}^{\infty} p'_{m-j}(y,\eta),$$

where $p'_{m-j}(y,\eta)$ is homogeneous in η of degree m-j. Thus P' in (2.4) must be considered as a *classical pseudo-differential operator* (Hörmander [5], Kohn and Nirenberg [1]). The lower order terms $p'_{m-1}(y,\eta)$, $p'_{m-2}(y,\eta)$, ... in (2.5) can be computed explicitly basing on the symbol $p(x,\xi)$ of P and on the amplitude $b(x,\eta)$ in (2.3), for which we assume in general an asymptotic expansion of the type (2.5)

$$b(x,\eta) \approx \sum_{j=0}^{\infty} b_{M-j}(x,\eta),$$

with $b_M(x,\eta)\neq 0$. To obtain an easy expression for P' we may then take advantage of the choice of χ, acting on (2.2), and $b(x,\eta)$, possibly simpliflying the lower order terms.

How the properties of P, say hypoellipticity and local solvability, are connected to those of P'? A further difficulty in the study of the conjugation (2.4), with respect to an ordinary change of variables, comes from the fact that the canonical transformation χ is often defined only locally with respect to the dual variables, acting from a conic neighorhood Γ of a point $(x_0,\xi_0)\in T^*(\Omega)\backslash 0$ into a conic neighborhood Γ' of $\chi^{-1}((x_0,\xi_0))$ $=(y_0,\eta_0)\in T^*(\Omega)\backslash 0$. Hörmander [7] solves this problem by introducing the *wave front set* of the distribution f, which describes the singularities of f from a microlocal point of view (i.e. locally also in ξ). One may base consequently on definitions of *micro-hypoellipticity* and *micro-solvability*, which are invariant under conjugation by Fourier integral operators.

We shall not give here details, in this connection, but observe that *if we can prove that* P' *is not (micro)locally solvable at* (y_0,η_0), *we may transfer the result by means of* (2.4) *to* (x_0,ξ_0) *and conclude that* P *is not locally solvable at* x_0.

Another important remark, coming from the theory of the pseudo-differential operators, is that *hypoellipticity and local solvability are also invariant under multiplication by elliptic factors*, i.e., arguing on principal symbols, we are always allowed to replace $p_m(x,\xi)$ by $q_M(x,\xi)p_m(x,\xi)$, for any $q_M(x,\xi)\neq 0$.

Let us present some applications of the preceding arguments to the linear partial differential (or pseudo-differential) operators of principal type, i.e. the operators with principal symbol $p_m(x,\xi)$ satisfying

(2.6) $d_\xi p_m(x,\xi)\neq 0$ *for all* (x,ξ) *in the characteristic manifold* $\Sigma=\{(x,\xi)\in T^*(\Omega)\backslash 0;$ $p_m(x,\xi)=0\}$.

Theorem 2.1. *Let* P *be an operator of principal type, with real-valued principal symbol* $p_m(x,\xi)$, *in a conic neighborhood* Γ *of* $(x_0,\xi_0)\in\Sigma$; *then* P *is (micro)locally solvable at* (x_0,ξ_0).

A local version of this theorem was first proved by Hörmander [2] using other techniques. Here, assuming $\partial_{\xi_n} p_m(x,\xi)\neq 0$ in Γ, by the implicit function theorem we write first

$$p_m(x,\xi) = q_{m-1}(x,\xi)(\xi_n - a(x,\xi')),$$

where $q_{m-1}(x,\xi)$ is elliptic of order m-1 and $a(x,\xi')$ is of order 1, with $\xi'=(\xi_1,...\xi_{n-1})$. As we observed before, we may ignore the elliptic factor $q_{m-1}(x,\xi)$ and limit ourselves to the study of the operator

$$P' = D_{x_n} - a(x,D').$$

Let us prove that there exists a Fourier integral operator E such that

(2.7) $$E^{-1}P'E = D_{y_n};$$

the theorem will be then proved, since D_{y_n} is obviously a solvable operator. Actually, it will be sufficient to fix as phase function

$$\omega(x,\eta) = x_n\eta_n + \omega_0(x,\eta')$$

where $\omega_0(x,\eta')$ is solution of the Hamilton-Jacobi equation

$$\partial_{x_n}\omega_0 - a(x,d_x\omega_0) = 0,$$

$$\omega_0|_{x_n=0} = x'\eta'.$$

We obtain in particular for the canonical transformation χ generated by ω:

$$\eta_n = \xi_n - a(x,\xi').$$

In view of (2.2) we may then conclude that the principal symbol of $E^{-1}P'E$ is given by η_n; the lower order terms in (2.5) can be eliminated in this case by means of a suitable choice of the amplitude function $b(x,\eta)$ of E.

In Duistermaat and Hörmander [1] the identity (2.7) allows also a precise description of the singularities of the solutions u of the equation $Pu=f\in C^\infty$: these (micro)singularities propagate along the bicharacteristics of $p_m(x,\xi)$, i.e. the integral curves of the Hamilton field H_{p_m} on the characteristic manifold Σ (observe that the bicharacteristics are invariant under canonical transformations and multiplication by elliptic factors).

Let us now consider pseudo-differential operators P of principal type with complex-valued principal symbol

$$p_m(x,\xi) = Re\,p_m(x,\xi) + iIm\,p_m(x,\xi),$$

defined in a conic neighborhood Γ of $(x_0,\xi_0)\in\Sigma$. Possibly by multiplying by i and shrinking Γ, we may assume $d_\xi Re\,p_m\neq 0$ in Γ. Let us write γ_0 for the bicharacteristic of $Re\,p_m$ through (x_0,ξ_0).

Theorem 2.2. *Under the preceding assumptions on* p_m, *suppose that* $Im\,p_m$ *changes sign at* (x_0,ξ_0) *from - to + moving in the positive direction on* γ_0; *then* P *is not locally solvable at* x_0.

The proof is very easy using canonical transformations, if we assume further that Imp_m vanishes of finite order at (x_0,ξ_0) on γ_0 (suppose for example $p_m(x,\xi)$ is an analytic function). In fact then, on every bicharacteristic of Rep_m nearby there must be a zero (y_0,η_0) where the same change of sign occurs, and we may choose it so that the order of the zero is minimal. In a conic neighborhood Γ' of (y_0,η_0) we have:

(i) *in* Γ' *the characteristic manifold* Σ *is a smooth submanifold of codimension* 2;

(ii) *on* $\Sigma\cap\Gamma'$ *we have* $H^j_{Rep_m} Imp_m=0$ *for* $j<k$ *and* $H^k_{Rep_m} Imp_m\neq 0$, *where* k *is a fixed odd integer.*

Under (i) and (ii) it is easy to construct a canonical transformation χ such that

$$p_m(\chi(y,\eta)) = \eta_1 \pm i y_1^k \eta_2, \text{ modulo elliptic factors.}$$

The lower order terms in the expression of $E^{-1}PE$ can be eliminated by choosing a suitable amplitude for E. We are then reduced to prove the theorem for the Mizohata operator which, as we observed in the Introduction, is not locally solvable if k is odd. The case when Imp_m vanishes of infinite order at (x_0,ξ_0) is more delicate; let us refer to Hörmander [9] for the proof.

If P in Theorem 2.2 is a linear partial differential operator, then at the point $(x_0,-\xi_0)$ the principal symbol $p_m(x,\xi)$ satisfies the same assumptions, but with a change of sign from + to -. Note that same assumptions with a change from + to - are also satisfied by the formal adjoint ${}^tP$ at (x_0,ξ_0). Therefore we have the following

Corollary 2.3. *Let* P *be a linear partial differential operator of principal type. With the preceding notations, if* Imp_m *changes sign at* (x_0,ξ_0) *on* γ_0 *then* P, ${}^tP$ *are not locally solvable at* x_0.

In the opposite direction, if there are not changes of sign along the bicharacteristics, then P is locally solvable (see Nirenberg and Treves [3] and Beals and Fefferman [1]). However, it is not yet known if the converse of Theorem 2.2 is valid (but in two dimensions Lerner [1] has proved that the answer is positive).

As a conclusion of this Section let us just quote, for sake of completeness, some other works dealing with these topics: Beals and Fefferman [2], Menikoff [1], Egorov and Popivanov [1], Treves [7], Moyer [1], Hörmander [8], Lu [1].

3. OPERATORS WITH MULTIPLE CHARACTERISTICS

In the former section we saw how the very early seventies marked a fixed point in the theory of local solvability. On one hand the works of Nirenberg, Treves and Egorov gave satisfactory results for operators of principal type; on the other hand, the technique of Fourier integral operators introduced by Hörmander fournished a powerful tool, capable of wide applications. There were then essentially two great open problems: to establish the converse of Theorem 2.2 of Nirenberg and Treves and to explore the field of operators with multiple characteristics. The first problem appeared soon very hard and will be not discussed here: see the last lines of the former Section and the Introduction.

As regards operators with multiple characteristics, the huge amount of papers related with them prevent us from giving an account of them all, especially for those dealing with sufficient conditions. Then, as already said in the Introduction, we limit ourselves to refer mainly about *necessary* conditions, considering in this section also some works concerned with operators with simple characteristics, but which are not of principal type in the previous sense. The proofs of the results which follow will be not reported; in most cases they are variants of Hörmander's [3] method (see Introduction).

All pseudo-differential operators considered below are assumed to be classical. They will be usually denoted with P; p will be then the symbol of P.

Probably the first example of an unsolvable operator with multiple characteristics is due to Grushin [3] (see also Ivrii [1] for results based on Hörmander's [3] condition), who proves that the operator, in $\mathbf{R}^2$,

(3.1) $$\partial_t^2 + t^2\partial_x^2 + i\lambda\partial_x$$

is hypoelliptic and locally solvable if and only if

(3.2) $$\lambda \neq \pm(2n+1), \quad n=0,1,2,\ldots.$$

Thus, while in the principal type case solvability is decided only by the principal part of the operator, when the characteristics are higher the lower order terms may play a fundamental rôle. For a bit more general operators the same result was found independently by Gilioli and Treves [1].

Theorem 3.1 (Gilioli and Treves [1]). *Let* P *be the differential operator*

(3.3) $$P(t,D_t,D_x) = (D_t - iat^kD_x)(D_t - ibt^kD_x) + ct^{k-1}D_x,$$

where $(t,x)\in \mathbf{R}^2$, a,b,c *are real numbers*, k *is an odd integer.* P *is not locally solvable at* (0,0) *if and only if one of the following conditions is satisfied:*

(i) $ab>0$;

(ii) $ab\le 0$, $a\neq 0$ *and, for some integer* $n\ge 0$, $c/(a-b) - n(k+1)=0$ *or* 1;

(iii) $ab\le 0$, $b\neq 0$ *and, for some integer* $n\ge 1$, $c/(a-b) + n(k+1)=0$ *or* 1.

With a=-b=1, c=λ-1, k=1, from (3.3) we recover (3.1). To understand the meaning of condition (3.2) (and of (ii), (iii) in the theorem above), we must discuss a bit Grushin[1], [3] approach to differential operators with polynomial coefficients, at least in a particular case. Denote by P the operator

$$(3.4)\qquad P(t,D_t,D_x) = \sum_{|\alpha+\beta|\le m;\ |\gamma|\le m\delta} a_{\alpha\beta\gamma} t^{\gamma} D_t^{\alpha} D_x^{\beta},$$

where $a_{\alpha\beta\gamma}$ are complex numbers, δ is real positive such that mδ is integer, $t\in \mathbf{R}^{n-1}$, $x\in \mathbf{R}$. We assume that the following conditions of quasihomogeneity and ellipticity are satisfied:

$$(3.5)\qquad P(t/\lambda,\lambda\tau,\lambda^{1+\delta}\xi) = \lambda^m P(t,\tau,\xi) \quad \textit{for every } \lambda>0;$$

$$(3.6)\qquad P^{\circ}(t,\tau,\xi) := \sum_{|\alpha+\beta|=m;\ |\gamma|=m\delta} a_{\alpha\beta\gamma} t^{\gamma} \tau^{\alpha} \xi^{\beta} \neq 0 \ \textit{for every } t\neq 0,\ \xi\in \mathbf{R},\ \tau\in \mathbf{R}^{n-1}:\ |\tau|+|\xi|>0.$$

The main result of Grushin [1] is then that, under these hypotheses, P is hypoelliptic if and only if

$$(3.7)\qquad \ker(P(t,D_t,\xi))\cap \mathcal{S}(\mathbf{R}^{n-1}) = 0 \quad \textit{for every } |\xi|=1.$$

The following standard application of this result occurs rather frequently: to prove that an operator P is locally solvable, one is reduced to show that (3.7) holds for the transposed operator.

Substituting ∂_x with $i\xi$ in (3.1) we get, when $|\xi|=1$, the ordinary differential operator

$$d^2/dt^2 - t^2 - \lambda \operatorname{sign}\xi.$$

Now it is known that the eigenvalues (in $L^2(\mathbf{R})$) of $d^2/dt^2 - t^2$ are $\lambda_n=-(2n+1)$, whose eigenfunctions (Hermite functions) are in $\mathcal{S}(\mathbf{R})$. Hence the operator in (3.1) is not hypoelliptic if $\lambda=\pm(2n+1)$, $n\in \mathbf{Z}_+$. The unsolvability for the values of λ in (3.2) is easily proved directly, by contradicting (a variant of) inequality (1.6). This gives, a fortiori, nonhypoellipticity for $\lambda=\pm(2n+1)$, $n\in \mathbf{Z}_+$.

In general, when the condition in (3.7) does not hold we have the following more recent result.

Theorem 3.2 (Popivanov [8]). *Let* P *be the operator in* (3.4), *satisfying* (3.5), (3.6). *If* $\ker(P(t,D_t,\xi))\cap\mathcal{S}(\mathbf{R}^{n-1})\neq 0$ *for some real* ξ, $|\xi|=1$, *then* ${}^tP$ *is not locally solvable at* 0.

This result, jointly with Grushin [1], [3], contains Theorem 3.1.

When k in (3.3) is an even integer one gets similar result of unsolvability; a bit more generally let us suppose that a,b,c are complex numbers.

Theorem 3.3 (Menikoff [2]). *Let* P *be the operator in* (3.3), *with* k *an even integer,* a,b,c *complex numbers,* $Re\,a\cdot Re\,b\neq 0$. *Then* P *is not locally solvable at* (0,0) *if and only if* $Re\,a\cdot Re\,b<0$ *and* $c/(a-b) - n(k+1)=1/2$ *for some integer* n.

Moreover Menikoff proves that, for these operators, local solvability is equivalent to hypoellipticity. The proofs are inspired by the above works of Grushin, Gilioli and Treves.

In Menikoff [3] are studied pseudo-differential operators which can be reduced microlocally to operators P of the following type, generalizing those in (3.3):

(3.8) $P(t,x,D_t,D_x)=(D_t+it^k\alpha(t,x,D_x))(D_t+it^k\beta(t,x,D_x))+t^l\gamma(t,x,D_x)+h(t,x,D_t,D_x)D_t$.

Here α,β,γ are first-order elliptic pseudo-differential operators in a conic neighborhood of $\theta=(0,0;\xi_0)$, h is of order zero, $t\in\mathbf{R}$, $x\in\mathbf{R}^n$, $Re\,\alpha\neq 0$, $Re\,\beta\neq 0$; k,l are nonnegative integers. While the papers of Gilioli and Treves [1] and Menikoff [2] were concerned with the case l=k-1, in Menikoff [3] the case l<k-1 is taken into account. The unsolvability result there proved, expressed for operators in the canonical form (3.8), is the following.

Theorem 3.4 (Menikoff [3]). *Let* P *be as in* (3.8) *with* $l\leq 2$ *and assume that* $Re\,\alpha(t,x,\xi)\cdot Re\,\beta(t,x,\xi)<0$ *in a conic neighborhood of* θ. *Define*

$$D(s) = ((\alpha(\theta)-\beta(\theta))^2 s^{2k} + 4\gamma(\theta)s^l$$

and suppose that

(i) $D(s)\in\mathbf{R}$, *for every* $s\in\mathbf{R}$;

(ii) *there exists* $s_0\in\mathbf{R}$ *such that* $D(s_0)<0$.

Then the operator P *is not locally solvable at* 0.

The meaning of D(s) and of the conditions (i),(ii) becomes more clear if we make an *asymptotic change of variables* (a *symplectic dilatation* in Hörmander's terminology).

This technique is introduced in Ivrii [1] and plays a fundamental rôle in the work of Ivrii and Petkov [1], as well as in many other papers on local unsolvability. Let us assume, for simplicity, that $n=1$, $\xi_0=1$, and $\alpha(t,x,D_x)=\alpha D_x$, $\beta(t,x,D_x)=\beta D_x$, $\gamma(t,x,D_x)=\gamma D_x$, with α,β,γ real numbers. Let ρ be a large positive parameter and set $s=\rho^{\lambda}t$, $y=\rho^{\mu}x$, with $\lambda=1/(2k-2l-2)$, $\mu=\lambda(l+2)$ (we make this choice of λ and μ in order to "weight" the first-order term containing γ as the second-order principal part). Call Q the adjoint operator of P after this change of variables and look for an asymptotic solution u of the homogeneous equation $Qu=0$ under the form $u=v_\rho\cdot\exp(i\rho y+i\rho^{1/2}w)$. Then

$$Qu=-\rho^{(2k-2l)\lambda}u[(w_s)^2-is^k(\alpha+\beta)w_s+s^l\gamma+s^{2k}\alpha\beta]+o(\rho^{(2k-2l)\lambda}). \tag{3.9}$$

If we want that the coefficient of $\rho^{(2k-2l)\lambda}$ vanishes, we must choose w in order that w_s is a root of the second-order polynomial in brackets in (3.9): its discriminant is just $-D(s)$.

The statement of the above result in terms of operators not already reduced to the canonical form (3.8) is a bit cumbersome; for it and for other results concerning constructions of parametrices and hypoellipticity, we refer the reader to the quoted paper (see also Yamamoto [1]). As an example let us consider the operator, in $\mathbf{R}^2$,

$$R = D_t^2 + t^{2k}D_x^2 + \lambda t^l D_x,$$

with $l<k-1$. If $Im\lambda\neq 0$ then R is hypoelliptic (and then locally solvable). If $\lambda\in\mathbf{R}\setminus\{0\}$ then R is not hypoelliptic, and if moreover $l\leq 2$ then it is not locally solvable at 0.

At last let us remark that the heavy restriction $l\leq 2$ in Theorem 3.4 is merely technical: it reflects the lack of complete asymptotic expansions for solutions of some ordinary differential equations.

Nonsolvability for operators as in (3.8), but with $Re\alpha$, $Re\beta$ having the same sign, is studied in Yamasaki [1].

Theorem 3.5 (Yamasaki [1]). *Let* P *be a pseudo-differential operator as in* (3.8), *with* α,β,γ *satisfying the same hypotheses. Assume* $l<k-1$ *and*

(i) $Re\alpha(\theta)>0$, $Re\beta(\theta)>0$;

(ii) $\gamma(\theta)\in C^\infty(\mathbf{R}_-\cup\{0\})$.

Then P *is not locally solvable at* 0.

The proof is inspired by the paper of Cardoso and Treves [1], which we are going now to describe.

Let P be a classical pseudo-differential operator, defined in a neighborhood Ω of a

point x_0 in $\mathbf{R}^n$. Let us assume that there exists some $\xi_0 \in \mathbf{R}^n \setminus \{0\}$ and a conic neighborhood Γ in $T^*(\Omega)\setminus\{0\}$ of (x_0,ξ_0) such that the principal symbol of P can be factorized in Γ in the following way:

(3.10) $\quad p_m(x,\xi) = q(x,\xi)((l(x,\xi))^2.$

Here q and l are supposed to be positively homogeneous of degree m-2 and 1, respectively, and q is elliptic in Γ. On l we make the following hypotheses:

(3.11) $\quad l(x_0,\xi_0) = 0;$

(3.12) $\quad d_\xi l(x_0,\xi_0) \neq 0.$

Unless multiplying l by i and shrinking Γ, from (3.12) we can assume that

(3.13) $\quad d_\xi Re\, l$ *does not vanish in* Γ;

let us then call γ_0 the null bicharacteristic strip of $Re\, l$ through (x_0,ξ_0). The result of Cardoso and Treves is then the following.

Theorem 3.6 (Cardoso and Treves [1]). *Let* P *be as above, satisfying* (3.10)-(3.13). *Assume furthermore that*

(3.14) $\quad Im\, l$ *has a finite odd order zero along* γ_0 *at* (x_0,ξ_0);

(3.15) $\quad$ *the change of sign at* (x_0,ξ_0) *of the restriction of* $Im\, l$ *to* γ_0 *is from* + *to* -.

Then ${}^tP$ *is not locally solvable at* x_0.

Let us point out that, under the hypotheses of the previous theorem, the operator ${}^tL$ is not locally solvable at x_0: see Theorem 2.2. Remark that this unsolvability result is completely independent of the lower order terms. When P is a *differential* operator, the theorem above holds with weaker hypotheses, as it has been already observed in Corollary 2.3: (3.15) is not needed.

The proof of Theorem 3.6 is rather long, though the method is standard: one looks for an asymptotic solution u of the form $u=v_\rho\cdot\exp(i\rho w_\rho)$ of the equation Pu=0, with v a smooth amplitude function with compact support and w_ρ a complex phase depending on the parameter ρ. However the construction of u is not immediate. Firstly, there is required a very careful control on the remainder term in the asymptotic expression of Pu. This is achieved by Cardoso and Treves by giving first the asymptotic formula when the symbols are analytic in the ξ variables; then, with a finite Taylor expansion, they can drop this assumption estimating the Taylor remainder term. Secondly, the lower order terms play an important rôle in the proof, though they do not appear explicitly in the statement. More precisely one has to distinguish two cases, depending on their "strong" or "little"

influence in the determination of the phase function. When in (3.14) the vanishing order is one the result is contained also in Sjöstrand [1].

Theorem 3.6 is generalized by Goldman to operators with higher order characteristics. Let us assume, in the notations above, that

$$p_m(x,\xi) = q(x,\xi)(l(x,\xi))^r,$$

with $r\geq 3$. Denote the subprincipal symbol of P by

$$p^s_{m-1}(x,\xi) = p_{m-1}(x,\xi) + \frac{i}{2}\sum_{1\leq j\leq n}(\partial^2 p_m/\partial x_j\partial\xi_j)(x,\xi);$$

we refer to Duistermaat and Hörmander [1] for its properties. Then, under hypotheses (3.11)-(3.15) and assuming further $p^s_{m-1}(x_0,\xi_0)\neq 0$, Goldman [1] proves nonsolvability at x_0 for the transposed operator ${}^tP$. Remark that, in this case, $p^s_{m-1}=p_{m-1}$.

What happens when $|l|$ has an *even* order zero along γ_0, in the notations of the previous theorem, it is conjectured for a simple operator in the survey work of Treves [6]. Let P be the following second-order differential operator in $\mathbf{R}^2$:

(3.16) $$P(t,x,D_t,D_x) = ((D_t+ia(t)D_x)^2+b(t)D_x+c(t),$$

where a,b,c are smooth functions; we suppose that a is real-valued and $a(t)=t^k a^o(t)$, with $a^o(0)\neq 0$. Here k is an even integer.

Theorem 3.7 (Treves [6]). *For the local solvability at* (0,0) *of the operator* P *in* (3.16) *it is necessary and sufficient that, in a neighborhood of* t=0, *it results* $|b(t)|\leq C|t|^{k-1}$, *for some positive constant* C.

In the paper of Treves the proof is missing, but it can be recovered through the preceding Theorem 3.5 of Yamasaki and the works of Okaji [1], [4].

The now quoted work of Okaji [1] contains a rather precise analysis of the vanishing orders in the lower order terms of differential operators in $\mathbf{R}^2$ in order to have or not to have local solvability. More precisely there are considered operators of the following type:

(3.17) $$(D_t+ia(t,x)D_x)^m + \sum_{i+j\leq m-1} b_{ij}(t,x)D_t^iD_x^j,$$

where a,b_{ij} are smooth functions; $a(t,x)=t^k a^o(t,x)$, with a^o real-valued and nonvanishing at (0,0); k is a positive integer. Let us write now $b_{ij}(t,x)=t^{k_{ij}}\,b^o_{ij}(t,x)$, where the functions b^o_{ij} do not vanish identically in a neighborhood of (0,0). When m=2 or 3, Okaji provides then

some necessary and some sufficient conditions for the local solvability of operators of this type, in terms of the exponents k and k_{ij}. For lack of space we refer for the statements to the above mentioned work.

In Okaji [5] it is considered also an example of an operator of the form (3.17) but with an infinite order of vanishing. It is the operator

$$P(t,x,D_t,D_x) = \left(D_t+iexp(-|t|^{-n})D_x\right)^2+b(t)\left(D_t+iexp(-|t|^{-n})D_x\right)+c(t,x)D_x+d(t).$$

When $c(t,x)=exp(-A|(t,x)|^{-l})$, the transposed operator ${}^tP$ is not locally solvable at (0,0) if $l<n$ or $l=n$ and $0<A<1$. We have unsolvability for ${}^tP$ also when $c(t,x)=\gamma t^k$, with γ a nonzero complex number and k a nonnegative integer.

As we have seen in Theorem 3.6, for a double characteristic operator whose principal symbol is factorized as in (3.10), with q elliptic and l a principal type *complex* operator, one may give suitable conditions on l in order that local unsolvability does not depend on the lower order terms. This is not the case when l is *real*. Changing slightly notations we shall consider pseudo-differential operators P of order 2m, having as principal symbol

(3.18) $\qquad p_{2m}(x,\xi)=(p_m(x,\xi))^2,$

where p_m is a symbol of real principal type, positively homogeneous of degree m. Let us assume that there exists a point (x_0,ξ_0), $\xi_0\neq 0$, such that $p_m(x_0,\xi_0)=0$ (and therefore $d_\xi p_m(x_0,\xi_0)\neq 0$). Then we have the following result.

Theorem 3.8 (Popivanov [1]). *Let* P *be an operator as above and suppose that*

(i) $Re\,p^s_{2m-1}(x_0,\xi_0)<0$;

(ii) $Im\,p^s_{2m-1}$ *has a finite odd order zero at* (x_0,ξ_0) *along the null bicharacteristic strip of* p_m *through* (x_0,ξ_0).

Then P *is not locally solvable at* x_0.

When P is a differential operator it is sufficient, as above, to require that $Re\,p^s_{2m-1}$ does not vanish at (x_0,ξ_0). The proof of Theorem 3.8 consists in the construction of an asymptotic solution u of the form $u=v_\rho\cdot exp(i\rho^2\phi+i\rho\psi)$ of the homogeneous equation ${}^tPu=0$, contradicting (1.6). The first phase function ϕ is real and it is the (unique) solution of the characteristic equation $p_m(x,grad\phi)=0$ under the conditions $grad\phi(x_0)=\xi_0$, $\phi(x)=|x|^2$ on, say, $x_1=0$. The second phase function ψ is complex; the hypotheses (i) and (ii) are used in order to have that $Im\psi$, though vanishing at x_0, is strictly positive nearby. The

amplitude function v_ρ is sought as usual under the form $\sum_{j=0}^{N} v_j \rho^{-j}$, for N sufficiently large, where the v_j are determined by the transport equations.

About Theorem 3.8 we refer also to Ivrii [1] for a condition as in Hörmander [3], Rubinstein [1] for second-order differential operators, Wenston [2] for general differential operators, Menikoff [4] for a proof involving a reduction via Fourier integral operators, Popivanov and Georgiev [1] for microlocal solvability (see Hörmander [10] for the definition of microlocal solvability).

As regards sufficient conditions for the local solvability of operators whose principal symbol is factorized as in (3.18), with p_m of real principal type, Popivanov [1] proves local solvability when $Re\, p^s_{2m-1} \geq 0$ or $Im\, p^s_{2m-1} > 0$ in a neighborhood of x_0 on the characteristic manifold of p_m. In Menikoff [4] there is proved that one has local solvability also when $Re\, p^s_{2m-1} < 0$, provided that $Im\, p^s_{2m-1}$ does not change sign (maybe having some zeros) in a neighborhood of (x_0,ξ_0), and does not vanish identically on any interval of the null bicharacteristic strips of p_m (see also Wenston [4] for differential operators). In the same work there are given also some sufficient conditions for the solvability in the case when the subprincipal symbol vanishes.

Theorem 3.8 has been generalized in many directions; we begin by considering operators whose principal symbols are a sum or a difference of terms as above. More precisely, let us assume that the principal symbol p_{2m} of a pseudo-differential operator P can be written, in a conic neighborhood of $(x_0,\xi_0) \in T^*(\Omega)\backslash 0$, $\Omega \subset \mathbf{R}^n$ a neighborhood of x_0, as follows:

$$(3.19) \qquad p_{2m}(x,\xi) = \sum_{j=1}^{k} \varepsilon_j (p_{m,j}(x,\xi))^2.$$

The symbols $p_{m,j}$ are supposed to be positively homogeneous of degree m, of real principal type; $k<n$ and $\varepsilon_j = \pm 1$. Let us denote $N=\{(x,\xi) \in T^*(\Omega)\backslash 0;\ p_{m,j}(x,\xi)=0,\ 1\leq j\leq k\}$. Then we have the following result.

Theorem 3.9 (Popivanov [3]). *Let* (x_0,ξ_0) *be in* N; *suppose that the* k *vectors* $\mathrm{grad}_\xi p_{m,j}(x_0,\xi_0), 1\leq j\leq k$, *are linearly independent, and that the Poisson brackets* $\{p_{m,j},p_{m,h}\}$, $1\leq j,h\leq k$, *vanish identically on* N. *Assume moreover that there exists an index* j_0 *such that*

(i) $\varepsilon_{j_0} Re\, p^s_{2m-1}(x_0,\xi_0) < 0$;

(ii) Imp^s_{2m-1} *has a first order zero at* (x_0,ξ_0) *along the null bicharacteristic strip of* p_{m,j_0} *through* (x_0,ξ_0).

Then P *is not locally solvable at* x_0.

As usual, dealing with differential operators, (i) can be weakened by requiring only that Rep^s_{2m-1} does not vanish at (x_0,ξ_0).

When in (ii) the order of vanishing is more generally odd, or even infinite, some further hypotheses, similar to those considered in Egorov and Popivanov [1] for the principal type case, permit to obtain again a result of unsolvability. In particular, when Imp^s_{2m-1} has an infinite order zero, it is assumed that Imp^s_{2m-1} does not vanish identically on any interval of the null bicharacteristic strips, lying on N, of some $p_{m,j}$. For lack of space we refer the interested reader to the paper of Popivanov [3], where he can find also some sufficient conditions for solvability concerning the case $\varepsilon_j=1$, for every $j=1,\dots,k$; they are on the lines of those stated under Theorem 3.8.

Recently Popelyukhin [2] has succeeded in proving a result of nonsolvability for the operators in Theorem 3.8, allowing Imp^s_{2m-1} to vanish identically on the null bicharacteristic strip of p_m through (x_0,ξ_0).

Theorem 3.10 (Popelyukhin [2]). *Let* P *be a pseudo-differential operator with principal symbol factorized as in* (3.18); p_m *is of real principal type and* $p_m(x_0,\xi_0)=0$. *Let* $\{\gamma(t);\ a\le t\le b\}$ *be an arc of the null bicharacteristic strip of* p_m *through* $(x_0,\xi_0)=\gamma(0)$. *Assume that*

(i) $Rep^s_{2m-1}(\gamma(t))<0$, *for* $a\le t\le b$;

(ii) $Imp^s_{2m-1}(\gamma(a))\cdot Imp^s_{2m-1}(\gamma(b))<0$.

Then P *is not locally solvable in any subset of* $\mathbf{R}^n$ *containing the projection of the arc* $\{\gamma(t);\ a\le t\le b\}$.

The proof is given by using a canonical transformation and a related Fourier integral operator to reduce the operator to a second-order one.

Another generalization of Theorem 3.8 can be found in Popivanov [4], where there are studied pseudo-differential operators with real principal symbol p_m such that $p_m(x_0,\xi_0)=\operatorname{grad}_{x,\xi}p_m(x_0,\xi_0)=0$, but $\partial^2 p_m/\partial\xi_{j_0}^2(x_0,\xi_0)\neq 0$ for some j_0.

In Wenston [1] is taken into account the "stability" of local solvability under

perturbations. Let P be a locally solvable differential operator of order m; an operator R of order l<m is called an *admissible lower order perturbation* of P if P+ϕR is still solvable for every smooth function ϕ. In this work there is given a sufficient condition for admissibility for operators P of the type

$$P(x,D)=Q_1^{j_1}(x,D)\dots Q_k^{j_k}(x,D),$$

where the operators Q_i are homogeneous, satisfying the condition of solvability of Nirenberg and Treves, and their characteristic manifolds are disjoint. As regards necessary conditions for admissibility of a perturbation R, when P is as above, it is shown that if every Q_i has real (or constant) coefficients, $\max_{1\le i\le k}\{j_i\}=2$ and l=m-1, then every double characteristic root ξ of P_m must be a root of R_{m-1} too.

We pass now to expose briefly some results concerning operators whose principal symbol is a power of order higher than two of some symbol of real principal type. In Wenston [3] are taken into account operators which can be written in the canonical form

$$P=D_t^{2p+1}+a(t,x,D_x),$$

where a is a first-order pseudo-differential operator (we use here the variables $t\in\mathbf{R}$, $x\in\mathbf{R}^n$). Referring to this canonical form it is proved there that P is not locally solvable at (0,0) when the following two conditions are satisfied for some ξ_0 in $\mathbf{R}^n\setminus\{0\}$:

(i) $Re\,a_1(0,0;\xi_0)\neq 0$;

(ii) $Im\,a_1(t,0;\xi_0)$ *has a finite odd order zero at* t=0.

In the same work there are given also some sufficient conditions for the local solvability of this class of operators. Let us assume that a_1 does not depend on x, $Re\,a_1(0,\xi)\neq 0$, $Im\,a_1(0,\xi)\equiv 0$ and the sign of $Im\,a_1(t,\xi)$ depends only on ξ. In this case, for instance, we have local solvability. Other sufficient conditions about a class of operators with odd order real characteristics can be found in Roberts and Wenston [1].

We mention also the work of Popivanov and Popov [2], which deals with operators P having as principal symbol $(p_m(x,\xi))^3$, where p_m is real, positively homogeneous of degree m. It is assumed that p_m is of "principal type" in the weak sense that for some (x_0,ξ_0), $\xi_0\neq 0$, $p_m(x_0,\xi_0)=0$ but $\mathrm{grad}_{x,\xi}p_m(x_0,\xi_0)\neq 0$. Then P is not locally solvable at x_0 if the following conditions are satisfied:

(i) $p^s_{3m-1}(x_0,\xi_0)\neq 0$:

(ii) $Im\,p^s_{3m-1}(x_0,\xi_0)=0$, $H_{p_m}(Im\,p^s_{3m-1})(x_0,\xi_0)>0$.

The former paper of Popivanov and Popov [1] was concerned with the case m=1. For a

generalization to higher order characteristics see Corli [2] and the next section.

Let us quote also the work of Okaji [2], where there is characterized the local solvability for the operator $D_t^3+at^kD_x^n$, n a positive integer.

Let us now go on considering double characteristic operators whose principal symbols are factorized into two *different* symbols of real principal type. Definitive results about a significative class of these operators have been obtained by Mendoza and Uhlmann [1], [2]. Let P be a classical pseudo-differential operator on an open subset Ω in $\mathbf{R}^n$, with principal symbol p_m factorized as follows:

(3.20) $\quad p_m = p_{m_1}p_{m_2}$.

The operators p_{m_1}, p_{m_2} are positively homogeneous of degree m_1, m_2, respectively, both of real principal type; the factorization (3.20) holds true near every point in $T^*(\Omega)\backslash 0$. The following hypotheses on the principal symbol p_m are assumed:

(3.21) *the doubly characteristic set* $\Sigma=\{(x,\xi)\in T^*(\Omega)\backslash 0;\ p_m(x,\xi)=\mathrm{grad}_{x,\xi}p_m(x,\xi)=0\}$ *is an involutive submanifold of codimension 2; i.e.* $\{p_{m_1}, p_{m_2}\}=0$ *on* Σ;

(3.22) *the Hamilton vector fields* H_{p_1}, H_{p_2} *and the radial direction* $\sum_{j=1}^n \xi_j\partial/\partial\xi_j$ *are linearly independent on* Σ.

In this framework we state now the main condition:

(3.23) Imp^s_{m-1} *does not change sign at* (x_0,ξ_0) *along the null bicharacteristic strip of* p_{m_1} *and* p_{m_2} *through* (x_0,ξ_0).

Theorem 3.11 (Mendoza and Uhlmann [1]). *Let* P *be a pseudo-differential operator satisfying* (3.20)-(3.22); *then* (3.23) *is a necessary condition for the microlocal solvability at* (x_0,ξ_0).

The proof consists in reducing, with a canonical transformation, to the second-order operator $D_{x_1}D_{x_2}+B(x,D_x)$, where B is of order one; then one proceeds along the lines of Hörmander [9].

On the other hand, Mendoza and Uhlmann [2] prove that for these operators there is local solvability when Imp^s_{m-1} does not vanish on Σ. For a more precise statement, as well as for results on the propagation of singularities, we refer to the last mentioned paper.

Local solvability for operators with double involutive characteristics, with principal symbol not necessarily factorized as in (3.20), is studied also in Popivanov [6], [7], [10]

and in Popelyukhin [1].

In the early seventies, just after the papers of Nirenberg and Treves, Rubinstein [2] found two examples of unsolvable operators, which are not of principal type and do not belong to any of the classes until now discussed. Both of them were the starting point for some further researches, as it happened, for instance, for the pioneering work of Gilioli and Treves.

Theorem 3.12 (Rubinstein [2]). *The operators, in* $\mathbf{R}^2$,

(3.24) $D_t^2+t^nD_x^2+(1+it^m)D_x$, m *odd*, n>4m+2,

(3.25) $D_t-it^nD_x^2+it^mD_x$, n *even*, n>2m+1,

are not locally solvable at (0,0).

Proofs are given following the standard pattern of Hörmander [3], using a suitable asymptotic change of variables.

These examples show once more the fundamental rôle played be the lower order terms when the operators are not of principal type. In fact, concerning for instance the operator in (3.24), the principal part $D_t^2+t^nD_x^2$ is locally solvable, but the "perturbative" term $(1+it^m)D_x$, provided it is sufficiently "strong" in a neighborhood of t=0, causes the unsolvability of the operator. On the other hand, the above result may be explained, roughly speaking, as follows: the operator $D_t^2+(1+it^m)D_x$ is not solvable when m is odd (see Theorem 3.8), and the "perturbation" $-it^nD_x^2$, if sufficiently "weak", does not affect the unsolvability. Analogously one may argue on the operator in (3.25): remark that the first order part is just Mizohata operator. Let us notice furthermore that the operator in (3.25) is solvable when n<2m+1.

Operators as in (3.24) were already taken into account in the paper of Ivrii [1], where unsolvability in the cases n=5,6, m=1, was proved. This shows that the "threshold" n>4m+2 given by Rubinstein is not the best one in order to have nonsolvability. This appears also from the work of Karatopraklieva [1] (see also Popivanov [8]), where there is studied the slightly more general operator

$$D_t^2+at^nD_x^2+(\alpha+i\beta t^m)D_x, \qquad m \textit{ odd}, n\geq 4m+2;$$

a,α,β are real constants, $a\beta\neq 0$. By employing the method of Grushin [1] it is proved there that this operator is not locally solvable at (0,0) when $\alpha\neq 0$. On the contrary, if $\alpha=0$, $a>0$ and n is even, then the operator is solvable. The same results hold for the transposed operator.

Let us now go on considering some classes of operators related to the example in (3.25). Kannai [1] proved that the operator $D_t\pm itD_x^2$, in $\mathbf{R}^2$, is not locally solvable at t=0 if the sign + is chosen, while it becomes so in the other case. Proofs depend heavily on the rather particular feature of the operator: there are used techniques dealing with the heat operator. In Popivanov [4] there are given necessary and sufficient conditions for the solvability of some operators, in $\mathbf{R}^2$, of the form $D_t+P(t,D_x)$, where P is a fourth-order differential operator with polynomial coefficients. Okaji [3] (see also Okaji [2]) studies the same problems for the more general case $D_t+P(t,x,D_x)$, with P of order m and (t,x) in a neighborhood of 0 in $\mathbf{R}^{1+n}$. In this work the conditions for solvability are given by means of a careful analysis of the order of vanishing at 0 of the coefficients of P (see Okaji [1] for the same approach to different problems). As an example there is completely established the solvability at $0\in\mathbf{R}^2$ of the operator $D_t+iat^lD_x^n+ibt^kD_x^m$, where n>m and a,b are real numbers. At the end of the paper, Okaji proposes the following definition of semi-local solvability.

Definition 3.13 (Okaji [3]). *Let* $P=P(t,x,D_t,D_x)$ *be an operator with* C^∞ *coefficients in a neighborhood of the origin* (0,0) *in* $\mathbf{R}^{1+n}$; P *is said to be semi-locally solvable at* (0,0) *with respect to* $t>0$ (*resp.* $t<0$) *if there exists a neighborhood* U *of* (0,0) *such that for every* $f\in C_0^\infty(U_+)$ (*resp.* $f\in C_0^\infty(U_-)$) *there exists some* $u\in\mathcal{D}'(U_+)$ (*resp.* $u\in\mathcal{D}'(U_-)$) *satisfying* $Pu=f$ *in* U_+(*resp.* U_-). *Here* $U_\pm=U\cap\{\pm t>0\}$.

With this terminology, Lewy operator, $D_t+iD_x+i(t+ix)D_y$, in $\mathbf{R}^3$, is not semi-locally solvable at the origin neither with respect to $t>0$ nor with respect to $t<0$, whereas Mizohata operator, $D_t+it^kD_x$, k odd, is semi-locally solvable at the origin in $\mathbf{R}^2$ with respect to both sides. On the other hand, the fourth order operator $D_t-it^nD_x^4+it^mD_x^2$, in $\mathbf{R}^2$, with n even, m odd, $n>2m+1$, is semi-locally solvable at (0,0) with respect to $t>0$ but not with respect to $t<0$.

Shananin has studied intensively solvability for operators of *quasi-principal type;* let us recall briefly what this term means. Let $P(x,D)=\sum_{|\alpha|\le d} a_\alpha(x)D^\alpha$ be a differential operator in $\mathbf{R}^n$, and let $m=(m_1,\dots m_n)$ denote an n-ple of positive integers: m is called a *weight set.* The *weighted order* M of P, with respect to the weight set m, is defined as

$$M=\max\{m\cdot\alpha:=\sum_{1\le j\le n} m_j\alpha_j;\ a_\alpha \textit{ does not vanish identically}\},$$

and the *weighted principal symbol* of P is

$$P^0_M(x,\xi)=\sum_{m\cdot\alpha=M} a_\alpha(x)\xi^\alpha.$$

A variable x_k is then called a *fundamental variable* if $m_k=\min_{1\le j\le n}\{m_j\}$; we shall denote by x' the set of fundamental variables.

Definition 3.14. *A differential operator* P *is called of quasi-principal type (with respect to the weight set* m*) if* $\operatorname{grad}_\xi P^0_M(x,\xi)\neq 0$ *when* $P^0_M(x,\xi)=0$, *for every* $\xi\neq 0$.

We have then the following solvability result.

Theorem 3.15 (Shananin [1]). *Let* P *be a differential operator of order* d, *defined in an open subset* Ω *in* $\mathbf{R}^n$; *let us suppose that* P *is of quasi-principal type with real weighted principal symbol and* M=d.
Then P *is locally solvable at each point in* Ω.

Remark that the condition M=d implies that there exists at least one fundamental variable. This result fails when the assumption M=d does not hold, as it is shown in Shananin [2] for the operator in $\mathbf{R}^2$

$$P(x,D)=D^3_{x_1}+D^2_{x_2}+6ix_1D_{x_1}D_{x_2}+6D_{x_2}.$$

It is of quasi-principal type with respect to the weight set m=(2,3) and the weighted principal symbol, $P^0_6(x,\xi)=\xi_1^3+\xi_2^2$, is real; however P is not locally solvable at the origin.

As one may easily understand from this example, the behaviour of the weighted lower order terms is essential when condition M=d is not satisfied.

Several necessary conditions for the local solvability of operators of quasi-principal

type have been given in Shananin [3],[4],[5]; a sufficient condition for a class of related operators is in Volevich and Gindikin [1].

As we pointed out several times, the results we exposed until here were otained by using mainly Hörmander method, or variants of it; in some other cases Grushin technique was used. We must now recall a quite different approach to the problem of the local solvability, which may be applied also in the study of hypoellipticity. It is the method of *concatenations*, introduced in Treves [5]. In this work second-order abstract evolution operators P are studied (but it is taken into account also the first order case) of the form

(3.26) $P=(\partial_t-a(t,A)A)(\partial_t-b(t,A)A)-c(t,A)A.$

The linear operator A in (3.26) is densely defined in a Hilbert space H; it is unbounded but self-adjoint, positive-defined with bounded inverse A^{-1}. Just to fix ideas, we may think $H=L^2(\Omega)$ and A as some self-adjoint extension of $-\Delta$. Moreover we have denoted with a,b,c some series in nonnegative powers of A^{-1} with coefficients in $C^\infty(I)$, where I is an open subset of $\mathbf{R}_t$; that is, for instance, $a(t,A)=\sum_{j=0}^{n}a_j(t)A^{-1}$. The convergence of these series is meant in the space of bounded linear operators on H, uniformly (on compact subsets of I) with respect to t and to every t-derivative. Under suitable hypotheses, some solvability and hypoellipticity results are then given in terms of the concatenation associated to P, that is, by means of a sequence of operators constructed from P. Since it does not seem possible to us to give in few lines a satisfactory account of the results obtained by Treves, we refer the reader to the quoted paper; nevertheless we would like to give an example of how his method works.

As we mentioned at the beginning of this section, Grushin operator

(3.27) $G_\lambda=D_t^2+t^2D_x^2+\lambda D_x, \qquad (t,x)\in\mathbf{R}^2,$

is not locally solvable at the origin when $\lambda=\pm(2n+1)$, n a positive integer. We want to sketch a proof of this result by the method of concatenations. More precisely, we shall show the microlocal unsolvability of G_λ at $\theta=(0,0;0,1)$ for $\lambda=-(2n+1)$.

Take $\lambda=-1$; then we can write

$$G_{-1}=(D_t+itD_x)(D_t-itD_x).$$

In this expression, the Mizohata operator appears as left hand factor; since it is not microlocally solvable at θ, the same must be true for G_{-1}. Take now $\lambda=-3$ and write

$$(D_t-itD_x)G_{-3}=(D_t+itD_x)P,$$

where P is some second-order operator. The first factor on the left hand side is microlocally solvable at θ, with solutions in C^∞. On the other hand, the left hand side is

not microlocally solvable at that point, arguing as before; then G_{-3} is not microlocally solvable at θ. The general case may be easily handled in the same way by induction.

The method of concatenations has been used in Gilioli and Treves [1] and several other papers. In Rodino [1] (see also Mascarello Rodino [1]) are taken into account degenerate pseudo-differential operators P, in $\mathbf{R}^{1+n}$, of the form

$$P(t,x,D_t,D_x)=\tau(D_t-r_M t|D_x|)\dots(D_t-r_2 t|D_x|)(D_t-r_1 t|D_x|)+\sum_{h+k\le M-1} c_{hk}t^h D_t^k|D_x|^{(M+h-k)/2},$$

where τ, c_{hk} $(0\le h,k\le M-1)$, r_j $(1\le j\le M)$ are complex constants, $\tau\neq 0$. Here $|D_x|$ denote the pseudo-differential operator with symbol $|\xi|$. Under the assumption

$$Im\, r_1>0, \quad Im\, r_j<0 \text{ for } j=2,3,\dots,M,$$

the hypoellipticity of P is equivalent to the local solvability of ${}^tP$, which in turn is equivalent to a Grushin type condition on the operator.

The same method is used also in Kwon [1], where there are studied double characteristic pseudo-differential operators whose principal symbol is nonnegative and the characteristic manifold is symplectic of codimension two (Grushin type operators, for instance). Some generalizations of the results of Treves [5] may be found in Gilioli [1], Oganesjan [1], Shananin [6].

An example of a highly degenerate unsolvable first-order operator is given in Elschner and Lorenz [1]. It is the operator, in $\mathbf{R}^2$,

$$(x^2+y^2)^k(x\partial_x+y\partial_y)+i(x^2+y^2)^k(x\partial_y-y\partial_x)+1,$$

which can be written in polar coordinates (r,ϕ) as

$$r^{2k+1}\partial_r+ir^{2k}\partial_\phi+1.$$

It is proved there that this operator is hypoelliptic but not locally solvable at (0,0). Generalizations to higher order operators are contained in the papers of Lorenz [1], [2]. In particular in Lorenz [2] there is considered the operator

$$r^{2q+2}\Delta+\mu(r)r^{2p+1}\partial_r+\lambda(r), \qquad x\in\mathbf{R}^n,\ r=|x|,$$

where μ,λ are real C^∞ functions, $\mu(0)\lambda(0)\neq 0$ and q,p are positive integers, $q>2p$. Under these assumptions this operator is not locally solvable at 0 if and only if $\lambda(0)>0$ and $\mu(0)>0$. For this kind of problems see also Felix [1].

All the results until here exposed were concerned with operators having indefinitely differentiable coefficients. We must however quote the paper of Colombini and Spagnolo [1] where, among other things, there is proved that for some function $A(t,x)\in C^{0,\alpha}(\mathbf{R}^2)$

for all $\alpha<1$, satisfying $C^{-1}\le A(t,x)\le C$ for some positive constant C, the equation

$$u_{tt}-(A(t,x)u_x)_x=x$$

has no C^1 solution in a neighborhood of (0,0).

4. LOCAL SOLVABILITY IN GEVREY CLASSES

The aim of this section is to give an account of some recent results on local solvability in Gevrey classes. It is convenient, before stating the reason for this kind of problem, to recall the definition of these classes of functions; a standard reference is Komatsu [1].

Let Ω be an open subset in $\mathbf{R}^n$, K a compact subset of Ω, h,s real numbers with $h>0$, $s\ge 1$; we define $G^{s,h}(K)$ as the space of the functions $\phi\in C^\infty(K)$ such that

$$\sup_{x\in K}|D^\alpha\phi(x)| \le Ch^{|\alpha|}\alpha!^s, \qquad |\alpha|=0,1,2\dots$$

for some positive constant C. The space $G^{s,h}(K)$ is a Banach space under the norm

$$|\phi;G^{s,h}(K)| = \sup_\alpha \sup_{x\in K} h^{-|\alpha|}\alpha!^{-s}|D^\alpha\phi(x)|.$$

The Gevrey function classes are then defined as follows:

$$(4.1)\qquad G^{(s)}(\Omega) = \operatorname{projlim}_{K\subset\subset\Omega}\operatorname{projlim}_{h\to 0} G^{s,h}(K)$$

$$(4.2)\qquad G^{\{s\}}(\Omega) = \operatorname{projlim}_{K\subset\subset\Omega}\operatorname{indlim}_{h\to\infty} G^{s,h}(K),$$

with the usual topologies of inductive or projective limits of locally convex spaces. More precisely, we shall refer to $G^{(s)}(\Omega)$ as *projective* Gevrey classes and to $G^{\{s\}}(\Omega)$ as *inductive* Gevrey classes (of order s). The strong dual spaces of $G^{(s)}(\Omega)$, $G^{\{s\}}(\Omega)$, $s>1$, are called spaces of *ultradistributions*; they will be denoted by $G^{(s)'}(\Omega)$, $G^{\{s\}'}(\Omega)$, respectively.

Remark that $G^{(1)}(\Omega)$ is nothing else than the space of the analytic functions on Ω; moreover from the definitions (4.1), (4.2) it is clear that

$$(4.3)\qquad G^{(s)}(\Omega)\subset G^{\{s\}}(\Omega)\subset C^\infty(\Omega), \quad \textit{for every } s\ge 1;$$

$$(4.4)\qquad G^{\{s\}}(\Omega)\subset G^{(s+\varepsilon)}(\Omega), \quad \textit{for every } s\ge 1,\ \varepsilon>0;$$

$$(4.5)\qquad G^{(s)}(\Omega)\subset G^{(t)}(\Omega),\ G^{\{s\}}(\Omega)\subset G^{\{t\}}(\Omega), \quad \textit{for every } 1\le s<t.$$

All above inclusions are strict.

Gevrey classes arise in a natural way when dealing with partial differential equations. So, for instance, the solutions u of the homogeneous heat equation $\partial_t u-\partial_x^2 u=0$ are analytic in x and of class $G^{\{2\}}$ with respect to t; and the Cauchy problem for a weakly hyperbolic equation, though not well posed in C^∞, becomes so in some Gevrey classes (see Ivrii

[3]).

We can now explain which is the problem we shall be concerned with. Let P be a differential operator and let us suppose that it is not solvable at some point, that is, for some $f\in C^\infty$ there are no distributions u solving the equation $Pu=f$ near that point. We may ask whether, by restricting the class of the data f, and allowing "worse" (i.e., in some classes wider than $\mathcal{D}'$) solutions than before, we can solve our equation. We expect that the answer may be affirmative: Cauchy-Kovalevsky theorem gives solvability (in nondegenerate cases) when s=1, though in neighborhoods depending on the data. It is then natural to give the following definition.

Definition 4.1. *Let* P *be a (pseudo-)differential operator in* Ω; P *is said to be locally* (s)*-solvable (resp.* {s}*-solvable) at* $x_0\in\Omega$ *if there exists a neighborhood* U *of* x_0 *such that for every* $f\in G^{(s)}(\Omega)$ $(f\in G^{\{s\}}(\Omega))$ *there exists* $u\in G^{(s)'}(\Omega)$ $(u\in G^{\{s\}'}(\Omega))$ *satisfying* $Pu=f$ *in* U.

Thus, if an operator P is s-solvable (we omit the parentheses when dealing with both cases at the same time), then it is t-solvable for every $t<s$ (by (4.5)); on the other hand, local s-unsolvability, for some s, implies unsolvability in C^∞ (by (4.3)).

Definition 4.1 has been given, in the (s)-case (really, in a more general case), by Björck [1], who proved that Hörmander's [3] condition is necessary also for the local (s)-solvability of first-order operators with analytic coefficients. His proof is a modification of Hörmander's one. For what concerns the definition for the projective case, it may look a bit astonishing that is was proposed only twenty years later by Rodino [2],[3], and this although at the same time related problems were studied (for example, hypoellipticity in Gevrey classes). Let us remark however that every {s}-unsolvability result, for s in an open interval of the type $(s_0,+\infty)$, $s_0>1$, implies immediately an analogous (s)-unsolvability result for s in the same interval, in view of (4.4); and vice-versa. So we shall state all results below for the {s}-case; also the proofs, as usual not reported here, are given in this case.

Let us begin by considering operators of principal type.

Theorem 4.2 (Rodino [2],[3]). *Let* P *be a classical analytic pseudo-differential operator, satisfying the hypotheses stated before Theorem 2.2. Suppose that* Imp_m *has a finite odd order zero at* (x_0,ξ_0) *when it is restricted to* γ_0 *and changes its sign from - to +*

moving in the positive direction of γ_0. *Then* P *is not locally* {s}*-solvable at* x_0 *for every* s>1.

The proof is like that of Theorem 2.2; this time however one must use Fourier integral operators with analytic phase and amplitude functions. For details we refer to Rodino [3].

Also Grushin operators G_λ (see (3.27)) are not locally {s}-solvable for any s>1; the proof outlined above works equally well in this case (see again Rodino [3]).

Let us now take into account a wider class of differential operators with double characteristics, precisely those considered in Theorem 3.6.

Theorem 4.3 (Cardoso [1]). *Let* P *be an analytic differential operator with principal symbol* $p_m(x,\xi)=q(x,\xi)(l(x,\xi))^2$, *where* q *and* l *are differential operators of orders* m-2,1, *respectively. Let us assume that* q *is elliptic and* l *is of principal type. Moreover let there exist some* (x_0,ξ_0), $\xi_0\neq 0$, *such that* $l(x_0,\xi_0)=0$; *we can suppose then that* $d_\xi Re\, l(x_0,\xi_0)\neq 0$. *If*

> *Im*l *has a finite odd order zero at* (x_0,ξ_0) *when it is restricted to the null bicharacteristic strip of Re*l *through* (x_0,ξ_0),

then P *is not locally* {s}*-solvable at* x_0, *for every* s>1.

The proof of this result is along the lines of the proof of Theorem 3.6, with the modifications needed in the Gevrey case: see Corli [1], and below for some more informations.

All the operators until now considered remained {s}-unsolvable for every s>1. However, operators which become {s}-solvable for some sufficiently small s>1 do exist; for instance the operator in R^2

$$(D_t^2+t^2D_x)^2-D_x$$

is locally {s}-solvable at (0,0) if and only if $1<s\leq 2$ (see Rodino [3]).

Let us now go on considering some necessary conditions for the local {s}-solvability of differential operators with multiple characteristics. The following result specifies Theorem 3.8.

Theorem 4.4 (Corli [1]). *Let* P *be an analytic differential operator, with principal symbol* $p_{2m}(x,\xi)=(p_m(x,\xi))^2$, *where* p_m *is of real principal type. Let us assume that*

$p_m(x_0,\xi_0)=0$ *for some* (x_0,ξ_0), $\xi_0\neq 0$, *and*

(i) $Rep^s_{2m-1}(x_0,\xi_0)\neq 0$;

(ii) Imp^s_{2m-1} *has a finite odd order zero at* (x_0,ξ_0) *along the null bicharacteristic strip of* p_m *through* (x_0,ξ_0).

Then P *is not locally* {s}*-solvable at* x_0 *when* s>2.

Though the proof of this theorem is similar to the proof of Theorem 3.8, two important differences deserve to be pointed out. First of all, it is needed an inequality, which is necessary for local {s}-solvability, of the type given in (1.6). The topology of the spaces $G^{\{s\}}(\Omega)$ does not permit to simply rephrase Hörmander's proof (whereas it is so in the $G^{(s)}$ case): however one may proceeds as it is done in Ivrii and Petkov [1], introducing moreover a sequence of spaces defining $G^{\{s\}}(\Omega)$ which is slightly different from the one used above. Secondly, rather precise estimates of the asymptotic solution are required, like those needed for the Cauchy problem by Ivrii [2].

In Corli [1] there is given also a necessary condition for {s}-solvability of a class of operators with higher order characteristics, making more precise a former result of Wenston [3].

Let us consider now differential operators whose principal symbols are powers of order larger than two of an operator of real principal type. More precisely let P be an analytic differential operator with principal symbol

$$p_{mr}(x,\xi)=(p_m(x,\xi))^r \qquad r\geq 3,$$

where p_m is of real principal type; let us suppose that $p_m(x_0,\xi_0)=0$.

Theorem 4.5 (Corli [2]). *Let* P *be as above. We make the following assumptions:*

(i) $Rep^s_{mr-1}(x_0,\xi_0)\neq 0$;

(ii) Imp^s_{mr-1} *has a finite odd order zero at* (x_0,ξ_0) *along the null bicharacteristic strip of* p_m *through* (x_0,ξ_0); *if r is odd we require furthermore that the ensuing change of sign is from - to +;*

Under these hypotheses the operator P *is not locally* {s}*-solvable at* x_0 *when* s>r/(r-1).

As we mentioned in the former section, this theorem was proved in the C^∞ category by Popivanov and Popov [2] when r=3; for r>3 this theorem provides then a result of C^∞

unsolvability which was not previously known. We refer again to Corli [2] for a result of the same kind in the case when the subprincipal symbol vanishes on the characteristic manifold. Both proofs deal with the techniques mentioned above.

For what concerns sufficient conditions for {s}-solvability for these classes of operators, when s is sufficiently near to one, let us mention the paper of Rodino and Zanghirati [1]. There is proved that classical analytic pseudo-differential operators, having for canonical model the operator

$$D_{x_1}^m + R(x,D),$$

where R is of order m-1, are microlocally {s}-solvable at $\theta=(0;0,\xi_0')$ when $1<s<m/(m-1)$, without any condition on R. On the other hand if we assume $Im\, r_{m-1}(\theta)\neq 0$, denoting by r_{m-1} the principal symbol of R, then we have microlocal {s}-solvability also in the case $s\geq m/(m-1)$ (see Liess and Rodino [1]). We refer to these papers for more details (see also Rodino [3], [4]).

At last, let us quote Gramchev [1] for sufficient conditions for {s}-solvability of operators in $\mathbf{R}^2$ of the form

$$(D_t+ia(t,x)D_x)^m + \sum_{0\leq j\leq m-1} B_j(t,x,D_x)(D_t+ia(t,x)D_x)^{m-j-1}$$

where a is a real analytic function vanishing of finite even order at t=0 and B_j are analytic differential operators of order j; we mention also Cattabriga and Zanghirati [1] for the surjectivity in $G^{(s)}(\mathbf{R}^2)$ (but not in $G^{(s)}(\mathbf{R}^3)$) of the Mizohata operator (1.7) with h even, and Ehrenpreis [1] for some other related results on surjectivity in Gevrey classes.

REFERENCES

BEALS R., FEFFERMAN C.:[1] On local solvability of linear partial differential equations; Ann. of Math. 97 (1973), 482-498.

— [2] Spatially inhomogeneous pseudodifferential operators, I; Comm. Pure Appl. Math. 27 (1974), 1-24.

BJÖRCK G.:[1] Linear partial differential operators and generalized distributions; Ark.

Mat. 6 (1966), 351-407.

CARATHEODORY C.:[1] Variationsrechnung und partielle Differentialgleichungen erster Ordnung; Teubner, Berlin 1935, Leipzig 1956[2]. English translation: Holden-Day, San Francisco 1965.

CARDOSO F.:[1] A necessary condition of Gevrey solvability for differential equations with double characteristics; Preprint n. 158, Universidade Federal de Pernambuco, 1988.

CARDOSO F., TREVES F.:[1] A necessary condition of local solvability for pseudo-differential equations with double characteristics; Ann. Inst. Fourier (Grenoble) 24:1 (1974), 225-292.

CATTABRIGA L., ZANGHIRATI L.:[1] Analytic and Gevrey global surjectivity of the Mizohata operator $D_2+ix_2^{2k}D_1$; Preprint.

COLOMBINI F., SPAGNOLO S.:[1] Some examples of hyperbolic equations without local solvability; Ann. École Norm. Sup. (4) 22 (1989), 109-125.

CORLI A.:[1] On local solvability in Gevrey classes of linear partial differential operators with multiple characteristics; Comm. Partial Differential Equations 14:1 (1989), 1-25.

— [2] On local solvability of linear partial differential operators with multiple characteristics; *to appear on* J. Differential Equations.

DIEUDONNE' J.:[1] La résolubilité des équations aux dérivées partielles linéaires; Bull. Soc. Math. Belg. Sér. A 35 (1983), 3-23. Liste d'errata à l'article de J. Dieudonné, ib. 35 (1983), 131.

DUISTERMAAT J.J., HÖRMANDER L.:[1] Fourier integral operators II; Acta Math. 128 (1972), 183-269.

EGOROV Yu. V.:[1] Subelliptic pseudo-differential operators; Dokl. Akad. Nauk SSSR 188:1 (1969), 20-22 (Russian); Soviet Math. Dokl. 10:5 (1969), 1056-1059.

— [2] Canonical transformations and pseudodifferential operators; Trudy Moskov. Mat. Obshch. 24 (1971) 3-28 (Russian); Trans. Moscow Math. Soc. 24 (1971), 1-28.

— [3] On necessary conditions for solvability of pseudodifferential equations of principal type; Trudy Moskov. Mat. Obshch. 24 (1971), 29-41 (Russian); Trans. Moscow Math. Soc. 24 (1971), 29-42.

— [4] On the solvability of differential equations with simple characteristics; Uspekhi Mat. Nauk 26:2 (1971), 183-198 (Russian); Russian Math. Surveys 26:2 (1971),

113-130.

— [5] Linear Differential Equations of Principal Type; Consultants Bureau, New York 1986.

EGOROV Yu.V., POPIVANOV P.R.:[1] Equations of principal type without solution; Uspekhi Mat. Nauk 29 (1974), 172-189 (Russian); Russian Math. Surveys 29 (1974), 176-194.

EHRENPREIS L.:[1] Lewy's operator and its ramifications; J. Funct. Anal. 68 (1986), 329-365.

ELSCHNER J., LORENZ M.:[1] An unsolvable hypoelliptic differential operator degenerating at one point; Wiss. Z. Tech. Hochsch. Karl-Marx-Stadt 23 (1981), 369-373.

FELIX R.:[1] Lokale Auflösbarkeit von Differentialoperatoren erster Ordnung in einem kritischen Punkt; Math. Z. 195:2 (1987), 291-300.

GARABEDIAN P.R.:[1] An unsolvable equation; Proc. Amer. Math. Soc. 25 (1970), 207-208.

GILIOLI A.:[1] A class of second-order evolution equations with double characteristics; Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 3:2 (1976), 187-229.

GILIOLI A., TREVES F.:[1]An example in the solvability theory of linear PDE's; Amer. J. Math. 96 (1974), 367-385.

GOLDMAN R.:[1] A necessary condition for the local solvability of a pseudodifferential equation having multiple characteristics; J. Differential Equations 19 (1975), 176-200.

GRAMCHEV T.V.:[1] Powers of Mizohata type operators in Gevrey classes; Preprint.

GRUSHIN V.V.:[1] On a class of hypoelliptic operators; Mat. Sb. 83:3 (1970), 456-473 (Russian); Math. USSR-Sb. 12:3 (1970), 458-476.

— [2] A differential equation without a solution; Mat. Zametki 10 (1971), 125-128 (Russian); Math. Notes 10 (1971), 499-501.

— [3] On a class of elliptic pseudodifferential operators degenerate on a submanifold; Mat. Sb. 84:2 (1971), 163-195 (Russian); Math. USSR-Sb. 13:2 (1971), 155-185.

HÖRMANDER L.:[1] On the theory of general partial differential operators; Acta Math. 94 (1955), 161-248.

— [2] Differential operators of principal type; Math. Ann. 140 (1960), 124-146.

— [3] Differential equations without solutions; Math. Ann. 140 (1960), 169-173.

— [4] Linear Partial Differential Operators; Springer-Verlag, Berlin 1963.

— [5] Pseudo-differential operators; Comm. Pure Appl. Math. 18 (1965), 501-517.
— [6] Pseudo-differential operators and non-elliptic boundary problems; Ann. of Math. 83:2 (1966), 129-209.
— [7] Fourier integral operators I; Acta Math. 127 (1971), 79-183.
— [8] Propagation of singularities and semiglobal existence theorems for (pseudo-) differential operators of principal type; Ann. of Math. 108 (1978), 569-609.
— [9] Pseudo-differential operators of principal type; *in:* H.G. Garnir (ed.): Singularities in Boundary Value Problems, Nato Adv. Study Inst., Maratea, 1980, 69-96, Reidel Publ. Co., Dordrecht 1981.
— [10] The Analysis of Linear Partial Differential Operators, vol. III & IV; Springer-Verlag, Berlin 1985.

IVRII V.Ya.:[1] Differential equations with multiple characteristics and with no solutions; Dokl. Akad. Nauk SSSR 198:2 (1971), 279-282 (Russian); Soviet Math. Dokl. 12:3 (1971), 769-772.
— [2] Condition for correctness in Gevrey classes of the Cauchy problem for weakly hyperbolic equations; Sibirsk. Mat. Zh. 17 (1976), 547-563 (Russian); Siberian. Math. J. 17 (1976), 422-435.
— [3] Correctness of the Cauchy problem in Gevrey classes for nonstrictly hyperbolic operators; Mat. Sb. 96:3 (1975), 390-413 (Russian); Math. USSR-Sb. 25:3 (1975), 365-387.

IVRII V.Ya., PETKOV V.M.:[1] Necessary conditions for the Cauchy problem for non-strictly hyperbolic equations to be well-posed; Uspekhi Mat. Nauk 29:5 (1974), 3-70 (Russian); Russian Math. Surveys 29:5 (1974), 1-70.

KANNAI Y.:[1] An unsolvable hypoelliptic differential operator; Israel J. Math. 9 (1971), 306-315.

KARATOPRAKLIEVA M.G.:[1] On local solvability and hypoellipticity of a class of differential operators with double characteristics; Serdica 8 (1982), 367-377 (Russian).

KIM J., KIM J.S., SHIN J.K.:[1] Unsolvability of the Mizohata operator; Bull. Korean Math. Soc. 18:1 (1981), 9-13.

KOHN J.J., NIRENBERG L.:[1] An algebra of pseudo-differential operators; Comm. Pure Appl. Math. 18 (1965), 269-305.

KOMATSU H.:[1] Ultradistributions, I. Structure theorems and a characterization; II. The kernel theorem and ultradistributions with support in a submanifold; III. Vector

valued ultradistributions and the theory of kernels; J. Fac. Sci. Univ. Tokyo Sect. IA Math. 20 (1973), 25-105; 24 (1977), 607-628; 29 (1982), 653-717.

KWON K.Y.:[1] Hypoellipticity and local solvability of operators with double characteristics; Comm. Partial Differential Equations 10:5 (1985), 525-542.

LERNER N.:[1] Sufficiency of condition (Ψ) for local solvability in two dimensions; Ann. of Math. 128:2 (1988), 243-258.

LEWY H.:[1] An example of a smooth linear partial differential equation without solution; Ann. of Math. 66:1 (1957), 155-158.

LIESS O., RODINO L.:[1] Inhomogeneous Gevrey classes and related pseudodifferential operators; Boll. Un. Mat. Ital. C (6) 3 (1984), 233-323.

LORENZ M.:[1] Unsolvable hypoelliptic differential operators with a totally characteristic point; Math. Nachr. 114 (1983), 151-161.

— [2] An elliptic differential operator degenerating at one point, which is hypoelliptic but not locally solvable; Math. Nachr. 120 (1985), 237-247.

LU L.J.:[1] The local solvability of a class of pseudo-differential equations of principal type; Acta Math. Sci. 6 (1986), 25-36.

MASCARELLO RODINO M.:[1] Sulla risolubilità locale di alcuni operatori pseudo differenziali con caratteristiche multiple; Rend. Sem. Mat. Univ. Politec. Torino 35 (1976-77), 27-34.

MENDOZA G.A., UHLMANN G.A.:[1] A necessary condition for local solvability for a class of operators with double characteristics; J. Funct. Anal. 52 (1983), 252-256.

— [2] A sufficient condition for local solvability for a class of operators with double characteristics; Amer. J. Math. 106 (1984), 187-217.

MENIKOFF A.:[1] On local solvability of pseudo-differential equations; Proc. Amer. Math. Soc. 43:1 (1974), 149-154.

— [2] Some examples of hypoelliptic partial differential equations; Math. Ann. 221 (1976), 167-181.

— [3] Pseudo-differential operators with double characteristics; Math. Ann. 231 (1977), 145-180.

— [4] On hypoelliptic operators with double characteristics; Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 4 (1977), 689-724.

MIZOHATA S.:[1] Solutions nulles et solutions non analytiques; J. Math. Kyoto Univ. 1:2 (1962), 271-302.

MOYER R.D.:[1] On the Nirenberg-Treves condition for local solvability; J. Differential

Equations 26 (1977), 223-239.

NIRENBERG L., TREVES F.:[1] Solvability of a first order linear partial differential equation; Comm. Pure Appl. Math. 16 (1963), 331-351.

— [2] On local solvability of linear partial differential equations. Part I: Necessary conditions; Comm. Pure Appl. Math. 23 (1970), 1-38.

— [3] On local solvability of linear partial differential equations. Part II: Sufficient conditions; Comm. Pure Appl. Math. 23 (1970), 449-510.

— [4] A Correction to: "On local solvability of linear partial differential equations. Part II: Sufficient conditions"; Comm. Pure Appl. Math. 24 (1971), 279-288.

OGANESJAN A.O.:[1] Local solvability and hypoellipticity for a class of second-order equations; Akad. Nauk Armjan. SSR Dokl. 70:2 (1980), 85-91 (Russian).

OKAJI T.: [1] The local solvability of partial differential operator with multiple characteristics in two independent variables; J. Math. Kyoto Univ. 20:1 (1980), 125-140.

— [2] On local solvability of the operator $D_t^3 + at^k D_x^n$; J. Math. Kyoto Univ. 22:1 (1982), 97-114.

— [3] On local solvability of some non-kowalevskian partial differential operators; J. Math. Kyoto Univ. 22:4 (1983), 619-642.

— [4] Gevrey-hypoelliptic operators which are not C^∞-hypoelliptic; J. Math. Kyoto Univ. 28:2 (1988), 311-322.

— [5] A class of pseudo-differential operators of logarithmic type and infinitely degenerate hypoelliptic operators; J. Math. Kyoto Univ. 28:2 (1988), 323-334.

PARENTI C., RODINO L.:[1] Parametrices for a class of pseudo differential operators I; Ann. Mat. Pura Appl. (4) 125 (1980), 221-278.

POPELYUKHIN A.S.:[1] The necessary conditions for local solvability of a class of pseudodifferential operators with double involutory characteristics; Vestnik Moskov. Univ. Ser. I Mat. Mekh. 43:1 (1988), 98-100 (Russian); Moscow Univ. Math. Bull. 43:1 (1988), 93-95.

— [2] On local solvability of pseudodifferential operators with double characteristics; Vestnik Moskov. Univ. Ser. I Mat. Mekh. 43:3 (1988), 75-77 (Russian); Moscow Univ. Math. Bull. 43:3 (1988), 76-78.

POPIVANOV P.R.:[1] On the local solvability of a class of pseudodifferential equations with double characteristics; Trudy Sem. Petrovsk. 1 (1975), 237-278 (Russian); Amer. Math. Soc. Transl. 118:2 (1982), 51-90.

— [2] Local properties of pseudo-differential equations with multiple characteristics; C. R. Acad. Bulg. Sci. 28:8 (1975), 1015-1018 (Russian).

— [3] Local solvability of pseudodifferential operators with characteristics of second multiplicity; Mat. Sb. 100:2 (1976), 217-241 (Russian); Math. USSR-Sb. 29:2 (1976), 193-216.

— [4] Local properties of linear pseudodifferential operators with multiple characteristics; C. R. Acad. Bulg. Sci. 29:4 (1976), 461-464 (Russian).

— [5] Sufficient conditions for the local solvability of a class of pseudodifferential operators of non-principal type; C. R. Acad. Bulg. Sci. 30:7 (1977), 981-984 (Russian).

— [6] Local properties of pseudodifferential operators with double involutive characteristics; Uspekhi Mat. Nauk 33:4 (1978), 223-224 (Russian); Russian Math. Surveys 33:4 (1978), 263-264.

— [7] On local properties of pseudodifferential operators with multiple characteristics; *in:* P.S. Aleksandrov, O.A. Oleinik (ed.): Reports (Trudy) of All-Union Conference in Partial Differential Equations, dedicated to the 75th Anniversary of I.G. Petrovski, Moscow, 193-199, Izd. Moskov. Univ., Moscow 1978 (Russian).

— [8] A class of differential operators with multiple characteristics which have not solutions; Pliska Stud. Math. Bulgar. 3 (1981), 47-60 (Russian).

— [9] A link between small divisors and smoothness of the solutions of a class of partial differential operators; Ann. Glob. Anal. and Geom. 1:3 (1983), 77-92.

— [10] Microlocal properties of pseudo-differential operators with double involutive characteristics; C. R. Acad. Bulg. Sci. 37:9 (1984), 1163-1167 (Russian).

POPIVANOV P.R., GEORGIEV Ch.:[1] Necessary conditions for local solvability of operators with double characteristics; Annuaire Univ. Sofia Fac. Math. Méc. 75 (1981), 57-71 (Bulgarian).

POPIVANOV P.R., POPOV G. S.:[1] Microlocal properties of a class of pseudodifferential operators with multiple characteristics; Serdica 6 (1980), 169-183 (Russian).

— [2] A priori estimates and some microlocal properties of a class of pseudodifferential operators; C. R. Acad. Bulg. Sci. 33:4 (1980), 461-463.

ROBERTS G.B., WENSTON P.R.:[1] Local solvability and hypoellipticity for operators with odd order characteristics; Comm. Partial Differential Equations 7:6 (1982), 715-741.

RODINO L.:[1] Nonsolvability of a higher-order degenerate elliptic pseudodifferential equation; Amer. J. Math. 102:1 (1980), 1-12.

— [2] Risolubilità locale in spazi di Gevrey; *in:* Atti del Meeting "Metodi di analisi reale nelle equazioni alle derivate parziali", Cagliari, 1985, 21-33, Tecnoprint, Bologna 1985.

— [3] Local solvability in Gevrey classes; *in:* F. Colombini, M.K.V. Murthy (ed.): Hyperbolic Equations, Proceedings of the Conference "Hyperbolic Equations and Related Topics", Padova, 1985, 167-185, Longman, Harlow 1987.

— [4] On linear partial differential operators with multiple characteristics; *in:* "Symposium Partial Differential Equations", Holzhau, 238-249, Teubner, Leipzig 1988.

RODINO L., CORLI A.:[1] Gevrey solvability for hyperbolic operators with constant multiplicity; *in:* L. Cattabriga, F. Colombini, M.K.V. Murthy, S.Spagnolo (ed.): Recent Developments in Hyperbolic Equations, Proceedings of the Conference "Hyperbolic Equations", Pisa, 1987, 290-304, Longman, Harlow 1988.

RODINO L., ZANGHIRATI L.:[1] Pseudo differential operators with multiple characteristics and Gevrey singularities; Comm. Partial Differential Equations 11:7 (1986), 673-711.

RUBINSTEIN R.:[1] Local solvability of the operator $u_{tt} + ia(t)u_x + b(t)u_t + c(t)u$; J. Differential Equations 14 (1973), 185-194.

— [2] Examples of nonsolvable partial differential equations; Trans. Amer. Math. Soc. 199 (1974), 123-129.

SCHAPIRA P.:[1] Une équations aux dérivées partielles sans solutions dans l'espace des hyperfonctions; C. R. Acad. Sci. Paris Sér. A 265 (1967), 665-667.

— [2] Solutions hyperfonctions des équations aux dérivées partielles du premier ordre; Bull. Soc. Math. France 97 (1969), 243-255.

SHANANIN N.A.:[1] On local solvability of equations of quasi-principal type; Mat. Sb. 97:4 (1975), 503-516 (Russian); Math. USSR-Sb. 26:4 (1975), 458-470.

— [2] An example of a locally unsolvable differential equation of quasiprincipal type with a real-valued weighted principal symbol; Mat. Zametki 19:5 (1976), 755-761 (Russian); Math. Notes 19 (1976), 447-451.

— [3] Necessary conditions for the local solvability of equations of quasi-principal type; Uspekhi Mat. Nauk 32:2 (1977), 235-236 (Russian).

— [4] On local unsolvability of differential equations with weighted derivatives; Mat.

Sb. 111:3 (1980), 465-477 (Russian); Math. USSR-Sb. 39:3 (1981), 417-428.

— [5] On local unsolvability and nonhypoellipticity of (pseudo-)differential equations with weighted symbols; Mat. Sb. 119:4 (1982), 548-563 (Russian); Math. USSR-Sb. 47:2 (1984), 541-556.

— [6] On local properties of a certain class of evolution equations in a Hilbert space; Uspekhi Mat. Nauk 42:2 (1987), 251-252 (Russian); Russian Math. Surveys 42:2 (1987), 295-296.

SJÖSTRAND J.:[1] Parametrices for pseudodifferential operator with multiple characteristics; Ark. Mat. 12 (1974), 85-130.

TREVES F.:[1] Topological Vector Spaces, Distributions and Kernels; Academic Press, New York 1967.

— [2] The local solvability of linear partial differential equations in two independent variables; Amer. J. Math. 92 (1970), 174-204.

— [3] On local solvability of linear partial differential equations; Bull. Amer. Math. Soc. 76:1 (1970), 552-571.

— [4] A link between solvability of pseudodifferential equations and uniqueness in the Cauchy problem; Amer. J. Math. 94 (1972), 267-288.

— [5] Concatenations of second-order evolution equations applied to local solvability and hypoellipticity; Comm. Pure Appl. Math. 26 (1973), 201-250.

— [6] On the local solvability of linear partial differential equations; Uspekhi Mat. Nauk 29:2 (1974), 252-281 (Russian); Russian Math. Surveys 29:2 (1974), 263-292.

— [7] Winding numbers and the solvability condition (Ψ); J. Differential Geom. 10 (1975), 135-150.

VOLEVICH L.R., GINDIKIN S.G.:[1] The Newton polyhedron and local solvability of linear partial differential equations; Trudy Moskov. Mat. Obshch. 48 (1985), 211-262 (Russian); Trans. Moscow Math. Soc. 48 (1986), 227-276.

WENSTON P.R.:[1] On local solvability of linear partial differential operators not of principal type; J. Differential Equations 22 (1976), 111-144.

— [2] A necessary condition for the local solvability of the operator $P_m^2(x,D)+P_{2m-1}(x,D)$; J. Differential Equations 25 (1977), 90-95.

— [3] A local solvability result for operators with characteristics having odd order multiplicity; J. Differential Equations 28 (1978), 369-380.

— [4] A sufficient condition for the local solvability of a linear partial differential

operator with double characteristics; J. Differential Equations 29 (1978), 374-387.

YAMAMOTO K.:[1] Parametrices for pseudo-differential equations with double characteristics I; Hokkaido Math. J. 5 (1976), 280-301.

YAMASAKI A.:[1] On a necessary condition for the local solvability of pseudo-differential operators with double characteristics; Comm. Partial Differential Equations 5:3 (1980), 209-224.

— [2] On the local solvability of $D_1^2+A(x_2,D_2)$; Math. Japonica 28:4 (1983), 479-485.

YOSHIKAWA A.:[1] On the hypoellipticity of differential operators; J. Fac. Sci. Univ. Tokyo Sect. IA Math. 14 (1967), 81-88.

Andrea Corli
Dipartimento di Matematica
Università di Ferrara
Via Machiavelli 35
I-44100 Ferrara, Italy

Luigi Rodino
Dipartimento di Matematica
Università di Torino
Via Carlo Alberto 10
I-10123 Torino, Italy

Constantin Carathéodory: An International Tribute (pp. 181–218)
edited by Th. M. Rassias

ENTROPY AND CURVATURE

JERRY DONATO

ABSTRACT. The necessary and sufficient conditions of reversible processes and irreversible processes are given using an invariant intrinsic covariant curvature tensor, R^{ℓ}_{ijk}. Two observations are made: entropy is curvature and disorder reflects path dependence. Equilibrium and stable processes are specified as $R^{\ell}_{ijk} = 0$ and non-equilibrium and unstable processes are specified as $R^{\ell}_{ijk} \gtrless 0$. Cartan's method of equivalence incorporates curvature, torsion and group properties into the analysis.

1. INTRODUCTION.

Section 2 presents the fundamental existence theorem for ordinary differential equations.

Section 3 restates the existence theorem of ordinary differential equations in a geometric form using the notions of vector fields and integral curves. The inclusion of a geometric object, denoted as M, in the procedure and the corresponding

composition of functions is presented. The theory of Lie groups and their interdependence between the geometrical interpretation of the integral curves of the vector field and the one-parameter group of transformation is mentioned.

Section 4 states the theorem that any Pfaffian differential equation in $\mathbf{R}^2$ admits an integrating factor. The exponential form of the integrating factor for a homogeneous linear first order differential equation is presented using a well-known ordinary differential equation. The general algebraic properties of linear exponential operators are mentioned. The following theorem is stated: if X is a skew-symmetrix matrix, then $\exp X$ is orthogonal and its relationship to the theory of Maurer-Cartan forms, Lie group theory and the first fundamental theorem of Lie is mentioned.

Section 5 presents the Frobenius integrability conditions for the Pfaffian equation. An example is given for $\mathbf{R}^3$ and the symmetry and cycle properties are noted. A recent paper by Lacomba and Hernández extensively illustrate these symmetry (reciprocity) conditions in many physical systems including thermodynamics; they also present the Inaccessibility Theorem of Constantin Caratheodory.

Section 6 presents Stokes' Theorem using differential forms.

Section 7 presents Poincare's Lemma and Its Converse and notes its relationship to the integrability conditions, that is, the order of taking partial derivatives commutes; this indicates path independent movements.

Section 8 describes a differential manifold.

Section 9 presents Gauss's Theorema Egregium and notes that intrinsic geometric properties of a surface should be investigated.

Section 10 briefly presents the subject of tensor calculus and develops the three-index symbols $\{ij, k\}$. The geometric notion of the parallel transport of a vector is developed into an invariant intrinsic covariant Riemannian-Christoffel curvature tensor, denoted as R^{ℓ}_{ijk}. Some algebraic properties of this tensor are mentioned. The following important theorem is stated: In order that a Riemannian space be

flat, it is necessary and sufficient that the components of its curvature tensor vanish identically.

Section 11 presents a brief review of some thermodynamic concepts. The inexact differentials of work, denoted as $đW$, and of heat, denoted as $đQ$ are presented. The first law of thermodynamics is presented and an example from a hydrostatic system is given where the integrability conditions of the differential equations are not met, that is, Poincare's Lemma and Its Converse are violated. Experimentation reveals that the integrating factor for systems is a thermodynamic temperature that can be defined provided that the second law of thermodynamic exists.

Section 12 describes what a reversible process is and what an irreversible process is. The following statements on thermodynamics are presented: Kelvin-Planck, Clausius and Carathéodory. The observation is made that natural processes are irreversible.

Section 13 presents the general geometric axioms of the necessary and sufficient conditions of reversible processes in terms of an invariant intrinsic covariant curvature tensor, R^{ℓ}_{ijk}. When R^{ℓ}_{ijk} exists, path dependent movements (irreversible processes) exists and when $R^{\ell}_{ijk} = 0$ path independent movements (reversible processes) exist. The inexact differentials $đW$ and $đQ$ are reflected in R^{ℓ}_{ijk}. The second law of thermodynamics is stated, that is, the notion of entropy is presented. The observation is made that entropy is reflected in the curvature tensor, R^{ℓ}_{ijk}. This observation suggests the following: entropy is curvature. If entropy is viewed as a disorder concept, then, geometrically, disorder reflects the path dependence of the processes.

Section 14 briefly presents the four equations corresponding to internal energy, enthalpy, Helmholtz function and Gibbs function and relates them to Maxwell's relations and their corresponding characteristic independent variables that are coupled through the Legendre transform. The surfaces associated with these functions (and also the PVT system) and their thermodynamic information are noted. Gauss's Theorema Egregium of 1827 and Riemann's presentation of 1854 points to the

mathematical investigations of the surfaces (manifolds) themselves. This suggests that the equilibrium/non-equilibrium processes and stable/unstable processes be investigated from a full geometric viewpoint. A non-equilibrium and an unstable concept are briefly described. The analysis suggests that equilibrium and stable processes can be specified as $R^{\ell}_{ijk} = 0$ and non-equilibrium and unstable processes can be specified as $R^{\ell}_{ijk} \gtrless 0$. This suggests that the observer of processes should use the postulates of non-Euclidean geometry.

Section 15 briefly presents Élie Cartan's method of equivalence that may be used in the analysis. Cartan's method of moving frames incorporates the mathematical concepts of curvature, torsion and group properties. Cartan's method considers the notions of the lifting of the linear group to G spaces and the related mapping of intrinsic torsion has to be considered in addition to curvature. If torsion is not constant and the group reduction and normalization processes are not constant, then many other possibilities and problems appear.

2. ORDINARY DIFFERENTIAL EQUATIONS.

The existence theorem for ordinary differential equations (ODE) is as follows:

THEOREM 2.1. ([4]) *Let $U \subset \mathbf{R}^n$ be an open set and I_ϵ, $\epsilon > 0$ denote the interval $-\epsilon < t < \epsilon$, $t \in \mathbf{R}$. Suppose $f^i(t, x^1, \ldots, x^n)$, $i = 1, \ldots, n$ be a function of class C^r, $r \geq 1$ on $I_\epsilon X U$. Then for each $x \in U$ there exists $\delta > 0$ and a neighborhood V of x, $V \subset U$ such that*

(I) For each $a = (a^1, \ldots, a^n) \in V$ there exists an n-tuple of C^r functions $x(t) = (x^1(t), \ldots, x^n(t))$ defined on I_δ and mapping I_δ into U which satisfy the system of first order differential equations

$$(*) \qquad \frac{dx^i}{dt} = f^i(x, t), \qquad i = 1, \ldots, n$$

and the initial conditions

$$(**) \qquad x^i(0) = a^i \qquad i = 1, \ldots, n\,.$$

For each a, the functions $x(t) = (x^1(t), \ldots, x^n(t))$ are uniquely determined in the sense that any other function $\overline{x}^1(t), \ldots, \overline{x}^n(t)$ satisfying (**) and (**) must agree with $x(t)$ on their common domain which includes I_δ.

(II) These functions being uniquely determined by $a = (a^1, \ldots, a^n)$ for every $a \in V$, can be written as $x^i(t, a^1, \ldots, a^n)$, $i = 1, \ldots, n$ in which case they are class C^r in all variables and hence determine a C^r map of $I_\delta \times V \to U$.

3. VECTOR FIELDS.

The hypotheses and conclusions of the fundamental existence theorem of ordinary differential equations can be restated in a coordinate free or geometric form using the concepts of vector fields and integral curves.

DEFINITION 3.1. ([3]) *Let $\varphi : I \to M$ be a differential mapping on an open interval of the t axis into M such that $\varphi(0) = x \in M$ and further let $x^i : M \to \mathbf{R}^n$ $(i = 1, \ldots, n)$ be a mapping from M into an admissible coordinate system $\mathbf{R}^n$. Hence the following procedure is*

$$I \xrightarrow{\varphi} M \xrightarrow{x^i} \mathbf{R}^n .$$

Now the ODE takes the following form

$$\frac{dx^i}{dt}\Big|_{t=0} = \frac{d}{dt}\Big|_{t=0} (x^i \circ \varphi)(i = 1, \ldots, n) .$$

Remark 3.2. Observe the inclusion of the geometric object M in the procedure and the corresponding composition of functions.

DEFINITION 3.3. ([23]) *Let $X \in \mathfrak{X}(M)$ be a vector field on M^n. A differentiable curve $\gamma : (a, b) \subseteq \mathbf{R} \to M$ is called an integral curve for X if*

$$\dot{\gamma}(t) = X(\gamma(t)) \quad \text{for } t \in (a, b)$$

or

$$\gamma_*\left(\frac{d}{dt}\right) = X\big|_{\gamma(a,b)}$$

where $X\big|_{\gamma(a,b)}$ means the restriction of X to $\gamma(a,b)$.

THEOREM 3.4. ([23]) *The curve $\gamma : t \to \gamma(t) = (x^i(t)), \quad 1 \le i \le n$ is an integral curve for the vector field*

$$X = \sum a^i(x)\frac{\partial}{\partial x^i}$$

if and only if $x^i(t)$ is a solution of the system of differential equations:

$$\frac{dx^i}{dt} = a^i(x(t)), \quad i = 1, \ldots, n\,.$$

DEFINITION 3.5. ([23]) *A tangent vector can be defined as*

$$X = \sum_{i=1}^{n} \frac{dx^i}{dt}\frac{\partial}{\partial x_i}$$

where $\frac{\partial}{\partial x_i}$ represents the basis in a given coordinate system.

DEFINITION 3.6. ([23]) *Let $x \in M$ and pick a tangent vector X_x at x. This assignment*

$$X : x \to X_x$$

is called a vector field on M. With respect to a local coordinate system $(x^1, \ldots, x^n)$, X_x can be expressed uniquely as

$$X_x = \sum_{i=1}^{n} \xi^i(x)\left(\frac{\partial}{\partial x_i}\right)_x .$$

The n functions $\xi^i (1 \le i \le n)$ are referred to as the components of X with respect to the given coordinate system. Only C^∞ vector fields are considered, that is, only vector fields whose components are C^∞ functions on a neighborhood of each point $x \in M$.

Remark 3.7. ([23]) The theory of Lie groups was developed by Sophus Lie in the study of systems of differential equations:

$$\frac{dx^i}{dt} = a^i(x(t)), \quad i = 1, \ldots, n$$

where C^∞-functions $a^i(x(t))$ are the components of a vector field $X = \sum_{i=1}^n a^i \frac{\partial}{\partial x^i}$. Hence the theory of Lie groups may be regarded as an interdependence between the geometrical interpretation of the integral curves of the vector field X as the solutions $x^i = x(t)$ of the system and the one-parameter group of transformation which is generated by $X = \sum_{i=1}^n a^i \frac{\partial}{\partial x_i}$ and which solves the system of differential equations.

4. INTEGRATING FACTORS.

THEOREM 4.1. ([23]) *Any Pfaffian differential equation in* $\mathbb{R}^2$

$$\omega = P dx + Q dy = 0\,, \qquad P = P(x,y);\ Q = Q(x,y)$$

admits an integrating factor.

Remark 4.2. ([21]) Consider the following homogeneous linear first order differential equation in its standard form

$$\frac{dy}{dx} + P(x)y = 0$$

multiply by a factor μ to get

$$\mu\frac{dy}{dx} + P\,\mu y = 0\,.$$

Apply Leibniz rule and divide by dx to get

$$\frac{d}{dx}(\mu y) = \mu\frac{dy}{dx} + y\frac{d\mu}{dx}$$

The last two equations will be equal if and only if

$$\frac{d\mu}{dx} = P\mu\,.$$

The last equation is a variables separable type, hence

$$\frac{d\mu}{\mu} = P\,dx$$

$$\ell n\,\mu = \int P\,dx$$

Therefore, the following function appears

$$\mu = e^{\int P\,dx}\,.$$

The explicit solution for the differential equation is

$$y = Ce^{-\int P\,dx}$$

where C is the constant of integration.

Example 4.3. ([5]) An application of the exponential form of the integrating factor to solve a well-known ODE can be viewed in the following way:

Consider

$$\frac{dx}{dt} + Ax = 0\,.$$

Multiply by e^{At} to give

$$e^{At}\frac{dx}{dt} + Ae^{At}x = 0\,.$$

Hence

$$d(xe^{At}) = 0$$

that is, xe^{At} is a constant. The above procedure can also be reversed; this suggest a type of inverse concept.

Example 4.4. (5,9,10]) Again, consider the following well-known ODE

$$\frac{dx}{dt} = Ax\,.$$

The standard scenario used in solving the above equation is as follows:

Divide both sides by x, multiply both sides by dt and integrate:

$$\int \frac{dx}{x} = A\int dt\,.$$

Integrating gives the following equation:

$$\ell n\, x = At + \text{ constant}$$

and the above equation can be rewritten as

$$x = e^{At} + e^{B} = e^{At}e^{B}$$

Geometrically, the above equation represents a one-parameter family of curves called integral curves. Each integral curve is the geometric representation of the corresponding solution of the differential equation. Specifying a particular solution means picking out a particular integral curve from the one-parameter family. This is usually done by prescribing a point normally referred to as an initial condition through which the integral curve must pass.

The solution of the ODE reveals that the exponential form of the constant of integration be multiplied by the exponential form of the parameter. In other words, the exponential product equals the sum of the exponentials, that is, multiplication commutes and addition is associative.

The solution of the ODE can also be viewed as a family of linear exponential operators which is a one-parameter group of linear transformations provided the operators, A and B, commute.

However, in general, the exponential operators do not commute, this means that addition is non-associative and multiplication is non-commutative. When this notion is applied to solving ODE, the exponential linear relationship between initial conditions and the parameter no longer holds. An obstruction to solving ODE has appeared. This obstruction reveals the existence of a geometric object called curvature; with such an object existing the analysis becomes path dependent.

THEOREM 4.5. ([13]) *If X is an orthogonal matrix whose elements are functions of any number of variables, then*

$$(dX)X^{-1}$$

is a skew-symmetric matrix of one-forms.

PROOF: ${}^tXX = I$ where t denotes a transpose.

$${}^tdXX + {}^tXdX = 0\,.$$

For an orthogonal matrix its inverse is its transpose, that is,

$$^tX = X^{-1} .$$

Hence

$$^tX^{-1}\,{}^tdX + dXX^{-1} = 0$$
$$^t(dXX^{-1}) + dXX^{-1} = 0 .$$

The following converse can also be established.

THEOREM 4.6. ([13]) *Suppose X is a matrix of functions defined on a domain U. Suppose X is orthogonal at a single point of U and that*

$$dX = AX$$

where A is a skew-symmetric matrix of one-forms. Then A is orthogonal on all of U.

PROOF: Let $C = {}^tXX$ then

$$dC = ({}^tdX)X + {}^tX(dX)$$
$$dC = (-{}^tXA)X + {}^tX(AX) = 0 .$$

Hence C is a constant matrix on U.

Now assume $C = I$ at one point of U, that is,

$$C = I \text{ on } U .$$

Then

$$^tXX = I \text{ on } U .$$

Hence X is orthogonal.

Remark 4.7. The form

$$(dX)X^{-1}$$

suggests the theory of Maurer-Cartan forms, Lie group theory and the first fundamental theorem of Lie.

THEOREM 4.8. ([13]) *If X is skew-symmetric matrix, then e^X is orthogonal where*

$$e^X = I + \sum_{n=1}^{\infty} \frac{X^n}{n!}$$

for the real matrix X.

5. INTEGRABILITY CONDITIONS.

THEOREM 5.1. ([23]) *The Frobenius condition $d\omega \wedge \omega = 0$ is called integrability condition for the Pfaffian equation $\omega = 0$.*

Example 5.2. ([2]) Given the following Pfaffian equation in $\mathbf{R}^3$ (which can be extended to $\mathbf{R}^n$) where $\omega = 0$

$$\omega = P(x,y,z)dx + Q(x,y,z)dy + R(x,y,z)dz = 0$$

then

$$d\omega \wedge \omega = \left[P\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + Q\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + R\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\right] dx \wedge dy \wedge dz\,.$$

Hence $\omega = 0$ is integrable if and only if the term in the square brackets vanishes.

Remark 5.3. Observe the cycle properties of P, Q and R, and x, y and z. These cycle properties appear in various forms in mathematical analysis especially from an algebraic viewpoint.

Remark 5.4. Observe also the symmetry properties. Symmetry also appears in various forms throughout the subject of mathematics and its applications. In a

recent paper ([17]) by E.A. Lacomba and D.B. Hernández titled "On the Role of Reciprocity Conditions in the Formulation of Conservation Laws and Variational Principles", they observed that certain symmetry conditions occur in many field theories such as thermostatics (in the form of Maxwell's relations), particle mechanics, Hamiltonian systems and electric circuits. Their analysis is based on the notions of differentiable manifolds and exterior algebra. The notion of reciprocity is extensively illustrated in their paper using the subject of thermodynamics; they cover the fundamental concepts of the subject and present the very famous *Inaccessibility Theorem of Constantin Caratheodory.*

Example 5.5. ([23]) Let $\omega \in F^1(\mathbf{R}^n)$ be a Pfaffian which does not vanish then what is the solution to the following:

Under what conditions are there functions $f, g : U \to \mathbf{R}$ (where U is a neighborhood of $\mathbf{R}^n$) satisfying

$$\omega = g\, df\ ?$$

To answer the question in general, the concept of an integral manifold is needed.

DEFINITION 5.6. ([23]) *An $(n-1)$-dimensional submanifold N of $\mathbf{R}^n$ given by $x^i = x^i(u^1, \ldots, u^{n-1})$ where $(u^1, \ldots, u^{n-1}) \in U \subset \mathbf{R}^{n-1}$ is called an integral manifold for the one-form ω if ω annihilates the tangent space $T_x(N)$ to N at every point $x \in N$, that is,*

$$\langle \omega_x, x \rangle = \omega_x(X_x) = 0\,, \quad \forall X_x \in T_x(N)\,.$$

THEOREM 5.7. ([23]) *A submanifold $N \subset \mathbf{R}^n$ is an integral manifold for the Pfaffian $\omega = \sum_{i=1}^n a_i dx^i \in F^1(\mathbf{R}^n)$ if and only if the system of partial differential equations*

$$\sum_{k=1}^{n} a_k(x) \frac{\partial x^k}{\partial x^\alpha} = 0\,, \quad 1 \le \alpha \le n-1$$

has a solution.

Example 5.8. ([23]) Given a one-form $\omega \in F^1(\mathbf{R}^n)$ which is nowhere zero in a neighborhood U of the Euclidean space $\mathbf{R}^n$. Does there exist a function $f : U \to \mathbf{R}$

such that $df \neq 0$ on U and such that the submanifold (hypersurfaces) of the type

$$N := \{x \in U | f(x) \text{ constant}, df(x) \neq 0\}$$

are integral surfaces for ω?

The answer is the following theorem.

THEOREM 5.9. ([23]) *A necessary condition (and also a sufficient condition) for a function $f : U \to \mathbf{R}$ to exist satisfying above example is the condition of Frobenius*

$$d\omega \wedge \omega = 0 .$$

Remark 5.10. ([13]) Consider the following equation in example 5.5

$$\omega = g\, df .$$

Using an inverse operation, rewrite as

$$df = g^{-1}\omega .$$

Apply the exterior algebraic operations to the first equation to get

$$d\omega = dg \wedge df$$

then substitute to get

$$d\omega = dg \wedge g^{-1}\omega = \frac{dg}{g} \wedge \omega .$$

Then $d\omega = \theta \wedge \omega$ where

$$\theta = g^{-1} dg = d\, \ell n|g| .$$

Hence

$$\omega \wedge d\omega = \omega \wedge \theta \wedge \omega = 0 .$$

6. STOKE'S THEOREM.

DEFINITION 6.1. ([20]) *An oriented 2-manifold with boundary in* $\mathbf{R}^3$ *is a surface with boundary in* $\mathbf{R}^3$ *whose boundary is a simple closed curve with orientation; an oriented 3-manifold in* $\mathbf{R}^3$ *is an elementary region in* $\mathbf{R}^3$ *whose boundary which is a surface, is given the outward orientation.*

THEOREM 6.2. ([20]) *Let* M *be an oriented* k*- manifold in* $\mathbf{R}^3 (k = 2 \text{ or } 3)$ *contained in some open set* K*. Suppose* ω *is a* $(k-1)$ *form on* K*, then*

$$\int_{\partial M} \omega = \int_{M} d\omega \, .$$

7. POINCARÉ'S LEMMA AND ITS CONVERSE.

LEMMA 7.1. ([23]) *If* ω *is a* p *form on* M *(manifold) for which there exists a* $(p-1)$ *for* α *such that* $d\alpha = \omega$*, then* $d\omega = 0$*.*

LEMMA - ITS CONVERSE 7.2. ([23]) *If* ω *is a* p*-form on an open set* $U \subset M$ *(which is contractible to a point) such that* $d\omega = 0$*, then there exists a* $(p-1)$ *form* α *such that* $\omega = d\alpha$*. (Exceptions: if* $p = 0$*, then* $\omega = f$ *and the vanishing of* df *means* f *is constant.*

Remark 7.3. ([10]) The above two lemmas establish the integrability conditions of differential equations; that is, the order of taking partial derivatives commutes and the conditions of path independent movements on a differential manifold. These notions are related to well-known theorems in calculus on path independent integrals and exact differentials.

8. DIFFERENTIABLE MANIFOLD.

DEFINITION 8.1. ([4]) *Begin with the definition of a topological manifold* M *of dimensions* n *which has the following properties: 1) it is a Hausdorff space*

with countable basis of open sets and 2) each point has a neighborhood homeomorphic to an open subset of $\mathbf{R}^n$. *Each pair* U, φ *where* U *is an open set of* M *and* φ *is a homeomorphism of* U *to an open subset of* $\mathbf{R}^n$, *is called a coordinate neighborhood; to* $q \in U$ *assign the* n *coordinates* $x^1(q), \ldots, x^n(q)$ *of its image* $\varphi(x)$ *in* $\mathbf{R}^n$ *where each* $x^i(q)$ *is a real valued function on* U, *the* i*th coordinate function. If* q *also lies in a second coordinate* V, ψ, *then it has coordinates* $\overline{x}^1(q), \ldots, \overline{x}^n(q)$ *in this neighborhood. Since* φ *and* ψ *are homeomorphisms, the following defines a homeomorphisms*

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V) .$$

The domain and range are the two open subsets of $\mathbf{R}^n$ *which correspond to the points* $U \cap V$ *by the two coordinate maps* φ *and* ψ *respectively. Similarly* $\varphi \circ \psi^{-1}$ *gives the inverse mapping.*

The fact that $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ are homeomorphisms and inverses to each other implies the continuity of the functions in coordinate form together with the corresponding identities. If $U \cap V$ is non-empty and U, φ and V, ψ are C^∞ compatible, then this implies that the change of coordinates is also C^∞; this is equivalent to requiring $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ to be diffeomorphism of the open subsets $\varphi(U \cap V)$ and $\psi(U \cap V)$ of $\mathbf{R}^n$. A differentiable of C^∞ (smooth) structure on a topological manifold M is a family $\mathcal{U} = \{\mathcal{U}_\alpha, \varphi_\alpha\}$ of coordinate neighborhoods such that: 1) the U_α cover M, 2) for any α, β the neighborhoods U_α, φ_α and U_β, φ_β are C^∞ compatible and 3) any coordinate neighborhood V, ψ compatible with every $U_\alpha, \varphi_\alpha \in \mathcal{U}$ is itself in $\mathcal{U}$.

9. GAUSS'S THEOREMA EGREGIUM.

DEFINITION 9.1. ([18]) *Let* $x = x(u, v)$ *be a coordinate patch on a surface of class* ≥ 1. *The First Fundamental Form of* $x = x(u, v)$ *denoted as* I, *is a homogeneous function of the second degree in* du *and* dv *with coefficients* E, F *and* G *called First Fundamental Coefficients which are functions of* u *and* v *and vary from point to point on the coordinate patch.*

Now suppose $x = x(u, v)$ is a patch on a surface of class ≥ 2. Then at each point on the patch there is a unit normal which is a function of u and v of Class C^1. Then the Second Fundamental Form of $x = x(u, v)$, denoted by II, is a homogeneous function of the second degree in du and dv with coefficients L, M and N called Second Fundamental Coefficients which are continuous functions of u and v.

Remark 9.2. ([18]) Given functions E, F, G, L, M and N of u and v of sufficiently high class, determine whether or not there exists a surface $x = x(u, v)$ for which E, F, G, L, M and N are the first and second fundamental coefficients. In general, the surface does not exist unless certain "compatibility" (integrability) conditions are satisfied. These conditions arise from the fact that if $x(u, v)$ is a function of class C^3, then the third order mixed partial derivatives of x are independent of the order of differentiation.

THEOREM 9.3. ([18]) *The Fundamental Theorem of Surfaces. Let E, F and G be functions of u and v of class C^2 and let L, M and N be functions of u and v of class C^1 all defined on an open set containing (u_0, v_0) such that for all (u, v),*

(i) $EG - F^2 > 0$, $E > 0$, $G > 0$

(ii) *E, F, G, L, M, N satisfy certain integratiblity conditions.*

Then there exists a patch $x = x(u, v)$ of class C^3 defined in the neighborhod of (u_0, v_0) for which E, F, G, L, M, N are the first and second fundamental coefficients. The surface represented by $x = x(u, v)$ is unique except for position in space.

DEFINITION 9.4. ([22]) *The Gaussian Curvature, K, is defined as follows:*

$$K = \frac{LN - M^2}{EG - F^2} = \frac{II}{I}.$$

THEOREM 9.5. ([18]) *The Theorema Egregium of Gauss is: The Gaussian curvature on a surface of class ≥ 3 is a function only of the coefficients of the first fundamental form and their derivatives.*

Remark 9.6. Guass showed that the geometry of a surface could be studied by concentrating on the surface itself, that is, the intrinsic geometric properties should be looked at.

10. PARALLEL TRANSPORT AND INTRINSIC COVARIANT CURVATURES.

Remark 10.1. ([16]) The subject of tensor calculus develops many properties of the transformation processes. The display of indices abound in the subject and the fundamental distinction between covariant tensors (indices in the lower position, subscripts) and contravariant tensors (indices in the upper position, superscripts) has to be kept in mind. The following well-known summation convention is used in the analysis: any expression involving a twice-repeated index (occurring twice as a subscript (a covariant tensor) and twice as a superscript (a contravariant tensor) or a subscript (a covariant tensor of rank one and once as a superscript (a contravariant tensor of rank one) shall automatically stand for its sum over the values $1, 2, 3, \ldots, n$ of the repeated index.

DEFINITION 10.2. ([16]) *The tangent vector (a differential) transforms as a contravariant tensor of rank one (upper index) in the following way:*

$$dx^i = \sum_{r=1}^{n} \frac{\partial x^i}{\partial \overline{x}^r}\, d\overline{x}^r$$

or rewritten as

$$A^i = \frac{\partial x^i}{\partial \overline{x}^r} \overline{A}^r$$

where the unbarred coordinate system is transformed to the barred coordinate system and the Jacobian $\frac{\partial x^i}{\partial \overline{x}^r}$ is a first order partial derivatives.

A corresponding covariant vector (a differential function) transforms as a covariant tensor of rank one (lower index) in the following way:

$$\frac{\partial}{\partial x_i} = \sum_{r=1}^{n} \frac{\partial \overline{x}^r}{\partial x_i} \frac{\partial}{\partial \overline{x}^r}$$

or rewritten as

$$A_i = \frac{\partial \overline{x}^r}{\partial x_i} \overline{A}_r$$

where the Jacobian $\frac{\partial \overline{x}^r}{\partial x_i}$ is first order partial derivatives.

Remark 10.3. ([12]) Observe that the first order partial derivatives of the covariant tensor of rank one are the inverses to the Jacobians of the contravariant tensor of rank one. The covariant vector of rank one is the gradient of an arbitrary differentiable function. The notions of inverses, functions, transformations (mappings) and Jacobians are interrelated.

DEFINITION 10.4. ([12]) *The contravariant tensor of rank two can be written as follows:*

$$A^{ij} = \frac{\partial x^i}{\partial \overline{x}_r}\frac{\partial x^j}{\partial \overline{x}_s}\overline{A}^{rs}$$

and the corresponding covariant tensor of rank two can be written as

$$A_{ij} = \frac{\partial \overline{x}^r}{\partial x_i}\frac{\partial \overline{x}^s}{\partial x_j}\overline{A}_{rs}\,.$$

There are also mixed tensors with indices in the upper and lower positions.

DEFINITION 10.5. ([12]) *The Jacobians considered above are of the second order. This suggests that Jacobians of the third order be considered. The geometric objects associated with this higher order Jacobian are called Christoffel symbols which observes the following transformation procedure:*

$$[ij,k] = \frac{\partial \overline{x}^r}{\partial x^i}\frac{\partial \overline{x}^s}{\partial x^j}\frac{\partial \overline{x}^t}{\partial x^k}\overline{[rs,t]} + \overline{A}_{rs}\frac{\partial^2 \overline{x}^r}{\partial x^i \partial x^j}\frac{\partial \overline{x}^s}{\partial x^k}$$

where

$$[ij,k] = \frac{1}{2}\left(\frac{\partial A_{ki}}{\partial x_j} + \frac{\partial A_{jk}}{\partial x_i} - \frac{\partial A_{ij}}{\partial x_k}\right)$$

are called Christoffel symbols of the first kind. The following Christoffel symbols of the second kind are:

$$[ij,k] = A_{kn}\{ij,n\}$$
$$\{ij,n\} = A^{nk}[ij,k]\,.$$

These symbols are assumed to be symmetric in the first two indices and they vanish if all the A_{ij}'s are constant. There are n^3 of these symbols where n is the space dimension.

The second term on the right hand side, that is

$$\overline{A}_{rs}\frac{\partial^2 \overline{x}}{\partial x_i \partial x_j}\frac{\partial \overline{x}^s}{\partial x_k}$$

makes these geometric objects non-invariant in the transformation process.

DEFINITION 10.6. ([12]) *The covariant derivative of A_i is described as follows*

$$A_{ij} = A_{ij} - \{ij,k\}A_k$$

and the covariant derivative of A^i is described as follows:

$$A^i_j = A^i_j + \{kj,i\}A^k\,.$$

Observe the differences between the above two equations.

DEFINITION 10.7. ([12]) *The following covariant tensor may also be formed*

$$A_{ijk} = A_{ijk} - \{ik,n\}A_{nj} - \{jk,n\}A_{ni}\,.$$

DEFINITION 10.8. ([12]) *The line integral can be described in tensorial notation as follows:*

$$\int_C A_i dx^i$$

where C is the curve.

DEFINITION 10.9. ([12]) *If A_{ij} is the covariant derivative of A_i, then*

$$A_{ij} - A_{ji} = 0$$

where the three-index symbols $\{ij,k\}A_k$ cancel out. The above operation is similar to the curling of a vector field.

DEFINITION 10.10. ([12]) *In tensorial notation, Stoke's theorem becomes*

$$\int_C A_i dx^i = -\frac{1}{2}\iint (A_{ij} - A_{ji})ds^{ij}$$

where the double integral is being taken over any surface bounded by the path of the single integral and where the factor $\frac{1}{2}$ is needed because each surface-element occurs twice.

Remark 10.11. ([8]) In Euclidean space there is no geometrical difference between a covariant tensor and a contravariant tensor, but in non-Euclidean space there is a difference.

DEFINITION 10.12. ([8]) *Consider the following contravariant tensor of rank one:*

$$A^i = \frac{\partial x^i}{\partial \overline{x}^j} \overline{A}^j .$$

Take the differential of the above equation to get

$$dA^i = \frac{\partial x^i}{\partial \overline{x}^j} d\overline{A}^j + \overline{A}^j d\left(\frac{\partial x^i}{\partial \overline{x}^j}\right)$$

$$dA^i = \frac{\partial x^i}{\partial \overline{x}^j} d\overline{A}^j + \overline{A}^j \frac{\partial^2 x^i}{\partial \overline{x}^k \partial \overline{x}^j} d\overline{x}^k .$$

When a cartesian coordinate system is assumed, the second term on the right hand side vanishes. When a curvilinear system of coordinates (a non- cartesian system) is used, the second term on the right hand side exists.

DEFINITION 10.13. ([8]) *In order to define a suitable generalization of the differential operator in a curvilinear coordinate system, the difference between two vectors must be performed at the same point. This means it is necessary to transport one of the two vectors from its position to the infinitesimally close position of the other. This transport operation is to be performed so that in the cartesian coordinate system this difference coincides with the usual differential dA^i. Because dA^i is the difference between the components of two infinitesimally close vectors, it follows that during the displacement in cartesian coordinates, the components of A^i are unchanged: this transport is then the displacement of a vector parallel to itself. However, in a curvilinear coordinate system, the components of a vector undergoing parallel transport are, in general, changed unlike the cartesian system case. Hence, if A^i are the components of a vector in x^i and $A^i + dA^i$, the components in $x^i + dx^i$*

the parallel transport of A^i from x^i to dx^i produces a variation of its components, δA^i. Hence, after the displacement the difference, DA^i, between the two vectors is given by

$$DA^i = dA^i - \delta A^i \, .$$

DEFINITION 10.14. ([8,12]) *The parallel displacement of a vector in a non-Euclidean reference coordinate system is generally path dependent. Hence displacing a vector along a closed curve, in general, results in the final vector not coinciding with the initial vector.*

The Riemannian tensor determines the variations of a vector A_i during its parallel displacement along an infinitesimal close contour.

The variation δA_i is described as follows:

$$\delta A_i = \frac{1}{2} \iint (A_{ijk} - A_{ikj}) dS^{jk} \, .$$

The quantity inside the brackets is described as follows:

$$A_{ijk} - A_{ikj} = A_\ell \left(\frac{\partial}{\partial x_j} \{ik, \ell\} - \frac{\partial}{\partial x_k} \{ij, \ell\} + \{ik, m\}\{mj, \ell\} - \{ij, m\}\{mk, \ell\} \right)$$

or rewritten as

$$A_{ijk} - A_{ikj} = A_\ell R^\ell_{ijk}$$

where the following Riemann-Christoffel Curvature Tensor is described as

$$R^\ell_{ijk} = \frac{\partial}{\partial x_j} \{ik, \ell\} - \frac{\partial}{\partial x_k} \{ij, \ell\} + \{ik, m\}\{mj, \ell\} - \{ij, m\}\{mk, \ell\} \, .$$

Hence the variation δA_i can be described as

$$\delta A_i = \frac{1}{2} R^\ell_{ijk} A_\ell dS^{jk} \, .$$

Remark 10.15. ([12]) The variation δA_i can be reduced to zero when R^ℓ_{ijk} is reduced to zero. The variation δA_i can be reduced to zero when the order of covariant differentiation commutes. The variation δA_i is reduced to zero when the vector A_i can be moved independent of the path taken. Hence the existence of R^ℓ_{ijk} reflects path dependent movements.

DEFINITION 10.16. ([12]) *The Riemann Christoffel Curvature tensor is an intrinsic covariant invariant tensor whose Jacobian is of the third order covariant (lower indices) and first order contravariant (upper index) of the following form:*

$$R^{\ell}_{ijk} = \frac{\partial \overline{x}^{r}}{\partial x_i} \frac{\partial \overline{x}^{s}}{\partial x_j} \frac{\partial \overline{x}^{t}}{\partial x_k} \frac{\partial x^{\ell}}{\partial \overline{x}_q} \overline{R}^{q}_{rst} .$$

DEFINITION 10.17. ([16]) *The Riemannian Curvature Tensor (RCT) in an n-dimensional space has n^n components. Many of the RCT components are dependent on other RCT components, hence, the number of independent RCT components will be substantially smaller than n^n. Without a metric, the number of independent RCT components is $\frac{n^2(n^2-1)}{3}$ and with a metric, the number of independent RCT components is $\frac{n^2(n^2-1)}{12}$.*

THEOREM 10.18. ([19]) *In order that a Riemannian space be flat, it is necessary and sufficient that the components of its curvature tensor vanish identically.*

11. THE INTEGRABILITY OF dQ and dW.

Remark 11.1. ([1,24]) Thermodynamics is concerned with energy and its transformation. When a sufficient number of thermodynamic states are specified, the internal state of a system is determined and its internal energy denoted as U, is fixed. An equilibrium state of a system exhibits a set of identifiable reproducible properties which are subject to precise mathematical descriptions. When the conditions for mechanical and thermal equilibrium are not satisfied, the states traversed by a system cannot be described in terms of thermodynamic coordinates referring to the system as a whole. When a system is displaced from equilibrium, it undergoes a process during which its properties change until a new equilibrium state is attained. Variables that express intensity of the system are zero order in mass and are called intensive variables. The intensive coordinates of a system such at temperature, T, and pressure, P, are independent of the mass. Variables that are related to mass are called extensive variables. The extensive coordinates of a system are proportional to the mass such as volume, V.

DEFINITION 11.2. ([1]) *The simplest thermodynamic system consists of a fixed mass of an isotropic fluid uninfluenced by chemical reactions or external fields. These systems are described in terms of the three measurable coordinates PVT, called a PVT system. Experiment shows that these three coordinates are not all independent and that fixing any two of them determines the third. Hence there must be an equation of state that interrelates these three coordinates for equilibrium states. This equation may be expressed in implicit functional form.*

DEFINITION 11.3. ([1,24]) *Work, denoted as W, in thermodynamics represents an exchange of energy between a system and its surroundings. Mechanical work occurs when a force acting on a system moves through a distance. This work is usually defined by an integral which can be described in a differential form. In thermodynamics the work done by a force is distributed over an area, that is, by a pressure acting through a volume, for example, a fluid pressure exerted on a piston. Note the combination of intensive and extensive variables in the description. There are other modes of thermodynamic work; different kinds of systems, described by other coordinates, are also important and they are subject to work done by forces other than pressure, for example, electrical work, work of magnetization, work of changing surface area, etc. When considering new types of thermodynamic systems, experiments determine the proper identification of the forces and displacements.*

The work done by a system depends not only on the initial and final states but also on the intermediate states, that is, on the path.

DEFINITION 11.4. ([24]) *Heat, denoted as Q, is internal energy in transit. Heat flows from one part of a system to another or from one system to another by virtue of only a temperature difference. Heat is not known during the process.*

DEFINITION 11.5. ([24]) *dW and dQ*

An infinitesimal amount of work is an inexact differential, that is, it is not the differential of an actual function of the thermodynamic coordinates.

There is no function of the thermodynamic coordinates representing the work

in a body. The phrase "work in a body" has no meaning. Work is an external activity or process that leads to a change in a body, namely, the energy in a body. To indicate that an infinitesimal amount of work is not a mathematical differential of a function W and to emphasize that it is an inexact differential, a line is drawn through the differential sign hence

$$đW .$$

The heat transferred to or from a system is not a function of the coordinates of the system but depends on the path by which the system was brought from the initial state to the final state. Hence, heat, Q, is not a function of the thermodynamic coordinates but depends on the path. Consequently an infinitesimal amount of heat is an inexact differential and is represented by the following symbol:

$$đQ .$$

DEFINITION 11.6. ([24]) *A mathematical differential form of the first law of thermodynamics is as follows:*

When a system whose surroundings are at a different temperature and on which work may be done undergoes a process, the energy transferred by nonmechanical means, equal to the difference between the internal-energy change and the work done, is called heat where heat is positive when it enters a system and negative when it leaves a system.

A process involving only infinitesimal changes in the thermodynamic coordinates of a system is known as an infinitesimal process. For such a process the above statement becomes

$$dU = đQ + đW .$$

If the infinitesimal process is quasi-static, then dU and $đW$ can be expressed in terms of thermodynamic coordinates only. An infinitesimal quasi-static process is one in which the system passes from an initial equilibrium state to a neighboring equilibrium state.

Remark 11.7. ([24]) The above mathematical formulation of the first law of thermodynamics has the following three related concepts: (1) the existence of an internal-energy function; (2) the principle of the conservation of energy and (3) the definition of heat as energy in transit by virtue of a temperature difference.

Example 11.8. ([24]) Consider any hydrostatic system contained in a cylinder equipped with a movable piston on which the system and the surroundings may act. Suppose that the cylinder has a cross- sectioned area, A, and that the pressure exerted by the system at the piston face is P and that the force is PA. The surroundings also exert an opposing force on the piston. If, under these conditions, the piston moves a distance dx in a direction opposite to that of the force PA, an infinitesimal amount of workdW can be described as follows:

$$dW = -PAdx\ .$$

But

$$Adx = dV\ .$$

Hence

$$dW = -PdV\ .$$

Consequently, the first law of thermodynamics becomes

$$dU = dQ - PdV$$

where U is a function of any two of the three thermodynamic coordinates and P is a function of V and T.

A similar equation may be written for other simple systems; for more complicated systems replace dW by two or more expressions.

Choosing T and V gives

$$dU = \left(\frac{\partial U}{\partial T}\right)_V dT + \left(\frac{\partial U}{\partial V}\right)_T dV\ .$$

Hence the first law of thermodynamics becomes

$$dQ = \left(\frac{\partial U}{\partial T}\right)_V dT + \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right] dV .$$

The subscripts next to the partial derivatives indicate that all the other independent variables are held constant except the one in the derivative being considered.

Most importantly, observe that the above equation is not exact, the conditions for an exact differential are not met, that is, the order of taking partial derivatives does not commute.

Remark 11.9. The last equation in Example 11.8 reveals the specific violation of Poincaré's Lemma and Its Converse presented in Lemma 7.1 and Lemma 7.2. Consequently, the integrability conditions of differential equations are not established; this means that the order of taking partial derivatives is not interchangeable. This also means that movements on a differential manifold are path dependent.

Remark 9.2 pointed to the integrability conditions that would have to be satisfied in order for a surface to exist; in that particular case, these conditions arise from the fact that if $x(u, v)$ is a function of class C^3, then the third order mixed partial derivatives of x are independent of the order of differentiation. Observe the inclusion of a third order differential function.

Remark 11.10. ([24]) Theorem 4.1 established that any Pfaffian differential equation in $\mathbb{R}^2$ admits an integrating factor. Consequently the inexact differential equation in Example 11.8 has an integrating factor; this is a mathematical property. Experimentation reveals that the integrating factor which is found for systems with any number of independent variables is an arbitrary function of the empirical temperature only, which is the same for all systems, hence, an absolute (or Kelvin) thermodynamic temperature can be defined provided that the second law of thermodynamics exist.

Remark 11.11. ([24]) In general, a Pfaffian differential form containing three differentials does not admit an integrating factor. Theorem 5.1 establishes that the Frobenius condition $d\omega \wedge \omega = 0$ is called integrability condition for Pfaffian equation $\omega = 0$ and Example 5.2 presents these conditions in R^3. The general integrability conditions also appear in Theorem 10.18 wherein it is established that in order for a Riemannian space to be flat, it is necessary and sufficient that the components of its curvature tensor vanish identically. In this latter case the concept of parallel transport or displacement of a vector along a closed curve was considered in the analysis and the observation was made that the final vector does not in general coincide with the initial vector, that is, the analysis is path dependent.

12. REVERSIBLE AND IRREVERSIBLE PROCESSES.

DEFINITION 12.1. ([24]) *A reversible process is one that is performed in such a way that, at the conclusion of the process, both the system and the local surroundings may be restored to their initial states without producing any changes in the rest of the universe.*

A process that does not fulfill these requirements is said to be irreversible.

Statement 12.2. ([24]) The Kelvin-Planck statement is as follows: No process is possible whose sole result is the absorption of heat from a reservoir and the conversion of this heat into work.

Statement 12.3. ([24]) The Clausius' statement is as follows: No process is possible whose sole result is the transfer of heat from a cooler to a hotter body.

Axiom 12.4. ([24]) Constantin Carthéodory axiom (The Inaccessibility Theorem of Carthéodory) is as follows: In the neighborhood, however close of any equilibrium state of a system of any number of thermodynamic coordinates, there exist states that cannot be reached, are inaccessible, by reversible adiabatic processes.

Remark 12.5. ([24]) When taking into account all the interactions that

accompany living processes such processes are irreversible; all natural spontaneous processes are irreversible.

13. GENERAL AXIOMS OF REVERSIBLE AND IRREVERSIBLE PROCESSES.

Axiom 13.1. The necessary and sufficient conditions for reversible paths (processes) is that the intrinsic covariant curvature tensor vanish.

Axiom 13.2. The necessary and sufficient conditions for irreversible paths (processes) is that the intrinsic covariant curvature tensor exists.

Remark 13.3. Axioms 13.1 and 13.2 are based on the following Theorem 10.18: In order that a Riemannian space be flat, it is necessary and sufficient that the components of its curvature tensor vanish identically. Section 10.13 described the difference, DA^i, of the displacement between two vectors in a curvilinear coordinate system as

$$DA^i = dA^i - \delta A^i$$

where δA^i is a variation of the vector components. Hence, the parallel displacement of a vector in a non-Euclidean reference coordinate system is generally path dependent.

Section 10.14 describes the variation δA_i as follows

$$\delta A_i = \frac{1}{2} R^{\ell}_{ijk} A_{\ell} dS^{jk}$$

where R^{ℓ}_{ijk} is the Riemannian Curvature Tensor. Hence, when R^{ℓ}_{ijk} exists, path dependent movements (irreversible processes) exist (Axiom 13.2) and when $R^{\ell}_{ijk} = 0$, path independent movements (reversible processes) exist (Axiom 13.1). The existence of an invariant geometric object called an intrinsic covariant curvature tensor reflects irreversible processes and the non-existence of these same geometric objects reflect a reversible process. The inexact differentials dQ and dW are reflected in the intrinsic covariant curvature tensor; when this tensor is reduced

to zero, the differentials become exact, that is, the general integrability conditions are fulfilled. Theorem 5.9 establishes that these conditions are $d\omega \wedge \omega = 0$. Recall also, from Remark 5.10, that the following equation was considered in the above result:

$$\omega = g df$$

and that

$$\theta = g^{-1} dg = d\,\ell n|g|\,.$$

The function g appears to relate ω and df and then recedes into the background to leave $d\omega \wedge \omega = 0$. Such an operation reflects the role played by an integrating factor. See also section 4 on integrating factors in $\mathbf{R}^2$ and Theorem 4.8 relating X as a skew- symmetric matrix and e^X as being orthogonal. Remark 4.7 also suggests the corresponding theory of Maurer-Cartan forms and Lie group theory.

DEFINITION 13.4. ([1]) *There exists a property called entropy, S, which is an intrinsic property of a system, functionally related to the measurable coordinates which characterize the system. For a reversible process, changes in this property is given by the following equation*

$$dQ = T\,dS\,.$$

Observe the relationship of the above equation to the following equation presented in Remark 5.10

$$\omega = g df$$

and see also Remark 13.3.

DEFINITION 13.5. ([1]) *The Second Law of Thermodynamics can be stated as follows: The entropy change of any system and its surroundings, considered together, is positive and approaches zero for any process which approaches reversibility.*

DEFINITION 13.6. ([17]) *The following equation incorporates the First and Second Laws of Thermodynamics:*

$$TdS = dU + PdV .$$

Remark 13.7. A reversible process is one in which the invariant intrinsic covariant curvature tensor is zero. An irreversible process is one in which this tensor is not zero. See Axiom 13.1 and Axiom 13.2. This suggests that the concept of entropy is reflected in an invariant geometric object called an intrinsic covariant curvature tensor. In other words: Entropy is Curvature.

Statement 13.8. ([15]) In 1865, Rudolph Clausius presented the two laws of thermodynamics in the following concise form: The energy of the universe is constant and the entropy of the universe tends to a maximum. J. Willard Gibbs used these words (in the original German) as a heading for his memoir "On the Equilibrium of Heterogeneous Substances."

Remark 13.9. ([24]) The entropy of a system or of a reservoir is a measure of the degree of molecular disorder existing in the system or reservoir. The disorder of a system is normally calculated by the theory of probability and is expressed by a quantity W known as the thermodynamic probability. The relation between entropy and disorder is then shown to be

$$S = \text{ constant } \ell n\, W .$$

By using this equation, a nonequilibrium state corresponds to a certain degree of disorder and hence to a definite entropy.

Remark 13.10. ([9]) The concept of disorder or entropy can be reflected in the geometric notion of curvature, that is, disorder is a path dependent concept based on an invariant intrinsic covariant curvature tensor.

14. INTRINSIC CURVATURE AND NONEQUILIBRIUM/UNSTABLE PROCESSES.

DEFINITION 14.1. ([15]) *The properties of a pure substance can be represented in terms of the following four functions: internal energy, U; enthalpy, $H = U + PV$; Helmholtz function, $F = U - TS$ and Gibbs function, $G = H - TS$. Any one of the eight quantities P, V, T, U, H, F, G and S may be expressed as a function of any two others.*

Example 14.2. ([1]) Consider a hydrostatic system undergoing an infinitesimal reversible process from one equilibrium state to another, then the following four equations appear

$$dU = TdS - PdV$$
$$dH = TdS + VdP$$
$$dF = -SdT - PdV$$
$$dG = -SdT + VdP$$

Since U, H, F and G are actual functions, their differentials are exact, that is, the order of taking partial derivatives commutes–the integrability conditions again. From these above equations, the well-known Maxwell relations can be easily established.

The four functions U, H, F and G with their corresponding characteristic independent variables SV, SP, TV, TP respectively, are coupled through the Legendre transform. The surfaces of these functions encode thermodynamic information in different ways and hence give the criteria for equilibrium through different geometrical treatments. The corresponding well-known three-dimensional coordinate system can be labeled USV, HSP, FTV and GTP. Observe that each function has a coordinate.

DEFINITION 14.3. ([1]) *A surface in a PVT coordinate system can also be constructed and the solid, liquid, gas and fluid regions can be presented using the sublimation, fusion and vaporization curves drawn on the surface. The important triple point can also be displayed.*

Remark 14.4. All the differentials are exact, that is, the analysis is path independent, this implies there is no curvature. Also, note the application of the Leibniz rule. Each function U, H, F and G has a coordinate associated with the corresponding pairs of coordinates SV, SP, TV and TP. Surfaces are then constructed and thermodynamic information is viewed.

As Remark 9.6 notes, Gauss's Theorema Egregium (October 8, 1827) suggests that the intrinsic geometric properties of a surface should be looked at, that is, the surface itself should be studied. Riemann (June 6, 1854) suggested (a) that Gauss's analysis could be extended to n dimensions, (b) that a quadratic differential is the structure to add to the notion of a surface (a manifold) and (c) that space and geometry are different. Definition 10.14 summarizes the above notions in the following Riemann- Christoffel Curvature Tensor

$$R^{\ell}_{ijk} = \frac{\partial}{\partial x_j}\{ik,\ell\} - \frac{\partial}{\partial x_k}\{ij,\ell\} + \{ik,m\}\{mj,\ell\} - \{ij,m\}\{mk,\ell\} \, .$$

When looking at surfaces as differential manifolds (see section 8), the fundamental concepts of ordinary differential equations can be restated in terms of vector fields and integral curves (see section 2 and section 3). This suggests that a full geometric viewpoint be incorporated into considering the conditions of equilibrium/nonequilibrium and stable/unstable processes.

DEFINITION 14.5. ([6,7]) *The field of non-equilibrium thermodynamics provides a general framework for the macroscopic description of irreversible processes. An unstable system can be viewed as one in which the disturbance of a system will grow in amplitude in such a way that the system progressively departs from the initial state and never reverts to it.*

Remark 14.6. ([9]) The underlying concept of non-equilibrium and unstable processes is path dependence. The mathematical construct of path dependence is the invariant intrinsic covariant curvature tensor R^{ℓ}_{ijk}. This suggests that non-equilibrium and unstable process can be specified as

$$R^{\ell}_{ijk} \gtrless 0$$

and that equilibrium and stable processes can be specified as

$$R^{\ell}_{ijk} = 0 \, .$$

Equilibrium and stable processes are path independent, that is, the integrability conditions are fulfilled. The three index symbols, the connection terms, are constant. The differential manifold is flat.

Non-equilibrium and unstable processes are path dependent, that is, the integrability conditions are not fulfilled. The three index symbols, the connection terms, are not constant. The differential manifold is not flat.

Consequently, assuming that the U, H, F and G functions in thermodynamics are exact differentials, implicitly suggests that the processes are in equilibrium and stable. The mathematical constructs assumed have limited the observer's view of the process.

Hence a much broader set of mathematical postulates should be used by the observer in order to incorporate equilibrium/non-equilibrium processes and stable/unstable processes. These postulates should be based on non-Euclidean geometry.

The Riemann Curvature Tensor, R^{ℓ}_{ijk} without a metric is a third order covariant and first order contravariant (with a metric, it is fourth order covariant). This suggests that third and fourth order terms should be included in the analysis.

A fundamental concept to be kept in mind is: intrinsic.

15. CARTAN'S METHOD OF EQUIVALENCE.

Remark 15.1. ([11]) The following equation

$$\omega = g \, df$$

used in Remark 5.10 suggests that Élie Cartan's method of equivalence may be used in the analysis. Using Cartan's method of moving frames, the mathematical notions of curvature, torsion and group properties can be incorporated.

Remark 15.2. The following presentation in this section is based on Robert B. Gardner's 1989 monograph, *The Method of Equivalence and Its Applications.*

Remark 15.3. ([14]) The purpose of the method of equivalence is to find the necessary and sufficient conditions so that geometric objects be equivalent, that is, the geometric objects should be mapped onto each other by a class of diffeomorphisms characterized as the set of solutions of a system of differential equations. The necessary and sufficient conditions are found in the form of differential invariants of the geometric object under the class of diffeomorphisms. The following presentation will be restricted to classes of diffeomorphisms which can be described as solutions of a first-order system of differential equations or equivalently by conditions on their Jacobians.

DEFINITION 15.4. ([14]) *The equivalence problem of Élie Cartan is as follows: Let* $\Omega_V = {}^t(\Omega_V^1, \ldots, \Omega_V^n)$ *be a coframe on an open set* $V \subset \mathbf{R}^n$ *and let* $\omega_U = {}^t(\omega_U^1, \ldots, \omega_U^n)$ *be a coframe on* $U \subset \mathbf{R}^n$*, and let* G *be a prescribed linear group in* $G\ell(n, \mathbf{R})$*, then find the necessary and sufficient conditions that there exists a diffeomorphism* $\Phi = U \to V$ *such that for each* $u \in V$

$$\Phi^* \Omega_V \big|_{\Phi(u)} = \gamma_{VU}(u) \omega_U \big|_u ,$$

where $\gamma_{VU}(u) \in G$.

Remark 15.5. ([14]) The ten lectures in Gardner's presentation were summarized as a flow chart. The following description is based on that flow chart:

Begin with a group, coframes and open set, compute the Maurer-Cartan forms and defining relations, perform a principal component decomposition and a Lie algebra compatible absorption, compute infinitesimal action on structure tensor, then determine if there is a trivial action.

If there is a trivial action, then determine if there is an identity structure; if there is an identity structure, then the problem is solved.

If there is no trivial action, then perform a normalization and group reduction procedure. Then determine if the procedure is a constant type.

If the procedure is a constant type, then change the group and coframe and begin the process again.

If the procedure is not a constant type, then other possibilities have to be considered.

If there is not identity structure, then determine if the system is in involution.

If the system is not in involution, perform a prolongation procedure and change the group, coframe and open set and begin the process again.

If the system is in involution, determine if the torsion is constant.

If the torsion is constant, then the problem is solved.

If the torsion is not constant, then "wild things" will appear.

Remark 15.6. When considering Cartan's method of equivalence, the notions of the lifting of the linear group to G spaces and the related mapping of intrinsic torsion have to be considered in addition to curvature. If torsion is not constant, and the group reduction and normalization processes are not constant, then many other possibilities and problems appear.

REFERENCES

1. Abbott, M.M. and Van Ness, H.C., *Thermodynamics*, McGraw- Hill Book Company, (1972), 1–140.
2. Abraham, R., Marsden, J.E. and Ratin, T., *Manifolds, Tensor Analysis, and Applications*, second edition, Springer-Verlag, New York, (1988), 392–448.
3. Arnold, V.I., *Ordinary Differential Equations*, MIT Press, Cambridge, MA, (1973), 1–47.
4. Boothby, W.M., *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, New York, (1975), 20–172.
5. Boyce, W.E. and DiPrima, R.C., *Elementary Differential Equations and Boundary Value Problems*, John Wiley and Sons, Inc., New York, (1965), 1–79.
6. Chandrasekhar, S. *Hydrodynamic and Hydromagnetic Stability*, Dover Publications, Inc., New York, (1961), 1–75.
7. DeGroot, S.R. and Mazur, P., *Non-equilibrium Thermodynamics*, Dover Publications, Inc., New York, (1984), originally published: Amsterdam: North-Holland Pub. Co., 1962, 1–42.
8. deSabbata, V. and Gasperini, M., *Introduction to Gravitation*, World Scientific Publishing Co., Singapore, (1985), 1–43.
9. Donato, J., *Donato's Research on Economics, Social Choice, Statistics and Control Theory through 1986*, MC Printing Company, Syracuse, New York, (April 1987), copyright 1987, Jerry Donato.
10. Donato, J., "Path Dependent Analysis", *Topics in Mathematical Analysis, A Volume Dedicated to the Memory of A.L. Cauchy*, Ed. Th. M. Rassias, World Scientific Publishing Company, Singapore, (1989), 210- -230.
11. Donato, J., "Cartan's Method and Plateau's Problem", *The Problem of Plateau: A Tribute to Jesse Douglas and Tibor Rado*, Ed. Th.M. Rassias, to appear.
12. Eddington, A.S. *The Mathematical Theory of Relativity*, third edition, Chelsea Publishing Company, New York, (1975), 43–75, first

published in 1923.

13. Flanders, H., *Differential Forms with Applications to the Physical Sciences*, Academic Press, Inc., New York, (1963), 1–111.
14. Gardner, R.B., *The Method of Equivalence and Its Applications*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, (1989).
15. Jolls, K.R. *The Art of Thermodynamics*, an unpublished manuscript presented at the B.F. Ruth Chemical Engineering Research Symposium VI in October, 1988 at Iowa State University and also presented in May 1989 in poster form at the Gibbs Symposium held at Yale University and to be included in the Proceedings of the Gibbs Symposium.
16. Kay, D.C., *Tensor Calculus*, McGraw-Hill Book Company, New York, (1988).
17. Lacomba, E.A. and Hernández, D.B., "On the Role of Reciprocity Conditions in the Formulation of Conservation Laws and Variational Principles", *Differential Geometry, Calculus of Variations and Their Applications*, (eds: G.M. Rassias and T.M. Rassias), Marcel Dekker Inc., 100, (1985), 305–334.
18. Lipschutz, M.M., *Differential Geometry*, McGraw-Hill Book Company, New York, (1969), 171–226.
19. Lovelock, D. and Rund, H., *Tensors, Differential Forms, and Variational Principles*, John Wiley and Sons, New York, (1975), 239– 297.
20. Marsden, J.E. and Tromba, A.J. *Vector Calculus*, third edition, W.H. Freeman and Company, New York, (1988), 490–584.
21. Protter, M.H. and Morrey, Jr., C.B., *Modern Mathematical Analysis*, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, (1964), 636–662.
22. Struik, D.J., *Lectures on Classical Differential Geometry*, second edition, Dover Publications, Inc., New York, (1961), 55– 104.
23. Von Westenholz, C., *Differential Forms in Mathematical Physics*, revised edition, North-Holland Publishing Company, (1981), 59– 256.

24. Zemansky, W.W. and Dittman, R.H., *Heat and Thermodynamics*, sixth edition, McGraw-Hill Book Company, (1981), 1–272.

Jerry Donato
917 Madison Street #213
Syracuse, New York 13210
U.S.A.

Constantin Carathéodory: An International Tribute (pp. 219–228)
edited by Th. M. Rassias

AXIOMATISATION OF THERMODYNAMICS

M. Dutta and T. Dutta

The paper consists of four sections. After the introduction in the first section, in the second section, as a tribute to Carathéodory, some distinctive features of the first of his papers in the subject, which remain unnoticed until now, is discussed very briefly. The third section contains a simple alternative proof of the well-known lemma of Carathéodory. Some concluding remarks are in the fourth section.

1. Introduction

In the beginning of the paper [1] of Carathéodory on thermodynamics, as developed in the usual traditional form, the following from his own remarks is to be noted: "There exists a physical quantity, which is not identical in nature with mechanical quantities, viz., mass, force, pressure, of which the characteristic properties can be determined through calorimetric measurements and which is named as 'heat'. The heat has the characteristics comparable with those of mechanical works in certain circumstances and further, always possesses the property that it follows from a hotter body to a colder body, when two bodies of different temperatures are in contact."

The main objective of the paper [1], as stated clearly, is to build up a theory in which the totality of the results is in agreement with experiences without any specific assumption about the physical nature of heat. For the purpose, a new quantity which has the characteristics depending on the instantaneous states of different bodies under considerations and consequently is different from *heat* is introduced.

In the long paper [1] there are many features of great interest and sig-

nificance in mathematics and also in physics. In the second section of the present paper after the introduction, a brief discussion has been made about how Carathéodory started from mechanics developed a new theory to derive the basic properties of heat from notions and results of mechanics through introduction of a new single coordinate besides the coordinates of mechanical nature. No attempt is made to discuss all aspects of Carathéodory's development in the paper [1], though they are of much interest and significance.

In the third section of the present paper, a simple alternative proof of the well-known lemma of Carathéodory which is the most interesting basic result of the paper [1] is given. In the alternative proof of the above lemma only simple notions and results of the abstract mathematics are used as in the papers [2], [3], [4].

The fourth section contains concluding remarks. They are followed by a bibliography of a few papers, relevant to the present paper.

In the other paper [5], written at the request of Max Planck, Carathéodory expressed his basic ideas and results of the paper [1], in a language easily understandable to physicists, familiar to usual developments of traditional thermodynamics. In this paper, no discussions is made of the paper [5].

2. Some Fine Points of the Paper [1]

For proper understanding of the significances and importance of the paper [1], it is necessary to recall some facts regarding its historical background, some significant points of mathematical formulation of the problem and then the main steps of the procedure used in proving Carathéodory's lemma, which is a very important result of mathematical interest.

2.1. *Historical Background*

In the beginning of last century after the studies of mathematical foundation of mechanics by Lagrange and of its successful applications by Laplace and others, mechanics was accepted as the most basic of all sciences and then attempts were made to explain all problems in science through mechanics. At that time, when the paper [1] was written by Carathéodory, the spirit was very much dominating. In the paper [1], Carathéodory deduced the basic results of thermodynamics as an extension of mechanics

through the introduction of a new coordinate of non-mechanical nature.

At the end of the thirties of the last century, from experience, 'heat' was identified finally as a form of energy and the universality of the law of conservation of energy in different forms, i.e., the first law of thermodynamics, was widely accepted. At the end of the forties of the last century, limitation to transformations of heat to work became evident through several experiments and observations and the second law of thermodynamics was formulated, of course, in different languages by different pioneers of physics. Within a few years, after different formulations, their equivalence was established. In the mid-sixties of the last century, an equivalent formulation of the second law, known as the entropy principle, was given by Clausius through inequality. As basic laws of physics are generally expressed through equalities where as the entropy principle is expressed through an inequality, attempts were made by Clausius, Boltzmann, Thomson and others to show its plausibility from mechanics. In the mathematical theory, developed in the paper [1], is a new successful attempt in the direction. Some interesting discussions may be seen in the paper [6] and also in the introductory part of the book [7] of Dutta.

2.2. *Some Relevant Essential Points*

In the paper [1] the main aim of the development is to introduce simultaneously temperature and entropy, very closely related to heat and to draw their basic properties from basic notions and results of mechanics. With the aim in view, in addition to coordinates of mechanical nature, only one additional coordinate of non-mechanical nature, satisfying only some simple obvious mathematical properties is introduced but nothing about the physical nature of the additional coordinate is assumed. The view-point is excellent and elegant epistemologically and also mathematically. Unfortunately, all the subsequent individuals who worked on this line including Born [8], Landé [9], Chandresekhar [10] failed to note this fine point and introduced emperical notions about temperature at the outset.

To develop a mathematical theory from the least number of postulates is considered as elegance of mathematics. In the paper [1] a mathematical theory of thermodynamics, explaining the experiences of non-mechanical nature, gathered from calorimetric experiments and observations, is developed by introduction of a single coordinate of non-mechanical nature with a minimum number of simple obvious postulates about its mathematical nature. The epistemological excellence of the paper [1] lies in building up

thermodynamics as a new mathematical theory, extended from the accepted mathematical theory of mechanics through the introduction of a single coordinate of non-mechanical nature. From this point of view, the theory, developed by Carathéodory in his paper [1] satisfies the criteria of 'naturalness' or 'Logical simplicity' and thus 'inner perfection' as elaborated by Einstein in his notes [11].

3. Carathéodory's Lemma

Before sketching the alternative proof of the lemma of the paper [1], which is now admitted as a basic result of the theory of partial differential equations, basic notions and postulates of Carathéodory, on which the present discussion is based, are described. After that, notion of processes, reversible processes, reversible restricted processes like reversible adiabatic processes are stated and deduced along with Carathéodory's lemma. The arguments used here are based on simple notions and results of equivalence relation of abstract algebra, of connectedness and compactness of general topology and of dimension theory.

3.1. *Basic Notions and Postulates*

The state of a thermodynamic system, the basic undefined object, is postulated to be specified by $(n+1)$ coordinates, $x_0, x_1, \ldots, x_n$, with the following characteristics:

i) The domain of each of the above coordinates is a one-dimensional continuum of real numbers, i.e., a real interval;

ii) Each of the n-coordinates, $x_1, \ldots, x_n$, is generally referred to as deformable coordinate as they specify the shape, the size and the like of the thermodynamic system and are controllable in the sense that each of them can be varied in the entire domain of variability (i.e., from any initial position to any other position in the interval of its definition) by mechanical means, i.e., by adjusting relevant external forces.

iii) x_0 also varies continuously in the domain of definition but not controllable in the above sense.

The product space of all the coordinates is a $(n+1)$ dimensional continuum and called the state space. The product space of all the deformable coordinates is an n-dimensional space and may be called the deformation

space, similar to the configurational space of mechanics.

Note. In most of the discussions of Carathéodory, as intervals are mainly considered in paper [1], connectedness plays the very basic role. Notion of simple connectedness of the state space is implicit in the entire paper [1]. In common with other branches of mathematical physics, treated as a continuum physics, in the paper [1], all functions are taken to be continuous and the function representing the energy of the system is taken to be continuous with continuous first partial derivatives.

Any change of states, represented by a curve joining a pair of points of the state space, is known as a process. Now it is well-known [1], the representation of a state by a point described in the above way is possible if the thermodynamic sytem is in equilibrium. So, any process, in which the intermediate states are not in equilibrium, cannot be fully represented by a continuous curve in the state space. A process in which all the intermediate states are in equilibrium (naturally, for very slow changes) is known as a quasi-static process and is represented by a continuous curve in the state-space. So, generally these processes are also reversible. Of course, Carathéodory [1] has discussed the possibility of the existence of some thermodynamic system in which quasi-static processes are not reversible. But by Carathéodory himself and after him by all other invesitgators, theories for the thermodynamic systems in which all quasi-static processes are reversible and converse are developed. So, here we take reversible processes only. Thus, a reversible process is a change of state where the end points are connected by a continuous curve in the state space.

It is simple to note that the relation between two points related by a reversible process is symmetric and transitive. If we consider the set of all points, related to one another by reversible processes, it is the entire state-space and thus study of consequences of this general relation as such leads to triviality. To get some results of interest, one should consider a set of reversible processes satisfying certain specified physical restriction. As for example, in the thermodynamics, the specific restriction is either 'adiabatic' or 'isothermal' or 'isochoric' or the like. A relation between points connected by reversible processes under certain specified restriction to be denoted by r.r.p (reversible restricted process), is regarded as a linear relation and is studied firstly here. Thus, r.a.p denotes reversible adiabatic processes.

3.1.1. *First law of thermodynamics*

In the infinitesimal form, the first law of thermodynamics is stated in the form:

$$dE(x_0, x_1, \ldots, x_n) + \sum_{i=1}^{n} X_i(x_0, x_1, \ldots, x_n)\, dx_i = 0 \qquad \text{(A)}$$

where $E(x_0, x_1, \ldots, x_n)$ is the energy of system, $X_i(x_0, x_1, \ldots, x_n)$ is generalised forced on the system corresponding to $x_i, i = 1, 2, \ldots, n$.

Note. As notion of time-derivatives of coordinates is not involved here, each point of the state space in the equation corresponds to a state of thermodynamic equilibrium. So, the variation in (A) is through states of thermodynamic equilibrium.

3.1.2. *Reversible adiabatic processes*

As already stated, any non-directed curve between two points in the state space represent a reversible process. A set of curves, which satisfies certain restriction by an equation or a number of equations, involving coordinates, are a set of reversible restricted processes (r.r.p). If the restriction is given by the Eq. (A), it is reversible adiabatic processes (r.a.p). The set of points related to a given point P by r.a.p is a r.a.s through P.

Proposition 1. Relation of being related by r.a.p. is an equivalence relation.

Proof. Evidently the relation is transitive and symmetric. It is reflexive trivially. So, it is an equivalence relation.

Corollary 1. Relation of being related by r.a.p is an equivalence relation.

Note. It is well-known that an equivalence relation induces a partition. Thus, the relation by r.a.p induces partitions in the state space, i.e., a pair of points not related by r.a.p can be in r.a. set and no two r.a. sets can

have a common point.

Proposition 2. A r.a. set is connected.

Proof. Any two points of a single r.a. set are connected by a number of arcs of r.a.p, so a r.a. set is arcwise connected. Then, by well-known result of general topology [12] it is connected.

Proposition 3. r.a. sets are compact when the set of values of each coordinate is bounded.

Proof. Now a set of values of any bounded coordinate is a compact set. As a finite Cartesian product of compact spaces is compact, r.a. sets are compact if a set of values of each coordinate is bounded.

Corollary. Thus, the r.a. set is locally compact.

Remark. Though the term 'continuum' is frequently used in analysis, geometry, physics and many other branches of mathematics, qualitative specification of the term is rarely found in abstract mathematics. In real analysis it is introduced by Cantor-Dedikin axiom based on the fact that a real line is a continuum. As real time is locally compact (not compact) connected, so a locally compact connected set may be termed as 'continuum'. Thus, we get the following proposition (cf. Hocknig and Young [13]).

Proposition. Each r.a. set is a continuum.

3.1.3. *Carathéodory's principle*

In a state space, every neighbourhood of a point P contains a point not accessible to P by r.a.p.

Remarks. Due to the above, each r.a. set has no interior, i.e., is a border set.

Proposition 4. The dimension of a r.a. set is less than $n+1$.

Proof. As every subset of the $(n+1)$-dimensional state space is of dimension $< (n=1)$ and as every $(n+1)$-dimensional subset of the state space has non-null interior [13], so, each r.a. set has dimension $\leq n$.

Proposition 5. It is clear that coordination is inherent in the entire discussion from the beginning. Now, the intersection of a small $(n+1)$-dimensional neighbour of a point P and the linear equation A is n-dimensional. If these intersections are taken as neighbourhood of P in a r.a. set, each of the r.a. sets is n-dimensional.

Proposition 6. The Pfaffian (A) admits a solution

$$F(x_0, x_1, \ldots, x_n) = 0\,.$$

Proof. In the coordinate geometry, a n-dimensional surface in $(n+1)$ dimensional space is given by an equation

$$F(x_0, x_1, \ldots, x_n) = 0$$

and conversely the equation $F(x_0, x_1, \ldots, x_n) = 0$ corresponds to a n-dimensional surface in $(n+1)$-dimensional space. Hence the proposition follows.

Remark. On the assumptions of $\frac{\partial E}{\partial x_0} \neq 0$, Carathéodory wrote the Pfaffian (A) as

$$dx_0 + \sum_{i=1}^{n} X_i'(x_0, x_1, \ldots, x_n)\, dx_i = 0 \qquad \text{(B)}$$

where $X_i'(x_0, x_1, \ldots, x_n) = X_i(x_0, \ldots, x_n) + \left(\frac{\partial E}{\partial x}\right) \Big/ \left(\frac{\partial E}{\partial x_0}\right)$ and proved that the Pfaffian (B) admits an integral. The result is referred to as Carathéodory's lemma.

4. Concluding Remarks

In the above, it is easily seen that a r.a. set is a n-dimensional hypersurface when the restriction is expressible by a single Pfaffian equation. Similarly it may be proved that a r.r. set is a $(n - s)$-dimensional hypersurface $(n \geq s)$, when the restriction is expressible by restriction as for example in isothermal isobaric (isochoric) cases, $s = 2$. Discussion of Carathéodory's lemma from set topology mainly, may be seen in papers [2], [3], [4] of Dutta. Of course, the original argument of the paper [1] appears to be much nearer to discussions of homotopy. The Carathéodory lemma may also be seen in standard books on partial differential equations [14].

There are many other interesting significances of the paper [1] not yet duly noticed. Here only few of them are discussed.

References

1. C. Carathéodory, *Untersuchungen Uber die Grundlagen der thermodynamics*, Math Ann. **67** (1909), 355–386.
2. M. Dutta, *On existence of reversible adiabatic surfaces*, Ann- Phys. **22** (1968), 321–28.
3. M. Dutta, *Sur lesintegrabilite des equations differentialles*, C. R., Acad. Sc., Paris, **250** (1970), 90–92.
4. M. Dutta and N. Shivaramakrishnan, *On some consequences of Carathéodory's principle*, Ind. J. Phys. **42** (1968), 753–56.
5. C. Carathéodory, *Uber die Bestimung der Energie and dei absolution Temperatur' mit revenille Prozessen*, Berlin Beriehte **29** (1925), 39–47.
6. M. Dutta, *Hundred years of entropy*, Physics Today **21** (1968), 75–86.
7. M. Dutta, *Bose Statistics*, World Press, Calcutta, 1974.
8. M. Born, *Kritische Betrachtungen zur traditionellea Darstellung der Thermodynamik*, Phys. Annalen **22** (1921), 281–224, 249–254, 281–286.
9. A. Landé, *Axiomatische Begrümdung der Thermodynamik durch Caratheodory*, Hand Buch der Phys., Bd. IX, ICpetel 1–4, 1926, pp. 281–311.
10. S. Chandrasekhar, Ch. I of *Stellar Structure*, Dover Series, USA (Original publication in 1938).
11. P. A. Schliff (ed.), *Albert Einstein-Philosopher, Scientist*, Tudor Publishing Co, 1959, pp. 199–242.
12. M. Dutta, T. K. Mukherjee and L. Debnath, *Elements of General Topology*, World Press, Calcutta, 1966.
13. J. G. Hocking and G. H. Young, *Topology*, Addisions Publication Co., U.S.A, 1961.

14. I. N. Sneddon, *Elements of Partial Differential Equations*, McGraw Hill, N.Y., 1957.

M. Dutta and T. Dutta
Satyendra Nath Bose School
for Mathematics and Mathematical Sciences
Calcutta Mathematical Society
AE-374, Salt Lake City, Calcutta- 700 064
India

Constantin Carathéodory: An International Tribute (pp. 229–242)
edited by Th. M. Rassias

DIFFERENTIABLE SOLUTIONS OF A GENERALIZED COCYCLE FUNCTIONAL EQUATION FOR SIX UNKNOWN FUNCTIONS

Bruce R. Ebanks

ABSTRACT. The general three-times continuously differentiable solutions of a functional equation for six unknown functions is presented. The equation contains as special cases many well-known functional equations such as Cauchy, Sincov, cocycle, and cyclic equations.

1. INTRODUCTION

The ultimate goal of this paper is the presentation of the general solution of the functional equation

$$F_1(x+y,z) + F_2(y+z,x) + F_3(z+x,y) + F_4(x,y) + F_5(y,z) + F_6(z,x) = 0 \qquad (1)$$

for all $x,y,z \in R$ (the reals), with F_1, F_2, and F_3 in the class $C^3(R^2)$. The general forms of all six unknown functions $F_i: R^2 \to R$ will be given.

Note that equation (1) contains several well-known functional equations as special cases. For instance, if $F_3 = -F_1$, $F_5 = F_4 = F_2 = 0$, and $F_6(z,x) = -F_1(x,z)$, then we obtain the Cauchy equation

$$F_1(x+y,z) = F_1(x,z) + F_1(y,z),$$

which means that F_1 is <u>additive</u> in its first variable. If $F_4 = F_5 = F_6$ and $F_1 = F_2 = F_3 = 0$, then we have the cyclic equation

$$F_4(x,y) + F_4(y,z) + F_4(z,x) = 0.$$

A different sort of cyclic equation (cf. (2) below) results from taking $F_6 = F_4 = F_5 = 0$ and $F_1 = F_2 = F_3$. Putting again $F_1 = F_3 = F_2 = 0$, $F_4 = F_5$, and $F_6(z,x) = -F_4(x,z)$ yields Sincov's equation

$$F_4(x,y) + F_4(y,z) = F_4(x,z).$$

To get the cocycle equation

$$F_1(x+y,z) + F_1(x,y) = F_1(x,y+z) + F_1(y,z),$$

take $F_3 = F_6 = 0$, $F_4 = F_1 = -F_5$, and $F_2(y,x) = -F_1(x,y)$.

Clearly, many generalizations of these well-known equations are included also in (1). We shall need to solve some of these equations on our way toward the solution of (1). We also make use of the following result concerning a particular sort of cyclic functional equation.

Lemma 1. (Ebanks[2]) The general solution $K: R^2 \to R$ of

$$K(x+y,z) + K(y+z,x) + K(z+x,y) = 0, \tag{2}$$

for $x,y,z \in R$, is given by

$$K(x,y) = A(x+y,2y-x), \qquad x,y \in R,$$

for an arbitrary function $A: R^2 \to R$ which is additive in its second variable.

Hence the general continuous solution of (2) is of the form

$$K(x,y) = (2y-x)h(x+y), \qquad x,y \in R,$$

for an arbitrary continuous map $h: R \to R$. (Here the condition "continuous" in the last statement could be replaced by weaker conditions such as "measurable.")

2. SOLUTION OF A THREE-FUNCTION GENERALIZATION OF (2).

Now we shall consider the cyclic equation

$$H_1(x+y,z) + H_2(y+z,x) + H_3(z+x,y) = 0, \tag{3}$$

supposed for all $x,y,z \in R$. We prove the following.

Theorem 1. The general solution $H_1, H_2, H_3: R^2 \to R$ of (3) is given by

$$\left.\begin{aligned} H_1(x,y) &= A(x+y,2y-x) - f(x+y) - g(x+y) \\ H_2(x,y) &= A(x+y,2y-x) + g(x+y) \\ H_3(x,y) &= A(x+y,2y-x) + f(x+y) \end{aligned}\right\} \tag{4}$$

for all $x,y \in R$, for some arbitrary maps $f,g: R \to R$ and an arbitrary $A: R^2 \to R$ which is additive in its second variable.

Consequently, the general continuous (or measurable) solution of (3) is given by (4) with

$$A(x,y) = yh(x), \qquad x,y \in R,$$

for arbitrary continuous (resp., measurable) maps $f,g,h: R \to R$.

Proof: First, put $x = 0$ in (3) to get

$$H_1(y,z) = -H_2(y+z,0) - H_3(z,y), \tag{5}$$

and use this to transform (3) into

$$H_2(y+z,x) - H_2(x+y+z,0) + H_3(z+x,y) - H_3(z,x+y) = 0. \tag{6}$$

Next, put $y = 0$ in (6), obtaining

$$H_2(z,x) - H_2(x+z,0) = H_3(z,x) - H_3(z+x,0), \tag{7}$$

and use this to modify (6) to

$$H_3(y+z,x) - H_3(y+z+x,0) + H_3(z+x,y) - H_3(z,x+y) = 0. \tag{8}$$

Now, putting $z = 0$ in (8) yields

$$H_3(y,x) - H_3(y+x,0) = -H_3(x,y) + H_3(0,x+y), \tag{9}$$

and with this (8) becomes

$$H_3(y+z,x) - H_3(y+z+x,0) + H_3(z+x,y) + H_3(x+y,z) - H_3(0,x+y+z) - H_3(z+x+y,0) = 0. \tag{10}$$

Now, defining $K: R^2 \to R$ by

$$K(x,y) := H_3(x,y) - 2/3\, H_3(x+y,0) - 1/3\, H_3(0,x+y), \tag{11}$$

equation (10) exactly expresses the fact that K satisfies equation (2). Thus, by Lemma 1, there exists a map $A: R^2 \to R$, additive in its second variable, for which

$$K(x+y) = A(x+y,2y-x), \qquad x,y \in R. \tag{12}$$

Backtracking, we see from (11) that H_3 is as expressed in (4), with $f: R \to R$ defined by $f(x) := 2/3\, H_3(x,0) + 1/3\, H_3(0,x)$ for all $x \in R$.

Substituting this form of H_3 from (4) into (7), we find that

$$H_2(z,x) = A(z+x,2x-z) + f(z+x) - A(z+x,-x-z) - f(z+x) + H_2(x+z,0) - A(z+x,2x-z) + g(z+x),$$

where $g: R \to R$ is defined by $g(x) := A(x,x) + H_2(x,0)$. (Here we have used the fact that $A(x,-y) = -A(x,y)$, which follows from the additivity.) Hence H_2 is as stated in (4).

Finally, putting H_2 and H_3 from (4) into (5) and simplifying, again using the additivity of A, we get precisely the assertion of (4) for H_1. Conversely, any $H_n (n = 1,2,3)$ given by (4) with A additive in its second variable, satisfy (3). (Note that $A(x,0) = 0$.) This establishes the first half of Theorem 1.

The second half of Theorem 1 is a simple corollary of the first half. Indeed, let $y = w - x$ in the representation for H_3 in (4). Then

$$H_3(x,w-x) = A(w,2w-3x) + f(w).$$

Since $x \to H_3(x,w-x)$ is continuous for each fixed w, the map A must be continuous in its second variable. But that means (by additivity) that A has the form $A(x,y) = yh(x)$, as asserted. Furthermore, the representation of H_3 takes the form

$$H_3(x,y) = (2y-x)h(x+y) + f(x+y).$$

Choosing now $x = 2y$, we have

$$H_3(2y,y) = f(3y), \qquad y \in R.$$

Thus the continuity of f follows from that of H_3. The continuity of h and g follow in a similar way, and we are done.

Remark 1. By Remark 3 in [2], the first half of Lemma 1 is correct also if $K: X^2 \to Y$ where X and Y are abelian groups which are divisible by 2 and by 3. It is clear from the proof above that the same is true of Theorem 1 here.

3. DIFFERENTIABLE SOLUTIONS OF A GENERALIZED CAUCHY-CYCLIC EQUATION

The equation of interest in this section is

$$\left.\begin{aligned}[G_1(x+y,z) - G_1(x,z) - G_1(y,z)] + [G_2(y+z,x) - G_2(y,x) \\ - G_2(z,x)] + [G_3(z+x,y) - G_3(z,y) - G_3(x,y)] = 0,\end{aligned}\right\} \quad (13)$$

for all $x,y,z \in R$, with $G_n: R^2 \to R$ $(n = 1,2,3)$. Obviously, if all G_n's are additive in their first variable, then they satisfy (13). Also, (13) has a certain cyclic nature. The next result gives the general solution of (13) in the class $C^3(R^2)$.

Theorem 2. The general solution G_1, G_2, G_3 in $C^3(R^2)$ of (13) is given by

$$\left.\begin{aligned} G_1(x,y) &= E[h](x,y) - D[p+q](x,y) + xr_1(y) \\ G_2(x,y) &= E[h](x,y) + D[q](x,y) + xr_2(y) \\ G_3(x,y) &= E[h](x,y) + D[p](x,y) + xr_3(y) \end{aligned}\right\} \qquad (14)$$

$(x,y \in R)$, for some $h,p,q \in C^3(R)$ and $r_n \in C^2(R)$ $(n = 1,2,3)$, where D (the Cauchy difference operator) and E are operators from $\{f: R \to R\}$ to $\{f: R^2 \to R\}$ defined for any $f: R \to R$ by

$$D[f](x,y) = f(x+y) - f(x) - f(y), \quad (x,y \in R)$$
$$E[f](x,y) = (2y-x)f(x+y) + xf(x) - 2yf(y).$$

Proof: Differentiating (13) once each with respect to x,y, and z, we have

$$G_1^{112}(x+y,z) + G_2^{211}(y+z,x) + G_3^{121}(z+x,y) = 0,$$

where the superscripts denote partial derivatives. But this is equation (3). Moreover, G_1^{112}, G_2^{211}, and G_3^{121} are continuous functions, by hypothesis. Therefore Theorem 1 yields their representation

$$G_1^{112}(x,y) = (2y-x)h_1(x+y) - f_1(x+y) - g_1(x+y),$$
$$G_2^{211}(x,y) = (2y-x)h_1(x+y) + g_1(x+y),$$
$$G_3^{121}(x,y) = (2y-x)h_1(x+y) + f_1(x+y)$$

for some continuous maps f_1, g_1, h_1. Integrating three times, we obtain

$$\left.\begin{aligned} G_1(x,y) &= (2y-x)h(x+y) - f(x+y) - g(x+y) + k_1(x) \\ &\quad + l_1(y) + xm_1(y), \\ G_2(x,y) &= (2y-x)h(x+y) + g(x+y) + k_2(x) + l_2(y) \\ &\quad + xm_2(y), \\ G_3(x,y) &= (2y-x)h(x+y) + f(x+y) + k_3(x) + l_3(y) \\ &\quad + xm_3(y) \end{aligned}\right\} \qquad (15)$$

$(x,y \in R)$ for some maps h, f, g, k_n, $l_n \in C^3(R)$ and $m_n \in C^2(R)$ $(n = 1,2,3)$.

Next, we substitute forms (15) into equation (13). After simplifying, we get

$$\left.\begin{aligned} &- [(x+z)h(x+z) + (y+z)h(y+z) + (x+y)h(x+y)] \\ &+ f(x+z) + g(y+z) - f(x+y) - g(x+y) + [k_1(x+y) \\ &- k_1(x) - k_1(y)] + [k_2(y+z) - k_2(y) - k_2(z)] \\ &+ [k_3(x+z) - k_3(x) - k_3(z)] - l_1(z) - l_2(x) - l_3(y) = 0 \end{aligned}\right\} \quad (16)$$

$(x,y,z \in R)$. Putting $x = y = 0$ in (16) yields

$$l_1(z) = -2zh(z) + f(z) + g(z) + a_1, \qquad z \in R, \quad (17)$$

for some constant a_1. Similarly, we obtain

$$l_2(x) = -2xh(x) - g(x) + a_2, \qquad x \in R, \quad (18)$$

$$l_3(y) = -2yh(y) - f(y) + a_3, \qquad y \in R, \quad (19)$$

for some constants a_2, a_3.

Now, putting $x = 0$ in (16) and using (17) and (19), we find, after some rearrangement of terms, that

$$\begin{aligned} &[-(y+z)h(y+z) + g(y+z) + k_2(y+z)] \\ &\quad = [-yh(y) + g(y) + k_2(y)] + [-zh(z) + g(z) + k_2(z)] \\ &\qquad + k_1(0) + k_3(0) - l_2(0) - a_1 - a_3. \end{aligned}$$

This means that

$$-ih + g + k_2 \text{ is affine,}$$

that is, additive plus a constant, where i is the identity map on R. Since all the maps are continuous (and even more, in $C^3(R)$), we have $(-ih + g + k_2)(x) = b_2x + c_2$, that is,

$$k_2(x) = xh(x) - g(x) + b_2x + c_2, \qquad x \in R, \quad (20)$$

for some constants b_2, c_2. Similarly,

$$k_3(x) = xh(x) - f(x) + b_3x + c_3, \qquad x \in R, \tag{21}$$

$$k_1(x) = xh(x) + f(x) + g(x) + b_1x + c_1, \qquad x \in R, \tag{22}$$

for constants b_1, b_3, c_1, c_3.

Finally, substituting (17)-(22) into (16) yields only

$$\sum_{n=1}^{3} (a_n + c_n) = 0. \tag{23}$$

With (17)-(23), representation (15) takes the form

$$\begin{aligned} G_1(x,y) &= (2y-x)h(x+y) + xh(x) - 2yh(y) \\ &\quad - [f(x+y) - f(x) - f(y)] - [g(x+y) - g(x) - g(y)] \\ &\quad + x[b_1 + m_1(y)] - (a_2 + c_2 + a_3 + c_3), \end{aligned}$$

$$\begin{aligned} G_2(x,y) &= (2y-x)h(x+y) + xh(x) - 2yh(y) \\ &\quad + [g(x+y) - g(x) - g(y)] \\ &\quad + x\,[b_2 + m_2(y)] + (a_2 + c_2), \end{aligned}$$

$$\begin{aligned} G_3(x,y) &= (2y-x)h(x+y) + xh(x) - 2yh(y) \\ &\quad + [f(x+y) - f(x) - f(y)] \\ &\quad + x\,[b_3 + m_3(y)] + (a_3 + c_3). \end{aligned}$$

Defining p, q, $r_n\colon R \to R$ $(n = 1,2,3)$ by

$$p(x) := f(x) - (a_3 + c_3), \quad q(x) := g(x) - (a_2 + c_2),$$

$$r_n(x) := b_n + m_n(x), \qquad x \in R,$$

and expressing the solutions with the aid of the operators D and E, we have precisely (14).

An easy verification of the converse completes the proof of Theorem 2.

Remark 2. The solutions of (13) can be expressed also as follows, in what may be a convenient form for some purposes. Note that

$$\begin{aligned} E[h](x,y) &= (2y-x)h(x+y) + xh(x) - 2yh(y) \\ &= 2y[h(x+y) - h(y)] - x[h(x+y) - h(x)] \\ &= 3y[h(x+y) - h(y)] - [(x+y)h(x+y) - xh(x) - yh(y)] \\ &= 3y[h(x+y) - h(y)] - D[ih](x,y), \end{aligned}$$

where $i: R \to R$ is the identity map,

$$i(x) := x, \qquad x \in R.$$

Hence, defining new maps $s, t, u: R \to R$ by

$$s := 3h, \quad t := p - ih, \quad u := q - ih,$$

we can express (14) as

$$\left.\begin{aligned} G_1(x,y) &= y[s(x+y) - s(y)] - D[t+u+is](x,y) + xr_1(y), \\ G_2(x,y) &= y[s(x+y) - s(y)] + D[u](x,y) + xr_2(y), \\ G_3(x,y) &= y[s(x+y) - s(y)] + D[t](x,y) + xr_3(y), \end{aligned}\right\} \qquad (14')$$

for all $x, y \in R$.

4. MAIN RESULT: DIFFERENTIABLE SOLUTIONS OF EQUATION (1)

We now turn to the main purpose of this paper, which is to give the general solution of (1) under the assumption $F_1, F_2, F_3 \in C^3(R^2)$.

Theorem 3. The general solution of (1) with F_1, F_2, F_3 from the class $C^3(R^2)$ is given by

$$\left.\begin{aligned}
F_1(x,y) &= ys(x+y) + xr_1(y) + f_1(x) + f_2(y) - (t+u+is)(x+y),\\
F_2(x,y) &= ys(x+y) + xr_2(y) + f_3(x) + f_4(y) + u(x+y),\\
F_3(x,y) &= ys(x+y) + xr_3(y) + f_5(x) + f_6(y) + t(x+y),\\
F_4(x,y) &= -xr_3(y) - yr_2(x) + f_7(x) + f_8(y) - f_1(x+y),\\
F_5(x,y) &= -xr_1(y) - yr_3(x) - (f_6+f_8)(x) - (f_2+f_9)(y) - f_3(x+y),\\
F_6(x,y) &= -xr_2(y) - yr_1(x) + f_9(x) - (f_4+f_7)(y) - f_5(x+y)
\end{aligned}\right\} \qquad (24)$$

for arbitrary maps $r_n \in C^2(R)$ $(n = 1,2,3)$, s, t, u, $f_j \in C^3(R)$ $(j = 1, \ldots, 6)$, and $f_k: R \to R$ $(k = 7,8,9)$. Clearly, we can take also f_7, f_8, and f_9 from class $C^3(R)$, if F_4, F_5, and F_6 are assumed to be in $C^3(R^2)$.

Proof: We begin by putting $x = 0$ in (1) and solve for F_5, getting

$$F_5(y,z) = -F_1(y,z) - F_2(y+z,0) - F_3(z,y) - F_4(0,y) - F_6(z,0). \qquad (25)$$

Using this in (1), we obtain

$$\begin{aligned}
&F_1(x+y,z) - F_1(y,z) + F_2(y+z,x) - F_2(y+z,0) + F_3(z+x,y)\\
&- F_3(z,y) + F_4(x,y) - F_4(0,y) + F_6(z,x) - F_6(z,0) = 0.
\end{aligned} \qquad (26)$$

Setting $y = 0$ in (26) and solving for F_6, we find that

$$\begin{aligned}
F_6(z,x) - F_6(z,0) = &-F_1(x,z) + F_1(0,z) - F_2(z,x) + F_2(z,0)\\
&-F_3(z+x,0) + F_3(z,0) - F_4(x,0) + F_4(0,0).
\end{aligned} \qquad (27)$$

With this, (26) reduces to

$$\begin{aligned}
&[F_1(x+y,z) - F_1(y,z) - F_1(x,z) + F_1(0,z)] + [F_2(y+z,x)\\
&- F_2(y+z,0) - F_2(z,x) + F_2(z,0)] + [F_3(z+x,y) - F_3(z,y)\\
&- F_3(z+x,0) + F_3(z,0)] + [F_4(x,y) - F_4(0,y) - F_4(x,0)\\
&+ F_4(0,0)] = 0.
\end{aligned} \qquad (28)$$

Now the substitution $z = 0$ in (28) yields

$$\begin{aligned}
F_4(x,y) - F_4(0,y) - F_4(x,0) + F_4(0,0)\\
= &- [F_1(x+y,0) - F_1(y,0) - F_1(x,0) + F_1(0,0)]\\
&- [F_2(y,x) - F_2(y,0) - F_2(0,x) + F_2(0,0)]\\
&- [F_3(x,y) - F_3(0,y) - F_3(x,0) + F_3(0,0)].
\end{aligned} \qquad (29)$$

Using (29) in (28), and defining $G_n: R^2 \to R$ $(n = 1,2,3)$ by

$$G_n(x,y) := F_n(x,y) - F_n(x,0) - F_n(0,y) + F_n(0,0), \quad x,y \in R, \tag{30}$$

we see that (28) can be written in the form

$$[G_1(x+y,z) - G_1(x,z) - G_1(y,z)] + [G_2(y+z,x) - G_2(y,x) - G_2(z,x)] + [G_3(z+x,y) - G_3(z,y) - G_3(x,y)] = 0,$$

which is (13).

Therefore, since the G_n's inherit the assumed regularity of the F_n's through (30), the forms of the G_n's are given explicitly by Theorem 2. We use the forms (14') provided in Remark 2 following the proof of the theorem. By (14') and (30), we have

$$\left.\begin{aligned} F_1(x,y) &= ys(x+y) - D[t+u+is](x,y) + xr_1(y) + g_1(x) + h_1(y), \\ F_2(x,y) &= ys(x+y) + D[u](x,y) + xr_2(y) + g_2(x) + h_2(y), \\ F_3(x,y) &= ys(x+y) + D[t](x,y) + xr_3(y) + g_3(x) + h_3(y), \end{aligned}\right\} \tag{31}$$

$(x,y \in R)$ for some maps $s,t,u,g_n,h_n \in C^3(R)$ and $r_n \in C^2(R)$ $(n = 1,2,3)$, where D is the Cauchy difference operator, and where $g_n(x) = F_n(x,0) - F_n(0,0)$ and $h_n(y) = F_n(0,y) - ys(y)$ $(n = 1,2,3)$.

Now, let us return to (29). With (31) and some simplification, we find that

$$F_4(x,y) = -D[u+t+is+g_1](x,y) - xr_3(y) - yr_2(x) + g_4(x) + h_4(y) \tag{32}$$

for some g_4, $h_4: R \to R$. (Recall that there were no regularity assumptions about F_4, F_5, F_6.) Next, turning back to (27) and substituting (32) and (31), we obtain

$$\begin{aligned} F_6(z,x) = {} & D[t](z,x) - zr_2(x) - xr_1(z) + g_6(z) \\ & -[xs(x) + g_1(x) + h_2(x) + g_4(x)] - g_3(z+x), \end{aligned} \tag{33}$$

for some map $g_6: R \to R$. And returning at last to (25), by (31)-(33) we have

$$F_3(y,z) = D[u+is](y,z) - yr_1(z) - zr_3(y) - [g_1(y) + h_3(y) + h_4(y)] - [h_1(z) + g_6(z)] - [g_2(y+z) + (y+z)s(y+z)]. \tag{34}$$

Finally, to express the solutions more simply, we define functions $f_1, \ldots, f_9: R \to R$ by

$$f_1 := g_1 + t + u + is, \qquad f_2 := h_1 + t + u + is,$$
$$f_3 := g_2 - u, \qquad f_4 := h_2 - u$$
$$f_5 := g_3 - t, \qquad f_6 := h_3 - t$$
$$f_7 := g_4 + f_1, \qquad f_8 := h_4 + f_1, \qquad f_9 := g_6 - t.$$

With these definitions, the representations (31)-(34) take the form (24).

Conversely, it is easy to check that functions of the form (24) indeed satisfy equation (1). This completes the proof.

Remark 3. The representation of F_1 in (24) can be simplified slightly by observing that $ys(x+y) - (is)(x+y) = -xs(x+y)$. Thus we can write

$$F_1(x,y) = -xs(x+y) + xr_1(y) + f_1(x) + f_2(y) - (t+u)(x+y).$$

This entails some loss of symmetry, however, in the presentation of the forms of F_1, F_2, and F_3.

5. CONSEQUENCES AND FURTHER REMARKS

Of course one can use these results to deduce the forms of solutions to equations which are special cases of the ones presented here. Although a straightforward approach may be more efficient for "simple" equations such as (2) and Sincov's equation, it may not be for more complicated equations.

As an illustration, we treat the Cauchy-cyclic functional equation

$$G(x+y,z) + G(y+z,x) + G(z+x,y) = G(x,z) + G(y,z) + G(y,x) + G(z,x) + G(z,y) + G(x,y), \tag{35}$$

which is a special case ($G_1 = G_2 = G_3$) of (13).

Corollary 1. The general solution G, among functions in $C^3(R^2)$, of (35) is given by

$$G(x,y) = E[h](x,y) + xr(y), \qquad x,y \in R, \tag{36}$$

for some $h \in C^3(R)$ and $r \in C^2(R)$, where E is the operator defined in Theorem 2.

Proof: By Theorem 2, G is of the form

$$G(x,y) = E[h](x,y) + D[p](x,y) + xr_1(y) \tag{37}$$

for some maps h, $p \in C^3(R)$ and $r_1 \in C^2(R)$. Substituting this form into (35) and simplifying, we find that p must satisfy

$$p(x+y+z) - [p(x+y) + p(y+z) + p(z+x)] + [p(x) + p(y) + p(z)] = 0.$$

That is, p is a generalized homogeneous polynomial of degree (at most) two. But since $p \in C^3(R)$, it is an actual homogeneous polynomial of degree at most two[1]. Thus,

$$p(x) = ax^2 + bx, \qquad x \in R,$$

for some constants a,b. Hence

$$D[p](x,y) = 2axy, \qquad x,y \in R.$$

Substituting this into (37) and defining $r \in C^2(R)$ by

$$r(y) := r_1(y) + 2ay, \qquad y \in R,$$

we get (36), and we are done.

In conclusion, let us observe that the differentiability assumption does not seem to be essential to the study of these functional equations. That hypothesis was used at only one key point, in the proof of Theorem 2. Moreover, most of the functions appearing in the solutions are arbitrary except for the smoothness condition. Recently, in joint work with C.T. Ng, we have succeeded in obtaining the general solutions of all functional equations treated here, with no regularity assumptions whatsoever. These results will be presented in a forthcoming paper.

Acknowledgment. This research was partially supported by an Arts and Sciences Research Grant of the University of Louisville.

REFERENCES

1. Aczél, J., Lectures on Functional Equations and Their Applications, Academic Press, New York and London, 1966.
2. Ebanks, B.R., "On Antisymmetric Bi-additive Functions and an Interesting System of Functional Equations," C.R. Math. Rep. Acad. Sci. Canada 10, 1-6 (1988).

Bruce R. Ebanks
Department of Mathematics
University of Louisville
Louisville, KY 40292
U.S.A.

Constantin Carathéodory: An International Tribute (pp. 243–277)
edited by Th. M. Rassias

AN ALTERNATIVE TO THE COMPLETE FIGURE OF CARATHÉODORY

by

Dominic G. B. Edelen

and

R. J. McKellar

ABSTRACT FAMILIES OF HORIZONTAL IDEALS OF CONTACT MANIFOLDS ARE STUDIED. EACH HORIZONTAL IDEAL IS SHOWN TO ADMIT AN Π-DIMENSIONAL MODULE OF CAUCHY CHARACTERISTIC VECTOR FIELDS THAT IS ALSO A MODULE OF ANNIHILATORS (IN THE SENSE OF CARTAN) OF THE CONTACT IDEAL. ANY COMPLETELY INTEGRABLE HORIZONTAL IDEAL IN THE FAMILY LEADS TO A FOLIATION OF THE CONTACT MANIFOLD BY SUBMANIFOLDS OF DIMENSION Π ON WHICH THE HORIZONTAL IDEAL VANISHES. EXPLICIT CONDITIONS ARE OBTAINED UNDER WHICH AN OPEN SUBSET OF A LEAF OF THIS FOLIATION IS THE GRAPH OF A SOLUTION MAP OF THE FUNDAMENTAL IDEAL THAT CHARACTERIZES A SYSTEM OF EULER-LAGRANGE EQUATIONS OF A MULTIPLE INTEGRAL VARIATIONAL PROBLEM. WE SHOW THAT EVERY SMOOTH SOLUTION MAP CAN BE OBTAINED IN THIS MANNER. CONDITIONS ARE ALSO OBTAINED UNDER WHICH EVERY LEAF OF THIS FOLIATION IS THE GRAPH OF A SOLUTION MAP OF THE FUNDAMENTAL IDEAL. A CLASSIFICATION OF SOLUTION MAPS IS OBTAINED BY STUDYING CERTAIN MODULES OF ISOVECTORS OF THE EULER-LAGRANGE IDEAL. A DIRECT SUM DECOMPOSITION OF THE COLLECTION OF ALL ISOVECTORS OF HORIZONTAL IDEALS IS OBTAINED, AND THE RESULTING DECOMPOSITION OF THE LIE ALGEBRA OF ISOVECTORS IS GIVEN. TRANSPORT PROPERTIES OF ISOVECTORS ARE USED TO EXAMINE MULTIPLE INTEGRAL VARIATIONAL PROBLEMS. THESE ARE SHOWN TO LEAD TO MOMENTUM-ENERGY COMPLEXES AND TRANSVERSALITY CONDITIONS IN A NATURAL MANNER.

1. INTRODUCTION

A major step forward in the geometric understanding of multiple integral variational problems obtained from the publication of Carathéodory's work on the construction of a complete figure [1]. More modern expositions that relax certain of the original requirements can be found in [2, 3]. Those familiar with the complete figure are well aware of the inherent difficulties and the implicit nature of this construction. The purpose of this paper is to present an alternative to the construction of the complete figure that often proves both useful and efficient.

The n-dimensional manifold of independent variables is denoted by M_n. We assume that M_n is orientable and that a system of local coordinates $(x^i \mid 1 \leq i \leq n)$ has been introduced. The volume element (basis n-form) of M_n will be denoted by μ. The conjugate basis for $(n-1)$-forms is given by $(\mu_i = \partial_i \rfloor \mu \mid 1 \leq i \leq n)$ with the properties (see [4], Section 3.5)

$$dx^j \wedge \mu_i = \delta_i^j \mu\,, \quad d\mu_i = 0\,. \tag{1.1}$$

Study of first order, multiple integral variational problems requires "place holders" for the dependent variables and their first derivatives. These are provided by the introduction of a *contact manifold* $K = M_n \times R^m$, with local coordinates $(x^i, q^\alpha, y_i^\alpha \mid 1 \leq i \leq n, 1 \leq \alpha \leq N)$, where N is the number of dependent variables,

$$m = N(n+1)\,, \tag{1.2}$$

and *contact 1-forms*

$$C^\alpha = dq^\alpha - y_i^\alpha dx^i\,, \quad 1 \leq \alpha \leq N\,. \tag{1.3}$$

Let $\mathfrak{D}$ be an open, connected subset of a copy of M_n. A map $\Phi : \mathfrak{D} \to K$ is said to be *regular* if and only if $\Phi^* \mu \neq 0$. The collection of all regular maps is denoted by R. If a regular map Φ is such that Φ^* annihilates each of the contact 1-forms, then the y's become the derivatives of the dependent variables with respect to the independent variables on the range of Φ (see [4], Chapter 6). The collection of all regular, annihilating maps of the contact 1-forms is denoted by

(1.4) $$RC = \{\Phi : \mathfrak{D} \to K \mid \Phi^*\mu \neq 0\,,\ \Phi^*C^\alpha = 0\}\,.$$

Let $L(x^i, q^\alpha, y_i^\alpha)$ be a given element of $\Lambda^0(K)$. The *action integral* associated with the Lagrangian L is defined by

(1.5) $$A[\Phi] = \int_{\mathfrak{D}} \Phi^*(L\mu)\,,$$

for any $\Phi \in RC$. The *Euler-Lagrange n-forms* associated with the action integral (1.5) are given by

(1.6) $$E_\alpha = \Lambda_\alpha \mu - d\Lambda_\alpha^i \wedge \mu_i\,,$$

where

(1.7) $$\Lambda_\alpha = \frac{\partial L}{\partial q^\alpha}\,, \quad \Lambda_\alpha^i = \frac{\partial L}{\partial y_i^\alpha}\,.$$

As is well known, a map $\Phi \in RC$ is a *critical point* of the action integral $A[\Phi]$ if and only if

(1.8) $$\Phi^*E_\alpha = 0\,,\ 1 \leq \alpha \leq N\,.$$

Cartan [5] has shown us that this information can be organized in a more efficient fashion by studying the closed *fundamental ideal*

(1.9) $$\mathcal{F} = I\{C^\alpha, dC^\alpha, E_\alpha, dE_\alpha \mid 1 \leq \alpha \leq N\}$$

of $\Lambda(K)$. The fundamental ideal obviously contains the *contact ideal*

(1.10) $$\mathcal{C} = I\{C^\alpha, dC^\alpha \mid 1 \leq \alpha \leq N\}$$

as a subideal. Further, an elementary calculation and the identity $dC^\alpha \wedge \mu_j = -dy_j^\alpha \wedge \mu$ show that $dE_\alpha \equiv 0 \bmod \mathcal{C}$, and hence the fundamental ideal assumes the simpler form

(1.11) $$\mathcal{F} = I\{C^\alpha, dC^\alpha, E_\alpha\} .$$

The collection of all *solution maps* of the multiple integral variational problem with action integral $A[\Phi]$ is given by

(1.12) $$S = \{\Phi : \mathfrak{D} \to K \mid \Phi^*\mu \neq 0, \ \Phi^*\mathcal{F} = 0\}.$$

The requirement $\Phi^*\mu \neq 0$ guarantees that the range of Φ in K projects onto M_n as an n-dimensional region; that is, the x's remain independent on the range of Φ. On the other hand, $\Phi^*\mathcal{F} = 0$ if and only if

(1.13) $$\Phi^*C^\alpha = 0 , \quad \Phi^*E_\alpha = 0 , \quad 1 \leq \alpha \leq N$$

because $\Phi^*\Omega = 0$ implies $\Phi^*d\Omega = 0$. The fundamental problem associated with a given action integral $A[\Phi]$ is to establish that the set of solution maps S is not vacuous. Of equal importance, at least from the practical viewpoint, is to obtain definite algorithms for explicit calculation of solution maps when they exist. The results reported below have been guided by Cartan's views. They may be looked upon as an alternative to the complete figure construct of Carathéodory.

2. CANONICAL SYSTEMS OF VECTOR FIELDS

We noted in the previous section that the contact ideal $\mathcal{C}$ is a closed subideal of the fundamental ideal. It turns out that much of the analysis can be based solely on this subideal. This is because the generators of the fundamental ideal $\mathcal{F}$ that characterize the specific variational problem under study are represented by the Euler-Lagrange n-forms $\{E_\alpha \mid 1 \leq \alpha \leq N\}$, rather than by 0-forms. Since $\Lambda(K)$ is a graded algebra, it proves useful to introduce the graded submodules of $\mathcal{C}$ over $\Lambda^0(K)$ by

(2.1) $$\mathcal{C}^k = \mathcal{C} \cap \Lambda^k(K) .$$

An essential aspect of the Cartan approach [5] is the construction of modules of vector fields on K that are annihilators of the fundamental ideal. Let $T(K)$ denote the Lie algebra of smooth vector fields on K.

Definition 2.1 Let $\mathcal{N}$ be an ideal of $\Lambda(K)$, let $\mathcal{N}^k = \mathcal{N} \cap \Lambda^k(K)$, and let $\mathcal{U}$ be a module of $T(K)$ over $\Lambda^0(K)$. $\mathcal{U}$ is a module of *Cartan annihilators* if and only if every k-tuple $(U_1, \cdots, U_k)$ of elements of $\mathcal{U}$, $1 \leq k \leq \dim(K)$, satisfies the conditions

$$U_k \rfloor U_{k-1} \rfloor \cdots \rfloor U_1 \rfloor \mathcal{N}^k = 0 .$$

Modules of Cartan annihilators can be constructed for the ideal $\mathcal{C}$, but the situation is significantly simpler because we will only have to achieve the construction for "normalized" bases. The reasons for this will become apparent in what follows. The following notation will be used for the elements of the natural basis for $T(K)$:

$$\partial_i = \partial/\partial x^i, \quad \partial_\alpha = \partial/\partial q^\alpha, \quad \partial^i_\alpha = \partial/\partial y^\alpha_i . \tag{2.2}$$

Definition 2.2 A system of n vector fields $\{V_i \mid 1 \leq i \leq n\}$ on K is said to be a *canonical system* (i.e., a basis for a module of Cartan annihilators of $\mathcal{C}$) if and only if the vectors satisfy the normalization conditions

$$V_i \rfloor dx^j = \delta^j_i , \tag{2.3}$$

and the Cartan annihilator conditions

$$V_{i_1} \rfloor V_{i_2} \rfloor \cdots \rfloor V_{i_k} \rfloor \mathcal{C}^k = 0 , \quad 1 \leq k \leq n . \tag{2.4}$$

Remark. Since there are only n vectors in a canonical system, the conditions (2.4) will necessarily be satisfied by a canonical system for all $k > n$.

Theorem 2.1 *A system of vector fields* $\{V_i \mid 1 \leq i \leq n\}$ *is a canonical system if and only if*

$$V_i = \partial_i + y^\alpha_i \partial_\alpha + A^\alpha_{ij} \partial^j_\alpha , \quad 1 \leq i \leq n , \tag{2.5}$$

where the A's *are any system of elements of* $\Lambda^0(K)$ *that satisfy the symmetry relations*

(2.6) $$A^{\alpha}_{ij} = A^{\alpha}_{ji} ,$$

and the Lie product of any two elements of (V_i) *has the evaluation*

(2.7) $$[\![V_i, V_j]\!] = \left\{ V_i \langle A^{\alpha}_{jk} \rangle - V_j \langle A^{\alpha}_{ik} \rangle \right\} \partial^{k}_{\alpha} .$$

The contact manifold K *thus admits an* $Nn(n+1)/2$*—fold infinity of canonical systems, and hence* $\mathfrak{C}$ *admits an* $Nn(n+1)/2$*—fold infinity of n-dimensional modules of Cartan annihilators.*

Proof. Any system of n vector fields on K that satisfies the normalization conditions (2.3) has the form

$$V_i = \partial_i + v^{\alpha}_{i} \partial_{\alpha} + v^{\alpha}_{ij} \partial^{j}_{\alpha} , \quad 1 \leq i \leq n .$$

The system $\{V_i \mid 1 \leq i \leq n\}$ is therefore a linearly independent system. Noting that $\mathfrak{C}$ is generated by the 1-forms C^{α} and the 2-forms dC^{α}, satisfaction of conditions (2.4) can be achieved if and only if

$$V_i \rfloor C^{\alpha} = 0 , \quad V_i \rfloor V_j \rfloor dC^{\alpha} = 0 , \quad 1 \leq \alpha \leq N .$$

The results given by (2.5), (2.6), and (2.7) then follow from elementary algebraic calculations. If $\{V_i \mid 1 \leq i \leq n\}$ is a canonical system, then the system $\{U_i = N^{j}_{i} V_j \mid 1 \leq i \leq n, N^{j}_{i} \in \Lambda^{0}(K), \det(N^{j}_{i}) \neq 0\}$ is also a system of Cartan annihilators of the contact ideal; that is, the system $\{U_i \mid 1 \leq i \leq n\}$ will satisfy the graded annihilator conditions (2.4). □

In order to clarify some of the properties of canonical systems, we recall several standard definitions.

Definition 2.3 A vector field U is a *Cauchy characteristic* of an ideal $\mathcal{N}$ of $\Lambda(K)$ if and only if

(2.8) $$U \rfloor \mathcal{N} \subset \mathcal{N} .$$

Definition 2.4 A vector field U is an *isovector* of an ideal $\mathcal{N}$ of $\Lambda(K)$ if and only if

(2.9) $$\pounds_U \mathcal{N} \subset \mathcal{N} .$$

Theorem 2.2 *Let* $\{V_i \mid 1 \le i \le n\}$ *be a canonical system for the contact manifold* K. *If* U *is any vector field in the linear span of* $\{V_i\}$ *over* $\Lambda^0(K)$, *then* U *is neither a Cauchy characteristic nor an isovector of the contact ideal* C.

Proof. We first use Theorem 2.1 to obtain $V_i \rfloor dC^\alpha = dy_i^\alpha - A_{ij}^\alpha dx^j$. Since any vector field U in the linear span of $\{V_i\}$ has the representation $U = n^i V_i$, $n^i \in \Lambda^0(K)$, we have

$$U \rfloor dC^\alpha = n^i (dy_i^\alpha - A_{ij}^\alpha dx^j) \notin C .$$

Noting that $\pounds_U C^\alpha = U \rfloor dC^\alpha + d(U \rfloor C^\alpha) = U \rfloor dC^\alpha$ for any U in the linear span of $\{V_i\}$, the previous calculation shows that $\pounds_U C^\alpha \notin C$. □

3. REPRESENTATIONS IN TERMS OF HORIZONTAL AND VERTICAL IDEALS

The fact that no vector in the linear span of a canonical system is either a Cauchy characteristic of the contact ideal or an isovector of the contact ideal shows that we have been dealing with the wrong ideal of $\Lambda(K)$. We therefore proceed to reformulate the problem.

Definition 3.1 The *vertical ideal* of $\Lambda(K)$ is the closed differential ideal that is defined by

$$\mathcal{V} = I\{dx^i \mid 1 \le i \le n\} . \tag{3.1}$$

Definition 3.2 A *horizontal ideal* of $\Lambda(K)$ is defined by

$$\mathcal{H}[A_{ij}^\alpha] = I\{C^\alpha, H_i^\alpha \mid 1 \le \alpha \le N, 1 \le i \le n\} , \tag{3.2}$$

with

$$H_i^\alpha = dy_i^\alpha - A_{ij}^\alpha dx^j \tag{3.3}$$

for each choice of $\{A_{ij}^\alpha \in \Lambda^0(K)\}$ that satisfies the symmetry conditions

$$A_{ij}^\alpha = A_{ji}^\alpha . \tag{3.4}$$

Definition 3.3 A horizontal ideal $\mathcal{H}[A_{ij}^{\alpha}]$ serves to define an associated *horizontal module* $\mathcal{H}^*[A_{ij}^{\alpha}]$ of $T(K)$ by

$$\mathcal{H}^*[A_{ij}^{\alpha}] = \{U \in T(K) \mid U \rfloor \mathcal{H}[A_{ij}^{\alpha}] \subset \mathcal{H}[A_{ij}^{\alpha}]\} ; \tag{3.5}$$

that is, $\mathcal{H}^*[A_{ij}^{\alpha}]$ is the module of Cauchy characteristic vector fields of $\mathcal{H}[A_{ij}^{\alpha}]$.

Theorem 3.1 *The horizontal module* $\mathcal{H}^*[A_{ij}^{\alpha}]$ *admits the canonical system*

$$V_i = \partial_i + y_i^{\alpha}\partial_{\alpha} + A_{ij}^{\alpha}\partial_{\alpha}^{j} , \quad 1 \le i \le n \tag{3.6}$$

as a basis. Hence $\mathcal{H}^*[A_{ij}^{\alpha}]$ *is a module of Cauchy characteristics of* $\mathcal{H}[A_{ij}^{\alpha}]$,

$$V_i \rfloor C^{\alpha} = 0 , \quad V_i \rfloor H_j^{\alpha} = 0 . \tag{3.7}$$

and $\mathcal{H}^*[A_{ij}^{\alpha}]$ *is a module of Cartan annihilators of the contact ideal.*

Proof. Since $\mathcal{H}[A_{ij}^{\alpha}]$ is generated by the 1-forms $\{C^{\alpha}, H_i^{\alpha}\}$, any element Ω of $\mathcal{H}[A_{ij}^{\alpha}]$ is of the form

$$\Omega = C^{\alpha} \wedge P_{\alpha} + H_j^{\alpha} \wedge Q_{\alpha}^{j}$$

with $\{P_{\alpha}, Q_{\alpha}^{j}\}$ elements of $\Lambda(K)$ of the same degree. We therefore have

$$U \rfloor \Omega = (U \rfloor C^{\alpha})P_{\alpha} + (U \rfloor H_j^{\alpha})Q_{\alpha}^{j} \mod \mathcal{H}[A_{ij}^{\alpha}] ,$$

and hence $U = u^i\partial_i + u^{\alpha}\partial_{\alpha} + u_i^{\alpha}\partial_{\alpha}^{i}$ can belong to $\mathcal{H}^*[A_{ij}^{\alpha}]$ if and only if

$$0 = U \rfloor C^{\alpha} = u^{\alpha} - y_i^{\alpha}u^i , \quad 0 = U \rfloor H_i^{\alpha} = u_i^{\alpha} - A_{ik}^{\alpha}u^k .$$

It thus follows that any $U \in \mathcal{H}^*[A_{ij}^{\alpha}]$ is of the form

$$U = u^i\{\partial_i + y_i^{\alpha}\partial_{\alpha} + A_{ij}^{\alpha}\partial_{\alpha}^{j}\} = u^i V_i \tag{3.8}$$

with $\{V_i\}$ given by (3.6). Since $A_{ij}^{\alpha} = A_{ji}^{\alpha}$, Theorem 2.1 shows that $\{V_i \mid 1 \le i \le$

$n\}$ is a canonical system. Thus, since the elements in a canonical system are independent, (3.8) shows that $\{V_i \mid 1 \leq i \leq n\}$ is a basis for $\mathcal{H}^*[A^\alpha_{ij}]$. □

This result is fundamental in what follows. It tells us how to construct a horizontal ideal $\mathcal{H}[A^\alpha_{ij}]$ of $\wedge(K)$, for any given module $\mathcal{H}^*[A^\alpha_{ij}]$ of Cartan annihilators of C, such that $\mathcal{H}^*[A^\alpha_{ij}]$ becomes a module of Cauchy characteristics of $\mathcal{H}[A^\alpha_{ij}]$. The extensive body of information associated with Cauchy characteristics is thus made available for the study of problems in the calculus of variations along the lines initiated by Cartan.

It is clear from the definition of the vertical and horizontal ideals of $\wedge(K)$ that $\wedge^1(K)$ admits the direct sum decomposition

$$\wedge^1(K) = \left\{\mathcal{V} \cap \wedge^1(K)\right\} \oplus \left\{\mathcal{H}[A^\alpha_{ij}] \cap \wedge^1(K)\right\}. \tag{3.9}$$

This leads to the following result that will be instrumental in what follows.

Theorem 3.2 *If* f *is any smooth function on* K, *then*

$$df = V_i\langle f\rangle dx^i + (\partial_\beta f)C^\beta + (\partial^j_\beta f)H^\beta_j, \tag{3.10}$$

and hence

$$df \equiv V_i\langle f\rangle dx^i \mod \mathcal{H}[A^\alpha_{ij}]. \tag{3.11}$$

Proof. For any $f \in \wedge^0(K)$, we have

$$df = (\partial_k f)dx^k + (\partial_\alpha f)dq^\alpha + (\partial^i_\alpha f)dy^\alpha_i.$$

However, $dq^\alpha = C^\alpha + y^\alpha_k dx^k$, $dy^\alpha_i = H^\alpha_i + A^\alpha_{ik}dx^k$, by (1.3) and (3.3), and hence an elimination of dq^α and dy^α_i by using these relations gives (3.10) and (3.11). □

Theorem 3.3 *For any horizontal ideal* $\mathcal{H}[A^\alpha_{ij}]$ *of* $\wedge^0(K)$ *we have*

$$dC^\alpha = -H^\alpha_i \wedge dx^i, \tag{3.12}$$

$$dH^\alpha_i = -\tfrac{1}{2}\left\{V_m\langle A^\alpha_{ki}\rangle - V_k\langle A^\alpha_{mi}\rangle\right\}dx^m \wedge dx^k \tag{3.13}$$

$$- (\partial_\beta A^\alpha_{ik}) C^\beta \wedge dx^k - (\partial^j_\beta A^\alpha_{ik}) H^\beta_j \wedge dx^k ,$$

where $\{V_i \mid 1 \leq i \leq n\}$ *is the canonical basis for* $\mathfrak{H}^*[A^\alpha_{ij}]$. *Hence* $\mathfrak{H}[A^\alpha_{ij}]$ *is a closed differential ideal of* $\wedge(K)$ *if and only if*

$$V_k \langle A^\alpha_{mi} \rangle = V_m \langle A^\alpha_{ki} \rangle . \tag{3.14}$$

Proof. It follows directly from $C^\alpha = dq^\alpha - y^\alpha_i dx^i$ that $dC^\alpha = -dy^\alpha_i \wedge dx^i$. Since $dy^\alpha_i = H^\alpha_i + A^\alpha_{ij} dx^j$ by (3.3), we obtain

$$-dC^\alpha = (H^\alpha_i + A^\alpha_{ij} dx^j) \wedge dx^i = H^\alpha_i \wedge dx^i$$

when we use the symmetry relations $A^\alpha_{ij} = A^\alpha_{ji}$. In like manner, (3.3) gives $dH^\alpha_i = -dA^\alpha_{ik} \wedge dx^k$. Use of Lemma 3.1 to evaluate dA^α_{ik} thus gives us (3.13) when we use the symmetry relations $A^\alpha_{ij} = A^\alpha_{ji}$. The relations (3.12) show that we will always have $dC^\alpha \equiv 0 \bmod \mathfrak{H}[A^\alpha_{ij}]$. On the other hand, (3.13) shows that we will have $dH^\alpha_i \equiv 0 \bmod \mathfrak{H}[A^\alpha_{ij}]$ if and only if (3.14) hold. Thus, since $\mathfrak{H}[A^\alpha_{ij}]$ is generated by the 1-forms $\{C^\alpha, H^\alpha_i\}$, $\mathfrak{H}[A^\alpha_{ij}]$ is a closed differential ideal of $\wedge(K)$ if and only if (3.14) hold. □

Theorem 3.4 *The following statements are equivalent:*

(*i*) $\mathfrak{H}[A^\alpha_{ij}]$ *is a closed differential ideal of* $\wedge(K)$,

$$d\mathfrak{H}[A^\alpha_{ij}] \subset \mathfrak{H}[A^\alpha_{ij}] ; \tag{3.15}$$

(*ii*) $\mathfrak{H}^*[A^\alpha_{ij}]$ *is involutive,*

$$[\![\mathfrak{H}^*[A^\alpha_{ij}], \mathfrak{H}^*[A^\alpha_{ij}]]\!] \subset \mathfrak{H}^*[A^\alpha_{ij}] , \tag{3.16}$$

and

$$[\![V_i, V_j]\!] = 0 ; \tag{3.17}$$

(*iii*) $\mathfrak{H}^*[A^\alpha_{ij}]$ *is a module of isovectors of* $\mathfrak{H}[A^\alpha_{ij}]$,

(3.18) $$\pounds_U \mathcal{H}[A^\alpha_{ij}] \subset \mathcal{H}[A^\alpha_{ij}] \qquad \forall\ U \in \mathcal{H}^*[A^\alpha_{ij}] .$$

The conditions that $A^\alpha_{ij} \in \wedge^0(K)$ *must satisfy in order for these results to hold are*

(3.19) $$A^\alpha_{ij} = A^\alpha_{ji}$$

and

(3.20) $$V_i\langle A^\alpha_{jk}\rangle = V_j\langle A^\alpha_{ik}\rangle .$$

Proof. Theorem 3.3 has shown that $\mathcal{H}[A^\alpha_{ij}]$ is a closed differential ideal if and only if the relations (3.14) hold. However, Theorem 2.1 shows that (3.14) are both necessary and sufficient in order for the Lie products $[\![V_i, V_j]\!]$ to satisfy (3.17). Conversely if the Lie products $[\![V_i, V_j]\!]$ satisfy (3.17), then (2.7) show that the relations (3.14) are satisfied. Now, any two vectors U_1 and U_2 in $\mathcal{H}^*[A^\alpha_{ij}]$ are of the form $U_1 = n^i_1 V_i$, $U_2 = n^j_2 V_j$ because $(V_i \mid 1 \le i \le n)$ is a basis for $\mathcal{H}^*[A^\alpha_{ij}]$, and a direct calculation shows that the Lie product has the evaluation

$$[\![U_1, U_2]\!] = n^j_1 n^j_2 [\![V_i, V_j]\!] + U_1\langle n^j_2\rangle V_j - U_2\langle n^i_1\rangle V_i .$$

Hence $[\![U_1, U_2]\!]$ belongs to $\mathcal{H}^*[A^\alpha_{ij}]$ if and only if $[\![V_i, V_j]\!]$ belongs to $\mathcal{H}^*[A^\alpha_{ij}]$. This shows that $\mathcal{H}^*[A^\alpha_{ij}]$ is involutive if and only if $[\![V_i, V_j]\!]$ belongs to $\mathcal{H}^*[A^\alpha_{ij}]$ for all $1 \le i < j \le n$. However, (2.7) show that $[\![V_i, V_j]\!]$ belongs to $\mathcal{H}^*[A^\alpha_{ij}]$ if and only if the relations (3.14) are satisfied. This establishes the equivalence of (i) and (ii). Since $\mathcal{H}[A^\alpha_{ij}]$ is generated by the 1-forms (C^α, H^α_i), any vector U in $\mathcal{H}^*[A^\alpha_{ij}]$ is an isovector of $\mathcal{H}[A^\alpha_{ij}]$ if and only if $\pounds_U C^\alpha$ and $\pounds_U H^\alpha_{ij}$ are in $\mathcal{H}[A^\alpha_{ij}]$. Now,

$$\pounds_U C^\alpha = U\rfloor dC^\alpha + d(U\rfloor C^\alpha) = U\rfloor dC^\alpha ,$$

$$\pounds_U H^\alpha_i = U\rfloor dH^\alpha_i + d(U\rfloor H^\alpha_i) = U\rfloor dH^\alpha_i ,$$

where we have used the fact that any $U \in \mathcal{H}^*[A^\alpha_{ij}]$ is a Cauchy characteristic of $\mathcal{H}[A^\alpha_{ij}]$ in order to obtain the second equalities. We now use the evaluations (3.12) and (3.13) to obtain

$$\pounds_U C^\alpha \equiv 0 \mod \mathcal{H}[A^\alpha_{ij}]$$

and

$$\pounds_U H^\alpha_i \equiv -\tfrac{1}{2}\left\{V_m<A^\alpha_{ki}> - V_k<A^\alpha_{mi}>\right\} U\rfloor(dx^m\wedge dx^k) \mod \mathcal{H}[A^\alpha_{ij}].$$

Thus, since any $U \in \mathcal{H}^*[A^\alpha_{ij}]$ can be written in the form $U = n^i V_i$, we have $U\rfloor(dx^m\wedge dx^k) = u^m dx^k - u^k dx^m \in \mathcal{V}$. Accordingly, $\pounds_U H^\alpha_i$ is in $\mathcal{H}[A^\alpha_{ij}]$ if and only if $V_m<A^\alpha_{ki}> = V_k<A^\alpha_{mi}>$, and these conditions are both necessary and sufficient for $\mathcal{H}[A^\alpha_{ij}]$ to be a closed differential ideal by Theorem 3.3. This establishes the equivalence of (i) and (iii). □

4. CLOSURE CONDITIONS AND THE RESULTING FOLIATION STRUCTURES

Each choice of the functions $A^\alpha_{ij}(x^k, q^\beta, y^\beta_k)$, satisfying the symmetry conditions $A^\alpha_{ij} = A^\alpha_{ji}$, leads to a horizontal ideal $\mathcal{H}[A^\alpha_{ij}]$ of $\Lambda(K)$ and to an associated horizontal module $\mathcal{H}^*[A^\alpha_{ij}]$ of $T(K)$ that is both a module of Cauchy characteristic vectors of $\mathcal{H}[A^\alpha_{ij}]$ and a module of Cartan annihilators of $\mathcal{C}$. Theorem 3.4 shows that $\mathcal{H}[A^\alpha_{ij}]$ is stable under Lie transport by any vector in $\mathcal{H}^*[A^\alpha_{ij}]$ (i.e., $\mathcal{H}^*[A^\alpha_{ij}]$ is a module of isovectors of $\mathcal{H}[A^\alpha_{ij}]$) and that $\mathcal{H}^*[A^\alpha_{ij}]$ is involutive if and only if $\mathcal{H}[A^\alpha_{ij}]$ is a closed differential ideal of $\Lambda(K)$. Since $\mathcal{H}[A^\alpha_{ij}]$ is generated by the 1-forms $\{C^\alpha, H^\alpha_i\}$, the Frobenius Theorem [4] tells us that $\mathcal{H}[A^\alpha_{ij}]$ is *completely integrable* if and only if $\mathcal{H}[A^\alpha_{ij}]$ is a closed differential ideal. We will therefore restrict our consideration from now on to horizontal ideals of $\Lambda(K)$ that are completely integrable.

Definition 4.1 The collection of all completely integrable horizontal ideals of $\Lambda(K)$ is denoted by $\mathfrak{H}(K)$; that is,

$$\mathfrak{H}(K) = \{\mathcal{H}[A^\alpha_{ij}] \mid d\mathcal{H}[A^\alpha_{ij}] \subset \mathcal{H}[A^\alpha_{ij}]\}\ . \tag{4.1}$$

Theorem 4.1 *A horizontal ideal* $\mathcal{H}[A^\alpha_{ij}]$ *belongs to* $\mathfrak{H}(K)$ *if and only if the* A's *satisfy*

$$A^\alpha_{ij} = A^\alpha_{ji}\ , \tag{4.2}$$

(4.3) $$V_i\langle A^\alpha_{jk}\rangle = V_j\langle A^\alpha_{ik}\rangle ,$$

where

(4.4) $$V_i = \partial_i + y^\alpha_i \partial_\alpha + A^\alpha_{ij}\partial^j_\alpha , \quad 1 \le i \le n$$

is the canonical basis for the associated horizontal module $\mathcal{H}^*[A^\alpha_{ij}]$, *and* (4.3) *are equivalent to*

(4.5) $$[\![V_i, V_j]\!] = 0 .$$

The set $\mathfrak{H}(K)$ *is not vacuous because*

(4.6) $$A^\alpha_{ij} = \partial_i\partial_j\xi^\alpha(x^k)$$

satisfies the conditions (4.2) *and* (4.3) *for every smooth choice of the functions* $(\xi^\alpha(x^k) \mid 1 \le \alpha \le N)$.

Proof. Theorem 3.4 and the Frobenius Theorem show that a horizontal ideal $\mathcal{H}[A^\alpha_{ij}]$ is completely integrable if and only if (4.2) and (4.3) hold. It is then a simple computation to see that the A's given by (4.6) satisfy the conditions (4.2) and (4.3), and hence $\mathfrak{H}(K)$ is not vacuous. □

If $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$, then $\mathcal{H}[A^\alpha_{ij}]$ is a closed differential ideal that is generated by $m = N(1+n)$ independent 1-forms $(C^\alpha, H^\alpha_i \mid 1 \le \alpha \le N, 1 \le i \le n)$. Since $\dim(K) = n + m$, the Frobenius Theorem implies that K is foliated by n-dimensional manifolds such that $\mathcal{H}[A^\alpha_{ij}]$ vanishes when restricted to any leaf of this foliation. By definition, $\mathcal{H}^*[A^\alpha_{ij}]$ is a module of Cauchy characteristics of $\mathcal{H}[A^\alpha_{ij}]$ that has the canonical basis $(V_i \mid 1 \le i \le n)$. Accordingly, any solution g of the system of n, simultaneous, involutive, linear partial differential equations

(4.7) $$V_i\langle g\rangle = 0 , \quad 1 \le i \le n$$

will be constant on any leaf of the foliation generated by $\mathcal{H}[A^\alpha_{ij}]$. Since $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$ if and only if $[\![V_i, V_j]\!] = 0$, and (4.4) show that none of the vector fields (V_i) have critical points, the fundamental existence theorem for the system (4.7) asserts the existence of m functionally independent primitive integrals $(g_\Sigma \in \Lambda^0(K)$

$| 1 \leq \Sigma \leq m)$. Thus, any leaf of the foliation generated by $\mathcal{H}[A_{ij}^{\alpha}]$ will satisfy

$$g_{\Sigma}(x^i, q^{\alpha}, y_i^{\alpha}) = k_{\Sigma}, \quad 1 \leq \Sigma \leq m \tag{4.8}$$

for some choice of the constants $(k_{\Sigma} | 1 \leq \Sigma \leq m)$. It thus follows that each leaf of the foliation generated by $\mathcal{H}[A_{ij}^{\alpha}]$ is transverse to the fibers of K because $V_i \rfloor dx^j = \delta_i^j$.

If we introduce collective coordinates (z^A) on K by

$$(z^A | 1 \leq A \leq n + m) = (x^i, q^{\alpha}, y_i^{\alpha} | 1 \leq i \leq n, 1 \leq \alpha \leq N), \tag{4.9}$$

then the elements of the canonical basis take the generic form

$$V_i = v_i^A(z^B)\partial_A, \quad 1 \leq i \leq n, \tag{4.10}$$

with

$$[\![V_i, V_j]\!] = 0. \tag{4.11}$$

Let $P_0:(z_0^A)$ be an arbitrarily chosen point of K. We can then define a map Ψ_1: $J_1 \subset \mathbb{R} \to K | z^A = Z_1^A(z_0^B; u^1)$ by solving the initial value problem

$$\frac{dZ_1^A}{du^1} = v_1^A(Z_1^B), \quad Z_1^A(z_0^B; 0) = z_0^A. \tag{4.12}$$

We can define a map Ψ_2: $J_2 \subset \mathbb{R}^2 \to K | z^A = Z_2^A(z_0^B; u^1, u^2)$ by solving the initial value problem

$$\frac{dZ_2^A}{du^2} = v_2^A(Z_2^B), \quad Z_2^A(z_0^B; u^1, 0) = Z_1^A(z_0^B; u^1). \tag{4.13}$$

Continuing in this fashion, we thus obtain a map $\Psi = \Psi_n : J_n \subset \mathbb{R}^n \to K | z^A = Z_n^A(z_0^B; u^1, u^2, \cdots, u^n)$ by solving the initial value problem

$$\frac{dZ_n^A}{du^n} = v_n^A(Z_n^B), \quad Z_n^A(z_0^B; u^1, \cdots, u^{n-1}, 0) = Z_{n-1}^A(z_0^B; u^1, \cdots, u^{n-1}). \tag{4.14}$$

Since all of the vector fields (V_i) commute, the map Ψ that results from this

sequential integration of the orbital equations for (V_i) *from the point* P_0 is independent of the order in which we select the basis vectors (V_i). It is then easily seen that Ψ maps $J_n \subset \mathbb{R}^n$ into the leaf $\mathcal{L}(P_0)$ of the foliation generated by $\mathcal{H}[A_{ij}^{\alpha}]$ that passes through the point $P_0 \in K$. We will thus refer to such maps Ψ as *leaf maps*. Since $\mathcal{H}[A_{ij}^{\alpha}]$ restricted to any leaf of the foliation generated by $\mathcal{H}[A_{ij}^{\alpha}]$ vanishes, we have

$$\Psi^* C^{\alpha} = 0 , \quad \Psi^* H_i^{\alpha} = 0 \tag{4.15}$$

for any leaf map Ψ. Noting that Ψ^* commutes with exterior differentiation and that $V_n \rfloor V_{n-1} \rfloor \cdots \rfloor V_1 \rfloor \mu = 1$, we obtain

$$\Psi^* C^{\alpha} = 0 , \quad \Psi^* dC^{\alpha} = 0 , \quad \Psi^* \mu \neq 0 . \tag{4.16}$$

Thus, any leaf map of the foliation generated by $\mathcal{H}[A_{ij}^{\alpha}] \in \mathfrak{H}(K)$ is a solving map of the contact ideal. These results are summarized in the following theorem.

Theorem 4.2 *For each* $\mathcal{H}[A_{ij}^{\alpha}]$ *in* $\mathfrak{H}(K)$, *the space* K *is foliated by manifolds of dimension* n *that are transverse to the fibers of* K *and* $\mathcal{H}[A_{ij}^{\alpha}]$ *vanishes when restricted to any leaf of this foliation. If* $(V_i \mid 1 \leq i \leq n)$ *is the canonical basis for* $\mathcal{H}^*[A_{ij}^{\alpha}]$, *then the leaves of the foliation are given in implicit form by*

$$g_{\Sigma}(x^j, q^{\beta}, y_j^{\beta}) = k_{\Sigma} , \quad 1 \leq \Sigma \leq m = N(1+n) , \tag{4.17}$$

where the functions $(g_{\Sigma} \mid 1 \leq \Sigma \leq m)$ *constitute a complete, independent system of primitive integrals of the commutative system of partial differential equations*

$$V_i \langle g \rangle = 0 , \quad 1 \leq i \leq n . \tag{4.18}$$

Sequential integration of the orbital equations of the system (V_i) *from a point* $P_0 \in K$ *gives the leaf map* $\Psi : J_n \subset \mathbb{R}^n \rightarrow K$ *that maps* J_n *into the leaf* $\mathcal{L}(P_0)$ *that passes through* P_0. *Any such leaf map* Ψ *is a solving map for both the horizontal ideal* $\mathcal{H}[A_{ij}^{\alpha}]$ *and the contact ideal* C.

5. AN EXISTENCE THEOREM FOR EULER–LAGRANGE EQUATIONS

Any $\mathcal{H}[A^{\alpha}_{ij}]$ in $\mathfrak{H}(K)$ has been shown to lead to a foliation of K by graphs of solution maps of the contact ideal $\mathcal{C}$. We also know that $\mathcal{C}$ is a subideal of the fundamental ideal $\mathcal{F} = I\{C^{\alpha}, dC^{\alpha}, E_{\alpha} \mid 1 \leq \alpha \leq N\}$, where the E's are the Euler-Lagrange n-forms given by (1.6). These facts prompt us to ask whether we can find a leaf of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$ that contains the graph of a solution map of the fundamental ideal. Since $\Psi^{*}\mathcal{C} = 0$ for any leaf map Ψ, the leaf map Ψ will annihilate the fundamental ideal if and only if $\Psi^{*}E_{\alpha} = 0$. On the other hand, we also know that $\Psi^{*}\mathcal{H}[A^{\alpha}_{ij}] = 0$, and hence the following lemma proves to be useful

Lemma 5.1 *Let $\mathcal{H}[A^{\alpha}_{ij}] \in \mathfrak{H}(K)$, let Ψ be any leaf map associated with the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$, and let $\{V_i \mid 1 \leq i \leq n\}$ be the canonical basis for $\mathcal{H}^{*}[A^{\alpha}_{ij}]$. We have*

$$E_{\alpha} \equiv F_{\alpha}\mu \mod \mathcal{H}[A^{\alpha}_{ij}] , \tag{5.1}$$

where the F's are elements of $\Lambda^{0}(K)$ with the evaluations

$$F_{\alpha} = \Lambda_{\alpha} - V_i\langle\Lambda^{i}_{\alpha}\rangle , \tag{5.2}$$

and hence

$$\Psi^{*}E_{\alpha} = \Psi^{*}(F_{\alpha}\mu) . \tag{5.3}$$

Proof. The Euler-Lagrange n-forms are defined by $E_{\alpha} = \Lambda_{\alpha}\mu - d\Lambda^{i}_{\alpha}\wedge\mu_i$. If we use (3.11) to evaluate the indicated exterior derivative $d\Lambda^{i}_{\alpha}$, we obtain

$$E_{\alpha} \equiv \Lambda_{\alpha}\mu - V_j\langle\Lambda^{i}_{\alpha}\rangle dx^{j}\wedge\mu_i \mod \mathcal{H}[A^{\alpha}_{ij}] .$$

The relations (5.1) and (5.2) then follow upon noting that $dx^{j}\wedge\mu_i = \delta^{j}_{i}\mu$. Thus, since any leaf map Ψ of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$ gives $\Psi^{*}\mathcal{H}[A^{\alpha}_{ij}] = 0$, we obtain (5.3). □

We now have all of the results that are necessary in order to establish the following existence theorem.

Theorem 5.1 *Let* $\mathcal{H}[A^{\alpha}_{ij}]$ *belong to* $\mathfrak{H}(K)$ *and let* $\mathfrak{F}[A^{\alpha}_{ij}]$ *be the point set in* K *that is defined by*

$$\mathfrak{F}[A^{\alpha}_{ij}] = \{P \in K \mid F_{\alpha} = 0\,,\ 1 \leq \alpha \leq N\}\,. \tag{5.4}$$

If Ψ *is the map associated with a leaf of the foliation generated by* $\mathcal{H}[A^{\alpha}_{ij}]$ *and the graph of* Ψ *intersects* $\mathfrak{F}[A^{\alpha}_{ij}]$ *in a point set that is pulled back to an open subset* $\mathfrak{D}$ *of* $\mathbb{R}^n$ *by* Ψ^*, *then the restriction of the domain of* Ψ *to* $\mathfrak{D}$ *defines a solution map of the fundamental ideal.*

Proof. Any map Ψ associated with a leaf of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$ is such that $\Psi^*\mu \neq 0$. Thus, Lemma 5.1 shows that $\Psi^*E_{\alpha} = 0$ if and only if $\Psi^*F_{\alpha} = 0$. Noting that the F's are 0-forms on K, $\Psi^*F_{\alpha} = 0$ can be satisfied only on those regions $\mathcal{R}$ of K where the graph of Ψ intersects the point set $\mathfrak{F}[A^{\alpha}_{ij}]$. Let $\mathfrak{D} = \Psi^*\mathcal{R}$, then $\mathfrak{D}$ can be the domain of a solution map only if $\mathfrak{D}$ is an open subset of $\mathbb{R}^n$. If $\Psi_{\mathfrak{D}}$ denotes the map that results from Ψ by restriction of the domain of Ψ to $\mathfrak{D}$, then $\Psi^*_{\mathfrak{D}}E_{\alpha} = 0$. □

The point set $\mathfrak{F}[A^{\alpha}_{ij}]$ is the set of simultaneous zeros of the N functions $F_{\alpha} = \Lambda_{\alpha} - V_i\langle\Lambda^i_{\alpha}\rangle$, and hence it depends on the choice of (A^{α}_{ij}) because (V_i) depend on the choice of (A^{α}_{ij}). This explains the notation $\mathfrak{F}[A^{\alpha}_{ij}]$. This notation is used in order to emphasize the fact that we have to test every possible choice of (A^{α}_{ij}) for which $\mathcal{H}[A^{\alpha}_{ij}]$ is contained in $\mathfrak{H}(K)$ in order to use Theorem 5.1 to obtain all solution maps of the fundamental ideal that are accessible by this method.

We have explicitly restricted our considerations to completely integrable horizontal ideals. It might therefore appear that this restriction could eliminate some or all solutions of the fundamental ideal (i.e., we could miss some of the solutions of the Euler-Lagrange equations under study). That this is not the case, at least for smooth solutions, is shown by the following result.

Theorem 5.2 *Any smooth* (C^2) *solution map of the fundamental ideal can be realized as an open, n-dimensional subset of a leaf of the foliation generated by a completely integrable horizontal ideal.*

Proof. Let $\Phi : J_n \subset \mathbb{R}^n \to K$ be a smooth (C^2) solution map of the fundamental ideal. Since $\Phi^*\mu \neq 0$, Φ has a local presentation

$$\Phi \mid x^i = u^i\,,\quad q^{\alpha} = \phi^{\alpha}(u^k)\,,\quad y^{\alpha}_i = \frac{\partial\phi^{\alpha}(u^k)}{\partial u^i}\,. \tag{5.5}$$

If we set

$$A^{\alpha}_{ij} = \frac{\partial^2 \phi^{\alpha}(u^k)}{\partial u^i \partial u^j} ,$$

then $\mathcal{H}[A^{\alpha}_{ij}] \in \mathfrak{H}(K)$, as is easily checked. Sequential integration of the orbital equations of the canonical basis $(V_i \mid 1 \leq i \leq n)$ for $\mathcal{H}^*[A^{\alpha}_{ij}]$ gives the leaf maps

$$x^i = x^i_0 + u^i , \quad q^{\alpha} = q^{\alpha}_0 + y^{\alpha}_{i0} u^i + \phi^{\alpha}(x^k_0 + u^k) , \tag{5.6}$$

$$y^{\alpha}_i = y^{\alpha}_{i0} + \frac{\partial \phi^{\alpha}(x^k_0 + u^k)}{\partial u^i} ,$$

where $(u^i \mid 1 \leq i \leq n)$ is a system of coordinates on a neighborhood J_n of $\mathbb{R}^n$ that contains the origin. It is then easily seen that the solution map with the local presentation (5.5) coincides with the leaf map given by (5.6) with all integration constants set equal to zero. □

In the simplest cases, $\mathfrak{F}[A^{\alpha}_{ij}]$ will be a submanifold of K of codimension N. It is well known, however, that the sets of simultaneous zeros of N smooth functions on a manifold K of dimension $n + N(1+n)$ can have a very complicated structure. The conditions of Theorem 5.1 further compound the problem by requiring us to determine intersections of $\mathfrak{F}[A^{\alpha}_{ij}]$ with the n-dimensional leaves of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$, and then to test whether any such intersection pulls back to $\mathbb{R}^n$ to give an open set. It is thus abundantly clear that these tests can fail, and we would be unable to establish existence of a solution map of the fundamental ideal. Further, if the tests associated with Theorem 5.1 are positive, it could happen that only one leaf of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$ will intersect $\mathfrak{F}[A^{\alpha}_{ij}]$ in a point set that is the image of an open set $\mathfrak{D} \subset \mathbb{R}^n$ under the leaf map Ψ. This is also not unexpected because systems of relatively simple Euler-Lagrange equations are known to have solution sets that do not foliate K. There is therefore an obvious question that presents itself at this point. Can we find restrictions on the choices of (A^{α}_{ij}) for which these intersection problems become simpler? In particular, can we find whole leaves of the foliation of K that are graphs of solution maps, and when is every leaf of the foliation the graph of a solution map? Some answers to these questions are presented in the next section.

6. REDUCTION BY ISOVECTORS OF THE EULER-LAGRANGE IDEAL

Many of the questions associated with the determination of the structure of the point set $\mathfrak{F}[A^\alpha_{ij}]$ can be answered by studying yet another ideal of $\Lambda(K)$.

Definition 6.1 The *Euler-Lagrange ideal* of $\Lambda(K)$, associated with the Euler-Lagrange n-forms $(E_\alpha \mid 1 \leq \alpha \leq N)$, is given by

$$\mathcal{E}[A^\alpha_{ij}] = I(C^\alpha, H^\alpha_i, E_\alpha) = I(C^\alpha, H^\alpha_i, F_\alpha \mu) \ . \tag{6.1}$$

Theorem 6.1 *If* $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$, *then the Euler-Lagrange ideal is a closed differential ideal of* $\Lambda(K)$.

Proof. If $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$, then $\mathcal{H}[A^\alpha_{ij}]$ is a closed differential subideal of $\mathcal{E}[A^\alpha_{ij}]$. Thus, in view of (6.1) it is sufficient to check that $dF_\alpha \wedge \mu$ belongs to $\mathcal{E}[A^\alpha_{ij}]$. We know, however, that $dF_\alpha \equiv V_j\langle F_\alpha \rangle dx^j \bmod \mathcal{H}[A^\alpha_{ij}]$, by Theorem 3.2, and hence $dF_\alpha \wedge \mu \equiv 0 \bmod \mathcal{H}[A^\alpha_{ij}]$. □

It is not hard to prove that $\mathcal{H}^*[A^\alpha_{ij}]$ is not a module of Cauchy characteristics of the Euler-Lagrange ideal $\mathcal{E}[A^\alpha_{ij}]$, so we will not labor the reader with the details. The important question is whether the canonical basis vectors for $\mathcal{H}^*[A^\alpha_{ij}]$ are isovectors of $\mathcal{E}[A^\alpha_{ij}]$. The following result is therefore useful.

Lemma 6.1 *If* $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$ *and* $(V_i \mid 1 \leq i \leq n)$ *is the canonical basis for* $\mathcal{H}^*[A^\alpha_{ij}]$, *then*

$$\pounds_{V_i} E_\alpha \equiv \pounds_{V_i}(F_\alpha \mu) \equiv V_i\langle F_\alpha \rangle \mu \bmod \mathcal{H}[A^\alpha_{ij}] \ . \tag{6.2}$$

Proof. Since $\mathcal{H}[A^\alpha_{ij}]$ is stable under transport by any element of $\mathcal{H}^*[A^\alpha_{ij}]$, and $E_\alpha \equiv F_\alpha \mu \bmod \mathcal{H}[A^\alpha_{ij}]$, we have

$$\pounds_{V_i} E_\alpha \equiv \pounds_{V_i}(F_\alpha \mu) \bmod \mathcal{H}[A^\alpha_{ij}] \ .$$

The result then follows because Lie differentiation acts as a derivation and

$$\pounds_{V_i} \mu = d(V_i \rfloor \mu) = d\mu_i = 0$$

for any canonical system $(V_i \mid 1 \leq i \leq n)$. □

Theorem 6.2 *If* $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$ *and if* $(V_i \mid 1 \leq i \leq n)$ *is the canonical*

basis for $\mathfrak{K}^*[A^\alpha_{ij}]$, *then every vector field in* $\mathfrak{K}^*[A^\alpha_{ij}]$ *is an isovector of the Euler-Lagrange ideal* $\mathcal{E}[A^\alpha_{ij}]$ *if and only if*

$$V_i\langle F_\alpha\rangle = L^\beta_{\alpha i} F_\beta\,, \quad 1 \le \alpha \le N \tag{6.3}$$

are satisfied for some choice of the nN^2 *elements* $\{L^\beta_{\alpha i}\}$ *of* $\wedge^0(K)$.

Proof. Since $\mathcal{E}[A^\alpha_{ij}] = I\{C^\alpha, H^\alpha_i, F_\alpha\mu\}$, $\mathfrak{K}[A^\alpha_{ij}] \in \mathfrak{H}(K)$, and $\mathfrak{K}^*[A^\alpha_{ij}]$ is a module of isovectors of the subideal $\mathfrak{K}[A^\alpha_{ij}]$ by Theorem 3.4, it suffices to show that

$$\pounds_{V_i}(F_\alpha\mu) \equiv 0 \mod \mathcal{E}[A^\alpha_{ij}]\,.$$

When the congruences given by Lemma 6.1 are used, we obtain the conditions

$$V_i\langle F_\alpha\rangle\mu \equiv L^\beta_{\alpha i} F_\beta\mu \mod \mathfrak{K}[A^\alpha_{ij}]\,,$$

and hence the conditions (6.3) must be satisfied for some choice of the elements $\{L^\beta_{\alpha i}\}$ of $\wedge^0(K)$. Now, any $U \in \mathfrak{K}^*[A^\alpha_{ij}]$ has the form $U = n^i V_i$, and hence

$$\pounds_U(F_\alpha\mu) = n^i \pounds_{V_i}(F_\alpha\mu) + dn^i \wedge (F_\alpha V_i \rfloor \mu)\,.$$

Thus, when we use $dn^i \equiv V_j\langle n^i\rangle dx^j \mod \mathfrak{K}[A^\alpha_{ij}]$ and (6.3), we obtain

$$\pounds_U(F_\alpha\mu) \equiv (n^i L^\beta_{\alpha i} + V_i\langle n^i\rangle \delta^\beta_\alpha) F_\beta\mu \mod \mathfrak{K}[A^\alpha_{ij}]\,.$$

The result then follows upon noting that the right-hand side belongs to $\mathcal{E}[A^\alpha_{ij}]$. □

Theorem 6.3 *Let* $\mathfrak{K}[A^\alpha_{ij}]$ *be an element of* $\mathfrak{H}(K)$ *such that the* A's *satisfy the conditions* (6.3), *let* $(V_i \mid 1 \le i \le n)$ *be the canonical basis for* $\mathfrak{K}^*[A^\alpha_{ij}]$, *and let* $\mathcal{L}(P_0)$ *be the leaf of the foliation of* K *that contains the point* P_0. *If* P_0 *is in* $\mathfrak{F}[A^\alpha_{ij}]$, *then* $\mathcal{L}(P_0)$ *is contained in* $\mathfrak{F}[A^\alpha_{ij}]$. *If* Ψ *is the map from* $J_n \subset \mathbb{R}^n$ *that is constructed by sequential integration of the orbital equations of* (V_i) *starting from* P_0, *then* Ψ *is a map from* J_n *to* $\mathcal{L}(P_0)$ *that is a solution map of the fundamental ideal.*

Proof. Since P_0 belongs to $\mathfrak{F}[A^\alpha_{ij}]$ by hypothesis, we have

$$F_\alpha(P_0) = 0\,, \quad 1 \le \alpha \le N\,. \tag{6.4}$$

If we sequentially integrate the orbital equations of $\{V_i\}$ starting with the point P_0, we obtain a map Ψ from $J_n \subset \mathbf{R}^n$ into K such that the image of the origin in $\mathbf{R}^n$ is the point P_0 and the range of Ψ is contained in the leaf $\mathcal{L}(P_0)$ that contains P_0. Accordingly, (6.4) give us the evaluations

$$\Psi^*(F_\alpha(P_0)) = (\Psi^* F_\alpha)\big|_{u^i=0} = 0 . \tag{6.5}$$

Noting that all V_i restricted to the range of Ψ are tangent to the range of Ψ, satisfaction of the conditions (6.3) imply that $\Psi^* F_\alpha$ satisfy

$$\frac{d(\Psi^* F_\alpha)}{du^i} = (\Psi^* L^\beta_{\alpha i})(\Psi^* F_\beta) , \tag{6.6}$$

where $\{u^i \mid 1 \leq i \leq n\}$ is a system of local coordinates on $J_n \subset \mathbf{R}^n$. Sequential integration of the system (6.6) on J_n subject to the initial data (6.5) thus gives

$$\Psi^* F_\alpha = 0 , \quad 1 \leq \alpha \leq N . \tag{6.7}$$

Thus, Ψ is a solution map of the fundamental ideal by Theorem 5.1. □

An examination of the conditions (6.3) shows that there are basically three ways in which they can be satisfied.

Definition 6.2 The subset $\mathfrak{H}_s(K)$ of $\mathfrak{H}(K)$, that obtains for those choices of $\{A^\alpha_{ij}\}$ for which

$$F_\alpha = \Lambda_\alpha - V_i \langle \Lambda^i_\alpha \rangle = 0 , \quad 1 \leq \alpha \leq N , \tag{6.8}$$

is termed *special*. Any $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}_s(K)$ and the associated $\mathcal{H}^*[A^\alpha_{ij}]$ will also be termed special. The subset $\mathfrak{H}_r(K)$ of $\mathfrak{H}(K) - \mathfrak{H}_s(K)$, that obtains from the choices of $\{A^\alpha_{ij}\}$ for which

$$V_i \langle F_\alpha \rangle = 0 , \quad 1 \leq \alpha \leq N , \tag{6.9}$$

is termed *restricted*. The subset $\mathfrak{H}_g(K)$ of $\mathfrak{H}(K) - \mathfrak{H}_s(K) - \mathfrak{H}_r(K)$, that obtains from the choices of $\{A^\alpha_{ij}\}$ for which

(6.10) $$V_i\langle F_\alpha\rangle = L^{\beta}_{\alpha i} F_\beta\,, \quad 1 \le \alpha \le N\,,$$

for some not identically zero choices of the functions $(L^{\beta}_{\alpha i})$, will be termed *general*.

Theorem 6.4 *If* $\mathcal{H}[A^{\alpha}_{ij}]$ *belongs to* $\mathfrak{H}_S(K)$, *then every leaf of the foliation generated by* $\mathcal{H}[A^{\alpha}_{ij}]$ *is the graph of a solution map of the fundamental ideal; that is,* K *is foliated by graphs of solution maps of the Euler-Lagrange equations* $E_\alpha = 0$. *The conditions that the A's must satisfy in these circumstances are*

(6.11) $$A^{\alpha}_{ij} = A^{\alpha}_{ji}\,, \quad V_i\langle A^{\alpha}_{jk}\rangle = V_j\langle A^{\alpha}_{ik}\rangle,$$

(6.12) $$V_i\langle \Lambda^{i}_{\alpha}\rangle = \Lambda_\alpha\,, \quad 1 \le \alpha \le N\,,$$

where

(6.13) $$V_i = \partial_i + y^{\alpha}_i \partial_\alpha + A^{\alpha}_{ij}\partial^{j}_{\alpha}\,, \quad 1 \le i \le n\,.$$

Proof. The definition of $\mathfrak{H}_S(K)$ shows that the F's vanish throughout K for any $\mathcal{H}[A^{\alpha}_{ij}] \in \mathfrak{H}_S(K)$. Hence every point in K is a point P_0 for which Theorem 6.3 is applicable. This shows that every leaf of the foliation of K generated by $\mathcal{H}[A^{\alpha}_{ij}]$ is the graph of a solution map of the fundamental ideal. The conditions (6.11) and (6.12) that the A's must satisfy in order that $\mathcal{H}[A^{\alpha}_{ij}] \in \mathfrak{H}_S[A^{\alpha}_{ij}]$ follow directly from previous results. □

Theorem 6.5 *If* $\mathcal{H}[A^{\alpha}_{ij}]$ *belongs to* $\mathfrak{H}_r(K)$, *then each* F_α *is constant in value on any leaf of the foliation of* K *generated by* $\mathcal{H}[A^{\alpha}_{ij}]$. *Thus, any leaf of this foliation that touches the point set* $\mathfrak{F}[A^{\alpha}_{ij}]$ *is the graph of a solution map of the fundamental ideal. The conditions that the* A's *must satisfy in these circumstances are* (6.11), (6.13) *and*

(6.14) $$V_i\langle F_\alpha\rangle = V_i\langle \Lambda_\alpha - V_j\langle \Lambda^{j}_{\alpha}\rangle\rangle = 0\,, \quad 1 \le \alpha \le N\,, \quad 1 \le i \le n\,.$$

Proof. By definition, $\mathcal{H}[A^{\alpha}_{ij}]$ belongs to $\mathfrak{H}_r(K)$ if and only if the A's are such that $V_i\langle F_\alpha\rangle = 0$, $1 \le \alpha \le N$, $1 \le i \le n$. Thus, each of the F's is a solution of the system of simultaneous, linear partial differential equations $\{V_i\langle g\rangle = 0 \mid 1$

$\leq i \leq n$). Since $[\![V_i, V_j]\!] = 0$, the known properties of solutions of such systems shows that we must have $F_\alpha = f_\alpha(g_\Sigma)$, where $(g_\Sigma \mid 1 \leq \Sigma \leq m)$ is a system of independent primitive integrals of the system $(V_i<g> = 0 \mid 1 \leq i \leq n)$. We have shown previously, however, that the leaves of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$ are given in implicit form by the system of relations $(g_\Sigma = k_\Sigma \mid 1 \leq \Sigma \leq m)$. This shows that the F's are constant in value on the leaves of the foliation of K generated by $\mathcal{H}[A^{\alpha}_{ij}]$. Theorem 6.3 then shows that any leaf of this foliation that touches the point set $\mathcal{F}[A^{\alpha}_{ij}]$ is contained in $\mathcal{F}[A^{\alpha}_{ij}]$, and hence any leaf of this foliation on which all of the F's vanish is the graph of a solution map of the fundamental ideal. The conditions that the A's must satisfy in order that $\mathcal{H}[A^{\alpha}_{ij}]$ belong to $\mathfrak{H}_r(K)$ follow directly from previously established results. □

Theorem 6.6 *If* $\mathcal{H}[A^{\alpha}_{ij}]$ *belongs to* $\mathfrak{H}_g(K)$, *then the* F's *are constant in value only on those leaves of the foliation generated by* $\mathcal{H}[A^{\alpha}_{ij}]$ *that intersect* $\mathcal{F}[A^{\alpha}_{ij}]$. *Thus, any leaf of this foliation that intersects* $\mathcal{F}[A^{\alpha}_{ij}]$ *is the graph of a solution map of the fundamental ideal. The conditions that the* A's *must satisfy in these circumstances are* (6.11), (6.13), *and*

$$V_i<F_\alpha> = L^{\beta}_{\alpha i} F_\beta \,, \quad 1 \leq \alpha \leq N \,, \tag{6.15}$$

for some collection $(L^{\beta}_{\alpha i})$ *of functions not all of which vanish throughout* K.

Proof. By definition, $\mathcal{H}[A^{\alpha}_{ij}]$ will belong to $\mathfrak{H}_g(K)$ if and only if $V_i<F_\alpha> = L^{\beta}_{\alpha i} F_\beta$ for some not identically zero choice of $(L^{\beta}_{\alpha i})$. The F's can thus be constant in value on a leaf $\mathcal{L}$ of the foliation generated by $\mathcal{H}[A^{\alpha}_{ij}]$ only when all of the F's vanish at some point on $\mathcal{L}$, in which case the F's vanish on the whole leaf. This shows that any leaf $\mathcal{L}$ that intersects $\mathcal{F}[A^{\alpha}_{ij}]$ is contained in $\mathcal{F}[A^{\alpha}_{ij}]$, and hence $\mathcal{L}$ is the graph of a solution map of the fundamental ideal by Theorem 6.3. The conditions that the A's must satisfy in order for $\mathcal{H}[A^{\alpha}_{ij}]$ to belong to $\mathfrak{H}_g(K)$ follow directly from previously established results. □

7. ISOVECTORS OF THE HORIZONTAL IDEAL

We are interested in studying the set of all isovectors, $\mathrm{ISO}[A^{\alpha}_{ij}]$, of the closed horizontal ideal $\mathcal{H}[A^{\alpha}_{ij}]$, i.e., those vector fields $U \in T(K)$ such that

$$\pounds_U \mathcal{H}[A^{\alpha}_{ij}] \subset \mathcal{H}[A^{\alpha}_{ij}] \,. \tag{7.1}$$

It has already been established by Theorem 3.4 that $\mathfrak{H}^*[A_{ij}^{\alpha}]$ is a module of isovectors of $\mathfrak{H}[A_{ij}^{\alpha}]$ over $\wedge^0(K)$ with the canonical system

$$V_i = \partial_i + y_i^{\alpha}\partial_{\alpha} + A_{ij}^{\alpha}\partial_{\alpha}^{j}, \quad 1 \le i \le n, \tag{7.2}$$

as its basis. Therefore, $\{V_i, \partial_{\alpha}, \partial_{\alpha}^{j}\}$ is an admissible basis for $T(K)$. Thus

$$U = \eta^i V_i + \eta^{\alpha}\partial_{\alpha} + \eta_i^{\alpha}\partial_{\alpha}^{j} \tag{7.3}$$

will be an isovector of $\mathfrak{H}[A_{ij}^{\alpha}] = I\{C^{\beta}, H_i^{\beta} \mid 1 \le \beta \le N, 1 \le i \le n\}$ if and only if

$$\pounds_U C^{\alpha} \equiv 0 \mod \mathfrak{H}[A_{ij}^{\alpha}] \tag{7.4}$$

and

$$\pounds_U H_i^{\alpha} \equiv 0 \mod \mathfrak{H}[A_{ij}^{\alpha}]. \tag{7.5}$$

By means of the identity

$$\pounds_U \Omega = U \rfloor d\Omega + d(U \rfloor \Omega) \tag{7.6}$$

for any differential form Ω, it is found that

$$\pounds_U C^{\alpha} \equiv \left\{V_i\langle\eta^{\alpha}\rangle - \eta_i^{\alpha}\right\} dx^i \mod \mathfrak{H}[A_{ij}^{\alpha}] \tag{7.7}$$

and

$$\pounds_U H_i^{\alpha} \equiv \left\{V_j\langle\eta_i^{\alpha}\rangle - \eta^{\beta}\partial_{\beta}\langle A_{ij}^{\alpha}\rangle - \eta_k^{\beta}\partial_{\beta}^{k}\langle A_{ij}^{\alpha}\rangle\right\} dx^j \mod \mathfrak{H}[A_{ij}^{\alpha}]. \tag{7.8}$$

Hence U is an isovector of $\mathfrak{H}[A_{ij}^{\alpha}]$ if and only if

$$V_i\langle\eta^{\alpha}\rangle = \eta_i^{\alpha}, \tag{7.9}$$

$$V_j\langle\eta_i^{\alpha}\rangle = (\eta^{\beta}\partial_{\beta} + \eta_k^{\beta}\partial_{\beta}^{k})\langle A_{ij}^{\alpha}\rangle. \tag{7.10}$$

Therefore, the complete set of isovectors, $\mathrm{ISO}[A^{\alpha}_{ij}]$, can be characterized in the following manner.

Theorem 7.1 *A vector field* $U \in T(K)$ *is an isovector of the horizontal ideal* $\mathfrak{H}[A^{\alpha}_{ij}] \in \mathfrak{H}(K)$ *if and only if it is of the form*

$$U = n^i V_i + \eta^{\alpha}\partial_{\alpha} + V_i\langle\eta^{\alpha}\rangle\partial^i_{\alpha} \tag{7.11}$$

for any choice of the n *functions* $\{n^i \in \wedge^0(K) \mid 1 \le i \le n\}$ *and for any choice of the* N *functions* $\{\eta^{\alpha} \mid 1 \le \alpha \le N\}$ *that satisfy*

$$V_j V_i\langle\eta^{\alpha}\rangle = (\eta^{\beta}\partial_{\beta} + V_k\langle\eta^{\beta}\rangle\partial^k_{\beta})\langle A^{\alpha}_{ij}\rangle\ . \tag{7.12}$$

It can be shown that (7.12) is an over-determined system whose integrability conditions are satisfied identically for any completely integrable horizontal ideal $\mathfrak{H}[A^{\alpha}_{ij}] \in \mathfrak{H}(K)$. Since (7.12) constitute a system of linear equations in the variables $\{\eta^{\beta} \mid 1 \le \beta \le N\}$, we see that

$$\mathfrak{W}[A^{\alpha}_{ij}] = \{W_{\eta} = \eta^{\alpha}\partial_{\alpha} + V_i\langle\eta^{\alpha}\rangle\partial^i_{\alpha} \quad \forall\ \eta^{\alpha} \text{ satisfying (7.12)}\} \tag{7.13}$$

is a vector subspace of $T(K)$, and that $\mathrm{ISO}[A^{\alpha}_{ij}]$ admits the direct sum decomposition

$$\mathrm{ISO}[A^{\alpha}_{ij}] = \mathfrak{H}^*[A^{\alpha}_{ij}] \oplus \mathfrak{W}[A^{\alpha}_{ij}] \tag{7.14}$$

as a vector space. We have seen that $\mathfrak{H}^*[A^{\alpha}_{ij}]$ is a module over $\wedge^0(K)$, but $\mathfrak{W}[A^{\alpha}_{ij}]$ is not. Nonetheless, if we restrict consideration to the associative algebra

$$\mathfrak{P}[A^{\alpha}_{ij}] = \{f \in \wedge^0(K) \mid V_i\langle f\rangle = 0,\ 1 \le i \le n\}\ , \tag{7.15}$$

the linearity of the system (7.12) in $\{\eta^{\alpha}\}$ shows that $\mathfrak{W}[A^{\alpha}_{ij}]$ becomes a module over $\mathfrak{P}[A^{\alpha}_{ij}]$. Accordingly, since $\mathfrak{P}[A^{\alpha}_{ij}]$ is a subalgebra of $\wedge^0(K)$, we have the following result.

Theorem 7.2 *The collection* $\mathrm{ISO}[A^{\alpha}_{ij}]$ *of all isovectors of* $\mathfrak{H}[A^{\alpha}_{ij}] \in \mathfrak{H}(K)$ *is a vector subspace of* $T(K)$ *that admits the direct sum decomposition*

(7.16) $$\mathrm{ISO}[A^{\alpha}_{ij}] = \mathfrak{K}^*[A^{\alpha}_{ij}] \oplus \mathcal{W}[A^{\alpha}_{ij}]$$

of submodules over the associative algebra $\mathcal{P}[A^{\alpha}_{ij}]$ *of all smooth functions that are annihilated by the action of all elements of* $\mathfrak{K}^*[A^{\alpha}_{ij}]$.

It is well known that T(K) forms a Lie algebra with product $[\![\ ,\ ]\!]$ given by the commutator. Since

(7.17) $$\pounds_{[\![U,\ V]\!]} = \pounds_U \pounds_V - \pounds_V \pounds_U ,$$

the definition of $\mathrm{ISO}[A^{\alpha}_{ij}]$ shows that

(7.18) $$[\![\mathrm{ISO}[A^{\alpha}_{ij}],\ \mathrm{ISO}[A^{\alpha}_{ij}]]\!] \subset \mathrm{ISO}[A^{\alpha}_{ij}] ,$$

and thus $\mathrm{ISO}[A^{\alpha}_{ij}]$ is a Lie subalgebra of T(K).

Due to the direct sum decomposition of $\mathrm{ISO}[A^{\alpha}_{ij}]$ established in Theorem 7.2, one is interested in the Lie algebraic properties of the submodules $\mathfrak{K}^*[A^{\alpha}_{ij}]$ and $\mathcal{W}[A^{\alpha}_{ij}]$. By virtue of the relations

(7.19) $$[\![n^i V_i,\ U]\!] = n^i [\![V_i,\ U]\!] - U\langle n^i \rangle V_i$$

and

(7.20) $$[\![V_i,\ U]\!] = V_i\langle u^j \rangle V_j + (V_i\langle u^{\alpha} \rangle - u^{\alpha}_i)\partial_{\alpha}$$
$$+ (V_i\langle u^{\alpha}_j \rangle - u^{\beta}\partial_{\beta}\langle A^{\alpha}_{ij} \rangle - u^{\beta}_k \partial^k_{\beta}\langle A^{\alpha}_{ij} \rangle)\partial^j_{\alpha} ,$$

where

(7.21) $$U = u^i \partial_i + u^{\alpha}\partial_{\alpha} + u^{\alpha}_i \partial^i_{\alpha}$$

is any element of T(K), we see that $\mathfrak{K}^*[A^{\alpha}_{ij}]$ is an ideal of $\mathrm{ISO}[A^{\alpha}_{ij}]$. Therefore,

(7.22) $$[\![\mathfrak{K}^*[A^{\alpha}_{ij}],\ \mathrm{ISO}[A^{\alpha}_{ij}]]\!] \subset \mathfrak{K}^*[A^{\alpha}_{ij}] .$$

Noting that $\mathrm{ISO}[A^{\alpha}_{ij}]$ is closed under the Lie product and that

$$[\![\, \mathcal{W}[A^{\alpha}_{ij}],\ \mathcal{W}[A^{\alpha}_{ij}]]\!] \cap \mathcal{H}^{*}[A^{\alpha}_{ij}] = \varnothing \,,$$

we have the following results.

Theorem 7.3 *If* $\mathcal{H}[A^{\alpha}_{ij}] \in \mathfrak{H}(K)$, *then the direct sum decomposition*

$$\mathrm{ISO}[A^{\alpha}_{ij}] = \mathcal{H}^{*}[A^{\alpha}_{ij}] \oplus \mathcal{W}[A^{\alpha}_{ij}] \tag{7.23}$$

induces the Lie algebra decomposition

$$[\![\, \mathcal{H}^{*}[A^{\alpha}_{ij}],\ \mathcal{H}^{*}[A^{\alpha}_{ij}]]\!] \subset \mathcal{H}^{*}[A^{\alpha}_{ij}] \,, \tag{7.24}$$

$$[\![\, \mathcal{H}^{*}[A^{\alpha}_{ij}],\ \mathcal{W}[A^{\alpha}_{ij}]]\!] \subset \mathcal{H}^{*}[A^{\alpha}_{ij}] \,, \tag{7.25}$$

$$[\![\, \mathcal{W}[A^{\alpha}_{ij}],\ \mathcal{W}[A^{\alpha}_{ij}]]\!] \subset \mathcal{W}[A^{\alpha}_{ij}] \,. \tag{7.26}$$

8. TRANSPORT PROPERTIES

As we shall now see, one of the reasons for studying isovectors of $\mathcal{H}[A^{\alpha}_{ij}]$ is that they provide mappings between solution maps of $\mathcal{H}[A^{\alpha}_{ij}]$. This is done by means of the transport operator

$$\mathcal{T}_{U}(s) = \exp(s\,U) \,, \tag{8.1}$$

for any $U \in \mathrm{ISO}[A^{\alpha}_{ij}]$, which is such that [4]

$$\mathcal{T}^{*}_{U}(s) = \exp(s\,\pounds_{U}) \tag{8.2}$$

Recall that Ψ is a solution map of $\mathcal{H}[A^{\alpha}_{ij}]$ if and only if $\Psi^{*}\mu \neq 0$ and $\Psi^{*}\mathcal{H}[A^{\alpha}_{ij}] = 0$. Transport of Ψ by $U \in \mathrm{ISO}[A^{\alpha}_{ij}]$ satisfies

$$(\mathcal{T}_{U}(s) \circ \Psi)^{*} = \Psi^{*} \circ \mathcal{T}^{*}_{U}(s) = \Psi^{*} \exp(s\,\pounds_{U}) \,. \tag{8.3}$$

Therefore

$$(\mathcal{T}_U(s) \circ \Psi)^* \mu = \Psi^* \exp(s \pounds_U) < \mu > \neq 0 \tag{8.4}$$

for all sufficiently small s near $s = 0$. Furthermore, we have

$$(\mathcal{T}_U(s) \circ \Psi)^* \mathcal{H}[A^\alpha_{ij}] = \Psi^* \left[\exp(s \pounds_U) < \mathcal{H}[A^\alpha_{ij}] > \right] = \Psi^* \mathcal{H}[A^\alpha_{ij}] = 0 \tag{8.5}$$

since $U \in \mathrm{ISO}[A^\alpha_{ij}]$. Thus, we have the following result.

Lemma 8.1 *If Ψ is a solution map of the horizontal ideal $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}[K]$, then $\mathcal{T}_U(s) \circ \Psi$ is a solution map of the horizontal ideal for all s in a sufficiently small neighborhood of $s = 0$.*

In our procedure the solution maps are obtained as leaves of the foliation generated by $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$. They are given by

$$g_\Sigma(x^i, q^\alpha, y^\alpha_i) = k_\Sigma , \quad 1 \leq \Sigma \leq m = N(1+n) , \tag{8.6}$$

where $\{g_\Sigma\}$ is any system of m independent elements of the associative algebra

$$\mathcal{P}[A^\alpha_{ij}] = \{ f \in \wedge^0(K) \mid V_i < f > = 0, \ 1 \leq i \leq n \} . \tag{8.7}$$

The submodule decomposition of $\mathrm{ISO}[A^\alpha_{ij}]$ gives rise to two different kinds of transport of the leaves. If $V = n^i V_i \in \mathcal{H}^*[A^\alpha_{ij}]$, then

$$\mathcal{T}_V(s) < f > = \exp(s V) < f > = f \tag{8.8}$$

for all $f \in \mathcal{P}[A^\alpha_{ij}]$. Thus transport by an element of $\mathcal{H}^*[A^\alpha_{ij}]$ takes any leaf of the foliation generated by $\mathcal{H}[A^\alpha_{ij}]$ into itself. On the other hand, if $W \in \mathcal{W}[A^\alpha_{ij}]$, then (7.20) yields $[\![V_i, W]\!] = 0$ and thus $[\![V_i, \mathcal{T}_W(s)]\!] = 0$. Hence,

$$V_i < \mathcal{T}_W(s) < f > > = [\![V_i, \mathcal{T}_W(s)]\!] < f > + \mathcal{T}_W(s) < V_i < f > > = 0$$

for all $f \in \mathcal{P}[A^\alpha_{ij}]$. Therefore, $g = \mathcal{T}_W(s) < f >$ is in $\mathcal{P}[A^\alpha_{ij}]$ and thus, in general, the action of any nontrivial element of $\mathcal{W}[A^\alpha_{ij}]$ will transport a leaf of the foliation generated by $\mathcal{H}[A^\alpha_{ij}]$ into another leaf of that foliation. Note that $\mathcal{T}_W(s)$

leaves the base manifold M_n invariant. These considerations lead to the following result.

Theorem 8.1 *If* $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$, *then* $\mathcal{T}_U(s)$ *is a map of* $\mathcal{P}[A^\alpha_{ij}]$ *into* $\mathcal{P}[A^\alpha_{ij}]$ *for all* $U \in ISO[A^\alpha_{ij}]$ *and for all* s *in a sufficiently small neighborhood of* $s = 0$. *Moreover, if* $V \in \mathcal{H}^*[A^\alpha_{ij}]$, *then* $\mathcal{T}_V(s)$ *is the identity map for* $\mathcal{P}[A^\alpha_{ij}]$.

9. CALCULUS OF VARIATIONS

Using our formalism, the action integral for a multiple integral problem in the calculus of variations is given by

$$A[\Psi] = \int_{\mathfrak{D}} \Psi^*(L\mu)\,, \tag{9.1}$$

where $L \in \Lambda^0(K)$ and $\mathfrak{D}$ is an n-dimensional, arcwise connected point set of M_n with boundary $\partial\mathfrak{D}$. The ideal of $\Lambda(K)$ that is naturally associated with this action integral is the closed Euler-Lagrange ideal

$$\mathcal{S}[A^\alpha_{ij}] = I(C^\alpha, H^\alpha_i, E_\alpha), \tag{9.2}$$

where E_α are the Euler-Lagrange n-forms

$$E_\alpha = \Lambda_\alpha \mu - d\Lambda^i_\alpha \wedge \mu_i\,, \tag{9.3}$$

and

$$\Lambda_\alpha = \frac{\partial L}{\partial q^\alpha}\,, \quad \Lambda^i_\alpha = \frac{\partial L}{\partial y^\alpha_i}\,. \tag{9.4}$$

Therefore, a solution map Ψ of the Euler-Lagrange ideal $\mathcal{S}[A^\alpha_{ij}]$ is a solution map of the horizontal ideal $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$ such that the Euler-Lagrange equations are satisfied, i.e.,

$$\Psi^* E_\alpha = 0\,, \quad 1 \le \alpha \le N\,. \tag{9.5}$$

As discussed in Section 8, transport of a solution map Ψ of the horizontal ideal $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$ by an isovector $U \in ISO[A^\alpha_{ij}]$ is given by $\Psi_U(s) = \mathcal{T}_U(s) \circ \Psi$. By means of (8.3) the resulting action integral becomes

(9.6) $$A[\Psi_U(s)] = \int_{\mathfrak{D}} \Psi^* \exp(s\pounds_U)<L\mu> .$$

Thus the *finite variation* of the action integral generated by $U \in ISO[A^\alpha_{ij}]$, defined by

(9.7) $$\Delta_U(s)A[\Psi] = A[\Psi_U(s)] - A[\Psi] ,$$

reduces to

(9.8) $$\Delta_U(s)A[\Psi] = \int_{\mathfrak{D}} \Psi^*\Big((\exp(s\pounds_U) - 1)<L\mu>\Big) .$$

Much information can be gained from the finite variation by studying the expression $\pounds_U<L\mu>$. An equivalent expression which facilitates calculations is obtained from the Cartan n-form

(9.9) $$Car_L = L\mu + J ,$$

where

(9.10) $$J = \Lambda^i_\alpha C^\alpha \wedge \mu_i \in \mathfrak{H}[A^\alpha_{ij}]$$

because $C^\alpha \in \mathfrak{H}[A^\alpha_{ij}]$ for all admissible choices of (A^α_{ij}). Thus, for any $U \in ISO[A^\alpha_{ij}]$, we have

(9.11) $$\pounds_U J \equiv 0 \mod \mathfrak{H}[A^\alpha_{ij}] .$$

Therefore, $\pounds_U(L\mu) \equiv \pounds_U Car_L \mod \mathfrak{H}[A^\alpha_{ij}] = \pounds_U(L\mu + J)$ for every $U \in ISO[A^\alpha_{ij}]$.

Theorem 7.1 shows that $U \in ISO[A^\alpha_{ij}]$ if and only if

(9.12) $$U = n^i V_i + \eta^\alpha \partial_\alpha + V_i<\eta^\alpha>\partial^i_\alpha ,$$

for any $n^i \in \Lambda^0(K)$ and for any $\eta^\alpha \in \Lambda^0(K)$ that satisfy

(9.13) $$V_j V_i<\eta^\alpha> = (\eta^\beta \partial_\beta + V_k<\eta^\beta>\partial^k_\beta)<A^\alpha_{ij}> .$$

In view of the identity

$$\pounds_U(L\mu + J) = U\rfloor d(L\mu + J) + d(U\rfloor(L\mu + J)) , \tag{9.14}$$

we start by finding that

$$d(L\mu + J) = \Lambda_\alpha C^\alpha \wedge \mu + d\Lambda^i_\alpha \wedge C^\alpha \wedge \mu_i . \tag{9.15}$$

Thus,

$$U\rfloor d(L\mu + J) \equiv \eta^\alpha(\Lambda_\alpha \mu - d\Lambda^i_\alpha \wedge \mu_i) \mod \mathfrak{H}[A^\alpha_{ij}] . \tag{9.16}$$

Since

$$d(U\rfloor(L\mu + J)) \equiv d(n^j L\mu_j + \eta^\beta \Lambda^j_\beta \mu_j) \mod \mathfrak{H}[A^\alpha_{ij}] , \tag{9.17}$$

we have

$$\pounds_U(L\mu + J) \equiv \eta^\alpha E_\alpha + d(n^j L + \eta^\beta \Lambda^j_\beta) \wedge \mu_j \mod \mathfrak{H}[A^\alpha_{ij}] . \tag{9.18}$$

This expression can be simplified to

$$\pounds_U(L\mu + J) \equiv \eta^\alpha E_\alpha + d(\xi^j L) \wedge \mu_j \mod \mathfrak{H}[A^\alpha_{ij}] , \tag{9.19}$$

if we define

$$\xi^i = n^i + \tfrac{1}{L}\eta^\beta \Lambda^i_\beta \tag{9.20}$$

provided $L \neq 0$. At the same time $U \in ISO[A^\alpha_{ij}]$ can be resolved with respect to a new basis $(V_i, T_\alpha, \partial^i_\alpha)$ as

$$U = \xi^i V_i + \eta^\alpha T_\alpha + V_i \langle \eta^\alpha \rangle \partial^i_\alpha , \tag{9.21}$$

where

(9.22) $$T_\alpha = \partial_\alpha - \tfrac{1}{L}\Lambda^i_\alpha V_i \ .$$

These considerations serve to establish the following result.

Theorem 9.1 *If* $U \in \mathrm{ISO}[A^\alpha_{ij}]$ *is given by*

(9.23) $$U = \xi^i V_i + \eta^\alpha T_\alpha + V_i<\eta^\alpha>\partial^i_\alpha \ ,$$

where $\{\xi^i, \eta^\alpha \mid 1 \le i \le n, 1 \le \alpha \le N\}$ *belong to* $\Lambda^0(K)$, *the* η*'s satisfy*

(9.24) $$V_j V_i<\eta^\alpha> = (\eta^\beta \partial_\beta + V_k<\eta^\beta>\partial^k_\beta)<A^\alpha_{ij}> \ ,$$

and $L \neq 0$, *then*

(9.25) $$\pounds_U(L\mu) \equiv \eta^\alpha E_\alpha + d(\xi^j L)\wedge\mu_j \mod \mathcal{H}[A^\alpha_{ij}] \ .$$

If the solution map Ψ of the horizontal ideal $\mathcal{H}[A^\alpha_{ij}]$ satisfies the Euler-Lagrange equations, i.e.,

(9.26) $$\Psi^* E_\alpha = 0 \ , \quad 1 \le \alpha \le N \ ,$$

and if transport is by a $U \in \mathrm{ISO}[A^\alpha_{ij}]$ such that $\xi^j = 0$, $1 \le j \le n$, then (9.25) shows that

(9.27) $$\Psi^* \pounds_U(L\mu) = 0 \ ,$$

and hence

(9.28) $$\Delta_U(s)A[\Psi] = 0 \ .$$

We therefore have

(9.29) $$A[\Psi_U(s)] = A[\Psi]$$

by (9.7). Therefore, if $\{A^\alpha_{ij}\}$ is such that every leaf of the foliation generated by $\mathcal{H}[A^\alpha_{ij}] \in \mathfrak{H}(K)$ gives rise to a solution of the Euler-Lagrange equations, i.e., $\mathcal{H}[A^\alpha_{ij}]$

$\in \mathfrak{H}_S(K)$, then transport of such a solution by an isovector preserves the value of the action integral provided $\xi^i = 0$, $1 \leq i \leq n$. The reader should note, however, that this transport process will, in general, alter the region $\mathfrak{D}$ of integration.

Some insight concerning this situation can be gained by resolving the isovector U with respect to the natural basis $(\partial_i, \partial_\alpha, \partial^i_\alpha)$ for $T(K)$,

$$U = u^i\partial_i + u^\alpha\partial_\alpha + (u^i A^\alpha_{ij} + V_j\langle u^\alpha - y^\alpha_i u^i\rangle)\partial^j_\alpha . \tag{9.30}$$

The conditions $\xi^j = 0$ then become

$$\Lambda^j_\alpha u^\alpha = \mathbf{H}^j_i u^i , \tag{9.31}$$

where $(\mathbf{H}^j_i)$ is the *Hamiltonian complex* given by

$$\mathbf{H}^j_i = y^\alpha_i \Lambda^j_\alpha - L\delta^j_i . \tag{9.32}$$

Thus $\xi^j = 0$ are the well-known conditions of *transversality* in the calculus of variations. We therefore have the following results.

Corollary 9.1 *If Ψ is a solution map of the Euler-Lagrange ideal and $L \neq 0$, then the action integral is constant in value under transport by transversal isovectors U of the horizontal ideal of a special system $\mathfrak{H}[A^\alpha_{ij}] \in \mathfrak{H}_S(K)$, i.e.,*

$$U = \eta^\alpha T_\alpha + V_i\langle\eta^\alpha\rangle\partial^i_\alpha \tag{9.33}$$

where

$$V_jV_i\langle\eta^\alpha\rangle = (\eta^\beta\partial_\beta + V_k\langle\eta^\beta\rangle\partial^k_\beta)\langle A^\alpha_{ij}\rangle . \tag{9.34}$$

Remark. Note that if $\mathfrak{H}[A^\alpha_{ij}] \in \mathfrak{H}(K) - \mathfrak{H}_S(K)$, then transport of a solution map Ψ by a transversal isovector will not result in solution maps for all values of s in a sufficiently small neighborhood of $s = 0$.

Remark. The basis $(V_i, T_\alpha, \partial^i_\alpha)$ of $T(K)$ and the transversality conditions both arise in the complete figure construction of Hamilton-Jacobi theory in the calculus of variations of multiple integrals [1, 2, 3]. We do not make use of that

construction, but it can be seen that the two approaches have several features in common.

Theorem 9.1 is also useful when considering the infinitesimal variation, which is defined by

$$\delta_U A[\Psi] = \lim_{s \to 0} \left\{ \frac{\Delta_U(s) A[\Psi]}{s} \right\} . \tag{9.35}$$

In view of (9.9), the infinitesimal variation can be expressed as

$$\delta_U A[\Psi] = \int_{\mathfrak{D}} \Psi^* \pounds_U(L\mu) . \tag{9.36}$$

Theorem 9.1 can then be used to establish the following result.

Corollary 9.2 *The infinitesimal variation of the action integral that is generated by the isovector* $U \in ISO[A^{\alpha}_{ij}]$, *i.e.*,

$$U = \xi^i V_i + \eta^\alpha T_\alpha + V_i \langle \eta^\alpha \rangle \partial^i_\alpha \tag{9.37}$$

where ξ^i, η^α *belong to* $\wedge^0(K)$,

$$V_j V_i \langle \eta^\alpha \rangle = (\eta^\beta \partial_\beta + V_k \langle \eta^\beta \rangle \partial^k_\beta) \langle A^\alpha_{ij} \rangle , \tag{9.38}$$

and $L \neq 0$, *is given by*

$$\delta_U A[\Psi] = \int_{\mathfrak{D}} \Psi^* (\eta^\alpha E_\alpha) + \int_{\partial\mathfrak{D}} \Psi^* (\xi^j L\, \mu_j) . \tag{9.39}$$

The simplest problem in the calculus of variations is where $\mathfrak{D}$ is a given fixed region and the dependent variables are specified on the boundary, $\partial\mathfrak{D}$, of that region. The first condition restricts the variations (transport by isovectors) to those that do not alter the independent variables, i.e., $n^i = 0$, while the second condition can be expressed as the Dirichlet data conditions

$$\Psi^* \eta^\alpha = 0 \quad \text{on } \partial\mathfrak{D} . \tag{9.40}$$

These conditions imply that $\xi^i = 0$, $1 \leq i \leq n$, on $\partial\mathfrak{D}$ and the boundary integral in (9.39) vanishes. Thus, we have

(9.41) $$\delta_U A[\Psi] = \int_{\mathfrak{D}} \Psi^*(\eta^\alpha E_\alpha) .$$

This same result also obtains under homogeneous Neumann data conditions

(9.42) $$\Psi^*(\Lambda^j_\beta \mu_j) = 0 \quad \text{on} \quad \partial\mathfrak{D}$$

(see (9.20)). In both situations, satisfaction of the Euler-Lagrange equations is a sufficient condition for the infinitesimal variation to vanish. We note, however, that the η's must be solutions of the system (9.38), and hence (9.41) shows that we can not obtain necessary conditions for the infinitesimal variation to vanish by means of these methods. Necessity of satisfaction of the Euler-Lagrange equations is easily shown, however, by considering isovectors of the contact ideal [3, 4].

One of the authors (R.J.K.) would like to express his gratitude to Lehigh University for the hospotality extended during his sabbatical year.

REFERENCES

1. Carathéodory, C., *Über die Variationsrechnung bei mehrfachen Integralen*, Acta Math. (Szeged) 4 (1929), 193.
2. Rund, H., *The Hamilton-Jacobi theory of the geodesic fields of Carathéodory in the calculus of variations of multiple integrals*, The Greek Mathematical Society, C. Carathéodory Symposium, September 3-7, 1973.
3. Edelen, D. G. B., Isovector Methods for Equations of Balance, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands 1980.
4. Edelen, D. G. B., Applied Exterior Calculus, Wiley-Interscience, New York, 1985.
5. Cartan, E., Les Systèmes Différentiels Extérieurs et leurs Applications Géométriques, Hermann, Paris, 1945.

Dominic G. B. Edelen
Lehigh University
Bethlehem, PA 18015
U.S.A.

R. J. McKellar
University of New Brunswick
Fredericton, N.B.
Canada E3B 5A3

Constantin Carathéodory: An International Tribute (pp. 278–288)
edited by Th. M. Rassias

ON SOME UNIVALENT INTEGRAL OPERATORS

Otto Fekete

Let A denote the set of functions $f(z)=z+a_2z^2+\ldots$ that are analytic in the unit disc, and let S denote the subset of A consisting of univalent functions. With some suitable conditions on the constants α and c and on $f,g,h\in A$, the author shows that the function F given by the integral operator

$$F(z)=\left[\frac{\alpha+c}{z^c}\int_0^z f^\alpha(t)g^{c-1}(t)h'(t)dt\right]^{1/\alpha}$$

is stralike in U or in other subclasses of S.

1. Introduction

In the theory of univalent functions, the class of functions with positive real part, known also as Carathéodory functions in honour of Constantin Carathéodory, who studied first the coefficients of this class ([1]), plays an important role. Let $\mathcal{P}$ denote the class defined by

$$\mathcal{P}=\{p|p(z)=1+c_1z+\ldots,\operatorname{Re}p(z)>0,z\in U\}, \tag{1}$$

where $U=\{z\big||z|<1\}$.

Let A denote the set of functions $f(z)=z+a_2z^2+\ldots$ that are analytic in U and let S denote the subset of A consisting of univalent functions. Let $CV(\alpha),ST(\alpha)$ and CC, $0\le\alpha<1$, denote respectively the classes of convex functions of order α, starlike functions of order α and close-to-convex functions, i.e.

$$CV(\alpha)=\{f\in A|\operatorname{Re}(1+zf''(z)/f'(z))>\alpha,z\in U\},$$
$$ST(\alpha)=\{f\in A|\operatorname{Re}(zf'(z)/f(z))>\alpha,z\in U\},$$
$$CC=\{f\in A|\exists g\in CV(0),\operatorname{Re}(f'(z)/g'(z))>0,z\in U\}.$$

In [5], S. S. Miller and P. T. Mocanu investigate the integral operator $J(f)$ defined by

$$J(f)(z) = \left[\frac{\beta+\gamma}{z^{\gamma}\Phi(z)}\int_0^z f^{\alpha}(t)\phi(t)t^{\delta-1}dt\right]^{1/\beta} \tag{2}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{C}$ and ϕ, Φ are analytic functions which map subsets of A into $ST(0)$, and extend and sharpen many of the previously obtained results.

In [7], T. N. Shanmugam considered the integral operator $I(f)$ defined by

$$F(z) = I(f)(z) = \left[\frac{\alpha+c}{z^c}\int_0^z f^{\alpha}(t)g^{c-1}(t)h'(t)dt\right]^{1/\alpha} = z + \dots \tag{3}$$

where $\alpha, c \geq 0, \alpha + 1 \geq c, f, g, h \in CV(0)$ and prove that $I(f) \in ST(0)$.

In this paper we extend the integral operator (3) for $\alpha, c \in \mathbb{C}$ and improve some results from [7], using te differential subordination technique and the method from [5].

2. Preliminaries

If f and g are analytic in U, then the function f is subordinate to g, written $f \prec g$, if g is univalent, $f(0) = g(0)$ and $f(U) \subset g(U)$. We need to introduce a special mapping from U onto a slit domain, called "open door" function.

Let $a \in \mathbb{C}, \operatorname{Re} a > 0$ and let

$$N = [|a|(1 + 2\operatorname{Re} a)^{1/2} + \operatorname{Im} a]/\operatorname{Re} a\,. \tag{4}$$

If H is the univalent function $H(z) = 2Nz/(1 - z^2)$ and $b = H^{-1}(a)$ then we define the "open door" function Q_a as follows

$$Q_a(z) = H[(z + b)/(1 + \bar{b}z)]\,,\ z \in U\,. \tag{5}$$

Q_a is univalent, $Q_a(0) = a$ and $Q_a(U) = H(U)$ is the complex plane slit along the half-lines $\operatorname{Re} w = 0, \operatorname{Im} w \geq N$ and $\operatorname{Re} w = 0, \operatorname{Im} w \leq -N$. For $a = 1$ we have

$$Q_1(z) = \frac{1+z}{1-z} + \frac{2z}{1-z^2} \tag{6}$$

Lemma 1. ([6]) Let Q_a be the function defined above and let P be an analytic function in U satisfying $P \prec Q_a$. If p is analytic in $U, p(0) = 1/c$ and p satisfies the differential equation

$$zp'(z) + P(z)p(z) = 1\,, \tag{7}$$

then $\operatorname{Re} p(z) > 0$ in U.

Lemma 2 ([4]) Let $\alpha, c \in \mathbb{C}$ with $\operatorname{Re} \alpha > 0$ and

$$-\operatorname{Re} c/\operatorname{Re}\alpha = \rho_0 \leqq \rho < 1\,. \tag{8}$$

If $G \in A$ and

$$\operatorname{Re}\left[\alpha z G'(z)/G(z)\right] \geqq \rho \operatorname{Re}\alpha$$

then the function

$$F(z) = \left[(\alpha + c)z^{-c}\int_0^z G^{\alpha}(t)t^{c-1}dt\right]^{1/\alpha} = z + \ldots \tag{9}$$

is analytic in U and satisfies

$$\operatorname{Re}\left[\alpha z F'(z)/F(z)\right] \geqq w(\rho)\operatorname{Re}\alpha = \operatorname{Inf}\{H(z) | z \in U\} \tag{10}$$

where

$$H(z) = \frac{(1-z)^{2(\rho-1)\operatorname{Re}\alpha}}{\int_0^1 t^{\alpha+c-1}(1+tz)^{2(\rho-1)\operatorname{Re}\alpha}\,dt} - c\,.$$

This result is sharp and $\rho \leqq w(\rho)$.

Remark 1. In the special case when α and c are real and $\rho \geqq (\alpha - c - 1)/2\alpha$, the value of $w(\rho)$ as given in (10) can be simplified. In particular, if (8) is replaced by

$$\operatorname{Max}\{(\alpha - c - 1)/2\alpha\,;\, -c/\alpha\} = \rho_0 \leqq \rho < 1 \tag{8'}$$

then (10) can be replaced by

$$\operatorname{Re}\frac{zF'(z)}{F(z)} \geqq w(\rho) = \frac{1}{\alpha}\left[\frac{(\alpha+c)2^{-2\alpha(1-\rho)}}{{}_2F_1(2\alpha(1-\rho), \alpha+c; \alpha+c+1; -1)} - c\right] \tag{10'}$$

where ${}_2F_1$ is the hypergeometric function. As in Lemma 2, $w(\rho) \geq \rho$ and (10′) is sharp.

Lemma 3. ([2]),[3]) Let $\psi : \mathbb{C}^2 \times U \to \mathbb{C}$ satisfy the condition

$$\operatorname{Re} \psi(ix, y; z) \leqq 0 \text{ for all } x, y \in R \text{ with } y \leqq -\frac{1}{2}(1+x^2) \tag{11}$$

and all $z \in U$. If p is analytic in $U, p(0) = 1$ and $\operatorname{Re}\ \psi(p(z), zp'(z); z) > 0, z \in U$, then $\operatorname{Re}\ p(z) > 0$ in U.

3. The Real Case

First we will prove that there exists a regular function F satisfying (3). Let

$$H(z) = \left(\frac{f(z)}{z}\right)^{\alpha} \left(\frac{g(z)}{z}\right)^{c-1} h'(z) = 1 + d_1 z + \dots$$

and choose the branches which equal 1 when $z = 0$. For

$$G(z) = f^{\alpha}(z) g^{c-1}(z) h'(z) = z^{\alpha+c-1} H(z)$$

we have

$$K(z) = \frac{\alpha + c}{z^{\alpha+c}} \int_0^z G(t)dt = 1 + \frac{\alpha + c}{\alpha + c + 1} d_1 z + \dots,$$

hence K is well defined and regular in U. Now let

$$F(z) = [z^{\alpha} K(z)]^{1/\alpha} = z[K(z)]^{1/\alpha}$$

where we choose the branch of $[K(z)]^{1/\alpha}$ which equals 1 when $z = 0$. Then F will be regular in U with $F(0) = 0$ and $F'(0) = 1$ and F will satisfy (3).

From (3) we see that

$$z^c F^{\alpha}(z) = (\alpha + c) \int_0^z f^{\alpha}(t) g^{c-1}(t) h'(t) dt$$

and differentiating we obtain

$$\alpha F^{\alpha-1}(z) F'(z) z^c + c z^{c-1} F^{\alpha}(z) = (\alpha + c) f^{\alpha}(z) g^{c-1}(z) h'(z). \tag{12}$$

For

$$p(z) = \frac{zF'(z)}{F(z)}, \tag{13}$$

differentiating (12) logarithmically we obtain

$$p(z) + \frac{zp'(z)}{\alpha p(z) + c} = \frac{zf'(z)}{f(z)} + \frac{c-1}{\alpha}\frac{zg'(z)}{g(z)} + \frac{1}{\alpha}\left(1 + \frac{zh''(z)}{h'(z)}\right) - \frac{c}{\alpha}. \tag{14}$$

Theorem 1. Let $\alpha > 0, c \geqq 0$ and $f \in ST(0)$. If

$$(c-1)\mathrm{Re}\,\frac{zg'(z)}{g(z)} + \mathrm{Re}\,\left(1 + \frac{zh''(z)}{h'(z)}\right) \geqq \begin{cases} c - \frac{\alpha}{2c}, & \text{for } \alpha \leqq c \\ c - \frac{c}{2\alpha}, & \text{for } \alpha > c \end{cases} \tag{15}$$

then F defined by (3) is in $ST(0)$.

Proof. From (14) we obtain

$$p(z) + \frac{zp'(z)}{\alpha p(z) + c} - \phi_1(z) = \frac{zf'(z)}{f(z)} \tag{16}$$

where

$$\phi_1(z) = \frac{c-1}{\alpha}P(z) + \frac{1}{\alpha}Q(z) - \frac{c}{\alpha}, \tag{17}$$

$$P(z) = \frac{zg'(z)}{g(z)} \text{ and } Q(z) = 1 + \frac{zh''(z)}{h'(z)}. \tag{18}$$

Let $\psi_1 : \mathbb{C}^2 \times U \to \mathbb{C}$ be defined by

$$\psi_1(u, v; z) = u + \frac{v}{\alpha u + c} - \phi_1(z).$$

We will show that ψ_1 satisfies the conditions of Lemma 3. From $f \in ST(0)$ and (16) we have $\psi_1(p(z), zp'(z); z) > 0$ for $z \in U$. We have

$$\begin{aligned} \mathrm{Re}\,\psi_1(ix, y; z) &= \mathrm{Re}(ix + \frac{y}{\alpha ix + c} - \phi_1(z)) \\ &= \frac{-[c + 2\alpha^2\ \mathrm{Re}\ \phi_1(z)]x^2 - [c + 2c^2\ \mathrm{Re}\ \phi_1(z)]}{2(c^2 + \alpha^2 x^2)} \end{aligned}$$

if $y \leqq -\frac{1}{2}(1 + x^2)$. Condition (11) is verified if

$$\begin{cases} c + 2\alpha^2\ \mathrm{Re}\ \phi_1(z) \geqq 0 \\ c + 2c^2\ \mathrm{Re}\ \phi_1(z) \geqq 0. \end{cases}$$

A simple calculation shows that the previous condition is equivalent to (11) and applying Lemma 3 we conclude that $F \in ST(0)$.

Theorem 2. Let $\alpha > 0, c \geq 0$ and $f \in CV(0)$. If

$$(c-1)\mathrm{Re}\,\frac{zg'(z)}{g(z)} + \mathrm{Re}\left(1+\frac{zh''(z)}{h'(z)}\right) \geq \begin{cases} c - \frac{\alpha}{2} - \frac{\alpha}{2c} & \text{for } \alpha \leq c \\ c - \frac{\alpha}{2} - \frac{c}{2\alpha} & \text{for } \alpha > c \end{cases} \tag{19}$$

then F defined by (3) is in $ST(0)$.

Proof. From (14) we have

$$p(z) + \frac{zp'(z)}{\alpha p(z) + c} - \phi_2(z) = \frac{zf'(z)}{f(z)} - \frac{1}{2}$$

where

$$\phi_2(z) = \frac{c-1}{\alpha} P(z) + \frac{1}{\alpha} Q(z) - \frac{1}{2}$$

and P and Q are given by (18). Using the fact that convex functions are starlike functions of order $1/2$ we have

$$\psi_2(p(z), zp'(z); z) > 0 \text{ in } U$$

where

$\psi_2 : \mathbb{C}^2 \times U \to \mathbb{C}$ is defined by

$$\psi_2(u, v; z) = u + \frac{v}{\alpha u + c} - \phi_2(z).$$

Condition (11) in Lemma 3 is verified if

$$\begin{cases} c + 2\alpha^2 \ \mathrm{Re}\ \phi_2(z) \geq 0 \\ c + 2c^2 \ \mathrm{Re}\ \phi_2(z) \geq 0 \end{cases}$$

Like in the proof of Theorem 1, a simple calculation shows that the previous condition is equivalent to (11) and from Lemma 3 we conclude that $F \in ST(0)$.

Remark 2. For $c \geq 0, \alpha > c+1$ and $f, g, h \in CV(0)$ we obtain Theorem 1 from [7].

Up to this point we have considered the function F given by (3) as a function defined by an integral operator $I(f)$. By using the same method,

we can consider the function F given by (3) as a function defined by an integral operator $I(g)$ or $I(h)$. In this case we can prove the following two theorems in a similar way as Theorem 1 and Theorem 2.

Theorem 3. Let $\alpha > 0, c \geqq 0$ and $h \in CV(0)$. If

$$\alpha \operatorname{Re} \frac{zf'(z)}{f(z)} + (c-1) \operatorname{Re} \frac{zg'(z)}{g(z)} \geqq \begin{cases} c - \frac{\alpha}{2c}, & \text{for } \alpha \leqq c \\ c - \frac{c}{2\alpha}, & \text{for } \alpha > c \end{cases} \tag{20}$$

then F defined by (3) is in $ST(0)$.

Theorem 4. Let $\alpha > 0, c > 1, g \in ST(0)$. If

$$\alpha \operatorname{Re} \frac{zf'(z)}{f(z)} + \operatorname{Re}\left(1 + \frac{zh''(z)}{h'(z)}\right) \geqq \begin{cases} c - \frac{c}{2\alpha} & \text{for } \alpha > c \\ c - \frac{\alpha}{2c}, & \text{for } \alpha \leqq c \end{cases} \tag{21}$$

then F defined by (3) is in $ST(0)$.

4. The Complex Case

We again consider the operator $I(f)$ defined by (3), but now allow α and c to be complex and f to be in more general subsets of A.

Theorem 5. Let $\alpha, c \in \mathbb{C}, \operatorname{Re}(\alpha + c) > 0, \alpha \neq 0$ and $f, g, h \in A$. If f satisfies

$$\alpha \frac{zf'(z)}{f(z)} + (c-1)\frac{zg'(z)}{g(z)} + \frac{zh''(z)}{h'(z)} + 1 \prec Q_{\alpha+c}(z) \tag{22}$$

where Q_a is defined by (5) and F is defined by (3) then $F \in A, F(z)/z \neq 0$ and

$$\operatorname{Re}\left[\alpha \frac{zF'(z)}{F(z)} + c\right] > 0 \text{ for } z \in U\,. \tag{23}$$

Proof. From (22) we see that $F(z)/z \neq 0$ in U. Let

$$G(z) = \frac{1}{zg^{c-1}(z)h'(z)f^{\alpha}(z)} \int_0^z f^{\alpha}(t) g^{c-1}(t) h'(t) dt \tag{24}$$

and

$$P(z) = (c-1)\frac{zg'(z)}{g(z)} + \alpha\frac{zf'(z)}{f(z)} + \frac{zh''(z)}{h'(z)} + 1\,. \tag{25}$$

We obtain

$$G(z) = \frac{1}{(\frac{g(z)}{z})^{c-1}(\frac{f(z)}{z})^{\alpha}z^{\alpha+c}}\int_0^z (\frac{f(t)}{t})^{\alpha}(\frac{g(t)}{t})^{c-1}h'(t)\cdot t^{\alpha+c-1}dt$$
$$= \frac{1}{\alpha+c} + \dots$$

analytic in U with $G(0) = \frac{1}{\alpha+c}$. By differentiating (24) we deduce that G satisfies the differential equation (7). By using (22) we see that G satisfies Lemma 1 and so we have $\operatorname{Re} G(z) > 0$ in U, hence $G(z) \neq 0$ in U. By using (24) and (3) we obtain

$$F(z) = \left[\frac{\alpha+c}{z^c}G(z)zg^{c-1}(z)h'(z)f^{\alpha}(z)\right]^{1/\alpha}$$
$$= z[(\alpha+c)G(z)h'(z)]^{1/\alpha}[g(z)/z]^{(c-1)/\alpha}[f(z)/z] = z + \dots,$$

and differentiating logarithmically, a simple computation shows that

$$\frac{zF'(z)}{F(z)} = \frac{1}{\alpha}\left[\frac{1}{G(z)} - c\right]$$

hence

$$\operatorname{Re}\left[\alpha\frac{zF'(z)}{F(z)} + c\right] > 0 \text{ for } z \in U\,,$$

F is analytic in U and $F(z)/z \neq 0$.

Theorem 6. Let $\alpha, c \in \mathbb{C}, \operatorname{Re}\alpha > 0, \operatorname{Re}(\alpha+c) > 0, f, g, h \in A$ and

$$\operatorname{Re}\left[\alpha\frac{zf'(z)}{f(z)}\right] \geq \sigma \operatorname{Re}\alpha \tag{26}$$

where $\sigma < 1$. If there exists a real number $\rho < 1$ such that

$$\operatorname{Re}\left[\sigma\alpha + \frac{zh''(z)}{h'(z)} + 1 + (c-1)\frac{zg'(z)}{g(z)}\right] \geq \operatorname{Re}(\rho\alpha + c) \geq 0 \tag{27}$$

for $z \in U$, then the function F defined by (3) is analytic in $U, f(z)/z \neq 0$ and

$$\operatorname{Re}\left[\alpha\frac{zF'(z)}{F(z)}\right] \geq w(\rho)\operatorname{Re}\alpha \tag{28}$$

where $w(\rho)$ is given by (10) or (10').

Proof. Conditions (26) and (27) imply

$$\alpha\frac{zf'(z)}{f(z)}+\frac{zh''(z)}{h'(z)}+1+(c-1)\frac{zg'(z)}{g(z)}\prec Q_{\alpha+c},$$

hence by Theorem 1, $F\in A$ and $F(z)/z\neq 0$. From (26) we deduce that $f(z)/z\neq 0$ and hence the function

$$G(z)=f(z)\left[\frac{h'(z)g^{c-1}(z)}{z^{c-1}}\right]^{1/\alpha}$$

is in A. From (26) and (27) we obtain

$$\begin{aligned}\operatorname{Re}\left[\alpha\frac{zG'(z)}{G(z)}\right]&=\operatorname{Re}\left\{\alpha\frac{zf'(z)}{f(z)}+\frac{zh''(z)}{h'(z)}+1+(c-1)\frac{zg'(z)}{g(z)}-c\right\}\\&\geq\rho\operatorname{Re}\alpha\end{aligned}$$

and from (27) we see that ρ satisfies (8). Hence G satisfies the conditions of Lemma 2 and from

$$F(z)=[(\alpha+c)z^{-c}\int_0^z G^{\alpha}(t)t^{c-1}dt]^{1/\alpha}=z+\dots$$

we obtain F analytic in U and $\operatorname{Re}\left[\alpha\frac{zF'(z)}{F(z)}\right]\geq w(\rho)\operatorname{Re}\alpha$.

Theorem 7. Let $\alpha,c\in\mathbb{R},\alpha>0,\alpha+c>0,\rho_0$ given in (8) or (8') and suppose that there exists $\rho\in[\rho_0,1]$ such that $0\leq w(\rho)$, where w is given by (10) or (10'). If $f\in ST(0),g,h\in A$ and

$$\operatorname{Re}\left\{\frac{zh''(z)}{h'(z)}+1+(c-1)\frac{zg'(z)}{g(z)}-c\right\}\geq\rho\alpha \tag{29}$$

then $F\in ST(0)$.

Proof. Condition (29) implies (27) and from $\alpha\in\mathbb{R}$ we have

$$\operatorname{Re}\alpha\frac{zf'(z)}{f(z)}=\alpha\operatorname{Re}\frac{zf'(z)}{f(z)}\geq 0\quad\text{for }\sigma=0.$$

Hence by Theorem 6, $F = I(f) \in A$ and $\operatorname{Re} \frac{zF'(z)}{F(z)} \geqq w(\rho) \geqq \rho > 0$ i.e., $F \in ST(0)$.

If we take ρ in Theorem 3 to be the critical values $\rho = -c/\alpha, \rho = (\alpha - c - 1)/2\alpha$ and $\rho = 0$, we obtain the following three corollaries.

Corollary 1. Let $\alpha > 0, \alpha + c < 1, \alpha + c > 0, f \in ST(0)$ and $g, h \in A$. If

$$\operatorname{Re}\left\{\frac{zh''(z)}{h'(z)} + 1 + (c-1)\frac{zg'(z)}{g(z)}\right\} \geqq 0 \tag{30}$$

and $w(-c/\alpha) \geqq 0$ then $F \in ST(w(-c/\alpha))$.

Corollary 2. Let $\alpha > 0, \alpha + c \geqq 1, \alpha - c \geqq 0, f \in ST(0)$, and $g, h \in A$. If

$$\operatorname{Re}\left\{\frac{zh''(z)}{h'(z)} + 1 + (c-1)\frac{zg'(z)}{g(z)}\right\} \geqq \frac{\alpha + c - 1}{2} \tag{31}$$

then $F \in ST\left(\frac{\alpha - c}{2}\right)$.

Corollary 3. Let $\alpha > 0, \alpha + c > 0, f \in ST(0)$ and $g, h \in$A. If

$$\operatorname{Re}\left\{\frac{zh''(z)}{h'(z)} + 1 + (c-1)\frac{zg'(z)}{g(z)}\right\} \geqq c$$

then $F \in ST(w(0))$.

Remark 3. For particular values of α and c real, the results of Sec. 3 are not included in the results of Sec. 4.

References

1. C. Carathédory, *Uber den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen*, Math. Ann. **64** (1907), 95–115.
2. S. S. Miller, P. T. Mocanu, *Differential subordination and univalent functions*, Michigan Math. J. **28** (1981), 157–171.
3. S. S. Miller, P. T. Mocanu, *General second order differential inequalities in the complex plane*, Seminar on Geometric Function Theory, Preprint nr. 4, 1982, Babes-Bolyai University Cluj, pp. 96–114.
4. S. S. Miller, P. T. Mocanu, *On a class of spirallike integral operators*, Rev. Roum. Math. Pures et Appl. **31** (1986), 225–230.

5. S. S. Miller, P. T. Mocanu, *Classes of univalent integral operator*, to appear.
6. P. T. Mocanu, *Some integral operators and starlike functions*, Rev. Roum. Math. Pures et Appl. **31** (1986), 231–235.
7. T. N. Shanmugam, *On some integral operators*, Bull. Calcutta Math. Soc. **79** (1987), 71–74.

Otto Fekete
Hauptstrasse 47
7426 Pfronstetten
West Germany

Constantin Carathéodory: An International Tribute (pp. 289–311)
edited by Th. M. Rassias

THE VARIATIONAL STRUCTURE OF GENERAL RELATIVITY

Marco Ferraris & Mauro Francaviglia

Abstract

The variational structure of General Relativity is revisited. After discussing conditions which ensure that second-order Lagrangians linear in the second-order derivatives generate second-order field equations and first-order Poincaré-Cartan forms, it is shown that the Hilbert Lagrangian of General Relativity satisfy these conditions. An equivalent first-order covariant Lagrangian formalism is shortly discussed.

1. Introduction

It is a widely accepted axiom in Mathematical Physics that field equations of physical field theories should follow from some variational principle. As is well known also the gravitational equations of General Relativity can be obtained as the Euler-Lagrange equations of some suitable Lagrangian (this was first proved by Hilbert [23] and slightly later by Einstein but in a more general physical situation [3]).

However, in spite of the fact that the variational character of Einstein's gravitational equations has been known for more than seventy years, it is our opinion that the "variational structures" associated to General Relativity (together with its generalizations) still deserve some attention, in order to better clarify a number of interesting features of Einstein's theory.

It is well known, in fact, that Einstein equations follow from a variational principle involving a second-order metric Lagrangian (the Hilbert Lagrangian $\mathcal{L}=R\sqrt{g}$), i.e., a Lagrangian containing a (Lorentzian) metric g together with its first and second-order derivatives. One should expect thence to obtain fourth-order field equations in g, while Einstein equations are in fact second-order ones, as if the Lagrangian which generates them be a first-order one. According to Calculus of Variations this indicates that the Hilbert Lagrangian is somehow "degenerate", its dependence on second-order derivatives being in fact linear and hidden in a full space-time divergence.

It has also been known for a long time that not only in respect of the order of field equations, but in a much wider range of problems, General Relativity behaves essentially as a first-order theory. This emerges clearly, for example, in all the various attempts to a consistent Hamiltonian formulation of its field equations, as well as in a variety of investigations about the conservation laws and the notions of energy and momentum in General Relativity (see [6, 7]).

This essentially first-order behaviour can be easily explained, as it was soon realized by Einstein himself, by reducing the Hilbert Lagrangian $R\sqrt{g}$ to an equivalent "first-order Lagrangian", by just dropping a full divergence containing the second-order derivatives of the metric field [4]. (Here two Lagrangians are said to be "equivalent" if they generate the same field equations). This procedure, however, is manifestly conflicting with one of the cornerstones of Einstein's construction, namely "full covariance", as the ensuing "first-order Lagrangian" cannot be properly considered as being a "Lagrangian" (at least in a fully meaningful sense) since it is not in fact a (covariant) scalar density! Although this new object, which has no satisfactory transformation properties under arbitrary changes of coordinates, still generates the correct (and covariant) field equations, its non-covariance is reflected in all kind of problems arising from its use. As an example, when using it (or some similar object) to generate conservation laws one necessarily generates pseudo-tensorial quantities (e.g., the pseudo-tensor of Einstein [4] or the so-called "Landau-Lifchitz complex" [27]) which do not obey satisfactory covariance properties and impose ad-hoc prescriptions or corrections to restore the necessary invariance under local diffeomorphisms of space-time. Analogously, when attempting a Hamiltonian formulation of General Relativity à la ADM [1], by using other first-order and non-covariant equivalent Lagrangians, full covariance breaks down and has to be restored a posteriori "by hands", e.g., by adding appropriate boundary terms to the action (see [22]).

Let us now mention that all variational principles can be conveniently formulated by means of the so-called "Poincaré-Cartan formalism", whereby an appropriate differential form $\Theta(\mathcal{L})$, the Poincaré-Cartan form, replaces the given Lagrangian $\mathcal{L}$. The possibility of replacing a Lagrangian by a Poincaré-Cartan form was long ago realized in the context of Classical Mechanics, and also guessed for the Lagrangians of "higher-order mechanics", while the extension of the method to field theory is more recent. A satisfactory geometrical formulation for first-order field theories appeared in 1968-1973 ([**15, 16, 19**]), while the general case of higher-order field theories was solved only in this decade, through the contributions of several authors (see, e.g., [**5, 17, 20, 24, 25**] and Refs. quoted therein). It turns out that for any given Lagrangian $\mathcal{L}$ there exists a whole family of Poincaré-Cartan forms, parametrized for example using an arbitrary linear connection in the base manifold M of the given Lagrangian field theory $(B,M,\pi;\mathcal{L})$; this family reduces to a single and unique Poincaré-Cartan form if the base M is one-dimensional (mechanics) or in the first-order case. However, in the other physically relevant situation, namely for second-order field theories, one can show the existence of a "canonical" Poincaré-Cartan form, so that it is in fact physically meaningful to speak of "the" Poincaré-Cartan form of any physical theory.

Now, the Poincaré-Cartan form $\Theta(\mathcal{L})$ turns out to be the most appropriate object to discuss a number of relevant notions for field theory, such as the notion of "regularity" of the Lagrangian (see e.g. [**17, 26, 29**]), the notion of phase space (see e.g. [**21**]) and the notion of conservation laws (see e.g. [**8, 11**]). In particular, if a Lagrangian $\mathcal{L}$ depends on derivatives of order k, its Euler-Lagrange equations should in principle be of order 2k, while the Poincaré-Cartan form should depend on derivatives of order 2k-1. However, if the Lagrangian is "degenerate", field equations will be of lower order and the degree of degeneracy of the Lagrangian will be reflected directly in the Poincaré-Cartan form, which will in this case depend on derivatives of a lower order s, with $k \leq s < 2k-1$.

From the above remarks, we see that the Poincaré-Cartan formalism is well suited to discuss the degree of degeneracy and the variational characterization of conservation laws for General Relativity. To our knowledge, the Poincaré-Cartan form for the Hilbert Lagrangian was first deduced by Szczyrba [30] by relying on an ad-hoc procedure, while a derivation from the general formula of second-order field theories may be found in [6]. As we said, a second-order theory should produce a third-order

Poincaré-Cartan form (and fourth-order field equations), while the Poincaré-Cartan form of the Hilbert Lagrangian turns out to be of first-order only (with second-order field equations). This fact further supports the idea that General Relativity behaves as being obtainable from a first-order variational principle.

Nevertheless, the first-order Lagrangian of Einstein suffers the aforementioned drawbacks about covariance and, moreover, it is classically known that no scalar density can be constructed out by just using a metric g together with its first derivatives, except of course the trivial one $\sqrt{g}$, which does not in fact depend on derivatives of g! This apparent contradiction has however a rather simple solution: one can show in fact the existence of a whole class of first-order global and covariant Lagrangians which are totally equivalent to the second-order Lagrangian of Hilbert. This class is again parametrized by the choice of an arbitrary linear connection in space-time M, hereafter referred to as "a background", which plays no dynamical role and may be considered, in a suitable sense, as a gauge fixing compatible with the dynamics of g (see [6]). A completely analogous result holds when attempting to reformulate General Relativity in terms of tetrads; in this case the counterpart of Einstein's first-order Lagrangian is not invariant under Lorentz rotations, while fixing a "background" allows to obtain an invariant first-order Lagrangian ([10, 28]).

Motivated by the above remarks, in this paper we shall shortly revisit the variational structure of General Relativity. We shall first discuss conditions under which a second-order Lagrangian which is linear in second-order derivatives produces field equations and Poincaré-Cartan forms of order lower than four and three respectively. The Hilbert Lagrangian will of course satisfy these conditions. We shall then briefly review the results of [6] about the existence and the properties of covariant first-order Lagrangians equivalent to the Hilbert Lagrangian.

2. Notation and Preliminaries

This paper will closely follow the notation introduced in our previous papers on the same subject, which may be found in [6] or [9]. A review on the geometric structure of Calculus of Variations and the Poincaré-Cartan formalism, together with its

notation, is given in the Lecture Notes **[13]**; notions from Differential Geometry are standard and we follow the notation of **[14]**.

As is well known, a *physical field theory* is usually based on the choice of a *fibration*

$$B \xrightarrow{\pi} M$$

together with a *Lagrangian* (of order $k \geq 1$), i.e., a bundle morphism:

$$(2.1) \qquad \mathcal{L} : J^k B \longrightarrow A^0{}_m(M) \qquad (m=\dim M)$$

from the k-th order jet prolongation J^kB onto the bundle of all skew-symmetric tensors of rank (0,m) over the base M. The Lagrangian $\mathcal{L}$ can also be viewed as a horizontal m-form of J^kB and locally one has:

$$(2.2) \qquad \mathcal{L} = L(x^\lambda, y^a, y^a{}_\mu, ..., y^a{}_{\mu_1 ... \mu_k})\mathbf{ds} \quad ,$$

where :

$$(2.3) \qquad \mathbf{ds} = dx^1 \wedge ... \wedge dx^m$$

is a local volume element in M. The quadruple $(B,M,\pi; \mathcal{L})$ is called a *variational principle* (of order k).

In the geometrical formulation of Calculus of Variations one frequently encounters the following concepts. The *structural forms* of B are the (vector-valued) differential 1-forms $\theta^a{}_{\mu_1 ... \mu_s} \in \Omega^1(J^{s+1}B) \otimes V^*(\pi)$ locally defined by :

$$(2.4) \qquad \theta^a{}_{\mu_1 ... \mu_s} = dy^a{}_{\mu_1 ... \mu_s} - y^a{}_{\mu_1 ... \mu_s \lambda} dx^\lambda \quad , \qquad 1 \leq s < \infty \quad ,$$

where $V^*(\pi)$ denotes the dual of the vertical bundle $V(\pi)=\mathrm{Ker}(T\pi) \subseteq TB$. A form $\omega \in \Omega^p(J^kB)$ is called a *contact form* iff $(j^k\sigma)^* \omega = 0$ for any section σ, i.e., if it factors through the structural forms above.

Moreover, one has an operator

$$D : C^\infty(J^kB) \longrightarrow \Omega^1{}_{hor}(J^{k+1}B)$$

for all orders $k \geq 1$, called the *formal differential* and intrinsically defined by the requirement

$$DF \circ j^{k+1}\sigma = d(F \circ j^{k}\sigma) \tag{2.5}$$

for any section $\sigma \in \Gamma(\pi)$. The local expression of DF is

$$DF \equiv (d_\mu F)\ dx^\mu \quad , \tag{2.6}$$

with $d_\mu F$ defined by:

$$d_\mu F \equiv \frac{\partial F}{\partial x^\mu} + y^b{}_\mu \frac{\partial F}{\partial y^b} + \ldots + y^b{}_{\mu\lambda_1\ldots\lambda_k} \frac{\partial F}{\partial y^b{}_{\lambda_1\ldots\lambda_k}} \quad . \tag{2.7}$$

One defines also the *vertical differential* $d_V F \in \Omega^1{}_{hor}(J^{k+1}B)$ by

$$DF = (\pi^{k+1}_k)^* dF - d_V F \quad , \tag{2.8}$$

which gives locally:

$$d_V F = \frac{\partial F}{\partial y^a}\,\theta^a + \sum_t \frac{\partial F}{\partial y^a{}_{\rho_1\ldots\rho_t}}\,\theta^a{}_{\rho_1\ldots\rho_t} \qquad (1 \leq t \leq k)\ ; \tag{2.9}$$

this is of course a contact 1-form.

For any compact domain $D \subseteq M$ the *action functional (s)* (over D) are defined by setting:

$$\mathfrak{a}[\sigma] = \int_D \mathcal{L}_\sigma = \int_D \mathcal{L} \circ j^k\sigma \quad ; \tag{2.10}$$

where $j^k\sigma \in \Gamma(\pi^k)$ is the *k-th order jet-prolongation* of σ, locally defined by :

$$y^a{}_{\mu_1\ldots\mu_s}(j^k\sigma) = \partial_{\mu_1\ldots\mu_s}\sigma^a(x^\lambda) \quad . \tag{2.11}$$

The section σ is called a *critical section* if the "first variation" of the action $\mathfrak{a}[\sigma]$ vanishes along all infinitesimal deformations of σ, i.e., all vertical vectorfields X^a (which respect some given boundary conditions). One finds

(2.12) $$\delta\mathcal{A} = \int_D [e_a(L)]X^a ds +$$

$$+ \int_{\partial D} (f_a{}^{\mu}X^a + f_a{}^{\mu\rho}X^a{}_{\rho} + \ldots + f_a{}^{\mu\rho_1\ldots\rho_{k-1}}X^a{}_{\rho_1\ldots\rho_{k-1}})\, ds_{\mu} \quad ,$$

where:

(2.13) $$ds_{\mu} = \partial_{\mu} \rfloor ds = (-1)^{\mu} dx^1 \wedge \ldots \wedge \hat{dx^{\mu}} \wedge \ldots \wedge dx^m \quad .$$

Here $e_a(L)$ is defined by:

(2.14) $$e_a(L) = p_a - d_{\mu}p_a{}^{\mu} + d_{\mu}d_{\rho}p_a{}^{\mu\rho} - \ldots + (-1)^k d_{\mu_1}\ldots d_{\mu_k} p_a{}^{\mu_1\ldots\mu_k}$$

and the "momenta" appearing in the boundary term are defined by recursive relations:

(2.15)
$$f_a{}^{\mu\rho_1\ldots\rho_{k-1}} = p_a{}^{\mu\rho_1\ldots\rho_{k-1}} \quad ,$$

$$f_a{}^{\mu\rho_1\ldots\rho_t} + d_{\rho}f_a{}^{\mu\rho_1\ldots\rho_t\rho} = p_a{}^{\mu\rho_1\ldots\rho_t} \qquad 0 \le t \le k-2 \quad ,$$

being

(2.16) $$p_a = \frac{\partial L}{\partial y^a} \quad , \quad p_a{}^{\mu} = \frac{\partial L}{\partial y^a_{\mu}} \quad , \ldots , \quad p_a{}^{\mu_1\ldots\mu_k} = \frac{\partial L}{\partial y^a_{\mu_1\ldots\mu_k}} \quad .$$

These are the leading coefficients of the differential $d\mathcal{L}$, which can be easily identified with a section of the *k-th order phase bundle* (see [8]):

(2.17) $$\mathbb{P}^k(\pi) \equiv \mathbb{P}(\pi^k) = A^0{}_m(M) \otimes V^*(\pi^k) \quad ,$$

where $\pi^k : J^kB \longrightarrow M$ is the natural projection. Requiring $\delta\mathcal{A} = 0$ with $j^{k-1}X$ vanishing on ∂D, we see that a section σ is critical if and only if it satisfies the local equations $e_a(L) = 0$, which are equations of the order $2k$ in σ. We have in fact a global *Euler-Lagrange morphism*

(2.18) $$e(L) : J^{2k}B \longrightarrow \mathbb{P}(\pi)$$

and the intrinsic form of Euler Lagrange equations is :

(2.19) $\qquad e(\mathcal{L}) \circ j^{2\kappa}\sigma = 0$.

The (local) expression (2.6) for the variation $\delta\alpha$ substantially states that *"the differential of the Lagrangian can be split into the sum of fields equations and a boundary term which is a pure divergence"*. This rough statement is in fact the naïve formulation of an important relation, called the *first variation formula*, whose global validity is equivalent to the existence of suitable forms called the *Poincaré-Cartan form(s)* of $\mathcal{L}$. These suitably generalize to higher-order field theory a concept classically known in Mechanics, whereby the Poincaré Cartan form is uniquely defined by

$$\Theta = L(t, q^a, \dot{q}^a)\, dt + \frac{\partial L}{\partial \dot{q}^a}\,(dq^a - \dot{q}^a\, dt) \quad .$$

For any given variational principle $(B,M,\pi,\mathcal{L})$ of order k we define in fact the *k-th order momentum bundle* of π by setting:

(2.20) $\qquad \mathbb{M}^k(\pi) = A^0{}_{m-1}(M) \otimes V^*(\pi^{k-1})$.

Local coordinates in $\mathbb{M}^k(\pi)$ are denoted by $(x^\lambda, y^a, f_a{}^\mu, \ldots, f_a{}^{\mu\rho_1\cdots\rho_{\kappa-1}})$ and they are defined by identifying $A^0{}_{m-1}(M)\otimes V^*(\pi^{k-1})$ with $\mathrm{Hom}(V(\pi^{k-1}), A^0{}_{m-1})$ and setting:

$$f(v) = (\, f_a{}^\mu v^a + f_a{}^{\mu\rho_1} v^a{}_{\rho_1} + \ldots + f_a{}^{\mu\rho_1\cdots\rho_{\kappa-1}} v^a{}_{\rho_1\ldots\rho_{\kappa-1}}\,) ds_\mu$$

for any element $f \in \mathbb{M}^k(\pi)$ and any vertical vector $v \simeq (v^a, v^a{}_{\rho_1}, \ldots, v^a{}_{\rho_1\ldots\rho_{\kappa-1}})$ in $J^{k-1}B$. We have then the following result: *For any linear connection* γ *in the base* M *there exists a global bundle morphism*

(2.21) $\qquad f(\mathcal{L},\gamma) \equiv f : J^{2k-1}B \longrightarrow \mathbb{M}^k(\pi)$

such that the following holds :

(2.22) $\qquad T\mathcal{L}(j^k v)|_\sigma = \langle e(\mathcal{L}) \circ j^{2k}\sigma \mid v \rangle + d\langle f \circ j^{2k-1}\sigma \mid j^{k-1} v \rangle$

for any $\sigma \in \Gamma(\pi)$ *and any vertical vectorfield* $v \in \mathcal{X}_V(\pi)$ *projecting over* σ (see [5]). Equation (2.16) is the *global first variation formula*.. Now, the morphism f defines uniquely a contact form $\tilde{f}$ (L,γ) by:

$$(2.23) \qquad \tilde{f}\,(L,\gamma) = (\tilde{f}_{a}^{\mu\rho_1\dots\rho_t})\,\theta^{a}_{\rho_1\dots\rho_t} \wedge ds_{\mu} \in \Omega^{m}(J^{2k-1}B)$$

The contact form $\tilde{f}$ defines then the *global Poincaré-Cartan form of* L (associated to the connection γ) as the m-form $\Theta(L,\gamma)$ over $J^{2k-1}B$ given by:

$$(2.24) \qquad \Theta(L,\gamma) = (\pi_{k}^{2k-1})^{*}L + \tilde{f}$$

where $(\pi_{k}^{2k-1})^{*}L$ is the pull-back to $J^{2k-1}B$ of the Lagrangian L. This form satisfies a few characteristic properties which form in fact the basis of the "axiomatic theory of Poincaré-Cartan forms" (see, e.g., [5]); besides some verticality conditions, these properties are mainly summarized by the requirement that *a section* $\sigma \in \Gamma(\pi)$ *is critical for* L *if and only if the following holds*

$$(2.25) \qquad (j^{2k-1}\sigma)^{*}\,[\,i_{\xi}\,d\Theta(L,\gamma)\,] = 0$$

for any vectorfield $\xi \in \mathcal{X}(J^{2k-1}B)$.

As we said, uniqueness of the Poincaré-Cartan form is lost in the higher-order case over a basis with dimension greater than one, whereby globality is achieved at the expense of introducing a linear connection in the basis M (as an extra parameter which helps in the globalization procedure). Of course, there are particular cases in which a global Poincaré-Cartan form can be obtained without having to resort to a suitable "globalization procedure" as above. Apart from the cases k=1, dim(M) arbitrary (=first-order field theory) and k arbitrary, dim(M)=1 (=higher-order Mechanics), a global Poincaré-Cartan form can be obtained at once if: (i) the basis M has a global frame (which can be used as a substitute for the connection γ); (ii) a linear connection γ already appears among the variables (i.e., if B is a fibered product of the bundle of connections C(M) with other fields). Notice that case (ii) covers a priori all the applications to relativistic theories, whereby a preferred linear connection is fixed by Physics itself. In any case, it is known that for a Lagrangian of the second-order there exists a *canonical* Poincaré-Cartan form, given by setting:

$$(2.26) \qquad \Theta \underset{\text{locally}}{\equiv} L + [\,\tilde{f}_{a}^{\mu}(L)\theta^{a} + \tilde{f}_{a}^{\mu\lambda}(L)\theta^{a}_{\lambda}\,] \wedge ds_{\mu} = L\,ds +$$

$$+ \ \tilde{f}_a{}^{\mu} (dy^a - y^a{}_{\rho}\, dx^{\rho}) \wedge ds_{\mu} + \tilde{f}_a{}^{\mu\lambda} (dy^a{}_{\lambda} - y^a{}_{\lambda\rho}\, dx^{\rho}) \wedge ds_{\mu} \ ,$$

with:

$$(2.27) \qquad \tilde{f}_a{}^{\mu}(\mathcal{L}) = p_a{}^{\mu} - d_{\rho}\, p_a{}^{\mu\rho} \quad , \qquad \tilde{f}_a{}^{\mu\lambda}(\mathcal{L}) = p_a{}^{\mu\lambda} \ .$$

For first-order theories this reduces to:

$$(2.28) \qquad \Theta(\mathcal{L}) = L\, ds + p_a{}^{\mu} (dy^a - y^a{}_{\rho}\, dx^{\rho}) \wedge ds_{\mu} \ .$$

3. Degenerate Second-Order Lagrangians

According to the general theory outlined above, a k-th order Lagrangian $\mathcal{L}$ will generate field equations of order 2k and a Poincaré-Cartan form of order (2k-1). In particular, a first-order Lagrangian should produce second-order field equations and a first-order Poincaré-Cartan form, while a second-order Lagrangian should produce, if completely "regular", fourth-order Euler-Lagrange equations and a third-order Poincaré-Cartan form. As is well known all the various notions of "regularity" can be suitably stated in terms of the Poincaré-Cartan forms; naïvely, one says that *a higher-order Lagrangian* $\mathcal{L} : J^k B \longrightarrow A^0{}_m(M)$ *is "degenerate"* (i.e., not "regular") *if the Euler-Lagrange equations* (which should in principle be of order 2k in <u>all</u> variables y^a) *are of a lower order* $p < 2k$. In fact, if a Lagrangian $\mathcal{L}$ of order k is degenerate, then its Poincaré-Cartan form has to live in a jet prolongation $J^s B$ of order s lower than 2k-1, and the order s gives a measure of the degree of degeneracy of $\mathcal{L}$ (see, e.g., **[18, 29]**). In particular, if the Poincaré-Cartan form $\Theta(\mathcal{L})$ is defined in some *odd* jet-prolongation $J^{2r-1}B$, with $r \leq k-1$, this might signal the existence of an *equivalent* <u>*regular*</u> Lagrangian $\mathcal{L}'$ of *lower* order $r \leq k-1$ (where by "equivalent" we mean such that the critical sections of $\mathcal{L}$ and $\mathcal{L}'$ are the same, although the corresponding variational principles have an apparently different order. A few the remarks on this still open problem may be found in in **[9]**).

In this Section we shall shortly address sufficient conditions which ensure that a second-order Lagrangian $\mathcal{L}$ is degenerate enough to produce field equations of the second-order (rather than fourth) and a Poincaré-Cartan form of the first-order (rather

than third). These conditions signal that there should be an equivalent first-order Lagrangian $\mathcal{L}'$, (locally) differing from $\mathcal{L}$ by a divergence.

Most of the Lagrangians governing field theories are (covariant) functions of affine combinations of the highest derivatives they contain (see [9] for a discussion on this point). In particular this is true of the (first-order) Lagrangian of Yang-Mills theory and of the Hilbert (second-order) Lagrangian of General Relativity. Let us first remark that a second-order Lagrangian linearly depending on second-order derivatives $y^a_{\mu\nu}$ cannot produce fourth-order equations. To see this, in fact, it is enough to observe that the momentum $p_a^{\mu\nu} \equiv \frac{\partial \mathcal{L}}{\partial y^a_{\mu\nu}}$ will no longer depend on second-order derivatives, so that $p_a^{\mu\nu}$ will be a function of $j^1y \equiv (y^a, y^a_\mu)$ only. Accordingly, the structure of the Euler Lagrange equations will be the following:

$$p_a(j^2y) - d_\mu[\, p_a^\mu(j^2y)] + d_\mu d_\nu[\, p_a^{\mu\nu}(j^1y)] = 0 \quad ,$$

and p_a will contribute to them only by j^2y, while the other two terms will at most generate third-order derivatives.

Let then $\mathcal{L} = L(x^\mu, y^a, y^a_\mu, y^a_{\mu\nu})ds$ be a second-order Lagrangian linearly depending on all the second-order derivatives $y^a_{\mu\nu}$. We have the following result:

Proposition 3.1 *The Euler Lagrange equations of a linear second-order Lagrangian*

$$\mathcal{L}(j^2y) = \{\, A_a^{\mu\nu}(j^1y)\, y^a_{\mu\nu} + B(j^1y) \,\}ds \tag{3.1}$$

are linear in the third-order derivatives $y^a_{\mu\nu\rho}$. *Moreover, they are of the second order if and only if the following condition is satisfied*:

$$\partial_{[a}{}^{(\rho}\, A_{b]}{}^{\mu\nu)} = 0 \quad . \tag{3.2}$$

Proof . Let us first calculate the naïve momenta $(p_a, p_a^\mu, p_a^{\mu\nu})$ of $\mathcal{L}$. We have:

$$\begin{aligned} p_a(j^2y) &= (\partial_a A_b{}^{\rho\sigma})\, y^b_{\rho\sigma} + \partial_a B \quad , \\ p_a{}^\mu(j^2y) &= (\partial_a{}^\mu A_b{}^{\rho\sigma})\, y^b_{\rho\sigma} + \partial_a{}^\mu B \quad , \\ p_a{}^{\mu\nu}(j^1y) &= A_a{}^{\mu\nu} \quad , \end{aligned} \tag{3.3}$$

where all the coefficients $A_a^{\mu\nu}$, $\partial_a A_b^{\rho\sigma}$, $\partial_a{}^{\mu} A_b^{\rho\sigma}$, $\partial_a B$ and $\partial_a{}^{\mu} B$ depend only on (j^1y). The corresponding Euler-Lagrange equations will have the following structure, where non-essential terms containing second-order derivatives are denoted by (....):

$$e_a(L) = (...) + (\partial_b{}^{\rho} A_a^{\mu\nu})\, y^b{}_{\rho\mu\nu} - (\partial_a{}^{\mu} A_b^{\rho\sigma})\, y^b{}_{\rho\sigma\mu} \quad .$$

Thus, from the symmetries of $y^b{}_{\rho\mu\nu}$ it follows:

$$e_a(L) = (\text{second order}) + (\partial_b{}^{\rho} A_a^{\mu\nu} - \partial_a{}^{\rho} A_b^{\mu\nu})\, y^b{}_{\rho\mu\nu} \quad . \tag{3.4}$$

From this, condition (3.25) follows. **(Q.E.D.)**

We shall now discuss how the Poincaré-Cartan form $\Theta(L)$ behaves under the hypothesis that L has the form (3.1). Calculating the momenta $\tilde{f}_a{}^{\mu}$ and $\tilde{f}_a{}^{\mu\nu}$ we find:

$$\tilde{f}_a{}^{\mu}(L) = p_a{}^{\mu} - d_\nu p_a{}^{\mu\nu} = (\partial_a{}^{\mu} A_b^{\rho\nu})\, y^b{}_{\rho\nu} + \partial_a{}^{\mu} B - d_\nu A_a^{\mu\nu} =$$

$$= (\partial_a{}^{\mu} A_a^{\rho\nu} - \partial_b{}^{\rho} A_a^{\mu\nu})\, y^b{}_{\rho\nu} + (\text{first order terms}) \quad ; \tag{3.5}$$

$$\tilde{f}_a{}^{\mu\nu}(L) \equiv p_a{}^{\mu\nu} \equiv A_a^{\mu\nu}(j^1y) \quad . \tag{3.6}$$

This shows that the Poincaré-Cartan form of any linear Lagrangian (3.1) never depends on third-order derivatives, since its coefficients depend at most on j^2y. Moreover, being $L = L\,ds$ and $\tilde{f}_a{}^{\mu}(L)$ both linear in the second-order derivatives $y^a{}_{\mu\nu}$, it follows that $\Theta(L)$ is in fact a form in $\Omega^2(J^2B)$ which is *linear* in the second-order derivatives $y^a{}_{\mu\nu}$. More precisely, one has:

$$\Theta(L) = L + [\,\tilde{f}_a{}^{\mu}(L)\,\theta^a + \tilde{f}_a{}^{\mu\lambda}(L)\,\theta^a{}_{\lambda}\,] \wedge ds_\mu =$$

$$= (A_a^{\mu\nu} y^a{}_{\mu\nu} + B)\,ds + A_a^{\mu\nu}\,(dy^a{}_{\nu} - y^a{}_{\rho\nu} dx^\rho) \wedge ds_\mu +$$

$$+ \{(\partial_a{}^{\mu} A_b^{\rho\nu} - \partial_b{}^{\rho} A_a^{\mu\nu})\, y^b{}_{\rho\nu} + \ldots\}(dy^a - y^a{}_{\rho} dx^\rho) \wedge ds_\mu =$$

$$= (A_a^{\mu\nu} y^a{}_{\mu\nu})\,ds + B\,ds + (\partial_a{}^{\mu} A_b^{\rho\nu} - \partial_b{}^{\rho} A_a^{\mu\nu})\, y^b{}_{\rho\nu}\,\theta^a \wedge ds_\mu +$$

$$+ A_a^{\mu\nu}\, dy^a{}_{\nu} \wedge ds_\mu - A_a^{\mu\nu}\, y^a{}_{\rho\nu} dx^\rho \wedge ds_\mu$$

and hence:

$$\Theta(\mathcal{L}) = B\,ds + (\partial_a{}^{\mu} A_b{}^{\rho\nu} - \partial_b{}^{\rho} A_a{}^{\mu\nu}) y^b{}_{\rho\nu}\, \theta^a \wedge ds_\mu \tag{3.7}$$

$$+ A_a{}^{\mu\nu}\, dy^a{}_\nu \wedge ds_\mu + (\text{ first order terms }) \quad .$$

Recalling that the coefficients $A_a{}^{\mu\nu}$ are symmetric, i.e. $A_a{}^{\mu\nu} = A_a{}^{\nu\mu}$, the following proposition is an immediate consequence of eqn. (3.7):

Proposition 3.2 *A linear second-order Lagrangian* (3.1) *admits a Poincaré-Cartan form* $\Theta(\mathcal{L})$ *which does not contain second-order derivatives* $y^a{}_{\mu\nu}$ *if and only if the following holds:*

$$\partial_a{}^{\mu} A_b{}^{\rho\nu} - \partial_b{}^{(\rho} A_a{}^{\nu)\mu} = 0 \quad . \tag{3.8}$$

We remark that the two conditions (3.2) and (3.8) are different, so that they define in fact different classes of second-order (linear) Lagrangians. However, they will be identically satisfied together if the coefficients $A_b{}^{\rho\nu}$ satisfy the simpler condition

$$\partial_a{}^{\mu} A_b{}^{\rho\nu} = 0 \qquad \forall\, a,b,\ \forall \mu,\rho,\nu \quad , \tag{3.9}$$

which corresponds to the simpler case of a Lagrangian of the form:

$$\mathcal{L} = [\, A_a{}^{\mu\nu}(x,y)\, y^a{}_{\mu\nu} + B(j^1 y)\,]\, ds \tag{3.10}$$

i.e., when $\mathcal{L}$ is linear in $y^a{}_{\mu\nu}$ with coefficients depending at most on the zero-th order derivatives of y itself. Many cases of physical interest fall in fact into this category.

In this case (3.3) are simplified to

$$\begin{aligned} p_a(j^2 y) &= (\partial_a A_b{}^{\rho\sigma})\, y^b{}_{\rho\sigma} + \partial_a B \quad , \\ p_a{}^{\mu}(j^1 y) &= \partial_a{}^{\mu} B \quad , \\ p_a{}^{\mu\nu}(j^0 y) &= A_a{}^{\mu\nu} \quad , \end{aligned} \tag{3.11}$$

and the corresponding Euler-Lagrange equations will be

$$e_a(\mathcal{L}) = (\text{first order}) + (\partial_a A_b{}^{\rho\mu} - \partial_b{}^{\rho}\partial_a{}^{\mu} B - \partial_b A_a{}^{\rho\mu})\, y^b{}_{\rho\mu} \quad , \tag{3.12}$$

i.e., they turn out to be *linear* in the second-order derivatives $y^b{}_{\rho\mu}$. For the "true" momenta one finds instead:

$$\tilde{f}_a{}^{\mu}(\mathcal{L}) = \tilde{f}_a{}^{\mu}(j^1 y) = \partial_a{}^{\mu} B - (\partial_b A_a{}^{\mu\nu}) y^b{}_{\nu} - \partial_{\nu} A_a{}^{\mu\nu} \quad ,$$

(3.13)

$$\tilde{f}_a{}^{\mu\nu}(\mathcal{L}) = A_a{}^{\mu\nu}(j^0 y) \quad .$$

Hence, the appropriate Poincaré-Cartan form $\Theta(\mathcal{L}) \in \Omega^m(J^1 B)$ is given by

$$\Theta(\mathcal{L}) = B(j^1 y)ds + \tilde{f}_a{}^{\mu}\, \theta^a \wedge ds_{\mu} + A_a{}^{\mu\nu}\, dy^a{}_{\nu} \wedge ds_{\mu} \quad . \tag{3.14}$$

and it is manifestly independent on second-order derivatives of y.

4. Variational Structure of General Relativity

As is well known Einstein's (vacuum) equations

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0 \quad , \tag{4.1}$$

have *variational character*, i.e., they are derivable from a variational principle. The dynamical field y is a *Lorentzian metric* g on the (4-dimensional) space-manifold M, i.e., a (local, smooth) section of the fiber bundle $\mathrm{Lor}(M) \longrightarrow M$ of all Lorentzian metrics over M. The appropriate functional space of field variables is $\mathcal{F} \equiv \Gamma^{\infty}_{\mathrm{loc}}(\mathrm{Lor}(M))$, which is a Fréchet manifold under the topology of C^{∞}-uniform convergence on all compact domains $D \subseteq M$. $\mathcal{F}$ turns out to be an open cone in the vector space $\mathbf{S}$ formed by all (local, smooth) sections of the bundle $S^0{}_2(M)$, i.e., of all (local, smooth) symmetric twice-covariant tensor fields over M. The Frechét space $\mathbf{S} \supseteq \mathcal{F}$ is thence the model of the manifold $\mathcal{F}$ itself, so that at each point $g \in \mathcal{F}$ the tangent space $T_g\mathcal{F}$ is isomorphic to $\mathbf{S}$ itself; this expresses the fact that all infinitesimal deformations of a Lorentzian metric g are symmetric tensors of the same rank. Locally we shall denote by $g \equiv g_{\mu\nu} dx^{\mu} \otimes dx^{\nu}$ any element of $\mathcal{F}$ and by $h \equiv h_{\mu\nu} dx^{\mu} \otimes dx^{\nu}$ any element of $\mathbf{S}$.

As it was first shown by Hilbert in 1915 [23], the functional which generates Einstein's vacuum equations is the integral of the scalar curvature R(g). The "Hilbert Lagrangian" is hence the horizontal form $\mathcal{L}_H = L_H ds \in \Omega^4{}_{\mathrm{hor}}\{J^2[\mathrm{Lor}(M)]\}$ locally defined by:

(4.2) $$L_H(j^2g) = R(g)\sqrt{g} = (g^{\mu\sigma}R_{\mu\sigma})\sqrt{g} \quad ,$$

which depends on the 2-jet j^2g through the Christoffel symbols and their first derivatives. The Hilbert action are the functional(s)

(4.3) $$\mathfrak{a}_H[g] = \int_D R(g)\sqrt{g}ds \quad ,$$

(for all compact domains $D \subset M$). The first variation of $\mathfrak{a}_H[g]$, i.e., its Gateaux derivative in any tangent direction h is known to be expressible as follows:

(4.4) $$D_g\mathfrak{a}_H \cdot h = -\int_D G^{\mu\sigma}h_{\mu\sigma}\sqrt{g}\, ds + \int_{\partial D} B^{\alpha}ds_{\alpha} \quad ,$$

where the boundary term is given by

(4.5) $$B^{\alpha} \equiv [\nabla_{\mu}h^{\mu\alpha} - g^{\mu\alpha}\nabla_{\mu}(g\cdot h)]\sqrt{g} \quad ,$$

with $g\cdot h \equiv g^{\mu\sigma}h_{\mu\sigma}$. Assuming that h vanishes on the boundary together with its first derivatives one finds that Euler-Lagrange equations are in fact (4.1). The structure of the boundary term (4.5) gives moreover some information about the structure of the Poincaré-Cartan form which will be discussed later.

Let us now analyze the above result in view of the "regularity" discussion of Section 3. The Hilbert Lagrangian (4.2) is second-order degenerate, because it is linear in the second-order derivatives of g. In fact, the Riemann-Christoffel tensor is linear in the second-order derivatives of g, being

(4.6) $$R^{\lambda}_{\mu\nu\sigma} = \{\, g^{\lambda\varepsilon}\delta^{\beta}_{\mu}\delta^{[\alpha}_{\nu}\delta^{\gamma]}_{\sigma} - g^{\lambda\alpha}\delta^{\gamma}_{\mu}\delta^{[\beta}_{\nu}\delta^{\varepsilon]}_{\sigma} \,\}\partial^2_{\alpha\beta}g_{\gamma\varepsilon} + (...) \quad ,$$

where (...) denotes terms which depend on j^1g only; from this it follows

(4.7) $$R = (g^{\varepsilon\alpha}g^{\beta\gamma} - g^{\beta\alpha}g^{\varepsilon\gamma})\partial^2_{\alpha\beta}g_{\gamma\varepsilon} + (...) \quad .$$

This last expression can in fact be recast as follows

(4.8) $$R = \frac{1}{2}\, G^{\alpha\beta\varepsilon\gamma}[\partial^2_{\alpha\beta} g_{\varepsilon\gamma} + g^{\mu\nu}\Gamma_{\alpha\beta,\mu}\Gamma_{\varepsilon\gamma,\nu}]$$,

where $\Gamma_{\alpha\beta,\mu} \equiv \{\alpha\beta,\mu\}$ are the Christoffel symbols of the first kind of g and

(4.9) $$G^{\alpha\beta\varepsilon\gamma} = g^{\alpha\varepsilon}g^{\beta\gamma} + g^{\alpha\gamma}g^{\beta\varepsilon} - 2g^{\alpha\beta}g^{\varepsilon\gamma}$$,

is the so-called "De Witt's metric" (see [2, 12]). Its "inverse" is given by:

(4.10) $$G_{\varepsilon\gamma\lambda\rho} = \frac{1}{2}\left(g_{\varepsilon\lambda}g_{\gamma\rho} + g_{\varepsilon\rho}g_{\gamma\lambda} - \frac{2}{3}\, g_{\varepsilon\gamma}g_{\lambda\rho}\right)$$

and satisfies

(4.11) $$G^{\alpha\beta\varepsilon\gamma}G_{\varepsilon\gamma\lambda\rho} = \delta^\alpha_\lambda\delta^\beta_\rho + \delta^\alpha_\rho\delta^\beta_\lambda$$.

Equation (4.8) tells us that the Lagrangian $\mathcal{L}_H(j^2g)$ is highly degenerate: it is in fact linear in the second-order derivatives $\partial^2_{\alpha\beta} g_{\gamma\varepsilon}$ and, moreover, the coefficients of these derivatives depend only on g and not on first-order derivatives, so that the Lagrangian $\mathcal{L}_H(j^2g)$ has the very particular form (3.10). Accordingly, the Euler-Lagrange equations have to be of the second-order only and linear in the second-order derivatives (i.e., "quasilinear second-order equations"). In fact, from (4.6) one sees immediately that the Einstein tensor $G_{\mu\nu}$ is indeed a linear function of $\partial^2_{\alpha\beta} g_{\gamma\varepsilon}$. Moreover, from our discussion of Section 3 it follows that the corresponding Poincaré-Cartan form $\Theta(\mathcal{L}_H)$, whose explicit expression will be discussed later, will live neither in $J^3[\mathrm{Lor}(M)]$ nor in $J^2[\mathrm{Lor}(M)]$, but more simply in $J^1[\mathrm{Lor}(M)]$, i.e., it will depend only on first-order derivatives of g. This should suggest the existence of a (covariant) Lagrangian of the first order in g, *equivalent* to the second-order Hilbert Lagrangian itself, in spite of the well known fact that there exist no scalar density $f(j^1g)\sqrt{g}$ containing only a metric tensor together with its first derivatives (but no other geometric object), except the trivial density $\sqrt{g}$ (which does not depend on first derivatives either!).

In order to obtain a *covariant* first-order Lagrangian $\mathcal{L}$ able to generate Einstein's equations one should therefore renounce the hypothesis that the Lagrangian contains only a metric, allowing a dependence on some extra geometric object. All "metric-affine" theories, whereby an extra linear connection Γ is allowed among the fields, satisfy a requirement of this kind, and it is known that metric-affine Lagrangians

exist which generate equations for both g and Γ which are a posteriori equivalent to (4.1). This, however, has the disadvantage of introducing an extra dynamical field. As it was shown in [6] there is however a further possibility, i.e., to let some extra object enter the Lagrangian as a fixed "background" (i.e., as a parameter but with no dynamics), allowing to define a whole class of *purely metric first-order covariant* Lagrangians equivalent to the Hilbert Lagrangian. To conclude this paper we shall shortly recall this result and we shall discuss the relations among the corresponding Poincaré-Cartan forms.

Setting for simplicity

$$\pi^{\mu\nu} \equiv \sqrt{g}\, g^{\mu\nu} \tag{4.12}$$

one can easily realize that the following decomposition holds

$$R\sqrt{g} = \partial_\lambda[A^\lambda(j^1 g)] + U_o(j^1 g) \quad , \tag{4.13}$$

where $U_o(j^1 g)$ and $A^\lambda(j^1 g)$ are given by :

$$A^\lambda(j^1 g) \equiv \pi^{\mu\sigma}\Gamma^\lambda_{\mu\sigma} - \pi^{\mu\lambda}\Gamma^\alpha_{\alpha\mu} \quad . \tag{4.14}$$

$$U_o = \pi^{\mu\sigma}(\Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\rho\sigma} - \Gamma^\rho_{\mu\sigma}\Gamma^\lambda_{\lambda\rho}) \quad . \tag{4.15}$$

This decomposition, first noticed by Einstein, provides a non-covariant first-order Lagrangian U_o which still generates Einstein equations (4.1), since it differs from $R\sqrt{g}$ by a pure divergence. It has of course the disadvantage of being non-covariant, a rather disturbing fact in General Relativity, which is known to have serious consequences for a coherent description of conservation laws (see [7]). To overcome this difficulty, one first defines, for any given symmetric connection $\Gamma^\lambda_{\mu\nu}$, the following object:

$$u^\lambda_{\mu\nu}(\Gamma) = \Gamma^\lambda_{\mu\nu} - \frac{1}{2}(\delta^\lambda_\mu \Gamma^\sigma_{\sigma\nu} + \delta^\lambda_\nu \Gamma^\sigma_{\sigma\mu}) \quad , \tag{4.16}$$

which satisfies the obvious symmetry $u^\lambda_{\mu\nu} = u^\lambda_{\nu\mu}$. This defines in fact a coordinate change in the bundle $\mathcal{C}_S(M)$ of all symmetric connections, its inverse being given by:

$$\Gamma^\lambda_{\mu\nu}(u) = u^\lambda_{\mu\nu} - \frac{1}{m-1}(\delta^\lambda_\mu u^\sigma_{\sigma\nu} + \delta^\lambda_\nu u^\sigma_{\sigma\mu}) \quad , \quad m=\dim(M) \quad . \tag{4.17}$$

Using these relations and taking $\pi^{\mu\nu}$ as independent variables, we can rewrite (4.13) as follows:

(4.18) $$R\sqrt{g} = \partial_\lambda[\pi^{\mu\sigma} u^\lambda{}_{\mu\sigma}(j^1\pi)] + U_o(j^1\pi) .$$

We fix now a background symmetric connection $\overset{*}{\Gamma}{}^\lambda{}_{\mu\nu}$ in (M,g) and we denote by $\overset{*}{u}{}^\lambda{}_{\mu\nu}$ its corresponding u-field. The following remarkable identity holds:

(4.19) $$q^\lambda{}_{\mu\nu} \equiv \{^\lambda{}_{\mu\nu}\}_g - \overset{*}{\Gamma}{}^\lambda{}_{\mu\nu} = \frac{1}{2} g^{\lambda\rho} (\overset{*}{\nabla}_\mu g_{\nu\rho} + \overset{*}{\nabla}_\nu g_{\rho\mu} - \overset{*}{\nabla}_\rho g_{\mu\nu}) \quad ,$$

where $\overset{*}{\nabla}$ denotes the covariant derivative with respect to $\overset{*}{\Gamma}$. As a consequence, (4.18) is transformed into the following:

(4.20) $$R\sqrt{g} = \partial_\lambda \{\pi^{\mu\sigma} [u^\lambda{}_{\mu\sigma}(j^1\pi)] - \overset{*}{u}{}^\lambda{}_{\mu\sigma}\} + L_{\overset{*}{\Gamma}}(x^\lambda; j^1\pi) \quad ,$$

with:

(4.21) $$L_{\overset{*}{\Gamma}} = \pi^{\mu\sigma} R_{\mu\sigma}(j^1\overset{*}{\Gamma}) + \pi^{\mu\sigma}(q^\rho{}_{\mu\lambda} q^\lambda{}_{\rho\sigma} - q^\rho{}_{\mu\sigma} q^\lambda{}_{\lambda\rho}) \quad .$$

Here $L_{\overset{*}{\Gamma}}$ is a <u>covariant</u> first-order Lagrangian which depends explicitly on space-time coordinates through the a priori assigned connection $\overset{*}{\Gamma}(x^\lambda)$ and an easy calculation shows that each Lagrangian of the family (4.21) still generates Einstein (vacuum) equations. Moreover, the background connection $\overset{*}{\Gamma}$ has no dynamics, since the first variation of $L_{\overset{*}{\Gamma}}$ with respect to $\overset{*}{\Gamma}$ is a total divergence.

If one wants now to express the Poincaré-Cartan form $\Theta(\mathcal{L}_H)$ in the basic variables $g_{\mu\nu}$, the relevant contact forms are the following:

(4.22) $$d_V g_{\mu\nu} \equiv dg_{\mu\nu} - (g_{\mu\nu,\rho})dx^\rho \quad ,$$

$$d_V g_{\mu\nu,\sigma} \equiv dg_{\mu\nu,\sigma} - (g_{\mu\nu,\sigma\rho})dx^\rho \quad .$$

and the coefficients of $\Theta(L_H)$ are given by

(4.23) $$\tilde{f}^{\mu\nu,\lambda} = \frac{\partial(R\sqrt{g})}{\partial g_{\mu\nu\,\lambda}} - d_\rho \frac{\partial(R\sqrt{g})}{\partial g_{\mu\nu\,\lambda\rho}} \quad ,$$

$$\tilde{f}^{\mu\nu,\lambda\rho} = \frac{\partial(R\sqrt{g})}{\partial g_{\mu\nu\,\lambda\rho}} \quad ,$$

which can be easily calculated with the aid of (4.8). We end up with the following implicit expression:

$$\Theta(\mathcal{L}_H) = R\sqrt{g}\,\mathbf{ds} + \{\tfrac{1}{2}\, G^{\alpha\beta\varepsilon\gamma}\sqrt{g}\, g^{\rho\sigma}\frac{\partial}{\partial g_{\mu\nu\,\lambda}}(\Gamma_{\alpha\beta,\rho}\Gamma_{\varepsilon\gamma,\sigma}) + \tag{4.24}$$

$$- \tfrac{1}{2}\, d_\rho(G^{\lambda\rho\mu\nu}\sqrt{g})\}(d_V g_{\mu\nu})\wedge \mathbf{ds}_\lambda + \tfrac{1}{2}\, G^{\lambda\rho\mu\nu}\sqrt{g}(d_V g_{\mu\nu,\rho})\wedge \mathbf{ds}_\lambda \quad .$$

which only apparently contains also second-order derivatives of $g_{\mu\nu}$, although these have to cancel out. To express $\Theta(\mathcal{L}_H)$ in an alternative way one can apply directly the relevant form of eqn. (3.10), where now B is the second addendum of (4.8), namely $\frac{1}{2}G^{\alpha\beta\varepsilon\gamma}\sqrt{g}g^{\rho\sigma}\Gamma_{\alpha\beta,\rho}\Gamma_{\varepsilon\gamma,\sigma}$, while the coefficients $A_a{}^{\mu\nu}$ are just $\frac{1}{2}G^{\lambda\rho\mu\nu}\sqrt{g}$. This gives soon the equivalent expression:

$$\Theta(\mathcal{L}_H) = \tfrac{1}{2}G^{\alpha\beta\varepsilon\gamma}\sqrt{g}\, g^{\rho\sigma}\Gamma_{\alpha\beta,\rho}\Gamma_{\varepsilon\gamma,\sigma} + \tag{4.25}$$

$$+ \tfrac{1}{2}\{G^{\alpha\beta\varepsilon\gamma}\sqrt{g}\, g^{\rho\sigma}\frac{\partial}{\partial g_{\mu\nu\,\lambda}}(\Gamma_{\alpha\beta,\rho}\Gamma_{\varepsilon\gamma,\sigma}) - d_\rho(G^{\lambda\rho\mu\nu}\sqrt{g})\}(d_V g_{\mu\nu})\wedge \mathbf{ds}_\lambda$$

$$+ \tfrac{1}{2}\,(G^{\lambda\rho\mu\nu}\, dg_{\mu\nu,\rho})\wedge \mathbf{ds}_\lambda \quad ,$$

which now does not contain j^2g but only j^1g .

This suggests to take immediately into account the fact that $\Theta(\mathcal{L}_H)$ has to live in $J^1[\mathrm{Lor}(M)]$ and change variables in this bundle, first from $(g_{\mu\nu}, g_{\mu\nu,\sigma})$ to $(\pi^{\mu\nu}, \pi^{\mu\nu}{}_\lambda)$ and then to $(\pi^{\mu\nu}, u^\lambda_{\mu\nu})$, which turn out to be in fact more convenient for this purpose. In terms of these new variables the relevant contact forms will be:

$$d_V\pi^{\mu\nu} \equiv d\pi^{\mu\nu} - (d_\rho\pi^{\mu\nu})\, dx^\rho \quad ,$$

$$d_V u^\lambda{}_{\mu\nu} \equiv du^\lambda{}_{\mu\nu} - (d_\rho u^\lambda{}_{\mu\nu})\, dx^\rho \quad , \tag{4.26}$$

and, after some manipulation, one ends up with the following expression:

$$\Theta(\mathcal{L}_H) = R\sqrt{g}\mathbf{ds} + (\pi^{\mu\nu}\, d_V u^\lambda{}_{\mu\nu})\wedge \mathbf{ds}_\lambda \quad . \tag{4.27}$$

which, using (4.26) and (4.18), can be finally rewritten as:

(4.28) $\Theta(\mathcal{L}_H) = (\pi^{\mu\nu}\, du^{\lambda}{}_{\mu\nu})\wedge ds_\lambda \;-\; \pi^{\mu\nu}\,(\,\Gamma^{\rho}{}_{\mu\lambda}\,\Gamma^{\lambda}{}_{\rho\nu} - \Gamma^{\rho}{}_{\mu\nu}\,\Gamma^{\lambda}{}_{\lambda\rho})\, ds$.

We notice that the coefficient of ds in (4.28) is minus the $(\Gamma\Gamma-\Gamma\Gamma)$-Lagrangian we called U_o. The expression (4.28) is in any case the local representation of a globally well-defined form on $J^1[Lor(M)]$, which, in each local chart, can also be written as follows:

(4.29) $\Theta(\mathcal{L}_H) = (\,-u^{\lambda}{}_{\mu\nu}\, d_V\pi^{\mu\nu})\wedge ds_\lambda + U_o(j^1\pi)ds \;+\; d(\pi^{\mu\nu}\, u^{\lambda}{}_{\mu\nu}\, ds_\lambda)$.

Although the decomposition (4.29) has only a local validity, it is nevertheless an interesting one. Formally, it expresses the Poincaré-Cartan form $\Theta(\mathcal{L}_H)$ as the sum of a non-covariant differential and the (first-order) Poincaré-Cartan form of the non-covariant (first-order) Lagrangian U_o. This local decomposition corresponds to the local (or non-covariant) decomposition (4.18) for $\mathcal{L}_H$. It can be covariantized in much the same way as we "covariantized" the decomposition (4.18) itself, i.e., by fixing a background connection $\overset{*}{\Gamma}$. In doing this, after some manipulation the following can be shown to hold:

(4.30) $\Theta(\mathcal{L}_H) = \Theta(\mathcal{L}_{\overset{*}{\Gamma}}) + d\Phi$,

where the global (m-1)-form $\Phi\in\Omega^{m-1}\{J^1[Lor(M)]\}$ is given by:

(4.31) $\Phi = \pi^{\mu\nu}(u^{\lambda}{}_{\mu\nu} - \overset{*}{u}{}^{\lambda}{}_{\mu\nu})\, ds_\lambda$,

and the Poincaré-Cartan form $\Theta(\mathcal{L}_{\overset{*}{\Gamma}})$ of the first-order Lagrangian $\mathcal{L}_{\overset{*}{\Gamma}} = L_{\overset{*}{\Gamma}} ds$ is given by:

(4.32) $\Theta(\mathcal{L}_{\overset{*}{\Gamma}}) = \{(\overset{*}{u}{}^{\lambda}{}_{\mu\nu} - u^{\lambda}{}_{\mu\nu})d_V\,\pi^{\mu\nu}\}\wedge ds_\lambda + L_{\overset{*}{\Gamma}}(x^\lambda; j^1\pi)ds$.

Equation (4.30) ensure us that the differentials $d\Theta(\mathcal{L}_H)$ and $d\Theta(\mathcal{L}_{\overset{*}{\Gamma}})$ are the same. This is enough to guarantee the dynamical equivalence of the two Lagrangians, since it ensures that they will give rise to the same critical sections in Lor(M). Moreover, the global decomposition (4.32) reflects the likewise global decomposition (4.20) at the Lagrangian level.

Applications of this first-order Lagrangian and of its Poincaré-Cartan form to the problem of gravitational energy can be found in **[6, 7]**.

REFERENCES

1. ARNOWITT R., DESER S., MISNER C.W., J. Math. Phys. **1**, 434 (1960); ARNOWITT R., DESER S., MISNER C.W., *The Dynamics of General Relativity,* in: «Gravitation: An Introduction to Current Research»; L. Witten ed.; Wiley (New York, 1962), pp. 227-265.

2. DE WITT B., Phys. Rev. **160**, 1113 (1967); Phys. Rev. **162**, 1195 (1967); Phys. Rev. **162**, 1239 (1967).

3. EINSTEIN A., Ann. der Physik **49**, 769 (1916).

4. EINSTEIN A., Sitzungsber. Preuss. Akad. Wiss. (Berlin) **2**, 778 (1915).

5. FERRARIS M., *Fibered Connections and the Global Poincaré-Cartan Form in Higher Order Calculus of Variations,* in: "Proceedings of the Conference on Differential Geometry and its Applications", Part II (Geometrical Methods in Physics); D. Krupka ed.; University J.E. Purkyne (Brno, Czechoslovakia, 1984), pp. 61-91.

6. FERRARIS M., FRANCAVIGLIA M., *New Superpotentials in General Relativity,* in: Proceedings of the International Symposium on "Space-time Symmetries" (U. of Maryland, College Park, 1988); S. Kim & W.W. Zachary eds.; Nucl. Phys. B (Proc. Suppl.) **6**, 405 (1989); North-Holland (Amsterdam, 1989). FERRARIS M., FRANCAVIGLIA M., *First-Order Lagrangians, Energy-Density and Superpotentials in General Relativity,* GRG Journ. (1990, in print).

7. FERRARIS M., FRANCAVIGLIA M., *The Lagrangian Approach to Conservation Laws in General Relativity,* in: "Mechanics, Analysys and Geometry: 200 Hundred Years after Lagrange"); M. Francaviglia & D.D. Holm eds.; Delta Series, North-Holland (Amsterdam, 1990, in print).

8. FERRARIS M., FRANCAVIGLIA M., *Energy-Momentum Tensors and Stress Tensors in Geometric Field Theories,* J. Math. Phys. **26**, 1243 (1985).

9. FERRARIS M., FRANCAVIGLIA M., MAGNANO G., J. Math. Phys. **31** (2), 378-387 (1990).

10. FERRARIS M., FRANCAVIGLIA M., MOTTINI M., *Energy Formulae for General Relativity in Tetrad Formalism,* (in preparation).

11. FERRARIS M., FRANCAVIGLIA M., ROBUTTI O., *Energy and Super-potentials in Gravitational Field Theories,* in: «Atti del 6° Convegno Nazionale di Relatività Generale e Fisica della Gravitazione (Firenze, 1984)»; M. Modugno ed., Pitagora Editrice (Bologna, 1986), pp. 137-150.

FERRARIS M., FRANCAVIGLIA M., ROBUTTI O., *On the Notion of Energy and the Existence of Superpotentials in Gravitational Theories,* in: "Géométrie et Physique", Proceedings of the "Journées Relativistes 1985" (Marseille, 1985); Y. Choquet-Bruhat, B. Coll, R. Kerner, A. Lichnerowicz eds.; Hermann (Paris, 1987), pp. 112-125.

12. FISCHER A.E., MARSDEN J., Journ. Math. Phys. **13**, 546 (1972).

13. FRANCAVIGLIA M., *Elements of Differential and Riemannian Geometry,* Monographs and Textbooks in Physical Sciences, Lecture Notes **4** (Proceedings Spring School on "Geometrical Methods in Theoretical Physics", Ferrara 1987), Bibliopolis (Napoli, 1988).

14. FRANCAVIGLIA M., *Relativistic Theories (the Variational Formulation),* Lecture Notes 13th Summer School of GNFM-CNR, Ravello 1989; Quaderni CNR (1990, in print).

15. GARCIA P.L., Collect. Math. (Barcelona) **19**, 73 (1968).

16. GARCIA P.L., *The Poincaré-Cartan Invariant in the Calculus of Variations,* in: "Symposia Mathematica, Vol. XIV" (INDAM), Academic Press (London, 1974), pp. 219-246.

17. GARCIA P.L., MUÑOZ J., *On the Geometrical Structure of Higher Order Variational Calculus,* in: "Modern Developments in Analytical Mechanics" (Proceedings IUTAM-ISIMM Symposium, Torino, 1982); Benenti S., Francaviglia M. & Lichnerowicz A. eds.; Acc. Sci. Torino (Torino, 1983), pp.127-147.

18. GARCIA P.L., MUÑOZ J., C.R. Ac. Sci. Paris **301** (1), 639 (1985).

19. GOLDSCHMIDT H., STERNBERG S., Ann. Inst. Fourier (Grenoble) **23**, 203 (1967).

20. GOTAY J.M., SHADWICK W.F., *An Exterior Differential Systems Approach to the Cartan Form in the Calculus of Variations,* (to appear, 1990).

21. GOTAY J.M., *A Multisymplectic Framework for Classical Field Theory and the Calculus of Variations,* in: "Mechanics, Analysys and Geometry: 200 Hundred Years after Lagrange"); M. Francaviglia & D.D. Holm eds.; Delta Series, North-Holland (Amsterdam, 1990, in print).

22. HANSON A., REGGE T., TEITELBOIM C., *Constrained Hamiltonian Systems;* Contributi del Centro Linceo Internazionale di Scienze Matematiche e Loro Applicazioni, n. 22; Accademia Nazionale dei Lincei (Roma, 1976).
REGGE T., TEITELBOIM C., Ann. Phys. (N. Y.) **88**, 286 (1974).

23. HILBERT D., Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl. 395 (1915).

24. KOLAR I., Journ. Geom. Phys. **1**, (2), 127 (1984);
HORAK M., KOLAR I., Czech. Math. J. **33**, 108 n.o 3, 467 (1983).

25. KRUPKA D., *Regular Lagrangians and Lepagean Forms,* in: "Differential Geometry and its Applications, Vol. I", Proceedings Brno 1986; Krupka D. S. Svec eds.; Reidel (Dordrecht, 1987), pp. 111-148.

26. KRUPKA D., *Lepagean Forms in Higher Order Variational Theory,* in: "Modern Developments in Analytical Mechanics" (Proceedings IUTAM-ISIMM Symposium, Torino, 1982); Benenti S., Francaviglia M. & Lichnerowicz A. eds.; Acc. Sci. Torino (Torino, 1983), pp.197-238.

27. LANDAU L.D., LIFCHITZ E.M., *Physique Théorique t. II: Théorie des Champs,* Third revised edition (MIR, Moscow, 1970).

28. MOTTINI M., Dr. Dissertation in Physics, Univ. of Milano (1990); unpublished.

29. SHADWICK W., Lett. Math. Phys. **5**, 409-416 (1982).

30. SZCZYRBA W., Commun. Math. Phys. **60**, 215 (1978); J. Math. Phys. **22**, 1926 (1981); J. Math. Phys. **28**, 146 (1987).

Acknowledgements

This work is partially supported by G.N.F.M.-C.N.R., I.N.F.N. and by the Research Project 40% "Geometria e Fisica" of M.U.R.S.T.

Marco Ferraris
Dipartimento di Matematica
Università di Cagliari
Via Ospedale 72
09100 CAGLIARI (ITALY)

Mauro Francaviglia
Istituto di Fisica Matematica "J.-L. Lagrange"
Università di Torino
Via C. Alberto 10
10123 TORINO (ITALY)

Constantin Carathéodory: An International Tribute (pp. 312–330)
edited by Th. M. Rassias

Ω- ADDITIVE FUNCTIONS ON TOPOLOGICAL GROUPS

Gian Luigi Forti - Luigi Paganoni *

ABSTRACT. In this paper we describe by means of local and global homomorphisms the solutions of the functional equations

$$f(x)f(y) = f(xy) \quad , \quad f_1(x)f_2(y) = f_3(xy) \qquad (x,y) \in \Omega \, ,$$

where f, f_1, f_2, f_3 are functions from a topological group X into a group S and Ω is a subset of X^2 with non-empty interior.

1. Introduction

In the last years Cauchy functional equations on restricted domains have been extensively studied : for a rich bibliography see Ref. 1, 4, 12, 13 (see also Ref. 5, 8, 14, 17, 18). Among the aims of these researches one is to establish conditions which guarantee that all solutions are homomorphisms. A field of

* Partially supported by M.P.I. : Research funds (60%).

research as much rich of results concerns alternative functional equations (see Ref. 1, 2, 4, 6, 7, 9, 10, 12, 13, 15) and the methods used to investigate both problems have many connections. In particular the search for the solutions of the alternative Cauchy equation

$$f(x)f(y) \neq f(xy) \qquad \text{implies} \qquad g(x)g(y) = g(xy)$$

requires a preliminary study about the solutions of the Cauchy equation on an open subset of a topological group.

In the present paper the following Cauchy and Pexider equations

$$f(x)f(y) = f(xy) \quad , \quad (x,y) \in \Omega \tag{1}$$

$$f_1(x)f_2(y) = f_3(xy) \quad , \quad (x,y) \in \Omega \tag{2}$$

are studied, where f, f_1, f_2, f_3 are functions from a topological group $(X,\cdot)$ into a group $(S,\cdot)$, under the fundamental assumption that the set Ω where the equations are satisfied has non-empty interior.

We describe, by using homomorphisms, the solutions of (1) and (2) on suitable projections of the interior of Ω. Starting point of this paper are some results due to L. Giudici (personally communicated to the authors) which have been presented to an international meeting on Functional Equations (see Ref. 16).

2. Notations and Preliminaries

Here we present the notations we shall use in the following.

$\mathcal{O}$ denotes the family of all open neighbourhoods of the identity element e in the topological group X and $\mathcal{U} \subset \mathcal{O}$ is a fundamental system of neighbourhoods of e.

If $Y \subset X$, Y° is the interior of Y.

By p_1, p_2, p_3 we denote the continuous and open functions from X^2 into X given by

$$p_1(x,y) = x \quad , \quad p_2(x,y) = y \quad , \quad p_3(x,y) = xy \, .$$

If $B \subset X^2$, we define

$$\pi(B) := p_1(B) \cup p_2(B) \cup p_3(B) \, .$$

Let $(x_0, y_0) \in X^2$ and $U \in \mathcal{O}$; define

$$\Gamma_{(x_0,y_0)}(U) := \{(x,y) \in X^2 : x \in x_0 U \,,\, y \in U y_0 \,,\, xy \in x_0 U y_0\}$$
$$= p_1^{-1}(x_0 U) \cap p_2^{-1}(U y_0) \cap p_3^{-1}(x_0 U y_0) \,.$$

$\Gamma_{(x_0,y_0)}(U)$ is an open neighbourhood of (x_0, y_0) . If $x_0 = y_0 = e$ we simply write $\Gamma(U)$ instead of $\Gamma_{(e,e)}(U)$.

A local homomorphism from $U \in \mathcal{O}$ into S is a function $a : U \to S$ such that

$$a(x)a(y) = a(xy) \qquad , \qquad (x,y) \in \Gamma(U) \,.$$

$\mathrm{Hom}_U(X,S)$ is the set of all local homomorphisms from U into S .

For a given group S , $\mathcal{R}(S)$ denotes the family of all topological groups X with the following property :

> there exists a fundamental system $\mathcal{U}$ of open neighbourhoods of e such that every $a \in \mathrm{Hom}_U(X,S)$, $U \in \mathcal{U}$, is the restriction of a suitable $b \in \mathrm{Hom}(X,S)$.

$\mathcal{R}_0(S)$ is the subset of $\mathcal{R}(S)$ of the groups without open proper subgroups.

A function $f : Y \to S$, $Y \subset X$, is called locally affine in the point $x_0 \in Y$ if $U \in \mathcal{O}$, $U \subset Y$, and $a \in \mathrm{Hom}_U(X,S)$ exist such that

$$f(x_0 t) = f(x_0)a(t) \qquad , \qquad t \in U \,.$$

Given a function $f : X \to S$ we define the following three sets :

$$E_f := \{x \in X : f \text{ is locally affine in } x\}$$
$$\Lambda_f := p_1^{-1}(E_f) \cap p_2^{-1}(E_f) \cap p_3^{-1}(E_f)$$
$$A_f := \{(x,y) \in X^2 : f(x)f(y) = f(xy)\} \,.$$

Lemma 1. *Assume f is locally affine in x_0 with $a \in Hom_U(X,S)$. Then*

$$f(tx_0) = b(t)f(x_0) \qquad \text{with} \qquad b \in Hom_{x_0 U x_0^{-1}}(X,S)$$

and, if $u_0 v_0 = x_0$,

$$f(u_0 t v_0) = f(u_0)\gamma c(t) f(v_0) \quad , \text{ with } \quad c \in Hom_{v_0 U v_0^{-1}}(X,S) \text{ and } \gamma \in S \,.$$

The converse is also true.

Proof. If $a \in \mathrm{Hom}_U(X,S)$ and $\sigma \in S$ then the function

$$x \longmapsto \sigma^{-1} a(x_0^{-1} x x_0)\,\sigma$$

is a local homomorphism on $x_0 U x_0^{-1}$. Let $x_0 \in E_f$ and $t \in x_0 U x_0^{-1}$, then

$$f(tx_0) = f(x_0 x_0^{-1} t x_0) = f(x_0) a(x_0^{-1} t x_0) = f(x_0) a(x_0^{-1} t x_0)[f(x_0)]^{-1} f(x_0) =$$
$$= b(t) f(x_0) \qquad , \qquad b \in \mathrm{Hom}_{x_0 U x_0^{-1}}(X,S)\,.$$

If $t \in v_0 U v_0^{-1}$,

$$f(u_0 t v_0) = f(u_0 v_0 v_0^{-1} t v_0) = f(u_0 v_0) a(v_0^{-1} t v_0) =$$
$$= f(u_0)\{[f(u_0)]^{-1} f(u_0 v_0)[f(v_0)]^{-1}\} f(v_0) a(v_0^{-1} t v_0)[f(v_0)]^{-1} f(v_0) =$$
$$= f(u_0)\,\gamma\, c(t) f(v_0) \quad , \quad c \in \mathrm{Hom}_{v_0 U v_0^{-1}}(X,S)\,.$$

The converse follows immediately.

Lemma 2. *The sets E_f and Λ_f are open.*

Proof. Let $x_0 \in E_f$, then there exist $U \in \mathcal{O}$ and $a \in \mathrm{Hom}_U(X,S)$ such that

$$f(x_0 t) = f(x_0) a(t) \qquad , \qquad t \in U\,.$$

If $t_0 \in U$ and we choose $V \in \mathcal{O}$ with $V \subset U$ and $t_0 V \subset U$, then

$$f(x_0 t_0 v) = f(x_0) a(t_0 v) = f(x_0) a(t_0) a(v) = f(x_0 t_0) a(v) \qquad , \qquad v \in V$$

and so $x_0 U \subset E_f$. Λ_f is open since p_1, p_2, p_3 are continuous.

Lemma 3. *Let X be a group without open proper subgroups. If $a, b \in \mathrm{Hom}(X,S)$ and $\alpha, \beta \in S$ (S is any group) are such that*

$$\alpha\, a(x) = \beta\, b(x)$$

on a non-empty open set O, then $\alpha = \beta$ and $a \equiv b$.

Proof. Fix $x_0 \in O$ and $U \in \mathcal{O}$ such that $x_0 U \subset O$. For every $t \in U$ we have

$$\alpha\, a(x_0)a(t) = \alpha\, a(x_0 t) = \beta\, b(x_0 t) = \beta\, b(x_0)b(t)$$

so $a(t) = b(t)$ for $t \in U$.

By our assumption on X we have $X = \bigcup_{n \geq 1} U^n$. If $x \in X$ then $x = t_1 t_2 \cdots t_n$ with $t_i \in U$, thus

$$a(x) = a(t_1 \cdots t_n) = a(t_1) \cdots a(t_n) = b(t_1) \cdots b(t_n) = b(t_1 \cdots t_n) = b(x)\,,$$

i.e. $a \equiv b$ and $\alpha = \beta$.

3. Main Results

This section contains the main results about local and global solutions of equations (1) and (2). Some results of this section extend to a more general setting known results obtained in the special case of euclidean spaces (see Ref. 1, 13).

Theorem 1. *Fix $(x_0, y_0) \in X^2$ and $U \in \mathcal{O}$. The functions*

$$f_1 : x_0 U \to S \quad , \quad f_2 : U y_0 \to S \quad , \quad f_3 : x_0 U y_0 \to S$$

are solutions of the equation

$$f_1(x) f_2(y) = f_3(xy) \quad , \quad (x, y) \in \Gamma_{(x_0, y_0)}(U) \tag{3}$$

if and only if they have the form

$$f_1(x_0 t) = \alpha\, a(t) \quad , \quad f_2(t y_0) = a(t)\, \beta \quad , \quad f_3(x_0 t y_0) = \alpha\, a(t)\, \beta \quad (t \in U) \tag{4}$$

where $a \in Hom_v(X, S)$ and $\alpha, \beta \in S$.

Proof. If f_1, f_2, f_3 have the form (4) then obviously they solve equation (3). Conversely assume that f_1, f_2, f_3 solve (3) and set

$$\alpha := f_1(x_0)\,,\ \beta := f_2(y_0)\,,\ a(t) := \alpha^{-1} f_3(x_0 t y_0)\beta^{-1} \quad , \quad t \in U\,.$$

Replacing in (3) (x,y) with (x_0t,y_0) and (x_0,ty_0), $t\in U$, we obtain

$$f_1(x_0t)\beta = f_1(x_0t)f_2(y_0) = f_3(x_0ty_0) = \alpha a(t)\beta$$

and

$$\alpha f_2(ty_0) = f_1(x_0)f_2(ty_0) = f_3(x_0ty_0) = \alpha a(t)\beta$$

whence

$$f_1(x_0t) = \alpha a(t) \quad , \quad f_2(ty_0) = a(t)\beta \ .$$

Furthermore

$(t,u)\in\Gamma(U)$ implies $\alpha a(t)a(u)\beta = f_1(x_0t)f_2(uy_0) = f_3(x_0tuy_0) = \alpha a(tu)\beta$,

that is $a(t)a(u) = a(tu)$. Thus $a\in \mathrm{Hom}_U(X,S)$.

The next topological theorem is the main tool for solving equation (1).

Theorem 2. *Let* $f: X\to S$. *Then*

$$A_f^\circ \subset \Lambda_f$$

and A_f° *is closed in* Λ_f *(with respect to the relative topology).*

Proof. Assume we are in the non-trivial case $A_f^\circ\neq\emptyset$. If $(x_0,y_0)\in A_f^\circ$ and we choose $U\in\mathcal{O}$ so that $\Gamma_{(x_0,y_0)}(U)\subset A_f^\circ$ then, by Theorem 1, we have

$$f(x_0t) = \alpha\, a(t) \quad , \quad f(ty_0) = a(t)\,\beta \quad , \quad f(x_0ty_0) = \alpha\, a(t)\,\beta \qquad (t\in U)$$

with $a\in \mathrm{Hom}_U(X,S)$. It follows

$$f(x_0y_0t) = f(x_0y_0ty_0^{-1}y_0) = \alpha\, a(y_0ty_0^{-1})\,\beta = \alpha\beta\,[\beta^{-1}a(y_0ty_0^{-1})\,\beta] = \alpha\beta\, b(t)$$

and

$$f(y_0t) = f(y_0ty_0^{-1}y_0) = a(y_0ty_0^{-1})\,\beta = \beta[\beta^{-1}a(y_0ty_0^{-1})\,\beta] = \beta\, b(t)$$

where $t \in y_0^{-1}Uy_0$ and $b \in \mathrm{Hom}_{y_0^{-1}Uy_0}(X,S)$.
Thus $x_0, y_0, x_0y_0 \in E_f$, i.e. $(x_0,y_0) \in \Lambda_f$, and so $A_f^\circ \subset \Lambda_f$.
Assume now A_f° not closed in Λ_f. Then there is a point $(x_0,y_0) \in \Lambda_f \backslash A_f^\circ$ each neighbourhood of which meets A_f°. By the definition of Λ_f and by Lemma 1 there exists $V \in \mathcal{O}$ such that, for all $t \in V$,

$$\begin{aligned} f(x_0ty_0) &= f(x_0)\gamma a_3(t)f(y_0) \\ f(x_0t) = f(x_0)a_1(t) \quad &, \quad f(ty_0) = a_2(t)f(y_0) \end{aligned} \tag{5}$$

with $a_1, a_2, a_3 \in \mathrm{Hom}_r(X,S)$ and $\gamma \in S$.
Moreover, since $\Gamma_{(x_0,y_0)}(V) \cap A_f^\circ \neq \emptyset$, there exist $(t_1,u_1) \in \Gamma(V)$ and $W \in \mathcal{O}$ such that $(x_0t_1, u_1y_0) \in A_f^\circ$ and $\Gamma_{(x_0t_1,u_1y_0)}(W) \subset \Gamma_{(x_0,y_0)}(V) \cap A_f^\circ$. By Theorem 1 there exists $a \in \mathrm{Hom}_w(X,S)$ such that, for all $t \in W$,

$$f(x_0t_1t) = f(x_0t_1)a(t) \quad , \quad f(uu_1y_0) = a(u)f(u_1y_0)\,. \tag{6}$$

From relations (5) and (6) we obtain $f(x_0t_1t) = f(x_0t_1)a(t) = f(x_0)a_1(t_1)a(t)$ and $f[x_0(t_1t)] = f(x_0)a_1(t_1t) = f(x_0)a_1(t_1)a_1(t)$.
Thus $a(t) = a_1(t)$ for all $t \in W$. Analogously $a(t) = a_2(t)$ on W.
We now confine our considerations to the set $\Gamma_{(x_0,y_0)}(W)$. The same construction as above gives a point $(t_2,u_2) \in \Gamma(W)$ and a set $W_1 \in \mathcal{O}$ such that

$$(x_0t_2, u_2y_0) \in A_f^\circ \quad \text{and} \quad \Gamma_{(x_0t_2,u_2y_0)}(W_1) \subset \Gamma_{(x_0,y_0)}(W) \cap A_f^\circ\,.$$

Therefore for every point $(t,u) \in \Gamma(W_1)$, by (5), we have

$$f(x_0)\gamma a_3(t_2tuu_2)f(y_0) = f(x_0t_2tuu_2y_0) = f(x_0t_2t)f(uu_2y_0)$$

and, since $t_2t, uu_2 \in W$,

$$\begin{aligned} f(x_0t_2t) &= f(x_0)a_1(t_2t) = f(x_0)a(t_2t) \quad , \\ f(uu_2y_0) &= a_2(uu_2)f(y_0) = a(uu_2)f(y_0)\,. \end{aligned}$$

It follows

$$a(t_2t)a(uu_2) = \gamma a_3(t_2tuu_2) \quad , \quad (t,u) \in \Gamma(W_1)\,.$$

Take $t = e$, $W_2 = (u_2^{-1}W_1u_2) \cap W_1$ and $u = u_2wu_2^{-1}$, $w \in W_2$. Then for every $w \in W_2$ we have

$$a(t_2)a(u_2w) = \gamma a_3(t_2u_2w)\,. \tag{7}$$

If we put in (7) $w = e$, we get

$$a(t_2)a(u_2) = a(t_2u_2) = \gamma\, a_3(t_2u_2)\ . \tag{8}$$

Then, by (7) and (8),

$$\gamma\, a_3(t_2u_2)a_3(w) = \gamma\, a_3(t_2u_2w) = a(t_2)a(u_2w) =$$
$$= a(t_2)a(u_2)a(w) = a(t_2u_2)a(w) = \gamma\, a_3(t_2u_2)a(w)$$

and from this it follows

$$a(w) = a_3(w) \quad , \quad w \in W_2\ . \tag{9}$$

Finally, by (7) and (9), we get $\gamma = e$ and so we can conclude that the point (x_0, y_0) belongs to A_f° , contrary to our assumption. By contradiction A_f° is closed in Λ_f .

Theorem 3. *Let $V \in \mathcal{O}$ with $\Gamma(V)$ connected. If a function $\varphi : V \to S$ is locally affine in every point of V then*

$$\varphi(x) = \alpha\, a(x)$$

where $a \in Hom_v(X, S)$ and $\alpha \in S$.

Proof. Define $f(x) := [\varphi(\epsilon)]^{-1}\varphi(x)$; f is a solution of equation (1) in a neighbourhood of (ϵ, e) , hence $A_f^\circ \neq \emptyset$; moreover by hypothesis we have $E_f = V$. By Theorem 2 the set $A_f^\circ \cap \Gamma(V)$ is open and closed in $\Gamma(V)$ thus, since $\Gamma(V)$ is connected, $A_f^\circ \cap \Gamma(V) = \Gamma(V)$. Then $f(x) = a(x)$ with $a \in \mathrm{Hom}_v(X, S)$ and $\varphi(x) = \alpha\, a(x)$, $\alpha = \varphi(\epsilon)$.

The following example shows that in Theorem 3 the connectedness of $\Gamma(V)$ cannot be dispensed with. Moreover the connectedness of $\Gamma(V)$ doesn't follow from that of V .

Example 1. Take $X = \mathbf{T}$ the unidimensional torus which we identify with the interval $[-\frac{1}{2}, \frac{1}{2}) \subset \mathbf{R}$. Let $V = (-\frac{1}{2}, \frac{1}{2})$, then

$$\Gamma(V) = \{(x, y) \in \mathbf{T}^2 : x, y, x + y \in V\}\ .$$

The set $\Gamma(V)$ can be identified with the square $(-\frac{1}{2},\frac{1}{2})^2$ without the two segments $x+y=\pm\frac{1}{2}$ and it is not connected.

Let $\varphi : V \to \mathbf{R}$ be the function defined by $\varphi(x)=x$. The function φ is obviously locally affine in every point of V; nevertheless $\varphi \notin \mathrm{Hom}_{\nu}(\mathbf{T},\mathbf{R})$ since for $x=\frac{1}{4}, y=\frac{1}{4}+\epsilon$ $(0<\epsilon<\frac{1}{4})$ we have $x+y=\epsilon-\frac{1}{2}$. So

$$\varphi(x+y)=\epsilon-\frac{1}{2}\neq\frac{1}{2}+\epsilon=\varphi(x)+\varphi(y)\,.$$

Corollary 1. *Let $V\in\mathcal{O}$ with $\Gamma(V)$ connected. Fix $x_0\in X$ and assume that $\varphi : x_0V\to S$ is locally affine in every point of x_0V. Then*

$$\varphi(x)=\alpha\, a(x_0^{-1}x)$$

where $a\in Hom_{\nu}(X,S)$ and $\alpha\in S$.

Proof. Apply Theorem 3 to the function $\psi : V\to S$ defined by $\psi(x):=\varphi(x_0x)$.

The following two theorems describe the solutions of equations (1) and (2).

Theorem 4. *Let f be a solution of equation (1) on a set Ω with $\Omega^{\circ}\neq\emptyset$. Then f is locally affine in every point of $\pi(\Omega^{\circ})$. Furthermore :*

a) if $e\in\pi(\Omega^{\circ})$ and $\Gamma(\pi(\Omega^{\circ}))$ is connected then $f\in Hom_{\pi(\Omega^{\circ})}(X,S)$;

b) if $\pi(\Omega^{\circ})=X$ and X is connected then $f\in Hom(X,S)$.

Proof. By Theorem 1 and Lemma 1 f is locally affine in every point of $\pi(\Omega^{\circ})$. By Theorem 3 we have

$$f(x)=\alpha\, a(x)\qquad,\qquad a\in \mathrm{Hom}_{\pi(\Omega^{\circ})}(X,S)\,.$$

If $(x_0,y_0)\in\Omega^{\circ}$, then $f(x_0y_0)=\alpha\, a(x_0y_0)=\alpha\, a(x_0)a(y_0)$ and $f(x_0y_0)=$ $=f(x_0)f(y_0)=\alpha\, a(x_0)\,\alpha\, a(y_0)$. So $\alpha=e$.

Example 2. Let $X = \mathbf{Q}$ and

$$\Omega = \{(x,y) \in \mathbf{Q}^2 : x < \sqrt{2}, y < \sqrt{2}, x+y < \sqrt{2}\} \cup \{(x,y) \in \mathbf{Q}^2; x > \sqrt{2}, y > \sqrt{2}\}.$$

In this case Ω is open , $\pi(\Omega) = \mathbf{Q}$ and $\mathbf{Q}$ is not connected. The function

$$f(x) = \begin{cases} 0 & , \quad x < \sqrt{2} \\ x & , \quad x > \sqrt{2} \end{cases}$$

is a solution of equation (1) on Ω but $f \notin \mathrm{Hom}(\mathbf{Q}, \mathbf{R})$.

Theorem 5. *Let (f_1, f_2, f_3) be a solution of equation (2) on a set Ω with $\Omega^\circ \neq \emptyset$.*

If $X \in \mathcal{R}_0(S)$ and at least two of the projections $p_1(\Omega^\circ), p_2(\Omega^\circ), p_3(\Omega^\circ)$ are connected then there exist $a \in Hom(X,S)$, $\alpha, \beta \in S$ such that

$$\begin{cases} f_1(x) = \alpha\, a(x) & , \quad x \in p_1(\Omega^\circ) \\ f_2(x) = a(x)\, \beta & , \quad x \in p_2(\Omega^\circ) \\ f_3(x) = \alpha\, a(x)\, \beta & , \quad x \in p_3(\Omega^\circ) . \end{cases}$$

The representation is unique.

Furthermore when two of the projections equal X the representation above holds if instead of $X \in \mathcal{R}_0(S)$ we require X connected.

Proof. Let $\mathcal{U}$ be a fundamental system of neighbourhoods of the identity of X related to the property $X \in \mathcal{R}(S)$. Assume $p_1(\Omega^\circ)$ and $p_2(\Omega^\circ)$ connected. For every $x \in p_1(\Omega^\circ)$ we choose y such that $(x,y) \in \Omega^\circ$. By Theorem 1, there exist $U_x \in \mathcal{U}$, $\overline{\alpha}_x \in S$ and $a_x \in \mathrm{Hom}_{U_x}(X,S)$ such that

$$f_1(xt) = \overline{\alpha}_x a_x(t) \quad , \quad t \in U_x . \tag{10}$$

Since $X \in \mathcal{R}_0(S)$, a_x can be uniquely extended to a global homomorphism that we still denote by a_x . We claim that a_x doesn't depend on the point $x \in p_1(\Omega^\circ)$. Indeed the subset of $p_1(\Omega^\circ)$ for which (10) holds with the same a_x is open ($f_1(xtv) = \overline{\alpha}_x a_x(tv) = = [\overline{\alpha}_x a_x(t)]\, a_x(v) = \overline{\alpha}_{xt} a_x(v)$); and since $p_1(\Omega^\circ)$ is connected we get

$$f_1(xt) = \overline{\alpha}_x a(t) \quad , \quad t \in U_x .$$

Fix $x_0 \in p_1(\Omega^\circ)$, then

$$f_1(x) = f_1(x_0 t) = \overline{\alpha}_{x_0} a(t) = \overline{\alpha}_{x_0} a(x_0)^{-1} a(x_0) a(t) =$$
$$= \alpha_{x_0} a(x_0 t) = \alpha_{x_0} a(x) \qquad , \qquad t \in U_{x_0} ,$$

so the subset of $p_1(\Omega^\circ)$ where the representation $f_1(x) = \alpha\, a(x)$ holds with the same α is open and, as above, from the connectedness of $p_1(\Omega^\circ)$ we deduce

$$f_1(x) = \alpha\, a(x) \qquad \text{for all } x \in p_1(\Omega^\circ) .$$

Analogously we get

$$f_2(x) = a(x)\beta \qquad \text{for all } x \in p_2(\Omega^\circ) ,$$

and, from equation (2), we get also

$$f_3(x) = \alpha\, a(x)\beta \qquad \text{for all } x \in p_3(\Omega^\circ) .$$

In the other cases the proof is analogous.
The uniqueness of the representation follows from Lemma 3.
Assume now $p_1(\Omega^\circ) = p_2(\Omega^\circ) = X$. By Theorem 1 the functions f_1 and f_2 are locally affine on X . Then, by Theorem 3, $f_1(x) = \alpha\, a(x)$ with $a \in \mathrm{Hom}(X, S)$. Analogously $f_2(x) = \beta\, b(x) = (\beta\, b(x)\, \beta^{-1})\beta = c(x)\,\beta$, $c \in \mathrm{Hom}(X, S)$. By Theorem 1 a and c coincide locally on X ; then, by Lemma 3, $a \equiv c$.

Remark 1. If S is a topological group we can ask for the continuous solutions of equations (1) and (2). It is immediately seen that all results of this section remain valid if local and global homomorphisms are always supposed continuous.

4. On the classes $\mathcal{R}(S)$ and $\mathcal{R}_0(S)$

Most of the results of the previous section give conditions under which a function is representable as a product of a constant time a local homomorphism. It is obviously interesting to know if every local homomorphism is, in a suitable

neighbourhood of the identity, the restriction of a global one, that is, following the notation introduced in Section 2, if $X \in \mathcal{R}(S)$. This property holds for some classes of topological groups (see Ref. 3 and 11); in this section we give some other conditions.

Note that, if $X \in \mathcal{R}_0(S)$, Lemma 3 implies that the global homomorphism is uniquely determined.

Theorem 6. *Let $(S,+)$ be a group (not necessarily commutative) and assume $(X,+)$ is a uniquely p-divisible abelian group. Suppose a subset $Y \subset X$ satisfies the following two conditions :*

i) if $x \in Y$ then $\frac{i}{p}x \in Y$, $1 \leq i \leq p$;

ii) for every $x \in X$ there exists $n_0 \in \mathbf{N}$ such that $\frac{1}{p^{n_0}}x \in Y$.

If $a : Y \to S$ satisfies

$$a(x+y) = a(x) + a(y) \qquad , \qquad x, y, x+y \in Y \tag{11}$$

then there exists $b \in Hom(X,S)$ such that $b_{|Y} = a$.

Proof. If $x \in Y$, by i) and (11) it follows $a(x) = a(p\frac{x}{p}) = p\,a(\frac{x}{p})$. Now, by induction, we have $a(x) = p^n a(\frac{x}{p^n})$, $n \geq 1$. Indeed if we assume this property true for n , then

$$p^{n+1}a(\frac{x}{p^{n+1}}) = p^n\left[pa(\frac{y}{p})\right] = p^n a(y) = p^n a(\frac{x}{p^n}) = a(x) .$$

Given $x \in X$ we define $n_0 := n_0(x) = \min\{n \in \mathbf{N} : \frac{x}{p^n} \in Y\}$ and $b(x) := p^{n_0}a(\frac{x}{p^{n_0}})$. By (ii) the function b is defined on the whole X and $b_{|Y} = a$. Let $z \in X$, then for every $n \geq n_0(z)$ we have

$$b(z) = p^{n_0}a(\frac{z}{p^{n_0}}) = p^{n_0}\left[p^{n-n_0}a(\frac{z}{p^n})\right] = p^n a(\frac{z}{p^n}) .$$

Given $x, y \in X$ let $\nu := \max\{n_0(x), n_0(y), n_0(x+y)\}$, then

$$b(x+y) = p^\nu a(\frac{x+y}{p^\nu}) = p^\nu a(\frac{x}{p^\nu} + \frac{y}{p^\nu}) = p^\nu\left[a(\frac{x}{p^\nu}) + a(\frac{y}{p^\nu})\right] =$$

$$= p^\nu a(\frac{x}{p^\nu}) + p^\nu a(\frac{y}{p^\nu}) = b(x) + b(y),$$

thus $b \in \mathrm{Hom}(X, S)$ (Note that in the previous chain of equalities we used the property that, by (11), $a(\frac{x}{p^\nu})$ and $a(\frac{y}{p^\nu})$ commute).

Corollary 2. *Let $(X, +)$ be a topological torsion-free divisible group and assume it has a fundamental system $\mathcal{U}$ of absorbing neighbourhoods of the identity, i.e.*

$$\text{for each } V \in \mathcal{U} \text{ and } x \in X \text{ there exists } r \in \mathbf{Q}^+ \text{ such that}$$
$$sx \in V \qquad \text{for all rationals } 0 < s \leq r .$$

Then $X \in \mathcal{R}(S)$ for each group S .

Proof. Each $V \in \mathcal{U}$ satisfies properties (i) and (ii) of Theorem 6.

Corollary 3. *Every topological vector space X over $\mathbf{Q}$ belongs, as an additive group, to $\mathcal{R}_0(S)$ for each group S .*

Proof. X has a fundamental system $\mathcal{U}$ of absorbing neighbourhoods of the origin, so each $V \in \mathcal{U}$ satisfies conditions (i) and (ii) of Theorem 6 and $X = \bigcup_{n \geq 1} nV$.

Remark 2 . The following example shows that if $X \in \mathcal{R}(S)$ and $\varphi : X \to Y$ is a continuous homomorphism, it is not generally true that $Y \in \mathcal{R}(S)$ as well.
Take $X = \mathbf{R} \in \mathcal{R}(\mathbf{R})$ (Corollary 3) and $Y = \mathbf{T}$, the unidimensional torus which we identify with the interval $[-\frac{1}{2}, \frac{1}{2}) \subset \mathbf{R}$. Let $V = (-\frac{1}{4}, \frac{1}{4})$ and $a : V \to \mathbf{R}$ given by $a(x) = x$. Obviously $a \in \mathrm{Hom}_V(\mathbf{T}, \mathbf{R})$ and it is continuous. Any $b \in \mathrm{Hom}(\mathbf{T}, \mathbf{R})$ extension of a must be continuos, but it is well known that this implies $b \equiv 0$ (see Ref. 3). Thus $\mathbf{T} \notin \mathcal{R}(\mathbf{R})$.

Remark 3 . The class $\mathcal{R}_0(S)$ is in general a proper subset of $\mathcal{R}(S)$ as we can see if we take $X = S = \mathbf{R}^*$ (the multiplicative group of the reals).

Obviously $\mathbf{R}^* \notin \mathcal{R}_0(\mathbf{R}^*)$ since $\mathbf{R}_+^*$ is an open subgroup of $\mathbf{R}^*$. Nevertheless, by Corollary 2, $\mathbf{R}_+^* \in \mathcal{R}(\mathbf{R}^*)$ and each $a \in \mathrm{Hom}(\mathbf{R}_+^*, \mathbf{R}^*)$ can be extended to $\bar{a} \in \mathrm{Hom}(\mathbf{R}^*, \mathbf{R}^*)$ by defining

$$\bar{a}(-x) = a(x) \quad , \quad x \in \mathbf{R}_+^* .$$

Note that the extension is not unique because we could also take

$$\bar{a}(-x) = -a(x) \quad , \quad x \in \mathbf{R}_+^* .$$

Theorem 7. *Let X be a topological group and assume there exists $U \in \mathcal{O}$ with the following properties :*

i) $\bigcup_{n \geq 1} U^n = X$;

ii) for each $x \in X$ and for each $n \geq 1$ the set $U^n \cap xU^{-1}$ is connected (possibly empty).

Then if $a \in Hom_U(X, S)$, where S is any group, there exists $b \in Hom(X, S)$ such that $a = b_{|U}$.

If there exists a fundamental system of neighbourhoods $\mathcal{U}$ such that every $U \in \mathcal{U}$ satisfies conditions (i) and (ii) then $X \in \mathcal{R}_0(S)$ for every group S.

Proof. Let $a \in \mathrm{Hom}_U(X, S)$, i.e.

$$a(xy) = a(x)a(y) \quad , \quad (x, y) \in \Gamma(U) .$$

We define inductively a sequence of functions $a_n : U^n \to S$ in the following way :

$$a_1(t) := a(t) ;$$
$$a_{n+1}(t) := a_n(x)a_1(y) \quad , \quad x \in U^n , \ y \in U , \ t = xy \in U^{n+1} .$$

Note that $a_{n+1|U^n} = a_n$. We must prove that the functions a_n, $n \geq 2$, are well defined. For each $n \geq 1$ and $t \in U^{n+1}$ set

$$X_t := \{(x, y) \in X \times X \ : \ xy = t\}$$
$$T_n(t) := \{(x, y) \in X_t \ : \ x \in U^n, y \in U\} .$$

We have $T_n(t) = \{(x, x^{-1}t) : x \in U^n \cap tU^{-1}\}$, so $p_1(T_n(t)) = U^n \cap tU^{-1}$ and by hypothesis it is connected . Since p_1 is a homeomorphism of X_t onto X , $T_n(t)$ is an open connected set in X_t (with the induced topology).
First we prove that a_2 is well defined. Fix $t \in U^2$ and let $(x_1, y_1) \in T_1(t)$, choose now $V \in \mathcal{O}$, $V \subset U$ symmetric and such that if $\epsilon \in V$ then

$$(x, y) = (x_1\epsilon, \epsilon^{-1}y_1) \in T_1(t) .$$

We have

$$a_1(x)a_1(y) = a_1(x_1\epsilon)a_1(\epsilon^{-1}y_1) = a_1(x_1)a_1(\epsilon)a_1(\epsilon^{-1})a_1(y_1) = a_1(x_1)a_1(y_1) .$$

This means that, for each $\lambda \in S$, $\{(x, y) \in T_1(t) : a_1(x)a_1(y) = \lambda\}$ is open in $T_1(t)$. Since $T_1(t)$ is connected, it follows that,

$$\text{for each pair } (x, y), (x_0, y_0) \in T_1(t) \quad , \quad a_1(x)a_1(y) = a_1(x_0)a_1(y_0) ,$$

that is a_2 is well defined. We now proceed by induction assuming $a_n, n \geq 2$, well defined. Let $t \in U^{n+1}$ and $(x_1, y_1) \in T_n(t)$, where $x_1 = uv$, $u \in U^{n-1}$, $v \in U$. Choose $V \in \mathcal{O}$, $V \subset U$ symmetric and such that, if $\epsilon \in V$ then

$$(x, y) = (x_1\epsilon, \epsilon^{-1}y_1) \in T_n(t) \text{ and } v\epsilon \in U .$$

We have

$$\begin{aligned} a_n(x)a_1(y) &= a_n(x_1\epsilon)a_1(\epsilon^{-1}y_1) = a_n(uv\epsilon)a_1(\epsilon^{-1}y_1) = a_{n-1}(u)a_1(v\epsilon)a_1(\epsilon^{-1}y_1) \\ &= a_{n-1}(u)a_2(vy_1) = a_{n-1}(u)a_1(v)a_1(y_1) = a_n(x_1)a_1(y_1) \end{aligned}$$

and, as for $n = 2$, we conclude that a_{n+1} is well defined. Let now $b : X \to S$ be the inductive limit of the sequence $\{a_n\}$. Obviously $b_{|U} = a$ and $b(xy) = b(x)b(y)$ for all $x \in X$ and $y \in U$. If we take $y \in X$ then, by (i), it is $y = y_1y_2 \cdots y_n$ for some n and with $y_i \in U$. So

$$b(xy) = b(xy_1y_2 \cdots y_n) = b(xy_1 \cdots y_{n-1})b(y_n) = \cdots = b(x)b(y) ,$$

that is $b \in \mathrm{Hom}(X, S)$.

The next result shows the possibility of constructing groups in $\mathcal{R}_0(S)$ starting from other groups in the same set. We recall here the definition of

topological semi-direct product of groups; for all details see Ref. 3. Let N and L be two topological groups with identities e' and e'' respectively and let $\mathcal{A}$ be the group of automorphisms of the (non-topological) group structure of N.

Let $\sigma : L \to \mathcal{A}$ be a homomorphism and suppose that the map

$$(x, y) \to \sigma_y(x)$$

of $N \times L$ into N is continuous. On $G = N \times L$ the following internal law of composition

$$(x, y)(x', y') = (x\sigma_y(x'), yy')$$

defines a group structure compatible with the product topology on $N \times L$. Moreover the canonical injections

$$j_1 : N \to G \qquad \text{and} \qquad j_2 : L \to G$$

are bicontinuous isomorphisms of N and L onto $j_1(N)$ and $j_2(L)$ respectively. The topological group G defined above is called semi-direct product of N and L relative to σ.

Theorem 8. *Let $X_1, X_2 \in \mathcal{R}_0(S)$ and let X be the topological semi-direct product of X_1 and X_2 relative to a given σ. Then $X \in \mathcal{R}_0(S)$.*

Proof. Let $\mathcal{U}_1$, $\mathcal{U}_2$ be fundamental systems of neighbourhoods of the identities $e' \in X_1$ and $e'' \in X_2$ related to the property $X_1, X_2 \in \mathcal{R}_0(S)$. Let U be a neighbourhood of the identity $(e', e'') \in X$ of the form $U = U_1 \times U_2$ with $U_1 \in \mathcal{U}_1$ and $U_2 \in \mathcal{U}_2$ symmetric. Take $a \in \mathrm{Hom}_U(X, S)$. Since $X_1 \times \{e''\}$ and $\{e'\} \times X_2$ are isomorphic to X_1 and X_2 respectively, then $a_{|U_2 \times \{e''\}}$ and $a_{|\{e'\} \times U_2}$ are local homomorphisms in X_1 and X_2, so there are unique $b_1 \in \mathrm{Hom}(X_1, S)$ and $b_2 \in \mathrm{Hom}(X_2, S)$ such that

$$b_{1|U_1} = a_{|U_1 \times \{e''\}} \qquad , \qquad b_{2|U_2} = a_{|\{e'\} \times U_2}$$

Define $b : X \to S$ as follows

$$b(x_1, y_1) := b_1(x_1)b_2(y_1) .$$

We have to prove that $b \in \mathrm{Hom}(X, S)$. If (x_1, y_1), $(x_2, y_2) \in X$ then

$$(x_1, y_1)(x_2, y_2) = (x_1\sigma_{y_1}(x_2), y_1y_2)$$

and

$$b\big((x_1,y_1)(x_2,y_2)\big) = b\big((x_1\sigma_{y_1}(x_2), y_1y_2)\big) = b_1(x_1\sigma_{y_1}(x_2))b_2(y_1y_2) =$$
$$= b_1(x_1)b_1(\sigma_{y_1}(x_2))b_2(y_1)b_2(y_2)\ .$$

To get our goal we have to prove that

$$b_1(\sigma_{y_1}(x_2)) = b_2(y_1)b_1(x_2)b_2(y_1^{-1})\ .$$

Since the function $(x,y)\longmapsto \sigma_y(x)$ is continuous on $X_1\times X_2$, we can find V_1 and V_2, neighbourhood of e' and e'' respectively, such that $(x,y)\in V_1\times V_2$ implies $\sigma_y(x)\in U_1$. Let now $x\in U_1\cap V_1$ and $y\in U_2\cap V_2$; then

$$b_1(\sigma_y(x)) = a(\sigma_y(x),e'') = a\big((\sigma_y(x),y)\cdot(e',y^{-1})\big) = a\big((e',y)(x,e'')\big)a(e',y^{-1}) =$$
$$= a(e',y)a(x,e'')a(e',y^{-1}) = b_2(y)b_1(x)b_2(y^{-1})\ .$$

For a fixed $y\in U_2\cap V_2$ the functions $\beta_1 : X_1\to S$ and $\beta_2 : X_1\to S$ defined by

$$\beta_1(x) := b_1(\sigma_y(x)) \qquad , \qquad \beta_2(x) := b_2(y)b_1(x)b_2(y^{-1})$$

are both homomorphisms of X_1 into S and agree on $U_1\cap V_1$.

Since $X_1\in\mathcal{R}_0(S)$, from Lemma 3, we have $\beta_1\equiv\beta_2$, i.e.

$$b_1(\sigma_y(x)) = b_2(y)b_1(x)b_2(y^{-1}) \tag{12}$$

for each fixed y in $U_2\cap V_2$ and for all $x\in X_1$. Since $X_2 = \bigcup_{n\geq 1}(U_2\cap V_2)^n$, to finish it is enough to prove by induction that (12) holds for every $y\in(U_2\cap V_2)^n$.

The property is true for $n=1$; assume (12) true for n and take $y\in(U_2\cap V_2)^{n+1}$. Then $y=y_1y_2$ with $y_1\in(U_2\cap V_2)^n$ and $y_2\in U_2\cap V_2$, thus we have

$$b_1\big(\sigma_y(x)\big) = b_1\big(\sigma_{y_1y_2}(x)\big) = b_1\big(\sigma_{y_1}(\sigma_{y_2}(x))\big) = b_2(y_1)b_1\big(\sigma_{y_2}(x)\big)b_2(y_1^{-1}) =$$
$$= b_2(y_1)b_2(y_2)b_1(x)b_2(y_2^{-1})b_2(y_1^{-1}) = b_2(y)b_1(x)b_2(y^{-1})\ .$$

Since it is easily seen that $\bigcup_{n\geq 1}U^n = X$, we have $X\in\mathcal{R}_0(S)$.

REFERENCES

1 Aczél J.; Dhombres, J., "Functional Equations Containing Several Variables", Encyclopedia of Mathematics and its Applications, **30**, Cambridge Univ. Press, (1988).

2 Borelli Forti, C.; Forti, G. L., *On a class of alternative functional equations of Cauchy type*, in "Topics in Mathematical Analysis", T. Rassias (Ed.), World Scientific Publ. Co. ,Singapore, 273–293, (1989).

3 Bourbaki, N., "General Topology, Part 1,2", Hermann, Paris, (1966).

4 Dhombres, J., "Some aspects of Functional Equations", Chulalongkorn University Press, Bangkok, (1979).

5 Dhombres, J.; Ger, R., *Conditional Cauchy equations*, Glasnik Mat., **13** (33), 39–62 (1978).

6 Forti, G. L., *La soluzione generale dell'equazione funzionale* $\{cf(x+y)-af(x)-bf(y)-d\}\{f(x+y)-f(x)-f(y)\}=0$, Le Matematiche **34**, 219–242 (1979).

7 Forti, G. L., *On an alternative functional equation related to the Cauchy equation*, Aequationes Math. **24**, 195–206 (1982).

8 Forti, G. L., *On some conditional Cauchy equations on thin sets*, Boll. Un. Mat. Ital. **6 2–B**, 391–402 (1983).

9 Forti, G. L., *The stability of homomorphisms and amenability, with applications to functional equations*, Abh. Math. Sem. Univ. Hamburg **57**, 215–226 (1987).

10 Forti, G. L.; Paganoni, L., *A Method for Solving a Conditional Cauchy Equation on Abelian Groups*,Ann. Mat. Pura Appl. (4) **127**, 79–99 (1981).

11 Hochschild, G., "La structure des groupes de Lie", Dunod, Paris, (1968).

12 Kuczma , M., *Functional Equations on Restricted Domains*, Aequationes Math. **18**, 1–34 (1978).

13 Kuczma , M., "An Introduction to the Theory of Functional Equations and Inequalities. Cauchy's Equation and Jensen's Inequality", Universytet Ślaski, Warszawa–Kraków (1985).

14 Paganoni, L., *Soluzione di una equazione funzionale su dominio ristretto*, Boll. Un. Mat. Ital. (5) **17–B**, 979–993 (1980).

15 Paganoni, L., *On an alternative Cauchy equation* Aequationes Math. **29**, 214–221 (1985).

16 Paganoni, L., Remark 23 in "The Twenty–sixth International Symposium on Functional Equations, April 24– May 3 , 1988", Aequationes Math. **37**, 111 (1989).

17 Paganoni, L. ; Paganoni Marzegalli, S. *Cauchy's functional equation on semigroups*, Fund. Math. **110**, 63–74 (1980).

18 Paganoni Marzegalli, S., *Cauchy's Equation on a Restricted Domain*, Boll. Un. Mat. Ital. (4) **14–A**, 398–408 (1977).

1980 *Mathematics subject classifications :* 39B20 , 39B30 , 39B50 , 39B70 .

Gian Luigi Forti — Luigi Paganoni
Dipartimento di Matematica
Università degli Studi di Milano
Via C. Saldini 50
20133 , Milano
Italia

Constantin Carathéodory: An International Tribute (pp. 331–343)
edited by Th. M. Rassias

THE BRST FORMALISM AND THE QUANTIZATION OF HAMILTONIAN SYSTEMS WITH FIRST CLASS CONSTRAINTS

J. GAMBOA
and
V.O. RIVELLES

After a brief review of the Batalin - Fradkin - Vilkovisky formalism (BFV) , we quantize the bosonic and fermionic relativistic particles. Several points not discussed in the literature are pointed out and we find the correct expressions for the Feynman propagator.

INTRODUCTION

Gauge invariance plays an important role in the present theoretical physics. In the past gauge invariance permited to solve important problems in quantum field theory and particle physics [1].

Along the hamiltonian lines the gauge symmetry appears when the theory under study has first class constraints [2].

We suppose that when some theory is given with $\phi_a(p,q) = 0$ then we say that $\phi_a(p,q)$ is a first class contraint iff,

$$[\phi_a, \phi_b] = C^c_{ab}\phi_c, \tag{1.1}$$

here [,] means Poisson braket.(1.1) is usually called in the physical literature "gauge open algebra",because generally (1.1) is not a closed algebra(in

the sense of ordinary Lie algebras).Here C_{ab}^{c} in general is not constant.

Algebras like (1.1) describe systems such as the relativistic particle, strings ,membranes,gravitation,etc(and of course, their supersymmetric relatives) .

The quantization of these theories is plagued with difficulties and for this reason it is necessary to study new quantization methods for which these systems can be studied.

In the last ten years, it has been discovered a general quantization method which permits to study these systems. This method, called BFV formalism is reviewed briefly in section 2. Section 3, is devoted to study some simple applications . Here the quantization of the relativistic particle is worked out in detail. Several points not discussed in the literature are pointed out and we find the correct expression for the Feynman propagator in both cases.Section 4 contains conclusions and an outlook.

2.- The BFV Formalism : Review

In this section we review the BFV formalism. As it was explained in the introduction, this method is a procedure for quantizing systems with first class constraints and is the most general method known today to treat this class of systems.

We consider a dynamical system described by a phase space F_1 whose coordinates are (p_i, q^i) ,i=1,2,3,...,N ;the canonical hamiltonian is H_0 and and the dynamical system is subject to M first class constraints ϕ_a satisfying the algebra (1.1).

The action for this system is taken to be:

$$S = \int_{t_1}^{t_2} dt(p_i\dot{q}^i - H_0 - \lambda^a\phi_a), \tag{2.1}$$

where the λ^a are lagrange multipliers. Then in the BFV formulation , we consider that the lagrange multiplier can be treated in the same foot as the canonical variables (p,q). This oblige us to introduce conjugate canonical momenta to λ^a , say π_a :

$$[\lambda_a, \pi^b] = \delta_a^b, \tag{2.2}$$

and,in order that the dynamics of the theory does not change ,they must

be imposed as new constraints,i.e,

$$\pi_a = 0. \tag{2.3}$$

In the BFV notation,the set of 2M constraints (ϕ_a, π^a) is denote by G_a and they obviously satisfy the gauge algebra

$$[G_a, G_b] = K^c_{ab} G_c. \tag{2.4}$$

In the algebraic sense,the procedure of treating the Lagrange multiplier on the same foot that as the coordinates (p,q), is equivalent to replace the old phase space F_1 by an other phase space F_2 ,such that :

$$(p^i, x_i) \longrightarrow (p^i, x_i) \oplus (\pi_a, \lambda^a). \tag{2.5}$$

The next step in BFV construction consists in incorporating a pair of canonically conjugated ghosts $(\eta_a, \mathcal{P}^a)$ (with opuest statistic) for each constraint,i.e,

$$\{\eta_a, \eta_b\} = 0, \{\mathcal{P}^a, P^b\} = 0,$$
$$\{\eta_a, \mathcal{P}^b\} = -\delta^b_a, \tag{2.6}$$

Thus the phase space is replaced by :

$$(p^i, x_i) \longrightarrow (p^i, x_i) \oplus (\pi_a, \lambda^a) \oplus (\mathcal{P}^a, \eta_a). \tag{2.7}$$

The hamiltonian structure (2.7) has remarkable properties .We would like to enumerate some of them: a)In (2.7) ,we have replaced the local gauge invariance by a global supersymmetry (BRST symmetry).This name is due to Becchi,Rouet, Stora and Tyutin ,who discovered a similar symmetry in the context of Yang-Mills theory [4 , 5].

The BRST symmetry is a name given by the physicists to a symmetry deeply rooted in cohomology theory[6].

b)the symmetry generator Q (usually called BRST charge) for a theory with the gauge algebra (2.4) , has the form :

$$Q = \eta_a G^a + \frac{1}{2} \mathcal{P}^a K^{bc}_a \eta_b \eta_c + \cdots, \tag{2.8}$$

(2.8) is anticommutative and is , by construction ,nilpotent,i.e.

$$\{Q, Q\} = 0. \tag{2.9}$$

c)At quantum level ,in the extended phase space (2.7) , there exist the following theorem proved by Fradkin and Vilkovisky [3] .

Theorem

Let a hamiltonian system with G_a constraints be described by the the effective action S_{eff} given by ,

$$S_{eff} = \int_{t_1}^{t_2} dt(p_i\dot{x}^i + \dot{\eta}_a\mathcal{P}^a + \pi_a\dot{\lambda}^a - H_0 - \{Q,\psi\}), \tag{2.10}$$

where Q is the BRST charge and ψ is an arbitrary function (gauge fixing function). Then , the path integral :

$$Z_\psi = \int D\mu \exp\,[iS_{eff}], \tag{2.11}$$

where $D\,\mu$ is a Liouville measure , is independent of the choice of ψ ,i.e.

$$Z_\psi = Z_{\psi'}.$$

This remarkable theorem is useful to prove the unitarity of theories and it permits to calculate off-shell propagators (generally a complicated problem).For a demostration of the theorem see the ref. [3].

3.- Aplications

A-Relativistic Particle

The massive relativistic particle is described by the following action :

$$S = \int_{t_1}^{t_2} d\tau(p^\mu\dot{x}_\mu - N\mathcal{H}), \tag{3.1}$$

where N is a Lagrange multiplier and $\mathcal{H}$ is a constraint defined by

$$\mathcal{H} = \frac{1}{2}(p^2 + m^2) = 0 \tag{3.2}$$

It is easy to verify using ,

$$[x_\mu, x_\nu] = 0 = [p_\mu, p_\nu],$$

$$[x_\mu, p^\nu] = \delta_\mu^\nu, \tag{3.3}$$

that the constraint algebra (3.2) is,

$$[\mathcal{H}, \mathcal{H}] = 0, \tag{3.4}$$

and by consequence , (3.4) is a first class algebra . Thus , to quantize the relativistic particle , we can use the BFV formalism developed in the section 2. The extended phase space (2.7) in this case is :

$$(p_\mu, x^\mu) \oplus (\pi_N, N) \oplus (\bar{\mathcal{P}}, \eta, \mathcal{P}, \bar{\eta}),$$

where π_N is the canonical momenta of N and the $\mathcal{P}'s$ and $\eta's$ are the anticommutative ghosts that in this case satisfy :

$$\{\eta, \bar{\mathcal{P}}\} = -1 = \{\bar{\eta}, \mathcal{P}\},$$

$$\{\eta, \bar{\eta}\} = 0 = \{\mathcal{P}, \bar{\mathcal{P}}\}.$$

The action in the extended phase space is now:

$$S = \int_{t_1}^{t_2} d\tau(\pi_N \dot{N} + \dot{\eta}\bar{\mathcal{P}} + \dot{\bar{\eta}}\mathcal{P} + p_\mu \dot{x}^\mu + \{Q, \psi\}), \tag{3.5}$$

using the prescription (2.8) , the BRST charge is :

$$Q = \eta\mathcal{H} + \mathcal{P}\pi_N, \tag{3.6}$$

and the gauge fixing function is chosen in the form :

$$\psi = N\bar{\mathcal{P}}. \tag{3.7}$$

The choice of ψ , according to the Fradkin-Vilkovisky theorem is arbitrary , nevertheless here it is convenient the election of (3.7) because it is equivalent to choose the proper time gauge $\dot{N} = 0$. This gauge choice is consistent with the reparametrization invariance .Using the Fradkin-Vilkovisky theorem , we obtain :

$$Z = \int DND\pi D\eta D\bar{\mathcal{P}} D\bar{\eta} D\mathcal{P} Dp^\mu Dx_\mu.$$

$$. \exp[i \int_{t_1}^{t_2} d\tau(\pi_N \dot{N} + \dot{\bar{\eta}}\bar{\mathcal{P}} + \dot{\bar{\eta}}\mathcal{P} + p_\mu \dot{x}^\mu + N\mathcal{H} + \mathcal{P}\bar{\mathcal{P}})]. \tag{3.8}$$

The integrals in (3.8) can be calculated imposing the following BRST invariant boundary conditions :

$$\begin{aligned} x(t_1) = x_1 \quad x(t_2) = x_2, \\ \eta_a(t_1) = 0 = \eta_a(t_2), \\ \eta(\bar{t}_1) = 0 = \eta(\bar{t}_2), \\ \pi_a(t_1) = 0 = \pi_a(t_2). \end{aligned} \tag{3.9}$$

Integrating π_N , we obtain the $\delta[\dot{N}]$ factor and the integration in the ghosts momenta give the usual expression for the trasition amplitude in the proper time gauge [7].

To integrate in x_μ and p_μ it is convenient to eliminate the zero mode associate to N(t) , we then write :

$$N(t) = N(0) + M(t), \tag{3.10}$$

where we have the following boundary condition for M(t)

$$M(0) = 0. \tag{3.11}$$

Using (3.10) the $\delta[\dot{N}]$ factor can be written as :

$$\delta[\dot{M}] = \int dN(0)\delta[M(t) - N(0)]det(\partial_\tau)^{-1}, \tag{3.12}$$

thus (3.8) becomes :

$$Z = \mathcal{N} \int dN(0) \int D\eta D\bar{\eta} Dx_\mu Dp^\mu det(\partial_\tau)^{-1}$$

$$\exp[i \int_{t_1}^{t_2} d\tau(p^\mu \dot{x}_\mu + N(0)\mathcal{H} + \dot{\eta}\dot{\bar{\eta}})]. \tag{3.13}$$

The determinant that appears in (3.13) is indeterminate and it can be taken out of the path integral as a factor absorbed by an overall normalization .

Following Teitelboim arguments [7] , the integral in N(0) can not be taken in the range $(-\infty, +\infty)$ because we are obliged to choose only one classical trajectory . This observation is physically very satisfactory and it is crucial to obtain the correct result .

Integrating on η and $\bar{\eta}$, we obtain $det(-\partial_\tau{}^2)$ .This expression can be calculated using the boundary condition (3.9) and ζ-function regularization . The result is $(t_2 - t_1)$ and the integral (3.12) is :

$$Z = \mathcal{N} \int_0^\infty dT \int Dx^\mu Dp_\mu \exp[i \int_{t_1}^{t_2} d\tau (p^\mu \dot{x}_\mu + N(0)\mathcal{H})], \tag{3.14}$$

where $T = N(0)(t_2 - T_1)$ and $\mathcal{N}$ is a normalization constant.The integration on p_μ gives :

$$Z = \mathcal{N} \int_0^\infty dN(0) \int Dx_\mu \exp[i \int_{t_1}^{t_2} d\tau (\frac{\dot{x}^2}{2N(0)} + \frac{}{2} m^2 N(0))], \tag{3.15}$$

Note that the effective action in (3.15) is precisely the einbein version of the relativistic particle.To integrate (3.15) we make the following change of variables :

$$x^\mu(t) = x_1^\mu + \frac{\Delta x^\mu}{\Delta t}(t - t_1) + y^\mu(t), \tag{3.16}$$

(3.16) is consistent with (3.8) iff:

$$y^\mu(t_1) = 0 = y^\mu(t_2). \tag{3.17}$$

Using (3.16) , (3.15) yields :

$$Z = N \int_0^\infty dT \; det(\frac{-\partial_\tau{}^2}{N(0)^2})^{\frac{-D}{2}} \exp[i(\frac{(\Delta x)^2}{2T} + \frac{m^2 T}{2})]. \tag{3.18}$$

The determinant in (3.18) can be calculated using ζ-function regularization and the boundary condition (3.17) and the result is :

$$det(\frac{\partial_\tau{}^2}{N(0)^2}) = T.$$

Thus , (3.18) is :

$$Z = N \int_0^\infty dT\ T^{\frac{-D}{2}} \exp[i(\frac{(\Delta x)^2}{2T} + \frac{m^2 T}{2}],$$

$$= \mathcal{N}' \int \frac{d^D p}{(2\pi)^D} \frac{\exp ip.(x_2 - x_1)}{p^2 + m^2 - i\epsilon}.$$

This expression is the Feynman propagator for the relativistic particle . Recently , two different derivations of this result has been obtained in the literature [8 , 9].Also Giannakis , Ordoñez ,Rubin and Zucchini have obtained similar results using the lagrangian formalism [10].

B-Spinning Particle

The massive spinning particle is described by the following constraints [11]:

$$\mathcal{H} = \frac{1}{2}(p^2 + m^2) = 0,$$

$$\mathcal{S} = \theta^\mu p_\mu + m\theta_5 = 0, \tag{3.19}$$

where θ_μ and θ_5 are grassmanian variables that obey the following algebra

$$\{\theta_\mu, \theta^\nu\} = i\delta_\mu^\nu,$$

$$\{\theta_5, \theta_5\} = i, \tag{3.20}$$

and the even variables satisfy the algebra (3.3).

Using (3.19) and (3.3) it is easy to verify that the constraint algebra is :

$$[\mathcal{H}, \mathcal{H}] = 0,$$

$$[\mathcal{H}, \mathcal{S}] = 0,$$

$$\{\mathcal{S}, \mathcal{S}\} = 2i\mathcal{H}. \tag{3.21}$$

It is easy to see using (2.8) that the BRST charge is :

$$Q = \eta\mathcal{H} + \mathcal{P}\pi_N + c\mathcal{S} + \pi_\lambda \mathcal{P}_c + i\bar{\mathcal{P}}cc, \tag{3.22}$$

where $(\eta, \bar{\eta}, \bar{\mathcal{P}}, \mathcal{P})$ are the coordinates and the ghost momenta (anticommutative) associated to $\mathcal{H}$ and $(c, \bar{c}, \bar{\mathcal{P}}_c, \mathcal{P}_c)$ are the coordinates and the ghost

momenta (commutative) associated to $\mathcal{S}$. The commutative ghost algebra is :

$$[c, \bar{\mathcal{P}}_c] = 1 = [\bar{c}, \mathcal{P}], \tag{3.23}$$

and zero in the other cases . π_λ is the canonical momenta of the fermionic Lagrange multiplier λ.

The fixing gauge function ψ is chosen as :

$$\psi = \bar{\mathcal{P}} N + \lambda \bar{\mathcal{P}}_c. \tag{3.24}$$

Using the Fradkin-Vilkovisky theorem we obtain :

$$Z = \int DN D\pi_N D\lambda D\pi_\lambda D\eta D\bar{\mathcal{P}} D\bar{\eta} D\mathcal{P} D\mathcal{P}_c D\bar{c} D\bar{\mathcal{P}}_c Dc D\theta_\mu D\theta_5 Dp^\mu Dx_\mu$$

$$\exp[i \int_{t_1}^{t_2} d\tau (\pi_N \dot{N} - \dot{\lambda}\pi_\lambda + \dot{\eta}\bar{\mathcal{P}} + \dot{\bar{\eta}}\mathcal{P} + \mathcal{P}_c \dot{c} + \mathcal{P}_c \dot{\bar{c}} + \frac{i}{2}\dot{\theta}^\mu \theta_\mu + \frac{i}{2}\dot{\theta}_5 \theta_5 +$$

$$p_\mu \dot{x}^\mu + N\mathcal{H} + \lambda \mathcal{S} + \mathcal{P}\bar{\mathcal{P}} - \mathcal{P}_c \bar{\mathcal{P}}_c - 2i\bar{\mathcal{P}} c\lambda)]. \tag{3.25}$$

In order to calculate (3.24) we impose the following BRST invariant boundary conditions :

$$x(t_1) = x_1, x(t_2) = x_2,$$

$$\eta(t_1) = \eta(t_2) = c(t_1) = c(t_2) = 0,$$

$$\bar{\eta}(t_1) = \bar{\eta}(t_2) = \bar{c}(t_1) = \bar{c}(t2) = 0,$$

$$\pi_N(t_1) = \pi_N(t_2) = \pi_\lambda(t_1) = \pi_\lambda(t_2) = 0,$$

$$\frac{1}{2}(\theta^\mu(t_1) + \theta^\mu(t_2)) = \gamma^\mu,$$

$$\frac{1}{2}(\theta_5(t_1) + \theta_5(t_2)) = \gamma_5. \tag{3.26}$$

Integrating over $\pi_N, \pi_\lambda, \mathcal{P}, \bar{\mathcal{P}}, \mathcal{P}_c$ and $\bar{\mathcal{P}}_c$, we obtain :

$$Z = \int DN D\lambda D\eta D\bar{\eta} D\bar{c} Dc D\theta_\mu D\theta_5$$

$$Dp^{\mu}Dx_{\mu}\delta[\dot{N}]\delta[\dot{\lambda}]$$

$$\exp[i\int_{t_1}^{t_2} d\tau(p_{\mu}\dot{x}^{\mu}+\frac{i}{2}\dot{\theta}^{\mu}\theta+\frac{i}{2}\dot{\theta}_5\theta_5+$$

$$+N\mathcal{H}+\lambda\mathcal{S}+\dot{\eta}\dot{\bar{\eta}}+2ic\lambda\dot{\bar{\eta}}+\dot{c}\dot{\bar{c}})]\ , \tag{3.27}$$

(3.27) is the hamiltonian expression for the path integral in the proper time gauge.

As in the bosonic relativistic particle case , we would like to eliminate the zero modes . For this reason we write the analogous of (3.10),

$$N(0)=N(0)+M(t),$$

$$\lambda(0)=\lambda(0)+\zeta(t), \tag{3.28}$$

where we have the following "boundary conditions" ,

$$M(0)=0,$$

$$\zeta(0)=0. \tag{3.29}$$

The equivalent of the equation (3.12) is

$$\delta[\dot{N}(0)]=\int dN(0)\delta[M(t)-N(0)]det(\partial_{\tau})^{-1},$$

$$\delta[\dot{\zeta}]=\int d\lambda(0)\delta[\zeta(t)-\lambda(0)]det(\partial_{\tau})^{+1}. \tag{3.30}$$

Such as in the relativistic particle case the determinants that appears in (3.30) are indetermined because we have not sufficient boundary conditions , nevertheless , in this case the bosonic and fermionic determinants are precisely cancelled . Replacing (3.30) in (3.27) and using the Teitelboim arguments to choose one classical trajectory , we obtain :

$$Z=\int_0^{\infty} dN(0)\int d\lambda(0)\int D\eta D\bar{\eta}D\bar{c}DcD\theta_{\mu}D\theta_5Dp^{\mu}Dx_{\mu}$$

$$\exp[i\int_{t_1}^{t_2} d\tau(p_{\mu}\dot{x}^{\mu}+\frac{i}{2}\dot{\theta}^{\mu}\theta_{\mu}+\frac{i}{2}\dot{\theta}_5\theta_5+$$

$$+N(0)\mathcal{H}+\lambda(0)\mathcal{S}+\dot{\eta}\dot{\bar{\eta}}+2ic\lambda(0)\dot{\bar{\eta}}+\dot{c}\dot{\bar{c}})], \tag{3.31}$$

(for the integration in $\lambda(0)$ we do not write the integration range because such concept not exist for the Berezin integral).

Using the boundary conditions (3.26) the ghosts integrals can be explicitly calculated . Integrating in p_μ :

$$Z=\mathcal{N}\int_0^\infty dN(0)\int d\lambda(0)\int D\theta_\mu D\theta_5 Dx_\mu$$

$$\exp[i\int_{t_1}^{t_2}d\tau(\frac{\dot{x}^2}{2N(0)}+\frac{m^2N(0)}{2}+\frac{i}{2}\dot{\theta}^\mu\theta_\mu+\frac{i}{2}\dot{\theta}_5\theta_5+\frac{\lambda(0)\theta_\mu\dot{x}^\mu}{N(0)}+$$

$$m\lambda(0)\theta_5] \tag{3.32}$$

In (3.32) the effective action is the one-dimensional supergravity action if N(0) and $\lambda(0)$ are interpreted as the graviton and the gravitino respectively .

Making the change of variables

$$x^\mu(t)=x_1^\mu+\frac{\Delta x^\mu}{\Delta t}(t-t_1)+y^\mu(t),$$

$$\theta^\mu(t)=\gamma_5\gamma^\mu+\psi^\mu(t), \tag{3.33}$$

$$\theta_5(t)=\gamma_5+\psi(t),$$

and using (3.26) consistency imply :

$$y^\mu(t_1)=0=y^\mu(t_2). \tag{3.34}$$

Using (3.33) and (3.34) in (3.32) and integrating in $\lambda(0),\psi^\mu(t)$,$\psi(t)$ and $y^\mu(t)$,we obtain :

$$Z=\mathcal{N}\int_0^\infty\frac{dT}{T^{\frac{D}{2}}}(\frac{\gamma_5\gamma_\mu\Delta x^\mu}{T}+m\gamma_5)\exp[i(\frac{\Delta x^2}{2T}+\frac{m^2T}{2}-i\epsilon)],$$

$$=\mathcal{N}\int\frac{d^Dp}{(2\pi)^D}\frac{e^{ip.\Delta x}}{p^2+m^2-i\epsilon}(\gamma_5\gamma^\mu p_\mu+m\gamma_5). \tag{3.35}$$

(3.35) is the Dirac propagator [8] .

Conclusions

In this paper we have studied the quantization of hamiltonian systems with first class constraints using the BFV formalism .

Using the two examples studied above we saw that the BFV formalism is a powerful method for quantizing theories with gauge fredom .

For more complicated theories , such as strings and membranes , the problem is not completely solved .The main difficulty is that at quantum level there are anomalies .

Using the BFV formalism this problem is not undertood at the path integral level.

Acnowledgments

One us (J.G) would like to thank V.Tapia and J.R. Gonçales for many valuable discussions .We would like also to thank miss Monica Pierri by pointing out some mistakes in a preliminary version of this paper. J.G was supported by a FAPESP postdoctoral fellowship and V.O.R was partially supported by CNPq.

References

[1] E.Abers and B.W.Lee, Gauge Theories , Phys. Rep. 9C(1973) 1

[2] P.A.M. Dirac , Lectures On Quantum Mechanics, (Yeshiva University ,1964).
A.J. Hanson , T. Regge And C. Teitelboim, Constrained Hamiltonian Systems ,
(Accademia Dei Lincei ,Roma 1976)

[3] E.S.Fradkin and G.Vilkovisky, Quantization Of Relativistic Systems With Constraints, Phys.Lett. 55B (1975) 224.
I.A.Batalin and G.Vilkovisky, Relativistic S-matrix Of Dynamical Systems With Bosons And Fermions Constraints, Phys. Lett.B69 (1977) 309.
M.Henneaux, Hamiltonian Form Of The Path Integral For Theories With Gauge Freedom, Phys. Rep. 126(1985)1.

[4] C. Becchi , A. Rouet and R. Stora,The Abelian Higgs-Kibble Model, Unitarity Of The S-operator, Phys. Lett. 52B (1974) 344.

[5] I.V. Tyutin ,Lebedev Report (unpublished) 1975.

[6] See, e.g., B.Doubronine ,S. Novikov and A. Fomenko , Geometrie Comtemporaine ,vol. 3, Mir ,Moscou 1984.

[7] C. Teitelboim , Quantum Mechanics Of The Gravitational Field, Phys.Rev. D25 (1982) 3152.

[8] V. Fainberg and A. Marshakov , Local Supersymmetry and Dirac Particle Propagator As A Path Integral, Nucl. Phys. B306 (1988) 659.

[9] A. Cohen , G. Moore , P. Nelson and J. Polshinski, An Off-Shell Propagator For String Theory, Nucl. Phys. B267 (1986) 143.

[10] I. Giannakis ,C. R. Ordoñez ,M. A. Rubin and R. Zucchini, Rockefeller Preprint RU88/b1/28 (1988).

[11] C. Teitelboim .Supergravity And Square Roots Of Constraints. Phys. Rev.Lett. 38 (1977) 1106.

J. Gamboa
Instituto de Fisica , Universidade de São Paulo
and
Centro de Estudios Cientificos de Santiago ,
Casilla 16443 ,Santiago 9 ,Chile

V.O. Rivelles
Instituto de Fisica , Universidade de São Paulo
CP 20516 ,CEP 01498 , São Paulo , Brasil .

Constantin Carathéodory: An International Tribute (pp. 344–373)
edited by Th. M. Rassias

INFINITE-DIMENSIONAL STOCHASTIC DIFFERENTIAL GEOMETRY IN MODERN LAGRANGIAN APPROACH TO HYDRODYNAMICS OF VISCOUS INCOMPRESSIBLE FLUID

Yuri E.Gliklikh

ABSTRACT. A class of stochastic processes on the group of diffeomorphisms such that their (in a certain sense) expectations, the curves in this group, are flows of viscous incompressible fluid is described. These processes satisfy a special stochastic analogue of the Newton's equation of motion written in geometrical terms. The corresponding equation on the tangent bundle is smooth and does not use any additional "internal" forces etc. The cases of fluid motion in the flat torus T^n and in a bounded domain in R^n are considered. The latter is represented by means of a special constraint system of the diffeomorphisms group of T^n.

1. Introduction.

In the volume dedicated to the memory of C.Carathéodory this paper is connected with the idea of geometrical and in some sense probabilistic description of the nature, the development of which was significantly affected by the work of C.Carathéodory.

We deal with the modern Lagrangian approach to hydrodynamics suggested in [1,9] where the geometry of the hydrodynamical configuration space, the infinite-dimensional manifold (group) of diffeomorphisms, was involved in the investigation of the fluid motion. In terms of this app-

roach the viscous incompressible fluid was considered in [9], but as compared with the perfect incompressible fluid[1,9], that theory did not possess completely natural geometrical properties: the additional force of the form $\nu\triangle$ (where $\triangle$ is Laplacian), which lost the derivatives, was taken into account. This gave some limitations, e.g. the diffeomorphisms of a too high differentiability were needed (belonging to the Sobolev class H^s for $s>\frac{n}{2}+5$, where n is dimension of the manifold M in which the fluid moved), only M without boundary was studied etc.

We introduce another way to the Lagrangian description of viscous incompressible fluid involving constructions of the stochastic differential geometry on the group of diffeomorphisms. The class of stochastic processes is determined which satisfy a certain stochastic analogue of the Newton's law of dynamics (of ordinary geodesics equation if the external force vanishes). The expectations (in a certain sense) of these processes are the curves in the group of diffeomorphisms which describe the motion of the viscous incompressible fluid. This way seems to be natural. It does not use additional forces, the corresponding equation on the tangent bundle deals with smooth vector fields only, the Sobolev class of the diffeomorphisms is H^s for $s>\frac{n}{2}+1$, etc.

We consider the motion of the fluid on the flat n-dimensional torus T^n (section 4) and in a bounded domain $\textcircled{M}$ with a smooth boundary in the Euclidean space R^n (section 5). The latter is investigated via the approach suggested in [2,3], where the fluid motion in $\textcircled{M}$ is considered as a constraint system on the group of diffeomorphisms of T^n so that the tangent bundle to the group is replaced by a certain subbundle Ξ.

Section 2 gives a brief survey of the geometry of groups of diffeomorphisms and Lagrangian approach to hyd-

rodynamics of perfect incompressible fluid which is a basis for further constructions. Section 3 is devoted to stochastic differential equations on manifolds. We study the relations between Ito equations and equations in mean derivatives introduced by E.Nelson (see e.g. [18]).

We assume the reader to be familiar with ordinary coordinates-free differential geometry (see e.g. [8,17]) and with the stochastic differential equations in linear spaces (see e.g. [12]). We should point out that all the necessary preliminaries are given in [13]; we also refer the reader to [9,19] for the details about the manifolds of diffeomorphisms and to [7,10,16,18] for the stochastics.

Some constructions of section 4 were announced in [13-15]

2. Survey of Modern Lagrangian Approach to Hydrodynamics.

Let M be n-dimensional compact oriented Riemanniam manifold without boundary, $\langle\ ,\ \rangle$ the Riemannian metric on M. Let $s>\frac{n}{2}+1$. Denote by $D^s(M)$ the set of all C^1-diffeomorphisms of M belonging to the Sobolev class H^s. Recall that when $s>\frac{n}{2}+K$, $K>0$, the space of Sobolev maps H^s is continuously imbedded in the space of C^K maps.

It is possible to define the structure of C^∞-smooth Hilbert manifold on $D^s(M)$ (see [9,19]). Here the tangent space $T_eD^s(M)$ at the point $e=id$ is a separable Hilbert space $H^s(TM)$ of all H^s-vector fields on M (the scalar product in $H^s(TM)$ is naturally generated by the Riemannian metric $\langle\ ,\ \rangle$ on M) and the tangent space $T_\eta D^s(M)$ at the point $\eta \in D^s(M)$ consists of all mapping $Y : M \longrightarrow TM$ such that $\pi Y = \eta$, where $\pi : TM \longrightarrow M$ is natural projection (i.e. $Y = X\circ\eta$ where $X \in H^s(TM) = T_eD^s(M)$).

$D^s(M)$ is a topological group, where the superposition $\circ$ is involved as a multiplication. For each $\eta \in D^s(M)$ the right-hand translation $R_\eta : D^s(M) \longrightarrow D^s(M)$, $R_\eta\theta = \theta\circ\eta$, is a C^∞-smooth mapping with the derivative $TR_\eta X = X\circ\eta$,

$X \in TD^s(M)$. The left-hand translation $L_\eta \theta = \eta \circ \theta$ is only continuous on $D^s(M)$, but when η is of the calss H^{s+1}, L_η is C^1-smooth with the derivative $TL_\eta X = T\eta \circ X$, $X \in TD^s(M)$.

Obviously one can define right-invariant vector fields on $D^s(M)$. Let $\overline{X}$ be such a field and X be a vector of this field belonging to $T_e D^s(M)$. The following property of $\overline{X}$ is very important for us: $\overline{X}$ is C^k - smooth on $D^s(M)$ iff the vector field X on M belongs to the class H^{s+k}, where $k = 1,2,\ldots,\infty$, $H^\infty = C^\infty$. Any right-invariant vector field $\overline{X}$ (C^1-smooth in the general case and continuous, when $s > \frac{1}{2}n + 2$) has the flow on $D^s(M)$. The integral curve $\eta(t)_e$ beginning at e is the flow of X on M; $\eta(t)_\theta = \eta(t)_e \circ \theta$, $\theta \in D^s(M)$.

Denote by $D^s_\mu(M)$ the submanifold in $D^s(M)$ consisting of all diffeomorphisms which preserve the form of Riemannian volume on M. $D^s_\mu(M)$ is also a subgroup in $D^s(M)$. The tangent space $T_e D^s_\mu(M)$ is the space of all zero-divergence H^s vector fields on M, $T_\eta D^s(M) = \{Y = X \circ \eta \mid X \in T_e D^s_\mu(M), \ \eta \in D^s_\mu(M)\}$. All the properties of right(left)-hand translations, right-invariant vector fields etc. mentioned for $D^s(M)$ are valid for $D^s_\mu(M)$.

Let $\eta \in D^s(M)$, $X, Y \in T_\eta D^s(M)$. Determine the scalar product $(\ ,\ )_\eta$ in $T_\eta D^s(M)$ by the formula

$$(X,Y)_\eta = \int_M \langle X(m), Y(m) \rangle_{\eta(m)} \mu(dm) \qquad (1)$$

where $\mu(dm)$ is the form of Riemannian volume. Notice that $X(m)$ and $Y(m)$ belong to $T_{\eta(m)}M$ and they are multiplied with respect to the metric tensor $\langle , \rangle_{\eta(m)}$ at $\eta(m)$. Using (1) for all $\eta \in D^s(M)$ we define the Riemannian metric on $D^s(M)$. Obviously this metric introduces the topology of the functional space $L_2 = H^0$ in the tangent spaces, which is weaker than the initial topology H^s.

The restriction of (1) to $TD^s_\mu(M)$ is a weakly Riemannian metric on $D^s_\mu(M)$ which is evidently right-invariant.

Consider the connector $K : TTM \longrightarrow TM$ of the Levi-Civita connection of the metric $\langle , \rangle$ (see e.g. [8,13]). Recall that the covariant derivative $\nabla_a b$ of the Levi-Civita connection for vector fields a and b on M is defined by the formula $\nabla_a b = K \circ Tb(a)$, and the covariant derivative of a vector field $a(t)$ along a smooth curve $m(t)$ is defined by the formula $\frac{D}{dt} a = K \circ \frac{d}{dt} a$ (see e.g. [13]). For vector fields X, Y on $D^s(M)$ define the covariant derivative $\overline{\nabla}_X Y$ by the formula

$$\overline{\nabla}_X Y = K \circ TY(X). \tag{2}$$

One can easily see that at each $\eta \in D^s(M)$ $TY(X)_\eta$ is a mapping of M into TTM, so (2) defines $\overline{\nabla}_X Y$ correctly. It is shown in [9] that $\overline{\nabla}$ is covariant derivative of the Levi-Civita connection of the metric $(,)$ on $D^s(M)$. The geodesic pulverization $\overline{Z}$ of this connection is described as follows:

$$\overline{Z}(X) = Z \circ X \tag{3}$$

for $X \in TD^s(M)$, where Z is the geodesic pulverization of the Levi-Civita connection on M (i.e. the vector field on TM). One can easily obtain from (3) the following statement: Z is $D^s(M)$-right-invariant and C^∞-smooth on $TD^s(M)$.

Recall the Hodge decomposition [9]

$$H^s(T M) = G^s \oplus E^s \oplus \ker \triangle \tag{4}$$

where G^s is the space of gradients of all H^{s+1} functions on M, E^s is the space of all H^s co-gradients on M, $\operatorname{Ker} \triangle$ is the space of all harmonic (i.e. both gradiant and co-gradient) vector fields on M and $\oplus$ denotes the orthogonal direct sum with respect to L_2 - scalar product (1) in $T_e D^s(M)$. By a co-gradient we mean a vector field corresponding to a co-exact form on M with respect to the Riemannian metric $\langle , \rangle$. Notice that $\operatorname{Ker} \triangle$ is a finite-dimensional space and consists of C^∞ smooth vector fields.

Denote by $P_e : T_eD^s(M) = H^s(TM) \longrightarrow E^s \oplus \operatorname{Ker} \triangle =$ $= T_eD^s_\mu(M)$ the $(\ ,\ )_e$ - orthogonal projection in (1.4). Consider the mapping $P : TD^s(M)|_{D^s_\mu(M)} \longrightarrow TD^s_\mu(M)$ determined for each $\eta \in D^s_\mu(M)$ by the formula

$$P_\eta = TR_\eta \circ P_e \circ TR_\eta^{-1}.$$

It is obvious that P is $D^s_\mu(M)$-right-invariant. There is an important and a rather complicated result (see [9]): P is a C^∞ -smooth mapping. Notice the important consequence of (4) and of the definition of P_e : for every $Y \in T_eD^s(M)$ we have

$$P_e(Y) = Y + \operatorname{grad} p \tag{5}$$

where p is a certain H^{s+1}-function on M unique to within the constants.

According to the standard construction of differential geometry now we may define the connector $\widetilde{K}$ and the covariant derivative $\widetilde{\nabla}$ of the Levi-Civita connection on $D^s_\mu(M)$ by the formulas

$$\widetilde{K} = P \circ K \tag{6}$$

$$\widetilde{\nabla}_X Y = P \circ \overline{\nabla}_X Y = K \circ TY(X) \tag{7}$$

where X,Y are vector fields on $D^s_\mu(M)$. Of course the Levi-Civita connection $\widetilde{H}$ itself is equal to $\operatorname{Ker} \widetilde{K} \subset TTD^s_\mu(M)$.

The geodesic pulverization S of this connection is a vector field on $TD^s_\mu(M)$ of the form

$$S = TP \circ \bar{Z} \tag{8}$$

It evidently follows from (8) that S is D^s_μ - right-invariant and C^∞ -smooth on $TD^s_\mu(M)$. Denote by $\widetilde{\exp}$ the corresponding exponential map of a neighbourhood of the zero section in $TD^s_\mu(M)$ onto $D^s_\mu(M)$; $\widetilde{\exp}$ is $D^s_\mu(M)$-right-invariant, C^∞ -smooth and covers some neighbourhood of each point in $D^s_\mu(M)$ (see [9,13]). According to the standard definition we determine the covariant derivatives $\frac{D}{dt}$ and

$\frac{\tilde{D}}{dt}$ of a vector field X(t) along a curve in $D^s(M)$ and $D^s_\mu(M)$ respectively by the formulas

$$\frac{\bar{D}}{dt} X(t) = K \circ \frac{d}{dt} X(t), \tag{9}$$

$$\frac{\tilde{D}}{dt} X(t) = P \circ \frac{\bar{D}}{dt} X(t) = \tilde{K} \circ \frac{d}{dt} X(t). \tag{10}$$

Let $F \in T_e D^s(M)$ and $\bar{F}$ be right-invariant vector field on $D^s_\mu(M)$ corresponding to F. Consider the mechanical system with the configuration space $D^s_\mu(M)$, with the kinetic energy $\mathcal{K}$ generated by $(\ ,\ )$ according to the usual formula $\mathcal{K}(X) = \frac{1}{2}(X,X)$, $X \in TD^s_\mu(M)$, see [13], and with the external force $\bar{F}$ (using the Riemannian metric we do not distinguish vectors and 1-forms). The Newton's law for this system has the form

$$\frac{\tilde{D}}{dt} \dot{\eta} = F. \tag{11}$$

The trajectories of this system describe the motion of perfect incompressible fluid on M. Indeed, let $\eta(t)$ be a solution to (11). Consider the vector $u(t) = TR_{\eta(t)}^{-1}(\dot{\eta}(t)) \in T_e D^s_\mu(M)$. As a consequence of (10), (6) and (5) the zero-divergence vector field u(t) on M satisfies the equation

$$\frac{\partial u}{\partial t} + \nabla_u u + \operatorname{grad} p = F \tag{12}$$

which is a well-known Euler equation of hydrodynamics. Here ∇ is Levi-Civita covariant derivative on M.

Notice that the curve of velocities $\dot{\eta}(t)$ on $TD^s_\mu(M)$ is an integral curve of the vector field

$$S + \bar{F}^\ell \tag{13}$$

where $\bar{F}^\ell$ is natural vertical lift of $\bar{F}$.

It is easy to prove the local existence and uniqueness of integral curve for (13) if $s > \frac{1}{2} n + 2$ (or if $F \in H^{s+2}$ when $s > \frac{1}{2} n + 1$) and consequently to obtain the local

existence and uniqueness of solutions to the Euler equation (12), see [9,13].

If $F = 0$, (11) turns into the equation of geodesics

$$\frac{\widetilde{D}}{dt}\dot{\eta}(t) = 0 \qquad (14)$$

and $\dot{\eta}(t)$ becomes an integral curve of S on $TD^s_\mu(M)$ the local existence and uniqueness for which is obvious because S is a C^∞ vector field.

For the case when a compact oriented Riemannian manifold M has a smooth boundary ∂M we should note that $D^s(M)$ and $D^s_\mu(M)$ are well-defined. Evidently any diffeomorphism of M maps ∂M into itself so that $T_e D^s(M)$ ($T_e D^s_\mu(M)$, respectively) consists of all H^s vector fields on M tangent to ∂M (zero-divergence vector fields, tangent to ∂M). Other properties of $D^s(M)$ and $D^s_\mu(M)$ for M with boundary and the description of fluid motion by means of a mechanical system on $D^s_\mu(M)$ can be found e.g. in [9].

Below we shall use another approach, suggested in [2,3], see [13] for details. In this approach the flow of perfect incompressible fluid on M with boundary is described as a constraint motion on $D^s_\mu(N)$ where N is an auxiliary manifold without boundary and the constraint is considered as a subbundle of $TD^s_\mu(N)$, cf. e.g. [11].

Let M be a compact oriented Riemannian manifold with boundary, N be an arbitrary compact oriented Riemannian manifold without boundary such that $\dim M = \dim N = n$, M is imbedded in N and the Riemannian metric on M is obtained as the restriction of the Riemannian metric of N (one may use N equal to double of M with the metric smoothly extended beyond the boundary). Let $s > \frac{1}{2}n + 1$.

Theorem 1.[2,3,13] There exist C^∞ smooth right-invariant subbundle Ξ^s in $TD^s_\mu(N)$ and C^∞-smooth right-invariant map $\bar{R} : TD^s_\mu(N) \longrightarrow \Xi^s$, the projection in fibres, which have the following properties:

(i) Consider the restriction operator $j : H^s(TN) \longrightarrow H^s(TM)$

of vector fields on N to M, and the fibre Ξ^s_e of Ξ^s in e. Then $j : \Xi^s_e \longrightarrow T_e D^s(M)$ is an isomorphism.
(ii) The subbundle Ξ^s is not integrable, its fibres are infinite-dimensional and have an infinite codimension in the fibres of $TD^s_\mu(N)$.
(iii) Consider the geodesic pulverization S on $TD^s_\mu(N)$ mentioned above. Let X(t) be an integral curve of the vector field $S^\Xi = T\bar{R} \circ S$ on Ξ^s with the initial condition $X(0) = Y \in \Xi^s_e$. The curve $\eta(t) = \tau X(t)$ in $D^s_\mu(N)$ consists of the diffeomorphisms mapping M into M and the restriction $\eta(t)|_M$ is a curve in $D^s_\mu(M)$ describing the motion of perfect incompressible fluid without external forces on M with the initial velocity $Y_0 = j\,Y$.

We should note that the restriction $\eta(t)|_{N \setminus M}$ does not describe the fluid motion on $N \setminus M$.

Corollary. Let $F \in \Xi^s_e$ be an external force. Replace in (iii) the field $S^\Xi = T\bar{R} \circ S$ by the field $T\bar{R}(S + \bar{F}^\ell) = S^\Xi + T\bar{R} \circ \bar{F}^\ell$. Then $\eta(t)|_M$ is a curve in $D^s_\mu(M)$ describing the motion of perfect incompressible fluid on M under the action of the external force $F_0 = j\,F$.

According to (i) F and Y are in one-to-one correspondence with F_0 and Y_0, respectively.

For the sake of further applications we should note that for each H^s-vector field Y on N the action of $\bar{R}_e : T_e D^s_\mu(N) \longrightarrow \Xi^s_e$ is described in terms of the restrictions of vector fields onto M as follows

$$\bar{R}_e(Y)|_M = Y|_M + \operatorname{grad} p \qquad (15)$$

(cf.(5)) where p is a unique (to within the constants) H^{s+1} function on M and grad p is orthogonal to the boundary ∂M (See [13]). It follows from the properties of right-invariant vector fields on $D^s_\mu(N)$ that for $Y \in H^{s+k}$, $k \geqslant 1$, $\bar{R}_e Y$ is also a H^{s+k} vector field on N.

3. Basic Constructions of Stochastic Differential Geometry and Equations in Mean Derivatives.

In what follows we shall consider stochastic processes with continuous time $t \in [0, \ell]$ defined on a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$ on which one can specify a Wiener process $w(t)$ assuming values in a certain space R^n. By E we denote the expectation and by E ($|\mathcal{B}$) the conditional expectation with respect to the σ-subalgebra $\mathcal{B}$ of the σ-algebra $\mathcal{F}$. (see e.g. [16]). The subalgebra $\mathcal{B}$ may be generated either by a random variable ζ (via inverse images of Borel sets) or by a certain condition u; the corresponding notations are as follows: E($|\zeta$) and E ($|u$). Any stochastic process $\xi(t)$ defines three families of σ-subalgebras of the σ-algebra $\mathcal{F}$: "the past" $\mathcal{P}_t^\xi$ generated by $\xi(s)$ for $s \leqslant t$, "the future" $\mathcal{F}_t^\xi$ generated by $\xi(s)$ for $s \geqslant t$, and "the present" $\mathcal{N}_t^\xi$ generated by $\xi(t)$. These families are assumed to be completed with all sets of zero probability.

By an Ito stochastic differential equation in a separable Hilbert space $\mathcal{H}$ one means the integral equation

$$\xi(t) = \xi_0 + \int_0^t a(\tau, \xi(\tau))d\tau + \int_0^t A(\tau, \xi(\tau))dw(\tau) \qquad (16)$$

where the first term in the right-hand side is the Lebesgue integral, the second term is the Ito integral; $a(t,x)$ is a vector field, $A(t,x) : R^n \longrightarrow \mathcal{H}$ is a field of linear operators. Equation (16) is usually written in the symbolic differential form

$$d\xi(t) = a(t, \xi(t))dt + A(t, \xi(t))dw(t).$$

The theory of stochastic integrals and differential equations has been presented in many monographs and textbooks (see, for example, [12]).

Definition 1. A process $\xi(t)$ is called non-anticipating with respect to the non-decreading family of σ-alge-

bras $\mathcal{B}_t$ if for every t $\xi(t)$ is measurable with respect to $\mathcal{B}_t$.

The theory of stochastic equations deals with two types of solutions, strong and weak. In this paper we need only strong solutions.

Definition.2. We say that equation (16) has a strong solution if for every Wiener process $w(t)$ there exists a stochastic process $\xi(t)$, defined on the same probabilistic space as $w(t)$ is, and non-anticipating with respect to ρ_t^w, so that for $\xi(t)$ and $w(t)$ equality (16) is valid almost surely (a.s.) for each t belonging to a certain interval.

Remark 1. Without loss of generality we may assume that for any strong solution $\xi(t)$ to (16) $\rho_t^\xi = \rho_t^w$ at each t. This property is shown as corollary to the classical strong solution existence theorem [4,12], cf. theorem 3 below.

Everywhere in this paper the process $\xi(t)$ itself will be also called the (strong) solution.

Definition 3. A process $\xi(t)$ is called a local strong solution to (16) if it is a strong solution until it leaves a certain neighbourhood of its initial position.

Note that every solution to (16) may be represented as a superposition of some local solutions (see [5]).

We recall that under smooth changes of coordinates equation (16) is not transformed according to a tensor law. Let $\mathcal{H} = R^n$ and let $\varphi : R^n \longrightarrow R^p$ be a smooth mapping, then $\varphi(\xi)$ satisfies the following equation (the Ito formula; see, for example, [5,12]):

$$d\varphi(\xi(t)) = (\varphi'(a(t,\xi)) + \frac{1}{2}\,\mathrm{tr}\,\varphi''(A,A))dt + \varphi'A(t,\xi))dw(t) \tag{17}$$

where $\mathrm{tr}\,\varphi''(A,A) = \sum_{i=1}^{n} \varphi''(A(\xi)e_i, A(\xi)e_i)$; $e_1, e_2, \ldots, e_n$ is an arbitrary orthonormal basis in R^n. The non-tensor

term $\frac{1}{2}\operatorname{tr}\varphi''(A, A)$ appears in (17) as a contribution due to the integral with respect to $(dw)^2$ while integrating the Taylor series of the function φ. Unlike the Lebesgue integral with respect to $(dt)^2$, the Ito integral with respect to $(dw)^2$ does not vanish.

In what follows in this section we consider a smooth manifold M modelled on a separable Hilbert space and equipped with a smooth exponential exp corresponding to a certain connection. Sometimes M will be a finite-dimensional Riemannian manifold; in this case we always use the Levi-Civita connection and its exponential map.

Let $a(t,m)$ be a vector field on M, $A(m)$ be a field of linear operators $A(m) : R^n \longrightarrow T_mM$, $w(t)$ be a Wiener process in R^n. Consider a class of stochastic processes $(a(t,m), A(m))$ in the tangent space T_mM which consists of solutions to stochastic differential equations.

$$X(t) = \int_0^t \mathring{a}(s,X(s))ds + \int_0^t \mathring{A}(s,X(s))dw(s) \qquad (18)$$

in T_mM, where $\mathring{a}(s,X)$ and $\mathring{A}(s,X)$ are Lipschitz, vanish outside a certain neighbourhood of the origin in T_mM, and are such that $\mathring{a}(s,0) = a(t,m)$ and $\mathring{A}(s,0) = A(m)$. Note that the solutions of (18) are strong.

The expression (see [5-7])

$$d\,\xi(t) = \exp_{\xi(t)}(a(t,\xi(t)), A(\xi(t))) \qquad (19)$$

is called the Ito equation in the form of Ya.I.Belopol'skaya - Yu.L.Daletskiĭ. It means that the process $\xi(t+\tau)$ for $\tau > 0$ belongs to the class $\exp_{\xi(t)}(a(t,\xi(t)), A(\xi(t)))$ until it leaves a certain neighbourhood of $\xi(t)$ (cf. definition 3).

We do not present the description of (19) in local coordinates but we should point out that it has the form of Ito equation in a linear space where the local connector of the connection is involved; this local description is

covariant under changes of coordinates.

Remark 2. Equation (19) is compatible with mappings of manifolds. Let $f : M \longrightarrow N$ be a C^2- mapping and there be a connection on the manifold N with the exponential map $\exp^N$ such that $f(\exp X) = \exp^N(Tf \circ X)$ for each $X \in TM$. One can easily show that for $\xi(t)$ on M satisfying (19) the process $f(\xi)$ on N satisfies the equation

$$df(\xi(t)) = \exp_{f(\xi(t))}(Tf \circ a(t,\xi(t)),\ Tf \circ A(\xi(t))).$$

A more detailed description and justification of this construction can be found in [5-7].

A sample trajectory of a stochastic process $\xi(t)$ is a.s. non-differentiable, i.e. the derivative $d\xi/dt$ does not exist. Following Nelson [18], we define the "mean forward derivative" by

$$D\xi(t)_m = \lim_{\Delta t \to +0} E\left(\frac{\xi(t+\Delta t) - \xi(t)}{\Delta t} \,\middle|\, \xi(t) = m \right) \qquad (20)$$

where $\Delta t \to +0$ means that $\Delta t \longrightarrow 0$ and $\Delta t > 0$. If $\xi(t)$ is a solution to (19) one can show that the limit (20) exists and $D\xi(t)_m = a(t, m)$. Here one must use the properties of Ito equations, of $w(t)$ and the fact that $\rho_t^{\xi} = \rho_t^{w}$ for strong solutions to (19), see remark 1.

The "mean backward derivative" is defined by

$$D_*\xi(t)_m = \lim_{\Delta t \to +0} E\left(\frac{\xi(t) - \xi(t-\Delta t)}{\Delta t} \,\middle|\, \xi(t) = m \right) \qquad (21)$$

It should be noted that in general $D\xi(t)_m \neq D_*\xi(t)_m$.

Following Nelson [18] we call $v(t,m) = \frac{1}{2}(D\xi(t)_m + D_*\xi(t)_m)$ and $u(t,m) = \frac{1}{2}(D\xi(t)_m - D_*\xi(t)_m)$ current and osmotic velocities of the process $\xi(t)$ respectively.

For the case when M is a n-dimensional Riemannian manifold let us assume that the field A in (19) has the form $A = \sigma A'$ where $\sigma > 0$ is a constant, $A'(m) : R^n \longrightarrow T_mM$ is an orthogonal operator (T_mM is considered as n-dimensional

Euclidean space with respect to Riemannian scalar product). Now the osmotic velocity can be described as follows. It is known that there exists a probability density $\rho(t,m)$ on $[0,\ell] \times M$ such that for any continuous function $f(t,m)$ on $[0,\ell] \times M$ we have

$$\int_{[0,\ell]\times M} f\rho \, d\nu = \int_{[0,\ell]\times\Omega} f(\xi(t)) \, d\mathbb{P}\, dt$$

where ν is Lebesgue measure on $[0,\ell]\times M$. Finally, one obtains $u = \sigma^2 \operatorname{grad} \log\sqrt{\rho}$. Using this fact one can show for a solution $\xi(t)$ to (19) that the limit in (21) exists and $D_*\xi(t)_m = a_*(t,m) = a(t,m) - 2u(t,m)$. The detailed presentation of these results can be found in [18].

Let $Y(t,m)$ be a smooth vector field on M. Define the "mean forward" and "mean backward" covariant derivatives of Y along $\xi(t)$ by the relations

$$\frac{D}{dt} Y|_m = K \circ \lim_{\Delta t \to +0} E\left(\frac{Y(t+\Delta t, \xi(t+\Delta t)) - Y(t,\xi(t))}{\Delta t} \,\middle|\, \xi(t)=m \right),$$

$$\frac{D_*}{dt} Y|_m = K \circ \lim_{\Delta t \to +0} E\left(\frac{Y(t,\xi(t)) - Y(t-\Delta t, \xi(t-\Delta t))}{\Delta t} \,\middle|\, \xi(t)=m \right) \tag{22}$$

involving the connector K of the given connection (cf. section 2).

When M is n-dimensional Riemannian manifold and $A = \sigma A'$ (see above) using the Ito formula (17) we obtain:

$$\frac{D}{dt} Y = \frac{\partial}{\partial t} Y + \nabla_a Y + \frac{\sigma^2}{2} \Delta Y,$$

$$\frac{D_*}{dt} Y = \frac{\partial}{\partial t} Y + \nabla_{a_*} Y - \frac{\sigma^2}{2} \Delta Y, \tag{23}$$

where Δ is Laplace-Beltrami operator $\nabla^* \circ \nabla$ (see [10, 13, 18]). $\frac{\sigma^2}{2}\Delta Y$ corresponds to $\frac{1}{2}$ tr $Y''(A,A)$ in Ito formula.

Formula (23) is easily verified in a normal neighbourhood of a point. See [18] for the details.

Let us fix smooth vector fields $a(t,m)$, $a_*(t,m)$ and a smooth field of operators $A(m)$. The problem arises: to describe by means of the Ito equations the processes which satisfy the equations

$$D\xi(t)_m = a(t, m) \tag{24}$$

$$D_*\xi(t)_m = a_*(t,m) \tag{25}$$

at each t,m for the given a, a_*, A. As it is mentioned above $\xi(t)$ satisfies (24) if it is a solution to (19). For equation (25) the answer is more complicated.

Consider a Wiener process $w(t)$ in R^n, $t\in[0,\ell]$ and for $t\in(0,\ell]$ define the process $D_*w(t)$ by the equality

$$D_*w(t) = \lim_{\Delta t\to+0} E\left(\frac{w(t)-w(t-\Delta t)}{\Delta t}\,\middle|\, w(t)\right). \tag{26}$$

It is obvious that for $x\in R^n$ $D_*w(t)_x = -2u^w(t,x)$, where $u^w(t,x)$ is the osmotic velocity of $w(t)$ at x. Recall that $u^w(t,x) = \operatorname{grad}\log\sqrt{\rho^w(t,x)}$, where the density $\rho^w(t,x) = (2\pi t)^{-(n/2)}\cdot e^{-(x\cdot x/2t)}$. The direct calculations give $\operatorname{grad}\log\sqrt{\rho^w} = -\frac{1}{2}\cdot\frac{x}{t}$. Thus $D_*w(t) = \frac{w(t)}{t}$.

Note that $\frac{w(t)}{t}$ is not determined for $t=0$ but the integral $\int_0^t \frac{w(s)}{s}ds$ is well-defined for $t\in[0,\ell]$ a.s. This follows from the estimate $E\left(\left|\int_0^t \frac{w(s)}{s}ds\right|\right) < C\cdot\sqrt{t}$ where the constant $C>0$ depends only on n.

Consider on M the following equation

$$d\xi(t) = \exp_{\xi(t)}(a_*(t,\xi(t))-A(\xi(t))D_*w(t),A(\xi(t))) \tag{27}$$

where $(a_*(t,m)-A(m)D_*w(t),A(m))$ means the class of stochastic processes in T_mM which consists of solutions to equations

$$X(\tau) = \int_0^\tau \mathring{a}_*(s,X(s))ds - \int_0^\tau \mathring{A}(s,X(s))\frac{w(s)}{s}ds + \int_0^\tau \mathring{A}(s,X(s))dw(s) \tag{28}$$

$\mathring{a}_*(s,X)$ and $\mathring{A}(s,X)$ are analogous to $\mathring{a}(s,X)$, $\mathring{A}(s,X)$ in (18). Note that the second integral in the right-hand side of (28) is well-defined because $A(s,X)$ is bounded.

Theorem 2. Let $\xi(t)$, $\xi(0) = m_0$, be a local strong solution to (27) in a certain neighbourhood $U \ni m_0$. Then $\xi(t)$ satisfies (25) for $m \in U$, $t > 0$.

Proof. For the sake of simplicity suppose that t and m are sufficiently close to 0 and m_0 respectively. Under this assumption consider $X \in T_{m_0}M$ such that $\exp X = m$. Let $(\tau) = \exp_{m_0} X(\tau)$, where

$$X(\tau) = \int_0^\tau \hat{a}_*(s,X(s))ds - \int_0^\tau \hat{A}(s,X(s))\frac{w(s)}{s}ds + \int_0^\tau \hat{A}(s,X(s))dw(s)$$

is a process in $T_{m_0}M$, $\tau \in [0,t]$. Such $X(\tau)$ exists because $\xi(t)$ is a solution to (27). It is sufficient to show that $D_*X(t)_X = \hat{a}_*(t,X)$.

The direct calculation gives

$$D_*X(t)_X = E(\hat{a}_*(t,X(t)) \mid X(t)=X) - E(\hat{A}(t,X(t))D_*w(t) \mid X(t)=X) +$$
$$+ \lim_{t\to+0} E\left(\frac{\int_0^t \hat{A}(s,X(s))dw(s) - \int_0^{t-\Delta t} \hat{A}(s,X(s))dw(s)}{\Delta t} \mid X(t)=X \right). \tag{29}$$

The first two summands in the right-hand side of (29) are obtained in such form because the corresponding processes have differentiable trajectories. It is easy to see that the sum of the last two summands in (29) is equal to zero. Indeed, $\int_0^t \hat{A}(s,X(s))dw(s) - \int_0^{t-\Delta t} \hat{A}(s,X(s))dw(s)$ is measurable with respect to $\mathcal{F}_t^w$. Then the assumption $\mathcal{P}_t^w = \mathcal{P}_t^X$ for a strong solution $X(t)$ of the Ito equation (see remark 1), the Markov property of $w(t)$ and the properties of the conditional expectation and of Ito integral give

$$\lim_{\Delta t\to+0} E\Big(\frac{\int_0^t \hat{A}(s,X(s))dw(s)-\int_0^{t-\Delta t}\hat{A}(s,X(s))dw(s)}{\Delta t}\Big| X(t)=X\Big) =$$

$$= \lim_{\Delta t\to+0} E(E(\frac{\int_0^t \hat{A}(x,X(s))dw(s)-\int_0^{t-\Delta t}\hat{A}(sX(s))dw(s)}{\Delta t}\Big|\rho_t^X\Big| X(t)=X)=$$

$$= E\Big(\lim_{\Delta t\to+0} E(\hat{A}(t,X(t))\frac{w(t)-w(t-\Delta t)}{\Delta t}\Big| w(t))\Big| X(t)=X\Big) =$$

$$= E(\hat{A}(t,X(t))D_*w(t)\big|X(t)=X). \quad \text{Thus } D_*X(t)_X =$$

$$= E\,(\hat{a}_*(t,X(t))\big|X(t)=X) = \hat{a}_*(t,X). \quad \text{Q.E.D.}$$

Theorem 3. Let $a_*(t,m)$ and $A(m)$ be at least C^1 - smooth in m and let $a_*(t,m)$ be continuous in t. Then for each $m_0\in M$ there exists a unique local strong solution $\xi(t)$ to (27) with the initial condition $\xi(0)=m_0$, and $\rho_t^\xi = \rho_t^w$ for each t for which $\xi(t)$ exists.

Theorem 3 is a consequence of the existence and uniqueness theorem [5], see also [10].

4. Stochastic-Geometrical Lagrangian Approach to Viscous Incompressible Hydrodynamics. General Construction.

Now we can transfer the developed stochastic machinery to the systems on the groups of diffeomorphisms. We shall do it for the special case of the groups of diffeomorphisms of the flat n-dimensional torus T^n, i.e. T^n with the Riemannian metric obtained from the Euclidean metric in R^n after factorization with respect to the integral lattice. For the sake of simplicity we suppose that the total Riemannian volume of T^n is equal to 1.

Consider on $D_\mu^s(T^n)$ the weakly Riemannian metric $(\ ,\ )$ (1), its Levi-Civita connection, exponential map $\widetilde{exp}$ and all other geometrical objects defined for the general case in section 2.

It is a well-known fact that all tangent spaces to T^n

are naturally isomorphic to R^n. Denote by $A(m) : R^n \longrightarrow T_m T^n$, $m \in T^n$, this natural isomorphism. Thus the field A of linear isomorphisms of R^n onto tangent spaces to T^n is constructed. Obviously for each given $y \in R^n$ the vector field $A \circ y : T^n \longrightarrow TT^n$ on T^n is constant (i.e. one may imagine that the same vector y is applied at every point of T^n) and consequently $A \circ y$ is a C^∞-smooth zero-divergent vector field. Moreover, the constant vector field $A \circ y$ is harmonic because evidently $d(A \circ y) = 0$.

Thus the field A may be considered as a linear operator $\tilde{A}(e) : R^n \longrightarrow T_e D^s_\mu(T^n)$. For $\eta \in D^s_\mu(T^n)$ denote by $\tilde{A}(\eta) : R^n \longrightarrow T_\eta D^s_\mu(T^n)$ the operator determined by the formula $\tilde{A}(\eta) \circ y = \left[\tilde{A}(e) \circ y\right] \circ \eta = \left[A \circ y\right] \circ \eta$. So the field $\tilde{A}$ of linear operators mapping R^n into tangent spaces to $D^s_\mu(T^n)$ is constructed. Obviously $\tilde{A}$ is right-invariant. Since the field A on T^n is C^∞-smooth, the right-invariant field $\tilde{A}$ on $D^s_\mu(T^n)$ is C^∞-smooth.

The construction of the field $\tilde{A}$ is a variant of the general construction of [10].

Fix a real constant $G > 0$. It is obvious that for a given vector field on $D^s_\mu(T^n)$ we may consider the stochastic equations of type (19) and (27) involving the exponential map $\widetilde{exp}$, the operator field $\tilde{A}$ and a Wiener process $w(t)$ in R^n.

For each $X \in T_e D^s_\mu(T^n)$ consider the natural decomposition $T_X TD^s_\mu(T^n) = \tilde{V}_X + \tilde{H}_X$ where $\tilde{V}_X = T_X T_{\pi X} D^s_\mu(T^n)$ is vertical subspace and $\tilde{H}_X$ is Levi-Civita connection at X. Define the weakly Riemannian metric $(\ ,\ )^T$ on $TD^s_\mu(T^n)$ determining the scalar products in $\tilde{V}_X$ and $\tilde{H}_X$ as inverse images of $(\ ,\ )$ with respect to $\tilde{K}$ and $T\pi$ respectively (recall that $T\pi : \tilde{H}_X \longrightarrow T_{\pi X} D^s_\mu(T^n)$ and $\tilde{K} : \tilde{V}_X \longrightarrow T_{\pi X} D^s_\mu(T^n)$ are isomorphisms) and assuming that $\tilde{V}_X$ and $\tilde{H}_X$ are orthogonal to each other. The application of calculus of variations shows that straight lines in tangent spaces to $D^s_\mu(T^n)$ and only they

are vertical Levi-Civita geodesics of $(\ ,\ )^T$ and all other such geodesics are mapped by $T\pi$ onto Levi-Civita geodesics of $(\ ,\ )$ on $D^s_\mu(T^n)$. Denote by $\widetilde{\exp}^T$ the exponential map of the Levi-Civita connection of $(\ ,\ )^T$ on $TD^s_\mu(T^n)$. It is obvious that for every $Y \in TTD^s_\mu(T^n)$ we have $\widetilde{\exp}^T Y = \widetilde{\exp}\, T\pi Y$ (cf. remark 2).

Define on $TD^s_\mu(T^n)$ the field of operators $\tilde{A}^T(X) = T\pi^{-1}\tilde{A}(\pi X)\big|_{H_X} : R^n \longrightarrow T_X TD^s_\mu(T^n)$. Let $F \in T_e D^s_\mu(T^n)$, $\bar{F}$ be the corresponding right-invariant vector field on $D^s_\mu(T^n)$, $\bar{F}^\ell$ the natural vertical lift of $\bar{F}$ on $TD^s_\mu(T^n)$. Let us consider a Wiener process $w(t)$ in R^n and determine the stochastic equations on $TD^s_\mu(T^n)$ as follows:

$$d\zeta(t) = \widetilde{\exp}^T_{\zeta(t)}(S(\zeta(t))\cdot\sigma\tilde{A}^T(\zeta(t))\circ D_* w(t), \sigma\tilde{A}^T(\zeta(t))) \qquad (30)$$

$$d\zeta(t) = \widetilde{\exp}^T_{\zeta(t)}(S(\zeta(t))+\bar{F}^\ell(\zeta(t))\cdot\sigma\tilde{A}^T(\zeta(t))\circ D_* w(t), \sigma\tilde{A}^T(\zeta(t))) \qquad (31)$$

where S is the geodesic pulverization of the Levi-Civita connection on $TD^s_\mu(T^n)$ (see sect.2).

Let $F \in H^{s+1}$ on T^n. Theorem 3 is valid for equations (30) and (31). Indeed, the fields S and $\tilde{A}^T$ are C^∞-smooth and the field $\bar{F}^\ell$ is C^1 - smooth on $TD^s_\mu(T^n)$. Let $\zeta(t)$ be a (local) strong solution to (30) or (31). Consider the process $\xi(t) = \pi\zeta(t)$ on $D^s_\mu(T^n)$ where $\pi: TD^s_\mu(T^n) \longrightarrow D^s_\mu(T^n)$ is the natural projection.

The main purpose of the rest of this section is to show that $\xi(t)$ is naturally connected with the motion of viscous incompressible fluid, namely the expectation (in a certain sense) of $\xi(t)$ is a flow on T^n of the fluid mentioned above.

Consider a solution $\zeta(t)$ to (30) or (31) with the initial condition $\zeta(0) = u_0 \in T_e D^s_\mu(T^n)$. For t, such that $\zeta(t)$ exists, and for $\omega \in \Omega$ $\zeta(t,\omega)$ is a vector in $T_{\xi(t,\omega)} D^s_\mu(T^n)$. Denote by $u(t,\omega)$ the vector $TR^{-1}_{\xi(t,\omega)}\zeta(t,\omega) \in T_e D^s_\mu(T^n)$, $u(t,\omega)$ is a random vector in $T_e D^s_\mu(T^n)$,

i.e. a random vector field on T^n. Denote by $u(t)$ the expectation of $u(t,\omega)$. We should point out that the vector $u(t)_m$ for each $m \in T^n$ is the expectation of the random vector $u(t,\omega)_m \in T_m T^n$.

Consider a certain value $\zeta(t,\omega)$, $\eta = \pi\zeta(t,\omega) \in D^s_\mu(T^n)$ and the random vector $E(\zeta(t) | \pi\zeta(t) = \eta) \in T_\eta D^s_\mu(T^n)$.

Lemma 1. $E(E(\zeta(t) | \pi\zeta(t) = \eta)) = T R_\eta u(t)$.

Indeed, $TR_\eta^{-1} E(E(\zeta(t) | \pi\zeta(t) = \eta) = E(E(TR_\eta^{-1}\zeta(t) | \pi\zeta(t) = \eta) = E(E(TR^{-1}_{\pi\zeta(t)}\zeta(t) | \pi\zeta(t) = \eta)) = E(TR^{-1}_{\pi\zeta(t)}\zeta(t)) = u(t)$.

Lemma 2. $D_*\xi(t)_\eta$ exists and it is equal to $TR_\eta u(t)$.

Proof. First consider the case when $\xi(t)$ is constructed from $\zeta(t)$ satisfying (30). It is obvious that $\zeta(t)$ has independent increments. The direct verification of the independence property shows that $\xi(t) = \pi\zeta(t)$ also has independent increments. Since $\xi(t)$ and $\xi(t) - \xi(t-\Delta t)$ are independent, $E(\xi(t) - \xi(t-\Delta t) | \xi(t) = \eta) = E(E(\xi(t) - \xi(t-\Delta t) | \xi(t) = \eta))$. As a consequrnce of theorem 2 one obtains the equality $D_*\zeta(t)_X = S(X)$ for $X \in TD^s_\mu(T^n)$. Recall that $T\pi S(X) = X$. So $D_*\xi(t)_\eta = E \lim_{\Delta t \to +0} E\left(\frac{\pi\zeta(t) - \pi\zeta(t-\Delta t)}{\Delta t} \middle| \pi\zeta(t) = \eta\right) = ET\pi\left(\lim_{\Delta t \to +0} E\left(\frac{\zeta(t) - \zeta(t-\Delta t)}{\Delta t} \middle| \pi\zeta(t) = \eta\right)\right) =$

$= ET\pi S(E(\zeta(t) | \pi\zeta(t) = \eta)) = E(E(\zeta(t) | \pi\zeta(t) = \eta))$.

The application of lemma 1 completes the proof. If $\zeta(t)$ satisfies (31), then by theorem 2 $D_*\zeta(t)_X = S(X) + \bar{F}^\ell(X)$ where $\bar{F}^\ell(X)$ is vertical. So $T\pi(S(X) + \bar{F}^\ell(X)) = X$ and the same arguments are valid for this case too. Q.E.D.

Denote by $b_*(t,\eta)$ the right-invariant vector field on $D^s_\mu(T^n)$ generated by $u(t) \in T_e D^s_\mu(T^n)$. It follows from lemma 2, equations (30) and (31), remark 2 and the relation between $\widetilde{\exp}$ and $\widetilde{\exp}^T$ that $\xi(t)$ satisfies the equation

$$d\xi(t) = \widetilde{\exp}_{\xi(t)}(b_*(t,\xi(t)) - \tfrac{1}{2}\widetilde{A}(\xi(t))D_*w(t), \tfrac{1}{2}\widetilde{A}(\xi(t))). \quad (32)$$

Using the Levi-Civita connection determine the mean backward $\frac{\widetilde{D}_*}{dt}$ covariant derivative along $\xi(t)$ on $D^s_\mu(T^n)$ according to the general formula (22) with connector $\widetilde{K}$ defined in (6).

Theorem 4. Let $\zeta(t)$ satisfy (30) ((31), respectively). Then $\xi(t)$ satisfies the equation

$$\frac{\widetilde{D}_*}{dt} D_* \xi(t)_\eta = 0 \quad (33)$$

(the equation

$$\frac{\widetilde{D}_*}{dt} D_* \xi(t)_\eta = F \quad (34)$$

respectively) at each point $\eta \in D^s_\mu(T^n)$.

Proof. Since the fields S, b_*, A are right-invariant, we may suppose that $\eta = e$ without loss of generality. Thus $D_*\xi(t)_e = u(t)$ by lemma 2. First consider $\zeta(t)$ satisfying (30). Using theorem 2 and according to (30) and (32) we obtain

$$E(\zeta(t-\Delta t)|\zeta(t)) = u(t) = \widetilde{\exp}^T(-S(u(t))\Delta t) + o(\Delta t),$$

$$E(\xi(t-\Delta t)|\xi(t) = e) = \widetilde{\exp}(-u(t)\Delta t) + o(t).$$

By definition of u(t) we have $u(t-\Delta t) = E(\zeta(t-\Delta t)\circ\xi^{-1}(t-\Delta t)) = E(E(\zeta(t-\Delta t)\circ\xi^{-1}(t-\Delta t)|\zeta(t) = u(t), \xi(t) = e) = \widetilde{\exp}^T(-S(u(t))\Delta t)\circ(\exp(-u(t)\Delta t))^{-1} + o(t)$. Then

$$E(b_*(t-\Delta t, \xi(t-\Delta t))|\xi(t) = e) = E(u(t-\Delta t)\circ\xi(t-\Delta t)|\xi(t) = e) = u(t-\Delta t)\circ E(\xi(t-\Delta t)|\xi(t) = e) = \widetilde{\exp}^T(-S(u(t))\Delta t) + o(\Delta t).$$

Thus

$$\frac{\widetilde{D}_*}{dt} D_* \xi(t)_e = \widetilde{K}\circ\lim_{\Delta t\to+0} E\left(\frac{u(t) - b_*(t-\Delta t, (t-\Delta t))}{\Delta t}|\xi(t) = e\right) = \widetilde{K}\circ S.$$

Since $S \in \widetilde{H}$, $\widetilde{K} \circ S = 0$ which proves (33).

If $\xi(t)$ satisfies (31) one should replace S by $S+\bar{F}^{\ell}$ in the above arguments, so $\frac{\widetilde{D}}{dt} D_* \xi(t)_e = \widetilde{K}(S+\bar{F}^{\ell}) = \bar{F}$. Q.E.D.

Note that (34) is one of stochastic analogues of the Newton's law (11), in particular (33) is an analogue of geodesics equation (14). If $\sigma = 0$, (34) and (33) turn into (11) and (14), respectively.

Let us find an analogue of Euler equation (12) for the case under consideration.

Theorem 5. Let $\xi(t)$ satisfy (30) ((31), respectively). Then the vector field $u(t)$ on T^n satisfies the Navier-Stokes equation

$$\frac{\partial}{\partial t} u + \nabla_u u - \nu \Delta u + \operatorname{grad} p = 0 \qquad (35)$$

(respectively

$$\frac{\partial}{\partial t} u + \nabla_u u - \nu \Delta u + \operatorname{grad} p = F\ ), \qquad (36)$$

where Δ is Laplace operator on T^n, $\nu = \frac{\sigma^2}{2}$.

Proof. Fix t such that $\xi(t)$ exists. Without loss of generality we may assume that $\xi(t) = e$. It is a consequence of the fact that S, b_*, $\widetilde{A}$ are right-invariant. The process $\xi(t)$ may be considered as a stochastic flow on T^n (see [10,13]). So for each $m \in T^n$ we can find $D_* \xi(t)_m = u(t,m)$. According to formula (23) we obtain $\frac{D_*}{dt} D_* \xi(t)_m = \frac{\partial}{\partial t} u + \nabla_u u - \nu \Delta u$. The construction of $\frac{\widetilde{D}}{dt}$ by general formula (22) with $\widetilde{K}$ defined by (8) and property (5) of P lead to $\frac{\widetilde{D}}{dt} D_* \xi(t)_e = P_e(\frac{\partial}{\partial t} u + \nabla_u u - \nu \Delta u) = \frac{\partial}{\partial t} u + \nabla_u u - \nu \Delta u + \operatorname{grad} p$. The application of formulas (33) and (34) completes the proof. Q.E.D.

Corollary 1. Under the conditions of theorem 5 the flow $g(t)$ of the vector field $u(t)$ on T^n is a curve in $D^s_\mu(T^n)$ describing the motion of viscous incompressible fluid on

T^n with the viscosity ν in the case of zero external force (under the action of the external force F, respectively).

Corollary 2. (i) Let $s > \frac{n}{2} + 1$, the zero-divergence vector field (external force) F on T^n belong to the class H^{s+1}, the zero divergence vector field (initial velocity) u_0 on T^n belong to H^s. Then for any $\nu > 0$ the unique solution $u(t) \subset T_e D^s_\mu(T^n)$ to the Navier-Stokes equation(36), $u(0) = u_0$, and the corresponding flow of viscous fluid $g(t)$ $D^s_\mu(T^n)$ exist on a certain interval $t \in [0, \varepsilon)$. (ii) The same is valid for $F \in H^s$, $u_0 \in H^s$ when $s > \frac{n}{2} + 2$.

Proof. Statement (i) is a consequence of the existence and uniqueness theorem 3 applied to equation (31) (see the beginning of this section) and of theorem 5. To prove (ii) note that under these assumptions the manifold $D^{s-1}_\mu(T^n)$ is well defined so that the right invariant vector field F on $D^{s-1}_\mu(T^n)$ is C^1 - smooth. Thus theorem 3 is valid for equation (31) on $TD^{s-1}_\mu(T^n)$. Q.E.D.

Analogously to the theory of stochastic differential equations in vector spaces we may call $g(t)$ the mathematical expectation of $\xi(t)$.

Theorem 6. Let F be a zero-divergence H^{s+k} vector field on T^n, $1 \leqslant k \leqslant \infty$, and u_0 be a zero-divergence H^{s+q} vector field, $1 \leqslant q \leqslant k$. Then $u(t)$ belongs to the class H^{s+q} for all t for which it exists in H^s.

Corollary 1. Under the conditions of theorem 6 the curve $g(t)$ lies in $D^{s+q}_\mu(T^n)$ for all t for which it exists in $D^s_\mu(T^n)$.

Corollary 2. If $F = 0$ then theorem 6 and corollary 1 are valid for all $q \geqslant 1$, in particular if $u_0 \in C^\infty$ then $u(t) \subset C^\infty$ and $g(t)$ lies in $D^\infty_\mu(T^n)$ for all t for which they exist in H^s.

The statements of theorem 6 and its corollaries are called the regularity properties.

Proof of theorem 6. Denote by $\gamma_Y(t)$ the solution to (31) with the initial condition $\gamma_Y(0) = Y$ and consider the random mapping $\gamma_t : TD^s_\mu(M) \longrightarrow TD^s_\mu(M)$ determined by the formula $\gamma_t(X) = \gamma_X(t)$, $X \in TD^s_\mu(T^n)$.

Lemma 3. The mapping γ_t is $D^s_\mu(T^n)$-right-invariant and C^k - smooth in square mean metric (s.m.C^k-smooth).

The right invariance is a consequence of the same property of $S, \tilde{A}^T, \bar{F}^\ell$. The s.m.$C^k$-smoothness of γ_t is a consequence of C^∞ - smoothness of S and $\tilde{A}^T$ and C^k - smoothness of the right-invariant vector field $\bar{F}$ (see section 1), i.e. $\bar{F}^\ell$ is also C^k - smooth. The arguments for the proving of smoothness here are completely analogous to those of [4,7]. A simple modification connected with the presence of the term $-\sigma \tilde{A}^T D_* w(t)$ is left for the reader as an exercise. Q.E.D.

Consider the right-invariant vector field $\bar{u}_o$ on $D^s_\mu(T^n)$ constructed from $u_o \in T_e D^s_\mu(T^n)$ by right-hand translations. Since u_o is H^{s+q} vector field on T^n, $\bar{u}_o$ is C^q - vector field on $D^s_\mu(T^n)$. It follows from lemma 3 that $\gamma_t \bar{u}_o$ is right-invariant random s.m.C^q-smooth vector field on $D^s_\mu(T^n)$. Thus $R_{\pi\gamma_t^{-1}} \bar{u}_o \circ (\gamma_t \bar{u}_o)$ is a random s.m. H^{s+q}-vector field on T^n and $u(t) = E(R_{\pi\gamma_t^{-1}} \bar{u}_o \circ (\gamma_t \bar{u}_o))$ is an H^{s+q} vector field on T^n. Q.E.D.

5. The Case of a Bounded Domain with Boundary.

In this section by combining the constructions of section 4 and of theorem 1 we consider the viscous fluid motion in a bounded domain in R^n with frictionless boundary. In the end of the section we say some words about the case of the fluid adhering to the boundary.

Let (H) be a bounded domain in R^n with a smooth boundary $\partial(H)$. Without loss of generality we may assume that (H) belongs to the unit cube in R^n, so after factorization of R^n with respect to the integral lattice (H) becomes

imbedded in a flat torus T^n. Let us apply the construction of theorem 1 to the case $M = \textcircled{H}$, $N = T^n$.

Consider the right-invariant C^∞ subbundle Ξ^s of $TD^s_\mu(T^n)$ and the projector $\bar{R} : TD^s_\mu(T^n) \longrightarrow \Xi^s$ which exist by theorem 1. Let $\tilde{H}$ be Levi-Civita connection on $TD^s_\mu(T^n)$. Then $H^\Xi = T\bar{R}\tilde{H}$ is a connection on the bundle Ξ^s where $\tilde{H}$ is considered at the points of Ξ^s. Using the connections $T\bar{R}\tilde{H}$ on Ξ^s and $\tilde{H}$ on $TD^s_\mu(T^n)$ we may construct, according to [5-7], the (affine) connection on the manifold Ξ^s. Denote by $\exp^\Xi : T\Xi^s \longrightarrow \Xi^s$ the exponential map of this connection.

Consider the vector fields $S^\Xi = T\bar{R}\, S$ and $T\bar{R}\bar{F}^\ell$ (where $F \in \Xi^s_e$) on Ξ^s introduced in theorem 1 and its corollary. Note that obviously $S^\Xi \subset T\bar{R}\tilde{H}$. We should also mention the evident property $T\pi S^\Xi(X) = X$ for every $X \in \Xi^s$, where $\pi : \Xi^s \longrightarrow D^s_\mu(T^n)$ is the natural projection.

At $X \in \Xi^s$ determine the linear operator $\tilde{A}^\Xi(X) = T\bar{R}\,\tilde{A}^T(X) : R^n \longrightarrow T_X\Xi^s$ where $\tilde{A}^T$ is described in section 4. It is evident that $T\pi\tilde{A}^\Xi \circ y = \tilde{A} \circ y$ (see section 4). Note that the fields S^Ξ, $T\bar{R}\,\bar{F}^\ell$, $\tilde{A}^\Xi$ on Ξ^s are $D^s_\mu(T^n)$-right-invariant, S^Ξ and $\tilde{A}^\Xi$ are C^∞-smooth and $T\bar{R}\,\bar{F}^\ell$ is C^k - smooth iff $F \in H^{s+k}$, $k > 0$, on T^n.

Thus we may consider on Ξ^s the stochastic differential equations

$$d\zeta(t) = \exp^\Xi_{\zeta(t)}(S^\Xi(\zeta(t)) - \sigma\tilde{A}^\Xi(\zeta(t)) \circ D_* w(t)\ ,\ \sigma\tilde{A}^\Xi(\zeta(t))), \tag{37}$$

$$d\zeta(t) = \exp^\Xi_{\zeta(t)}(S^\Xi(\zeta(t)) + T\bar{R}\,\bar{F}^\ell(\zeta(t)) - \sigma\tilde{A}^\Xi(\zeta(t)) \circ D_* w(t),\ \sigma\tilde{A}^\Xi(\zeta(t))) \tag{38}$$

(cf. (30) and (31)). Note that according to theorem 3 for each $X_0 \in \Xi^s$ there exists a unique local strong solution $\zeta(t)$ to (37) with the initial condition $\zeta(0) = X_0$,

because S^{Ξ} and $\tilde{A}^{\Xi}$ are C^{∞}-smooth and the same is valid for (38) if F is at least H^{s+1}-vector field on T^n, i.e. $T\bar{R}\,\bar{F}^{\ell}$ is C^1 on Ξ^s.

Let $\mathfrak{z}(t)$ be a local strong solution to (37) or (38) on Ξ^s. Consider its projection $\xi(t) = \pi\mathfrak{z}(t)$ on $D^s_{\mu}(T^n)$ and the vector $u(t) = E(TR^{-1}_{\xi(t)}\,\mathfrak{z}(t))$ which evidently belongs to Ξ^s_e.

Lemma 4. For each $\eta \in D^s_{\mu}(T^n)$ the vector $D_*\xi(t)_{\eta}$ exists and is equal to $TR_{\eta}\,u(t) \in \Xi^s_{\eta}$.

The proof of lemma 4 is analogous to the proof of lemma 2. Note that the analogue of lemma 1 also holds for this case.

Let us introduce the backward mean covariant derivative $\frac{D^{\Xi}_*}{dt}$ along $\xi(t)$ with respect to Ξ^s by the general formula (22) with the connector K^{Ξ} defined by the relation

$$K^{\Xi} = \bar{R} \bullet K . \tag{39}$$

By analogy with theorem 4 we obtain

Theorem 7. Let $\mathfrak{z}(t)$ on Ξ^s satisfy (37), ((38), respectively). Then $\xi(t) = \pi\mathfrak{z}(t)$ on $D^s_{\mu}(T^n)$ satisfies the equation

$$\frac{D^{\Xi}_*}{dt} D_* \xi(t) = 0 \tag{40}$$

(the equation

$$\frac{D^{\Xi}_*}{dt} D_* \xi(t) = F$$

respectively).

Theorem 8. Let $\mathfrak{z}(t)$ on Ξ^s satisfy (37) ((38), respectively), $\xi(t)$ and $u(t)$ be defined as it is mentioned above. The restriction $u(t)|_{\textcircled{M}}$ is a zero-divergence vector field on $\textcircled{M}$ tangent to $\partial\textcircled{M}$ and it satisfies in $\textcircled{M}$ the Navier-Stokes equation with the viscosity $\nu = \frac{G}{2}$ and the zero external force (external

force $F_{|\textcircled{H}}$, respectively).

Proof. Since $u(t)$ belongs to Ξ^s_e, $u(t)_{|\textcircled{H}}$ is zero-divergence vector field tangent to $\partial\textcircled{H}$ by definition of Ξ^s (see theorem 1). As in the proof of theorem 5 the process $\xi(t)$ in $D^s_\mu(T^n)$ may be considered as a stochastic flow on T^n; using the fact that $T\pi\tilde{A}^\Xi = \tilde{A}$ (see above), one can show that $\frac{D}{dt} D_* \xi(t)_m = \frac{\partial}{\partial t} u + \nabla_u u - \nu \triangle u$. Then the construction of $\frac{D^\Xi}{dt}$ by the general formula (22) with the connector K^Ξ defined by (39), the property (15) of $\bar{R}$ and formulas (40), (41) complete the proof. Q.E.D.

Corollary 1. The flow $g(t)$ of $u(t)$ on T^n may be restricted onto $\textcircled{H}$ and $g(t)_{|\textcircled{H}}$ is a curve in $D^s_\mu(\textcircled{H})$ describing the motion of incompressible fluid with viscosity ν in $\textcircled{H}$ with frictionless boundary $\partial\textcircled{H}$ under the action of the zero external force (external force $F_{|\textcircled{H}}$, respectively).

Recall that the words "frictionless boundary" mean that $u(t)_{|\textcircled{H}}$ is not necessarily equal to zero, but is tangent to $\partial\textcircled{H}$ only, i.e. the fluid does not adhere to the boundary.

Note that $F \in \Xi^s_e$ is uniquely determined by the external force $F_{|\textcircled{H}}$, see theorem 1 (i).

Corollary 2. (i) Let $s > \frac{n}{2} + 1$, the vector field (external force) $F \in T_e D^s_\mu(\textcircled{H})$ belong to the class H^{s+1} on $\textcircled{H}$. Consider the initial velocity vector field $u_0 \in T_e D^s_\mu(\textcircled{H})$. For any $\nu > 0$ the unique solution $u(t) \subset T_e D^s_\mu(\textcircled{H})$ of the Navier-Stokes equation with the force F and the corresponding flow $g(t) \subset D^s_\mu(\textcircled{H})$ of the viscous fluid exist on a certain interval $t \in [0, \varepsilon)$. (ii) The same is valid for $F \in H^s$, $u_0 \in H^s$ when $s > \frac{n}{2} + 2$.

Here the arguments in the proof are the same as for corollary 2 to theorem 5. Of course, $TD^s_\mu(T^n)$ and $TD^{s-1}_\mu(T^n)$ should be replaced by Ξ^s and Ξ^{s-1}, respectively.

Theorem 9. (Regularity theorem). Let $u_0, F \in \Xi^s_e$ such

that the external force $F_{|\textcircled{H}}$ is H^{s+k} vector field and the initial velocity $u_{o|\textcircled{H}}$ is H^{s+q} vector field on $\textcircled{H}$, $1 \leqslant q \leqslant k$. Then $u(t)_{|\textcircled{H}}$ belongs to the class H^{s+q} on the entire $\textcircled{H}$ including the boundary $\partial\textcircled{H}$ for all t for which it exists in H^s.

Proof. By theorem 1(i) F and u_o on the whole T^n belong to the same Sobolev classes as $F_{|\textcircled{H}}$ and $u_{o|\textcircled{H}}$ on $\textcircled{H}$. So we can use the arguments as in theorem 6 replacing the bundle $TD^s_\mu(T^n)$ by Ξ^s. In particular, $\bar{u}_o$ is C^q-smooth $D^s_\mu(T^n)$-right-invariant section of Ξ^s, hence $\gamma_t \bar{u}_o$ is s.m. C^q-smooth and $D^s_\mu(T^n)$-right-invariant random section of Ξ^s (the notations are similar to the proof of theorem 6). Since Ξ^s is a subbundle in $TD^s_\mu(T^n)$, sections of Ξ^s are vector fields on $D^s_\mu(T^n)$. The rest of the proof is the same as in theorem 6. Q.E.D.

Corollary 1. Under the conditions of theorem 9 the curve $g(t)_{|\textcircled{H}}$ lies in $D^{s+q}_\mu(\textcircled{H})$ for all t for which it exists in $D^s_\mu(\textcircled{H})$.

Corollary 2. If $F = 0$, then theorem 9 and corollary 1 are valid for all $q > 0$, in particular if $u_o \in C^\infty$ then $u(t) \subset C^\infty$ and $g(t)_{|\textcircled{H}}$ lies in $D^\infty_\mu(\textcircled{H})$ for all t for which they exist in H^s.

Note that the right-invariance with respect to the whole group $D^s_\mu(T^n)$ is necessary in the proof of theorem 9 when we are interested in obtaining the regularity on the entire $\textcircled{H}$ including $\partial\textcircled{H}$.

To consider the viscous fluid adhering to the boundary one can take into account the generalized friction force with the friction coefficient $-\delta_{\partial\textcircled{H}}$, minus surface delta function on $\partial\textcircled{H}$ (roughly speaking, the friction coefficient must be equal to zero in $T^n \setminus \partial\textcircled{H}$ and to minus infinity at the points of $\partial\textcircled{H}$). The corresponding right-invariant generalized force field on $D^s_\mu(T^n)$ should be constructed, lifted onto Ξ^s and added to $TR\,\bar{F}^\ell$ in (38).

We shall describe this idea in details elsewhere.

REFERENCES

1. Arnold V. "Sur la géometrie différentielle des groupes de Lie de dimension infinie et ses applications a l'hydrodynamique des fluides parfait", Ann. Inst. Fourier 16 (1966), 319-361.
2. Baranov Yu.S. and Gliklikh Yu.E. "About a certain non-integrable distribution on infinite-dimensional manifold of diffeomorphisms", Differentsial'naya geometriya mnogoobraziĭ figur, 17 (1986), 11-17 (Russian).
3. Baranov Yu.S. and Gliklikh Yu.E. "A certain mechanical constraint on the group of diffeomorphisms preserving the volume", Funct. anal. appl. 22, N 2,(1988), 61-62.
4. Belopol'skaja Ja.I. and Daleckiĭ Ju.L. "Diffusion processes in smooth Banach spaces and manifolds", Trans. Moscow Math. Soc. (Translation by AMS), Issue 1, (1980), 113-150.
5. Belopol'skaya Ya.I. and Daletskiĭ Yu.L. "Ito equations and differential geometry", Uspekhi mat. nauk, 37, N 3 (1982), 95-142 (Russian; translated in Russian Math. Surveys).
6. Belopolskaya Ya.I. and Dalecky Yu.L. "Stochastic equations and differential geometry", Kluwer, 1989.
7. Daletskiĭ Yu.L. "Stochastic differential geometry", Uspekhi mat. nauk, 38, N 3 (1983), 87-111 (Russian; translated in Russian Math. Surveys).
8. Dombrowski P. "On the geometry of the tangent bundle", Journal für die reine und angewandte Mathematik, 210 (1962), 73-88.
9. Ebin D.G. and Marsden J. "Groups of diffeomorphisms and the motion of an incompressible fluid", Annals of Math., 92, N 1 (1970), 102-163.

10. Elworthy K.D. "Stochastic differential equations on manifolds", Cambridge Univ. Press (1982).
11. Faddeev L.D. and Vershik A.M. "Differential geometry and Lagrangian mechanics with constraints", DAN SSSR, 202, N 3 (1972), 555-557 (Russian; translated in Soviet Math. Doklady).
12. Gihman I.I. and Skorohod A.V. "Stochastic differential equations", Springer (1973).
13. Gliklikh Yu.E. "Analysis on Riemannian manifolds and the problems of mathematical physics", Voronezh Univ. Press (1989) (Russian)
14. Gliklikh Yu.E. "Stochastic differential geometry on groups of diffeomorphisms and the hydrodynamics of viscous incompressible fluid", Uspekhi Mat. Nauk, 44, N 4 (1989), 224 (Russian; translated in Russian Math. Surveys).
15. Gliklikh Yu.E. "Stochastic differential geometry of the groups of diffeomorphisms and the motion of viscous incompressible fluid", Fifth International Vilnius conference on probability theory and mathematical statistics, Abstracts of communications, 1,(1989), 173-174.
16. Ikeda N. and Watanabe Sh. "Stochastic differential equations and diffusion processes", North-Holland (1981).
17. Lang S. "Differential manifolds", Springer (1985).
18. Nelson E. "Quantum fluctuations", Princeton Univ. Press (1985).
19. Rassias Th.M. "Foundations of global nonlinear analysis", Teubner (1986).

Yu.E.Gliklikh
Department of Mathematics
Voronezh State University
394693, Voronezh, USSR

Constantin Carathéodory: An International Tribute (pp. 374–380)
edited by Th. M. Rassias

APPLICATION OF C. CARATHÉODORY'S THEOREM TO A PROBLEM OF THE THEORY OF ENTIRE FUNCTIONS

A. A. Gol'dberg

A theorem of C. Carathéodory is used to construct the example which demonstrate that the distribution of a-values of nonvanishing entire functions may have pathologic character.

A particular case of C. Carathéodory's general theorem [1] which is considered can be formulated as follows.

Theorem C. Let $F_1 \supset F_2 \supset \ldots \supset F_\nu \supset \ldots$ be a sequence of simply connected Riemann surfaces, and the projection of F_1 on the finite complex plane C is a bounded domain. Let $F_\infty = \operatorname{int} \cap_{\nu=1}^{\infty} F_\nu \neq \emptyset, F$ is a connected component of F_∞. Let a be some point of F but not an algebraic branch point. Let the function f_ν be analytic in $D_R = \{z : |z| < R\}$ maps conformally D_R onto $F_\nu, f_\nu(z_0) = a, |z_0| < R, a \in F_\nu, \arg f'_\nu(z_0) = \theta$. Then $f_\nu \to f$ uniformly on compact subsets of D_R where f is an analytic function in D_R which maps conformally D_R onto the simply connected Riemann surface F so that $f(z_0) = a, \arg f'(z_0) = \theta$.

In order to formulate the problem of the theory of entire functions which we are going to solve let us give some notations. Let $\mathbb{C}^* = \mathbb{C}\backslash\{0\}$, and E^* be the class of nonvanishing entire functions with the function identically equal to zero added. By $n(r, a, f)$ we shall denote the counting function of a-values of the entire function f, by $A(r, f)$-the mean sheet number of the Riemann surface

$f(D_r)$ that is, [2]

$$A(r,f)=\frac{1}{\pi}\int_{D_r}\int\frac{|f'(z)|^2dxdy}{(1+|f(z)|^2)^2}\,,\ z=x+iy\,.$$

The main result of the paper is

Theorem 1. There exists a function $f\in E^*$ with the following properties:

$$\begin{aligned}
&1)\ \text{for all } a,b\in\mathbb{C}^*,\ a\neq b,\\
&\overline{\lim_{r\to\infty}}\ n(r,a,f)/n(r,b,f)=\infty\,,\\
&\underline{\lim_{r\to\infty}}\ n(r,a,f)/n(r,b,f)=0\,,\qquad (1)\\
&2)\ \text{for all } a\in\mathbb{C}^*\\
&\overline{\lim_{r\to\infty}}\ n(r,a,f)/A(r,f)=\infty\,,\\
&3)\ \text{for all } a\in\mathbb{C}\\
&\underline{\lim_{r\to\infty}}\ n(r,a,f)/A(r,f)=0\,.
\end{aligned}$$

The existence of an entire function (without demand of the lack of zeros) with properties 1)–3) where $\mathbb{C}^*$ is substituted by $\mathbb{C}$ was for the first time proved by the author [3], and only with the property 1) by S. Toppila [4] and independently by the author [3]. In [3] the Carathéodory's theorem was used. It was also used in [5] and [6] in this reach of questions.

In fact we shall prove some stronger result than Theorem 1. Let us define for the set E of entire functions the topology of uniform convergence on compact subsets of $\mathbb{C}$. It is well-known that one can metrize E so, that E become a complete metric space. The set E^* as closed subset of E is also a complete metric space.

The set is called a residual set [7] if its complement has first category in the sense of Baire. It follows from Baire's theorem ([7], theorem 9.1), that the residual set in E^* is non-empty. That is why it is obvious that Theorem 1 is contained in the following theorem.

Theorem 2. The set of functions from E^* satisfying the properties

1)–3) is residual in E^*.

Proof. Denote by E_n^* the set of functions from E^* satisfying the property $n), n = 1, 2, 3$. We shall denote by Ω the set $\{f \in E^* : f \equiv \text{const}\}$. For the determinacy let us assume that $f \in \Omega$ has none of the properties 1)–3). Because the intersection of a finite number of the residual sets is a residual set, to prove Theorem 2 it is sufficient to show that each of the sets $E_n^*, n = 1, 2, 3$, is a residual one.

Let us designate by $\pi A(F)$ the area of the Riemann surface F in the spherical metric. We shall denote by the symbol $\Rightarrow$ the uniform convergence on compact subsets. Let (K_n) be such a sequence of disks, $\overline{K_n} \subset \mathbb{C}^*$, that their radii tend monotonically to zero as $n \to \infty$, and every value of $\mathbb{C}^*$ is covered by infinite number of disks K_n', that are concentric to K_n but have twice smaller radii.

For each $m \in \mathbb{N}$ we shall designate by $n_0(m)$ such a natural number, that when $n \geq n_0(m)$

$$A(K_n) < (2m)^{-1} \tag{2}$$

holds.

Let us show at first that the set E_2^* is residual. Let us introduce $E_{mn}^2 = \{f \in E^* \backslash \Omega : (\exists a \in \overline{K_n'})(\forall r, m \leq r < \infty)[n(r, a, f) \leq mA(r, f)]\}$ when $m \in \mathbb{N}, n \in \mathbb{N}, n \geq n_0(m)$. It is not difficult to verify that

$$E_2^* = E^* \backslash \left(\cup_{m=1}^{\infty} \cup_{n=n_0(m)}^{\infty} (E_{mn}^2 \cup \Omega) \right) .$$

Let us show that the sets $E_{mn}^2 \cup \Omega$ are closed. Indeed, let $f_j \in E_{mn}^2 \cup \Omega$, and $f_j \Rightarrow f$. The case $f \in \Omega$ is trivial. Let $f \notin \Omega$. Then such $a_j \in \overline{K_n'}$ exist, that $n(r, a_j, f_j) \leq mA(r, f_j)$ when $r \geq m$. Without loss of generality we can assume that $a_j \to a \in K_n'$. Then $f_j - a_j \Rightarrow f - a$. If $f(z) \neq a$ when $|z| = r$, then

$$n(r, a_j, f_j) = \frac{1}{2\pi i} \int_{\partial D_r} \frac{f_j'(z)}{f_j(z) - a_j} dz$$
$$\to \frac{1}{2\pi i} \int_{\partial D_r} \frac{f'(z)}{f(z) - a} dz = n(r, a, f), j \to \infty ,$$

and when $j \geq j_0(r)$ then $n(r, a_j, f_j) = n(r, a, f)$. As $A(r, f_j) \to A(r, f)$ when $j \to \infty$, then $n(r, a, f) \leq mA(r, f)$ when $r \geq m$. But if the function

f has a-values on ∂D_r, then we obtain the same inequality by tending r' to r from the right-hand side in the inequality $n(r', a, f) \leq mA(r', f)$. Thus $f \in E^2_{mn}$ and the set $E^2_{mn} \cup \Omega$ is closed.

And now we are going to show that the set $E^2_{mn} \cup \Omega$ is nowhere dense. In other words for any function $f \in E^2_{mn} \cup \Omega$, for any $r, m < r < \infty$, and for any $\varepsilon > 0$ there exists a function $h \in E^* \backslash \{E^2_{mn} \cup \Omega\}$ such that $|f(z) - h(z)| < \varepsilon$ when $z \in D_r$. Let us regard at first that $f \in E^2_{mn}$. Let F_0 be a Riemann surface onto which f maps D_{r+1}. Let $z_0 \in D_1$ be such a point, that $w_0 = f(z_0) \in F_0$ is not an algebraic branch point of F_0. As $f \in E^*$ then $0 < q = \inf\{|w| : w \in f(D_{r+1}) \cup K_n\} < \sup\{|w| : w \in f(D_{r+1}) \cup K_n\} = Q < \infty$ (in this place we mean $f(D_{r+1})$ is an image of D_{r+1} in $\mathbb{C}^*$ but not the Riemann surface F_0). Let us designate by K^s_n a s-sheeted disk covering K_n with an algebraic branch point of the order $s-1$ over the centre of K_n. Let (S_ν) be a sequence of one-sheeted Jordan quadrilaterals $A_\nu B_\nu C_\nu D_\nu$ (the vertices are listed in the order of positive orientation of ∂S_ν) with the following properties:

a) $\{w : q/2 < |w| < 2Q\} \supset S_1 \supset S_2 \supset \ldots$,

b) arc $A_1B_1 \supset$ arc $A_2B_2 \supset \ldots$, arc $C_1D_1 \supset$ arc $C_2D_2 \supset \ldots$,

c) int $\cap_{\nu=1}^{\infty} S_\nu = \emptyset$,

d) on the boundary of the surface F_0 there is a sequence (arc $B'_\nu A'_\nu$), $\partial F_0 \supset$ arc $B'_1A'_1 \supset$ arc $B'_2A'_2 \supset \ldots$

the projection of arc $B'_\nu A'_\nu$ coincides with arc $A_\nu B_\nu$ but has opposite orientation. By sewing arc $A_\nu B_\nu \subset \partial S_\nu$ with arc $B'_\nu A'_\nu \subset \partial F_0$ and arc $C_\nu D_\nu \subset \partial S_\nu$ with arc $D'_\nu C'_\nu \subset \partial K^s_n$ we will obtain a simply connected Riemann surface $F(\nu, s)$. We will take s so large, that

$$s/(sA(K_n) + A(S_1) + A(F_0)) > (2A(K_n))^{-1} > m \tag{3}$$

(the second inequality follows from (2)). After having fixed s in such a way, set $F_\nu = F(\nu, s)$. It is easy to verify that the sequence (F_ν) satisfies all the assumptions of Theorem C with $F_\infty = F_0 \cup K^s_n, F = F_0, a = w_0$. Now let us define the sequence (f_ν) as in Theorem C, $\theta = \arg f'(z_0)$. Then, according to Theorem C, $f_\nu \Rightarrow f$ in D_{r+1}. Now let us fix such a value of ν, that $|f_\nu(z) - f(z)| < \varepsilon/2$ when $z \in D_r$. The pre-image under the mapping f_ν of the part of $K^s_n \subset F_\nu$ covering $\overline{K'_n}$ is contained in $D_{r_1}, r < r_1 < r+1$. Therefore $n(r_1, a, f_\nu) \geq s$ for all $a \in \overline{K'_n}$. On the other hand $A(r_1, f_\nu) \leq A(r+1, f_\nu) \leq A(F_0) + A(S_1) + sA(K_n)$. By (3)

$$n(r_1, a, f_\nu)/A(r_1, f_\nu) > m. \tag{4}$$

The function f_ν does not vanish in D_{r+1}. Therefore we can choose a branch of $\log f_\nu$ in D_{r+1}. One can find a sequence of polynomials (P_j) tending uniformly in $D_{r_2}, r_1 < r_2 < r+1$, to $\log f_\nu$ as $\nu \to \infty$. Then taking into account the boundedness of f_ν we obtain the sequence $(h_j), h_j = \exp P_j$, tending uniformly to f_ν in D_{r_2}. If j is large enough we will have $|h_j(z) - f_\nu(z)| < \varepsilon/2$ in D_r and on account of (4) for all $a \in \overline{K}_n'$

$$n(r_1, a, h_j)/A(r_1, h_j) > m \tag{5}$$

is valid. So let us fix this number j. The function $h = h_j$ is the sought for. Indeed $h \in E^*, |h(z) - f(z)| \le |h(z) - f_\nu(z)| + |f_\nu(z) - f(z)| < \varepsilon$ for $z \in D_r$ according to (5) $h \notin E_{mn}^2$ and obviously $h \notin \Omega$.

If $f \in \Omega$, i.e., $f \equiv c \in \mathbb{C}$, then instead of F_0 we take any one-sheeted disk $K, \overline{K} \subset \mathbb{C}^*$, lying in $(\varepsilon/2)$-neighbourhood of point c. When constructing (f_ν) we take (instead of f) any linear fractional function mapping D_{r+1} onto K. No other changes in the previous reasoning are needed.

So far as it has been shown that $E_{mn}^2 \cup \Omega$ is nowhere dense in E^*, then E_2^* is a residual set.

Let us show now, that the set E_3^* is a residual one. So far as the reasoning are similar to the proof above of the residuality of the set E_2^* we shall mark only the differences. Instead of E_{mn}^2 let us take

$$E_{mn}^3 = \{f \in E^* \setminus \Omega : (\exists a \in \overline{K}_n')(\forall r, m \le r < \infty) \\ [n(r, a, f) \ge (1/m) A(r, f)]\}, m, n \in \mathbb{N}$$

(here we shall not use the inequality (2), and we may take $n_0(m) = 1$). The property of being closed of $E_{mn}^3 \cup \Omega$ can be proved exactly in the same way as the property of being closed of $E_{mn}^2 \cup \Omega$. But the surfaces F_ν are constructed differently. Let $f \in E_{mn}^3$ maps $D_{r+1}, r > m$, onto the Riemann surface F_0. We take s-sheeted disk Q_n^s, that covers a disk $Q_n, \overline{Q}_n \subset \mathbb{C}^*, Q_n \cap K = \emptyset$, and sew it to F_0 via Jordan quadrilateral in the same way as we previously sewed a s-sheeted disk K_n^s to F_0. We obtain the Riemann surface F_ν. Let $n_0 = \max\{n(r+1, a, f) : a \in \overline{K}_n'\}$. Then for all $a \in \overline{K}_n', n(r+1-0, a, f_\nu) \le n_0 + 1$ is satisfied (summand 1 may appear if S_ν covers the point a). We choose number s so large that $(n_0 + 1)/(sA(Q_n')) < 1/m$, where Q_n' is the disk concentric with Q_n but with a twice smaller radius. It is clear that $A(r_1, f_\nu) \ge sA(Q_n')$ holds for some $r_1, r < r_1 < r+1$. Then

$$n(r_1, a, f_\nu)/A(r_1, f_\nu) \le (n_0 + 1)/(sA(Q_n')) < 1/m$$

for all $a \in \overline{K}_n'$. This inequality plays the same part as inequality (4) in the previous reasonings. All the rest is almost a literally repetition of the previous proof.

Finally we are going to show the residuality of the set E_1^*. First of all we will note that from the first inequality in (1) the second one follows, and on the contrary, because they are satisfied for all a and b, $a \neq b$. Therefore E_1^* can be defined as the set of functions from E^*, for which the first inequality from (1) holds. Similarly to the beginning of the proof of Theorem 2 let us define the sequences (K_n) and (K_n'), but we do not need (2) to be satisfied, and assume $n_0(m) = 1$. Let $(n,l) \in \mathbb{N}^2$. Let us say that $(n,l) \in P$, if $K_n \cap \overline{K}_l' = \emptyset$. Let us define

$$E_{mnl}^1 = \{f \in E^* \setminus \Omega : (\exists a \in \overline{K}_n')(\exists b \in \overline{K}_l')(\forall r, m \leq r < \infty) \\ [n(r,a,f) \leq mn(r,b,f)]\}, m \in \mathbb{N}, (n,l) \in P\,.$$

Then

$$E_1^* = E^* \setminus \cup_{(n,l)\in P} \cup_{m=1}^{\infty} \left(E_{mnl}^1 \cup \Omega\right)\,.$$

The proof of the residuality of E_1^* proceeds exactly in the same way as the proof of the residuality of the set E_2^*, but when constructing the sequence of the Riemann surfaces F_ν the number s is chosen differently. Let $n_0 = \max\{n(r+1,b,f) : b \in \overline{K}_l'\}$. Then $n(r+1-0,b,f_\nu) \leq n_0+1$ for all $b \in \overline{K}_l'$. We choose the number s so large that $s > m(n_0+1)$. We choose the number $r_1, r < r_1 < r+1$, in the same way as when proving that the set E_2^* is a residual one. Then $n(r_1,a,f_\nu) \geq s$ for all $a \in \overline{K}_n'$, and $n(r_1,b,f_\nu) \leq n_0+1$ for all $b \in \overline{K}_l'$. Then $n(r_1,a,f_\nu)/n(r_1,b,f_\nu) > m$. This inequality plays the same part as (4). In other aspects we repeat the proof of the residuality of E_2^*.

References

1. C. Carathéodory, *Untersuchungen über die konformen Abbildungen von festen und veränderlichen Gebieten*, Math. Ann., 72 (1912), 107–144.
2. R. Nevanlinna, *Analytic Functions*, Berlin – Heidelberg – N. Y., Springer, 1970.
3. A. A. Gol'dberg, *The counting functions for sequences of a-value of entire functions*, Sibirsk. Mat. Zh., 19 (1) (1978), 28–36 (Russian).
4. S. Toppila, *On the counting function for the a-values of a meromorphic function*, Ann. Acad. Sci. Fenn. Ser. AI, 2 (1976), 565–572.

5. A. E. Eremenko, *On the counting functions for sequences of a-values of functions holomorphic in the disk.* Teor. funkcii, funkc. analis i ih pril. **31** (1979), 59–62 (Russian).
6. A. A. Gol'dberg and N. V. Zabolockii, *On the a-values of functions meromorphic in the disk.* Sibirsk. Mat. Zh., **24** (3) (1983), 34–46 (Russian).
7. J. C. Oxtoby, *Measure and Category*, N. Y. – Heidelberg – Berlin, Springer, 1971.

A. A. Gol'dberg
University of Lvov
Faculty of Mathematics and Mechanics
290013, Lvov
Pushkin str., 38, app. 7
USSR

Constantin Carathéodory: An International Tribute (pp. 381–403)
edited by Th. M. Rassias

SIMPLY CONNECTED DOMAINS WITH FINITE LOGARITHMIC AREA AND RIEMANN MAPPING FUNCTIONS

A. Z. Grinshpan and I. M. Milin

Studying univalent functions in plane domains plays an important part in the geometric theory of functions of a complex variable (GTFCV). A function is called univalent in a domain B in the extended complex plane if it is meromorphic (in particular, regular) and one-to-one in B. The classic Riemann mapping theorem establishes a direct correspondence between univalent functions in the unit disk and simply connected domains. Some new properties of regular and univalent functions in the unit disk are obtained in the present paper.

Characteristic for GTFCV necessity of joint investigation of geometric and analytic properties of the functions under consideration is often connected with big difficulties. It was already revealed in the papers of the early investigators of our century L. Bieberbach, C. Caratheodory, P. Koebe, E. Lindelöf, C. Loewner and others. The Leningrad school of GTFCV — the authors are among its representatives — was founded by professor G. M. Goluzin. It was decisively influenced by his papers (the 30–40s.) as well as by his well-known book [1] first published in 1952. Both G. M. Goluzin's papers and his Leningrad followers' ones are exclusively aimed to overcome the abovementioned difficulties when analysing various univalent function properties and when solving the corresponding extremal problems.

1. Area and Logarithmic Area of a Domain

Let E denote the open unit disk in the complex z-plane. For any regular in E function

$$f(z) = \sum_{k=0}^{\infty} c_k z^k \quad \text{let} \quad \sigma(f) = \frac{1}{\pi} D(f),$$

where $D(f)$-Dirichlet's integral

$$\int_E \int |f'(z)|^2 d\sigma \quad (d\sigma \quad \text{area element}).$$

Hence

$$\sigma(f) = \sum_{k=1}^{\infty} k|c_k|^2.$$

If the function $f(z)$ is univalent in E, then the area of image $B = f(E)$ in the complex plane is equal to $\pi\sigma(f)$. By $\sigma(B)$ we denote also divided by π the area of any measurable plane set B. By polar coordinates (ρ, θ)

$$\sigma(B) = \frac{1}{\pi} \int_B \int \rho d\rho d\theta.$$

By the same coordinates the value

$$\int_B \int \rho^{-1} d\rho d\theta = \int_B \int |(\log W)'|^2 d\sigma$$

is the logarithmic area of a set B not containing 0 in the comlex W-plane. Let function $f(z), f(0) = 0$, is regular and univalent in E. For any $p > 0$ we introduce power coefficients $D_k(p) (k = 0, 1, \dots)$ of the function $f(z)$. Let

$$\left[\frac{f(z)}{z}\right]^p = \sum_{k=0}^{\infty} D_k(p) z^k, \tag{1}$$

fixing in case of ambiguity any branch of the function $[f(z)/z]^p$. Logarithmic coefficients β_k of the considered function $f(z)$ are defined by expansion

$$\log \frac{f(z)}{z} = \sum_{k=0}^{\infty} \beta_k z^k, \tag{2}$$

where

$$\beta_0 = \log f'(0)\,.$$

We have

$$\sigma\left(\log\frac{f(z)}{z}\right) = \sum_{k=1}^{\infty} k|\beta_k|^2\,.$$

This value independent of $f'(0)$ is usually called the logarithmic area of the funciton $f(z)$ and the set $B = f(E) \ni 0$. Note, though we do not need it here, that power coefficients (for $p > 0$) and logarithmic coefficients of a function $f(z)$ that does not vanish, are defined naturally as Taylor coefficients of functions $[f(z)]^p$ and $\log f(z)$ respectively.

When additionally normalized $c_1 = 1$, functions $f(z) = c_1 z + c_2 z^2 + \ldots$ regular and univalent in E form class S — the major object of univalent function theory. In this class the Koebe function $k_1(z) = z(1-z)^{-2}$ and its rotations $k_\chi(z) = \overline{\chi} k_1(\chi z), |\chi| = 1$, are of particular importance. It was the Koebe function that turned out to be extremal when traditional functionals were estimated on the class S. It was most brilliantly revealed in the famous Bieberbach conjecture of 1916: for $f(z) = z + c_2 z^2 + \ldots \in S$ and for each $n = 2, 3, \ldots$ the inequality $|c_n| \leq n$ holds with equality only for the functions $k_\chi(z), |\chi| = 1$, to which solution L. de Branges came in 1984 [2]. For $f(z) \in S$ logarithmic coefficients are generally represented as $\beta_k = 2\gamma_k$, that is,

$$\log\frac{f(z)}{z} = 2\sum_{k=1}^{\infty} \gamma_k z^k\,. \tag{3}$$

For $k_\chi(z)$ we have $\gamma_k = \chi^k/k$ and

$$D_k(p) = d_k(2p)\chi^k \quad (k = 1, 2, \ldots)\,,$$

where $d_k(\lambda)$ are binomial coefficients that are defined by the expansion

$$(1-z)^{-\lambda} = \sum_{k=0}^{\infty} d_k(\lambda) z^k\,,\; d_0(\lambda) = 1\,. \tag{4}$$

Various inequalities for coefficients of formal power series particular of exponential kind are of considerable importance in the theory of univalent functions [3; 4; 5, Ch. 2; 6; 7]. We need the following Theorem for formal series. Here and further for a regular function $g(z)$ in E and real p by

$$\int_{|z|=1} |g(z)|^p |dz|$$

we imply

$$\lim_{r\to 1-0}\int_{|z|=1}|g(rz)|^p|dz|\,.$$

Theorem 1. Let $\{A_k\}_1^\infty$ be an arbitrary consequence of complex numbers, $\sum\limits_{k=1}^{\infty} k|A_k|^2<\infty$, generating coefficients of formal series $D(z)$ by expansion

$$D(z)=\sum_{k=0}^{\infty}D_kz^k=\exp\left\{\sum_{k=1}^{\infty}A_kz^k\right\}.$$

Then for any $\lambda>0$ the inequality holds

$$\sum_{k=0}^{\infty}\frac{|D_k|^2}{d_k(\lambda)}\le\exp\left\{\frac{1}{\lambda}\sum_{k=1}^{\infty}k|A_k|^2\right\}, \tag{5}$$

where $d_k(\lambda)$ are defined by (4).

When $\lambda=1$ and 2 the stronger inequalities hold

$$\frac{1}{2\pi}\int_{|z|=1}(|D(z)|^2+|D(z)|^{-2})|dz|\le 1+|A_1|^2+\exp\{\sum_{k=1}^{\infty}k|A_k|^2\} \tag{6}$$

and

$$\begin{aligned}&\frac{1}{\pi}\int_E\int(|D(z)|^2+|D(z)|^{-2})d\sigma\\&\quad\le 1+\frac{|A_1|^2}{2}+\frac{|A_1^2-2A_2|^2}{12}+\exp\{\frac{1}{2}\sum_{k=1}^{\infty}k|A_k|^2\}\end{aligned} \tag{7}$$

respectively.

Equality in the inequalities (5), (6) (for $\lambda=1$) and (7) (for $\lambda=2$) occurs if and only if

$$A_k=\lambda\frac{\zeta^k}{k}\quad(k=1,2,\ldots),\zeta\in E\,.$$

First the inequality (5) in case $\lambda=1$ was obtained in [3]. For any $\lambda>0$ this inequality is proved in [5, Ch. 2]. It is obtained there as a special case of the more general inequality where an arbitrary entire function $\Omega(w)$ with nonnegative coefficients is taken as the generating function instead of

$\exp\{w\}$. Extension of these inequalities for norms of linear combinations of formal series in corresponding Hilbert spaces is given in [6]. The strengthened inequalities (6) and (7) are half-sums of inequalities for the generating functions

$$\Omega_1(w) = \exp(w) + \exp(-w)$$

and

$$\Omega_2(w) = \exp(w) - \exp(-w)\,.$$

We take into account that $D(z) \neq 0$ is regular in E. See more details of the proof of inequality (6) in [8]. The proof of inequality (7) is similar. The case of equality for any functions $\Omega(w)$ with strictly positive coefficients is given in [5, Ch. 2]. The case of equality for odd functions $\Omega(w)$ is proved in a similar way. Hence the conclusion on the sign of equality in inequalities (6) and (7) follows.

By Theorem 1 there proves the following effective Lemma for univalent functions practically earlier used in [9].

Lemma 1. Let $f(z) \in S$ and $r \in (0,1)$. Then the inequality

$$\log\left\{\frac{1}{2\pi r}\int_{|z|=1}|f(rz)||dz|\right\} \leq \sum_{k=1}^{\infty} k|\gamma_k|^2 r^{2k}$$

holds, where the coefficients γ_k are defined by (3).

The sign of equality in this inequality holds if and only if

$$f(z) = k_\chi(\rho z)/\rho, \rho \in (0,1], |\chi| = 1\,,$$

or

$$f(z) = z\,.$$

Proof. According to (I) we have

$$\begin{aligned}\frac{1}{2\pi}\int_{|z|=1}|f(rz)||dz| &= \frac{r}{2\pi}\int_0^{2\pi}\left|\left(\frac{f(re^{i\theta})}{re^{i\theta}}\right)^{1/2}\right|^2 d\theta \\ &= r\sum_{k=0}^{\infty}|D_k(1/2)|^2 r^{2k}\,, \quad D_0(1/2) = 1\,.\end{aligned} \tag{8}$$

Applying the inequality (5) for $\lambda = 1$ to the identity

$$\left[\frac{f(rz)}{rz}\right]^{1/2} = \exp\left\{\sum_{k=1}^{\infty} \gamma_k r^k z^k\right\}$$

by (8) it completes the proof.

Remark 1. The proof of the following inequalities for $f(z) = z + c_2 z^2 + \ldots \in S$

$$\frac{1}{2\pi}\int_{|z|=1}\left(\frac{|f(rz)|}{r} + \frac{r}{|f(rz)|}\right)|dz|$$
$$\leq 1 + |\frac{c_2 r}{2}|^2 + \exp\left\{\sum_{k=1}^{\infty} k|\gamma_k|^2 r^{2k}\right\}, r \in (0,1);$$
$$\frac{1}{2}\log\left\{\frac{1}{\pi}\int_E\int |\frac{f(z)}{z}|^2 d\sigma\right\} \leq \sum_{k=1}^{\infty} k|\gamma_k|^2;$$
$$\frac{1}{\pi}\int_E\int \left(|\frac{f(z)}{z}|^2 + |\frac{z}{f(z)}|^2\right) d\sigma$$
$$\leq 1 + \frac{|c_2|^2}{2} + \frac{|c_2^2 - c_3|^2}{3} + \exp\left\{2\sum_{k=1}^{\infty} k|\gamma_k|^2\right\}$$

via inequalities (6), (5) (for $\lambda = 2$) and (7) is similar to the proof of Lemma 1.

The assertion on the sign of equality in all of the three inequalities is the same as in Lemma 1. Except the case $\rho = 1$ in the last two inequalities, where we suppose the logarithmic area to the finite.

We note that the results of the investigation on estimating the mean modulus via some values generated by odd functions of the class S were published in [10; 11]. These results allowed their authors to predict the effective mean modulus estimate via the image area (see (25)).

Now we suppose that $w = f(z) \in S$ and $|f(z)| < M$ in E. By the logarithmic area definition we have for $\rho \in (0,1)$

$$\sigma\left(\log\frac{f(\rho z)}{z}\right) = \frac{1}{\pi}\int_{f(|z|=\rho)} \log\left|\frac{f(z)}{z}\right| d\arg f(z).$$

From here by the Green-Ostrogradski formula and evident nonnegativity of the logarithmic area of the set

$$\{w : |w| < M\}/f(E)$$

it follows the known inequality (see Sec. 3)

$$\sigma\left(\log\frac{f(z)}{z}\right) \le 2\log M\,.$$

Hence and from (3) turning to the function $f(rz)/r \in S, r \in (0,1)$, we obtain

$$\sum_{k=1}^{\infty} k|\gamma_k|^2 r^{2k} \le \frac{1}{2}\log\frac{M(r,f)}{r}\,, \tag{9}$$

where

$$M(r,f) = \max_{|z|=r}|f(z)|\,.$$

The inequality (9) is efficient in applications. In particular, by it and Lemma 1 via the Cauchy integral formula we have for $f(z) = z + c_2 z^2 + \ldots \in S$ and $r \in (0,1)$

$$|c_n| \le M^{1/2}(r,f)r^{1/2-n} \quad (n = 2,3,\ldots)\,. \tag{10}$$

From the estimate (10) which though is not the best result of that kind (see Remark 3) it follows immediately that

$$\lim_{n\to\infty}\frac{|c_n|}{n} = 0$$

if Hayman's index for the function $f(z)$ [12]

$$\alpha = \lim_{r\to 1-0} M(r,f)(1-r)^2 \tag{11}$$

is equal to 0. The known proofs of this fact are more bulky [12; 5, Ch. 3] though Hayman's asymptotics of coefficients c_n for $\alpha \neq 0$ is obtained relatively easy [5, Ch. 3]. On the other hand the inequality (9) does not allow us to obtain the following exact uniform estimate on the class S

$$\sum_{k=1}^{\infty} k|\gamma_k|^2 r^{2k} \le \log\frac{1}{1-r^2}\,,\ r \in (0,1)\,, \tag{12}$$

with the sign of equality only the functions $k_\chi(z), |\chi| = 1$. As it was first noted in [9] the proof of the inequality (12) known before as the Bazilevich conjecture is provided by the more deep properties of logarithmic coefficients, conjectured in [5, p. 72] and then proved by L. de Branges [2] for the Bieberbach conjecture solution. The properties are such that for $n = 1, 2, \ldots$ and $x_k = n + 1 - k$ $(k = 1, 2, \ldots, n)$ the functionals

$$\sum_{k=1}^{n} x_k k |\gamma_k|^2 \tag{13}$$

attain supremum on the class S only for the functions

$$k_\chi(z), |\chi| = 1 .$$

As this fact is of great stimulating importance, in [7] there considers the problem of defining all the admissible vectors, that is, such vectors $(x_1, \ldots, x_n)(n = 1, 2, \ldots)$ for which the Koebe function realizes supremum on S of functionals of the form (13). By the variation of Schaeffer-Spenser type there establishes [13; 9] that the admissible vectors necessarily satisfy the condition

$$\min_{\theta \in [0,\pi]} \sum_{k=1}^{n} x_k \sin(k\theta) = 0 , \tag{14}$$

which, however, in general case is not sufficient (A. Z. Grinshpan pointed out two-parameter family of disproving examples) [7]. Recently in [14] the condition (14) is proved again by the boundary Schiffer variation. The papers by P. Duren and Y. Leung and T. Koornwinder [15; 16] adjoin the problem on admissible vectors. The following conjecture is formulated in [17]: for functions $f(z) \in S$

$$\sum_{k=1}^{\infty} k |\gamma_k|^2 r^{2k} \leq \frac{1}{2} \log \frac{M(r^2, f)}{r^2} , r \in (0, 1) , \tag{15}$$

and in [17; 18] it is proved for some special cases. It is clear that the inequality (15) (if it holds) implies (12) as for each r $\sup M(r, f)$ on the class S is realized by the Koebe function and its rotations. In connection with the inequality (9) and the conjecture (15) we note that by the Cauchy inequality and Lemma 1 the inequality

$$\log \left[M(r^2, f) \frac{1 - r^2}{r^2} \right] \leq \sum_{k=1}^{\infty} k |\gamma_k|^2 r^{2k}$$

follows.

Research in the logarithmic areas of nonnormalized univalent functions and the area formulas in polar coordinates led in [19] to the introduction of A-measure concept for simply connected domains containing 0 and for Riemann functions mapping onto them. For A-measure a number of nontrivial properties were established there. In [19] it is actually proved that for simply connected domain A-measure is equivalent to the Teichmüller reduced logarithmic area introduced by him in 1938 [20]. Inequalities in terms of A-measure are more subtle than similar ones for the reduced moduli much applied (see also there and in [21]). In Sec. 2 we remind the definition and some properties of A-measure and consider the corresponding class A_S of Riemann mapping functions.

2. Simply Connected Domains with Nonpositive A-measure

Let function $f(z) = \sum_{k=1}^{\infty} c_k z^k$ be regular and univalent in E and $B = f(E)$. A-measure of the domain B or the function f is defined by the equality [19]:

$$A(B) = 2\log R + \sigma\left(\log \frac{f(z)}{z}\right),$$

where $R = |c_1|$ is the conformal radius of the domain B with respect to 0. Taking for any complex $t \neq 0$

$$B_t = \{zt : z \in B\},$$

we obtain $A(B_t) = 2\log|t| + A(B)$.

Therefore for any original domain B for which $A(B) < \infty$ at the expense of selecting multiplier t, value $A(B_t)$ can be made nonpositive. Such normalization is convenient in applications. The following properties of A-measure hold [19]:

1) $A(\{w : |w| < R\}) = 2\log R$;
2) A-measure is monotonic, that is, if D is a simply connected subdomain of B and $D \ni 0$ then

$$A(B) = A(D) + \frac{1}{\pi}\int\int_{B/D} \rho^{-1} d\rho d\theta,$$

in particular, for $D = \{w : |w| < R\}$ we get a formula for the Teichmüller reduced logarithmic area [20]

$$A(B) = 2\log R + \frac{1}{\pi}\int\int_{B/D} \rho^{-1} d\rho d\theta\,;$$

3) A-measure satisfies the inequality

$$A(B) \leq \log \sigma(B)$$

with the sign of equality if and only if B is a disk with the centre at the origin of coordinates, where a set of zero area is removed;

4) If for any points $v \in B$ and $w \in D$ the product $v \cdot w \neq 1$ then

$$A(B) + A(D) \leq 0$$

with the sign of equality if and only if

$$\sigma(C/B/\{w : w^{-1} \in D\}) = 0\,;$$

here $B \ni 0$ and $D \ni 0$ are simply connected domains in the complex plane.

Note that in 1960 E. Reich and S. E. Warschawski (see, for instance, [I, Addition]) proved the inequality for regular univalent and bounded functions in the disk with concentric circular slits which corresponds for these functions to the property (3) of A-measure.

Denote by A_S the class of regular and univalent in E functions $f(z) = c_1 z + \ldots$ with nonpositive A-measure. This class contains

- bounded functions: $|f| < 1$,
- functions with bounded image area: $\sigma(f) \leq 1$ (see the property (3) of A-measure),
- Bieberbach-Eilenberg functions (see Sec. 3) and others.

According to the definition of the class A_S, for any function $f(z) \in A_S$ we have

$$\sigma\left(\log \frac{f(z)}{z}\right) \leq 2\log \frac{1}{|f'(0)|}\,. \tag{16}$$

Hence we obtain immediately: $\sup\limits_{A_S} |f'(0)| = 1$ and all the extremal functions are of the form

$$f(z) = \chi z\,;\ |\chi| = 1\,.$$

The exact estimate of modulus of a function $f(z) \in A_S$ follows also directly from the inequality (16) [19] and extends Jenkin's inequality [22] for Bieberbach-Eilenberg functions (see Sec. 3) onto the class A_S.

Theorem 2. Let $f(z) \in A_S$ and $\zeta \in E$. Then the inequality

$$|f(\zeta)| \leq \frac{|\zeta|}{(1-|\zeta|^2)^{1/2}}$$

holds.
The sign of equality in this inequality holds if and only if

$$f(z) = c_1 z/(1-\bar{\zeta}z), |c_1| = (1-|\zeta|^2)^{1/2}.$$

Proof. From (2) and (16) it follows that

$$|f(\zeta)| = |\zeta f'(0)| \exp\left\{\mathrm{Re}\sum_{k=1}^{\infty}\beta_k\zeta^k\right\}$$
$$\leq \frac{|\zeta|}{(1-|\zeta|^2)^{1/2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{\infty}k|\beta_k - \frac{\bar{\zeta}^k}{k}|^2\right\}.$$

From this and the condition of equality in (16) we obtain Theorem's assertion.

To study further properties of the class A_S the following Lemma is useful.

Lemma 2. Let function $f(z) = c_1 z + \ldots \in A_S$. Then for any $p, \varepsilon > 0$ the inequality holds

$$\sum_{k=0}^{\infty}\frac{|D_k(p)|^2}{d_k(\varepsilon p)} \leq |c_1|^{2p(1-\varepsilon^{-1})}, \tag{17}$$

where D_k are defined by (1) for the function $f(z)$ and d_k are the binomial coefficients from (4).

The sign of equality in the inequality (17) holds if and only if

$$f(z) = c_1 z/(1-\zeta z)^{\varepsilon}, |c_1| = (1-|\zeta|^2)^{\varepsilon^2/2}, \tag{18}$$

where for $\varepsilon \leq 2$, ζ is any number in E and for $\varepsilon > 2$

$$|\zeta| \leq \frac{1}{\varepsilon - 1}.$$

Proof. From (1) and (2) for $f(z)$ we have the identity

$$\sum_{k=0}^{\infty} D_k(p) z^k = \exp\left\{\sum_{k=0}^{\infty} p\beta_k z^k\right\} = c_1^p \exp\left\{\sum_{k=1}^{\infty} p\beta_k z^k\right\}.$$

Assuming in the inequality (5) of Theorem 1, $A_k = p\beta_k$ $(k = 1, 2, \ldots)$ and $\lambda = \varepsilon p$ we obtain the inequality

$$\sum_{k=0}^{\infty} \frac{|D_k(p)|^2}{d_k(\varepsilon p)} \leq |c_1|^{2p} \exp\left\{p\varepsilon^{-1} \sum_{k=1}^{\infty} k|\beta_k|^2\right\}.$$

From this and (16) in view of (2), (17) follows.

For equality in (17) according to Theorem 1 it is necessary that

$$f(z) = c_1 z/(1 - \zeta z)^{\varepsilon}, \ \zeta \in E, \tag{19}$$

and besides

$$\sum_{k=1}^{\infty} k|\beta_k|^2 = 2 \log \frac{1}{|c_1|}.$$

It gives the function (18). Any function $f(z)$ of the form (19) for $\varepsilon \leq 1 + |\zeta|^{-1}$ is univalent and regular in E and maps E onto a starlike domain with respect to 0. It follows from the fact that its normalized logarithmic derivative $zf'(z)/f(z)$ belongs to the Caratheodory class of functions with positive real part in E [23] (see, for instance, [24, p. 41]). For $\varepsilon > 2$ and $|\zeta| > (\varepsilon - 1)^{-1}$ each function of the form (19) is obviously nonunivalent and therefore does not belong to the class A_S. It completes the proof of the lemma.

Theorem 3. For each function $f(z) = c_1 z + \ldots \in A_S$ and for any $p > 0$ the inequalities hold

$$\frac{1}{2\pi} \int_{|z|=1} |f(z)|^p |dz| \leq |c_1|^{p(1-p/2)}, \tag{20}$$

$$\frac{1}{\pi} \int_E \int \left|\frac{f(z)}{z}\right|^p d\sigma \leq |c_1|^{p(1-p/4)}. \tag{21}$$

The sign of equality in these inequalities occurs if and only if $f(z)$ is defined by (18) when $\varepsilon = 2/p$ in the inequality (20) and when $\varepsilon = 4/p$ in the inequality (21).

Proof. Like (8) from (1) we have

$$\frac{1}{2\pi}\int_{|z|=1}|f(z)|^p|dz| = \sum_{k=0}^{\infty}|D_k(p/2)|^2\,.$$

We obtain the inequality (20) by Lemma 2 when substituting in (17) p by $p/2$ and ε by $2/p$. From (1) and by integration we obtain the identity

$$\frac{1}{\pi}\int_E\int\left|\frac{f(z)}{z}\right|^p d\sigma = \sum_{k=0}^{\infty}\frac{|D_k(p/2)|^2}{k+1}\,. \tag{22}$$

From this and (17) when p is substituted by $p/2$ and ε by $4/p$ we obtain the inequality (21). The signs of equality in (20) and (21) follow from the proof.

It follows from Theorem 3 that the class A_S is subclass of the class Hp of regular functions in E, $f(z) = c_0 + c_1 z + \ldots$ which satisfy the condition

$$\frac{1}{2\pi}\int_{|z|=1}|f(rz)|^p|dz| \le 1\,,\ r\in(0,1)\,,$$

for $p = 2$.

Remark 2. Let functions $f(z) = c_1 z + \ldots$ and $\tilde{f}(z) = \tilde{c}_1 z + \ldots$ be regular and univalent in E, $B = f(E), \tilde{B} = \tilde{f}(E)$ and $A(B) + A(\tilde{B}) \le 0$. Denote for any $p, \varepsilon, \tilde{p}, \tilde{\varepsilon} > 0, p/\varepsilon = \tilde{p}/\tilde{\varepsilon}$, by I and $\tilde{I}$ the left side in (17) respectively for $f(z), p, \varepsilon$ and $\tilde{f}(z), \tilde{p}, \tilde{\varepsilon}$. Then like inequality (17) we prove the inequality

$$I\cdot\tilde{I} \le |c_1|^{2p(1-\varepsilon^{-1})}\cdot|\tilde{c}_1|^{2\tilde{p}(1-\tilde{\varepsilon}^{-1})} \tag{23}$$

with the sign of equality only for functions $f(z)$ and $\tilde{f}(z)$ of the form (19). The definition of A-measure and the class A_S might be extended to regular functions in $E, f(z) = c_1 z + \ldots$ satisfying the condition $f(z)/z \ne 0$ with the assertion of Theorems 2, 3 and that of Lemma 2 being unchanged, but the

exponent in formula (18) for the extremal functions can be any nonnegative number for any $\zeta \in E$.

Further in this paragraph it is convenient to formulate the results for the normalized regular and univalent functions in E, $f(z) = z + c_2 z^2 + \dots$ that is, for the class S. It is clear that for $f(z) \in S$, $\sigma(f) \geq 1$ with the sign of equality only for the function $f(z) = z$. The property (3) of A-measure gives much stronger inequality

$$\sigma(f) \geq \exp\left\{\sigma\left(\log\frac{f(z)}{z}\right)\right\}$$

with the sign of equality if and only if $f(E)$ is disk with the centre at the origin of coordinates and with slits of zero area. The latter inequality allows us of course to strengthen the inequality (9). Add to it that D. Aharonov and H. Shapiro [25] investigated $\inf \sigma(f)$ for the functions of the class S with $|c_2|$ fixed. In all the abovementioned cases and many others it is worth speaking only about the functions with finite image area. Using interconnection between such functions and the class A_S we prove the following Lemma.

Lemma 3. For power coefficients from (1) of each function $f(z) \in S$ ($\sigma(f) < \infty$), for binomial coefficients from (4) and for any $p > 0$ the inequality holds

$$\sum_{k=0}^{\infty} \frac{|D_k(p)|^2}{d_k(2p)} \leq \sigma^{p/2}(f) \tag{24}$$

with sign of equality only for the function $f(z) = z$. The exponent $p/2$ in (24) cannot be decreased simultaneously for all the functions of the class S.

Proof. As $\sigma(f) < \infty$ then the function

$$g(z) = \sigma^{-1/2}(f) f(z) \in A_S\,.$$

Indeed, via the property (3) of A-measure we obtain

$$A(g(E)) \leq \log \sigma(g) \leq 0\,.$$

Applying Lemma 2 for $\varepsilon = 2$ to the function $g(z) = \sigma^{-1/2}(f)z + \dots$, we find

$$\sum_{k=0}^{\infty} \frac{|D_k(p,g)|^2}{d_k(2p)} \le \sigma^{-p/2}(f),$$

where $D_k(p,g)(k = 0,1,\dots)$ are power coefficients from (1) of the function $g(z)$ unlike the values $D_k(p) = D_k(p,f)$ in (24). It follows from the equality (1) that

$$D_k(p,g) = \sigma^{-p/2}(f)D_k(p,f) \quad (k = 0,1,\dots).$$

In view of it, from the latter inequality we obtain (24). For equality in (24) it is necessary that the function $g(z)$ to be defined from (18), where $\varepsilon = 2$. Hence $f(z) = z/(1-\zeta z)^2$ and $\sigma(f) = (1-|\zeta|^2)^{-4}$ where $\zeta \in E$. It is possible only if $\zeta = 0$ and therefore $f(z) = z$. The assertion on the exponent in the inequality (24) is confirmed by the family of functions

$$\varphi_r(z) = k_1(r^{1/2}z)r^{-1/2} \in S, \quad r \in (0,1).$$

Indeed, for $\varphi_r(z)$ we have

$$\sigma(\varphi_r) = \frac{1+r+4r^2}{(1-r)^4},$$

$$I(r) = \sum_{k=0}^{\infty} \frac{|D_k(p,\varphi_r)|^2}{d_k(2p)} = (1-r)^{-2p}.$$

It follows from this that as $r \to 1$: $I(r)/\sigma^x(\varphi_r) \to \infty$ if $x < p/2$. This completes the proof of the Lemma.

Theorem 4. For each function $f(z) \in S$ with finite image area the inequalities hold

$$\frac{1}{2\pi}\int_{|z|=1} |f(z)||dz| \le \sigma^{1/4}(f), \tag{25}$$

$$\frac{1}{\pi}\int_E\int \left|\frac{f(z)}{z}\right|^2 d\sigma \le \sigma^{1/2}(f). \tag{26}$$

Equality both in (25) and (26) occurs if and only if $f(z) = z$. The exponent $1/4$ in (25) cannot be decreased simultaneously for all the functions of the class S. It is similar for (26).

Proof. In view of the formula (8) for the mean modulus as $r \to 1$ via the inequality (24) for $p = 1/2$ we obtain (25). By the identity (22)

for $p = 2$ and the inequality (24) for $p = 1$ we have (26). Assertions on equality and exponent both in (25) and (26) follow from Lemma 3.

Remark 3. The inequalities (25) and (26) can be strengthened by the inequalities (6) and (7) of Theorem 1 (see Remark 1). Namely for functions $f(z) = z + c_2 z^2 + \ldots \in S$ $(\sigma(f) < \infty)$ the inequalities hold

$$\frac{1}{2\pi}\int_{|z|=1}(|f(z)| + |f(z)|^{-1})|dz| \le 1 + \left|\frac{c_2}{2}\right|^2 + \sigma^{1/4}(f),$$

$$\frac{1}{\pi}\int_E\int\left(\left|\frac{f(z)}{z}\right|^2 + \left|\frac{z}{f(z)}\right|^2\right)d\sigma \le 1 + \frac{|c_2|^2}{2} + \frac{|c_2^2 - c_3|^2}{3} + \sigma^{1/2}(f).$$

Equality for these inequalities is realized only by the function $f(z) = z$.

The inequality (25) was formulated by N. A. Lebedev and I. M. Milin as a conjecture in the joint paper [11] in 1951. This conjecture remained so far unproved. It follows from (25) that

$$|c_n| < \sigma^{1/4}(f) \quad (n = 2, 3, \ldots). \tag{27}$$

The exponent $1/4$ in (27) as well as that in (25) cannot be decreased. Indeed for the functions applied in the proof of Lemma 3

$$\varphi_r(z), r = 1 - n^{-1} \quad (n = 2, 3, \ldots)$$

we have

$$\sigma(\varphi_r) \sim 6n^4 \quad \text{and} \quad |\{\varphi_r(z)\}_n| \sim e^{-1/2}n$$

as $n \to \infty$.

Note that the nonstrict inequality (27) is obtained by another way as well, that is, by the Fitzgerald inequality for the class S [26]:

$$|c_n|^4 \le \sum_{k=1}^{n} k|c_k|^2 + \sum_{k=n+1}^{2n-1}(2n-k)|c_k|^2 \quad (n = 2, 3, \ldots),$$

where the right side $\le \sigma(f)$.

Though the estimate (27) like (10) is nonsharp, it is applied [27]. In connection with the inequality (25), it is necessary to note that I. E. Bazilevich in 1959 using a number of his estimates (see, for instance, bibliography

in [I, Addition]) proved the following result. If $f(z) \in S$ and $|f(z)| < M$ in E then the mean modulus does not exceed the value

$$\frac{8}{3\pi} M^{1/2} + a,$$

where a is an absolute constant. The main term in this estimate as $M \to \infty$ is sharp as it is realized by the functions

$$M k_\chi^{-1}(k_\chi(z)/M), \ |\chi| = 1.$$

Assume now for $f(z) \in S$ and $r \in (0,1)$, $\sigma_r = \sigma(f(rz))$ and $\beta = \overline{\lim}_{r \to 1-0} \sigma_r^{1/2}(1-r)^2$. From this and (11) we have for the Hayman index of function $f(z), \alpha \geq \beta$. On the other hand it is known [12] that

$$\lim_{r \to 1-0} (1-r^2) \frac{1}{2\pi} \int_{|z|=1} |f(rz)||dz| = \alpha.$$

Therefore (25) gives $\alpha \leq 2\beta^{1/2}$. Thereby for the functions $f(z) \in S$ the inequalities hold

$$\left(\frac{\alpha}{2}\right)^2 \leq \beta \leq \alpha.$$

For the Koebe function we have $\alpha = 1$ and $\beta = (3/8)^{1/2}$. The inequality (26) is equivalent to the following inequality in terms of the Taylor coefficients of the function $f(z) = z + c_2 z^2 + \ldots \in S$

$$\left(\sum_{k=1}^{\infty} \frac{|c_k|^2}{k}\right)^2 \leq \sum_{k=1}^{\infty} k|c_k|^2.$$

By the identity

$$1 + \left(2 + \frac{|c_2|^2}{2}\right)\frac{|c_2|^2}{2} + \left(2 + 2\frac{|c_2|^2}{2} + \frac{|c_3|^2}{3}\right)\frac{|c_3|^2}{3} + \ldots = \left(\sum_{k=1}^{\infty} \frac{|c_k|^2}{k}\right)^2$$

we see that this inequality follows from the sharp coefficient estimates $|c_n| \leq n$ for all n.

It is interesting to compare the inequality (26) with (21) and with the Goluzin inequality [28]

$$\frac{1}{\pi} \int_E \int |f(z)|^2 d\sigma \leq 1$$

for the functions of the class H_1.

3. Bieberbach-Eilenberg Domains

Let a simply connected domain $B, 0 \in B$, be Bieberbach-Eilenberg domain, if for any points $v, w \in B$ the product $v \cdot w \neq 1$. Let R be the class of functions $f(z) = c_1 z + \dots$ regular in E and such that $f(E)$ belongs to some Bieberbach-Eilenberg domain. For functions of the class R which are called Bieberbach-Eilenberg functions, W. Rogozinsky in 1939 [29] introduced the conjecture: $|c_n| \leq 1$ $(n = 1, 2, \dots)$ with equality only for the functions

$$f(z) = \chi z^n\,, \ |\chi| = 1\,.$$

W. Rogozinsky himself proved his conjecture for $n = 1$ and 2. This conjecture was completely proved by N. A. Lebedev and I. M. Milin in 1948. Its short solution was published in 1949 [10] and the more detailed one appeared in 1951 [11]. In fact in [10; 11] it was proved that the class R is a subclass of H_1, that is, for $f(z) \in R$ the inequality holds

$$\frac{1}{2\pi} \int_{|z|=1} |f(z)||dz| \leq 1\,. \tag{28}$$

In 1961 N. A. Lebedev [30] obtained a stronger inequality than (28) for functions $f(z)$ of the class R

$$\frac{1}{2\pi} \int_{|z|=1} |f(z)|^2 |dz| \leq 1\,, \tag{29}$$

that is, he showed that the class $R \in H_2$. In [30] there established all the extremal functions for the inequality (29) (see Corollary 1 of Theorem 5). N. A. Lebedev deduced this inequality as a corollary of the inequality for the product of two integrals for univalent functions without common values (see [30] and [I, Addition]). In its turn he established this inequality for two univalent functions by means of his generalized theorem of areas for any finite number of functions without common values [30]. The inequality equivalent to the property (4) of A-measure can be deduced from the Lebedev generalized Theorem of areas as well. Hence it follows immediately that the class R^* which is the subclass of univalent functions of the class R is subclass of the class A_S. In other words Bieberbach-Eilenberg domains

have nonpositive A-measure. In particular, it implies the above-mentioned in Sec. 1, sharp estimate of logarithmic area for bounded functions of the class S. The important fact that for functions $f(z) = c_1 z + \ldots \in R^*$ the inequality (16) holds was proved several times in the 60–70s by the various authors (see bibliography e.g. in [5; 24] and the short proof in [31]). The interest in the class R^* and the inequality (16) is connected with the exponential Lebedev-Milin inequalities appeared in 1965–67 [3; 4]. The matter is that the combination of the inequality (16) and one or another exponential inequality gives easily the inequality (29), its extensions, sharp in the sense of the order of growth of coefficient estimate, sharp modulus function estimate etc. Such kind of applications of the inequality (16) for Bieberbach-Eilenberg functions was first published independently by Z. Nehari [32] and D. Aharonov [33] in 1970 and then by A. Z. Grinshpan (see [19] and [5, Ch. 3]). It is clear that the inequalities for the class R^* proved by (16) hold for the whole class A_S. The results for two functions without common values are conveniently formulated in terms of the class m of pairs $\{f, \tilde{f}\}$ of functions

$$f(z) = c_1 z + \ldots \quad \text{and} \quad \tilde{f}(z) = \tilde{c}_1 z + \ldots$$

regular in E and such that for any points $z, \tilde{z} \in E$ the product $f(z)\tilde{f}(\tilde{z}) \neq 1$. Such successful generalization of Bieberbach-Eilenberg functions was introduced by D. Aharonov [34]. We denote by m^* the subclass of pairs of univalent functions of the class m. Let a pair of functions

$$\{f, \tilde{f}\} \in m^*\,,\ B = f(E)\,, \tilde{B} = \tilde{f}(E)\,.$$

It follows from the definition of the class m^* and the property (4) of A-measure that $A(B) + A(\tilde{B}) \leq 0$. Thus the conditions of Remark 2 are satisfied. Therefore for any pair of functions of the class m^* the inequality (23) holds. In particular, if $f \equiv \tilde{f}$ then $f \in R^*$ and by (23) we obtain the corresponding inequality for the class R^* which gives (17). Point out particularly two cases for Bieberbach-Eilenberg functions $\varepsilon = 1$ and 2.

Theorem 5. Let $f(z) = c_1 z + \ldots \in R^*$ and $p > 0$. Then the inequality holds

$$\sum_{k=0}^{\infty} \frac{|D_k(p)|^2}{d_k(p)} \leq 1\,,$$

where D_k are power coefficients defined by (1) for the function $f(z)$ and d_k are binomial coefficients from (4), with equality in this inequality if and only if

$$f(z) = \pm i \frac{z(1-|\zeta|^2)^{1/2}}{1-\zeta z} e^{i \arg \zeta}, \zeta \in E. \tag{30}$$

Proof. In view of $f(z) \in A_S$ we obtain the desired inequality by (17) for $\varepsilon = 1$. The assertion on equality follows from (18) and the definition of the class R^*.

Corollary 1. For $f(z) \in R$ the inequality (29) holds with equality in it if and only if $f(z)$ is defined by (30).

The assertion of Corollary 1 follows from Theorem 5 for $p = 1$ and in view of the properties of subordinate functions [29; 35].

Corollary 2. For $f(z) \in R^*$ the inequality holds

$$\frac{1}{\pi} \int_E \int \left| \frac{f(z)}{z} \right|^4 d\sigma \leq 1, \tag{31}$$

with equality in (31) if and only if $f(z)$ is defined by (30).

The assertion of Corollary 2 follows from Theorem 5 for $p = 2$ and from the identity (22) for $p = 4$.

Theorem 6. In terms of thorem 5 the inequality holds

$$\sum_{k=0}^{\infty} \frac{|D_k(p)|^2}{d_k(2p)} \leq |c_1|^p. \tag{32}$$

The exponent in the right side of (32) cannot be increased simultaneously for all the functions of the class R^*.

Proof. The inequality (32) follows from (17) for $\varepsilon = 2$. Apply the example as in [19]: for each $n = 1, 2, \ldots$ the function $g_n(z) = z[n - (n-1)z]^{-2} = zn^{-2} + \ldots \in R^*$, its power coefficients $D_k(p, g_n)$ are equal to

$$n^{-2p} d_k(2p)(1 - n^{-1})^k \quad (k = 0, 1, \ldots).$$

Therefore for the function $g_n(z)$ the left side in (32) is equal to $(2n-1)^{-2p}$. It implies that already for all the terms of sequence $g_n(z)(n = 1, 2, \dots)$ the exponent in the right side of (32) cannot be increased simultaneously.

This completes the proof of the Theorem.

Corollary 1. For $f(z) = c_1 z + \dots \in R^*$ and for $n = 2, 3, \dots$ the inequalities

$$|c_n| \leq \frac{1}{2\pi} \int_{|z|=1} |f(z)||dz| \leq |c_1|^{1/2} \tag{33}$$

hold. The assertion of Corollary 1 follows from Theorem 6 for $p = 1/2$, the identity (8) as $r \to 1$ and the Cauchy integral formula. From (33) (by the properties of subordinate functions) the inequality (28) for the class R follows. The example like that of Theorem 6, as shown in [19] for the inequality weaker than (33), verifies that in the inequality $|c_n| \leq |c_1|^{1/2}$ the exponent at $|c_1|$ cannot be increased simultaneously for all functions of the class R^* and $n = 2, 3, \dots$.

Corollary 2. [19] For $f(z) = c_1 z + \dots \in R^*$ the inequality

$$\frac{1}{\pi} \int_E \int \left| \frac{f(z)}{z} \right|^2 d\sigma \leq |c_1|$$

holds. The exponent at $|c_1|$ in this inequality cannot be increased simultaneously for all functions of the class R^*.

The assertion of Corollary 2 follows from Theorem 6 for $p = 1$ and the identity (22) for $p = 2$.

References

1. G. M. Goluzin, *Geometric Theory of Functions of a Complex Variable*, Nauka: Moscow, 1966 (in Russian).
2. L. De Branges, *A proof of the Bieberbach conjecture*, Acta Math. **154** (1985), 137–152.
3. N. A. Lebedev and I. M. Milin, *An inequality*, Vestnik Leningrad. Univ. **20**, N 19 (1965), 157–158 (in Russian).
4. I. M. Milin, *On the coefficients of univalent functions*, Dokl. Akad. Nauk SSSR **176** (1967), 1015–1018 (in Russian).

5. I. M. Milin, *Univalent Functions and Orthonormal Systems*, Nauka: Moscow, 1971 (in Russian).
6. A. Z. Grinshpan, *Univalent functions and regularly measurable mappings*, Sibirsk. Mat. Z. **27**, N 6 (1986), 50–64 (in Russian).
7. A. Z. Grinshpan, *Exponentiation method for univalent functions*, in The Theory of Functions and Approximations (Trudy 4th Saratov Zim. Scholi 25.01–05.02.1988) SGU: Saratov p. 2 (1990), pp. 72–74 (in Russian).
8. I. M. Milin, *Some applications of theorems on logarithmic coefficients*, Sibirsk. Mat. Z., to appear (in Russian).
9. I. M. Milin and A. Z. Grinshpan, *Logarithmic coefficients means of univalent functions*, Complex Variables **7** (1986), 139–147.
10. I. M. Milin and N. A. Lebedev, *On the coefficients of certain classes of analytic functions*, Dokl. Akad. Nauk SSSR **67** (1949), 221–223 (in Russian).
11. N. A. Lebedev and I. M. Milin, *On the coefficients of certain classes of analytic functions*, Math. sb. **28** (70) (1951), 359–400 (in Russian).
12. W. K. Hayman, *Multivalent Functions*, Cambridge Univ. Press, 1958.
13. A. Z. Grinshpan, *On the power stability for the Bieberbach inequality*, Zap. Nauch. Sem. Leningrad. Otdel Mat. Inst. Steklov (LOMI), **125** (1983), 58–64 (in Russian).
14. V. V. Andreev and P. L. Duren, *Inequalities for logarithmic coefficients of univalent functions and their derivatives*, Ind. Univ., Math. journal **37**, N 4 (1988), 721–733.
15. P. L. Duren and Y. J. Leung, *Logarithmic coefficients of univalent functions*, J. Analyse Math. **36** (1979), 36–43.
16. T. H. Koornwinder, *Squares of Gegenbauer polynomials and Milin type inequalities*, Report PM–R 8412, Centre for Math. and Computer Science, Amsterdam (1984).
17. I. M. Milin, *On a property of the logarithmic coefficients of univalent functions*, in Metric Questions in the Theory of Functions, Naukova Dumka: Kiev (1980), pp. 86–90 (in Russian).
18. I. M. Milin, *On one conjecture for the logarithmic coefficients of univalent functions*, Zap. Nauch. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI), **125** (1983), 135–143 (in Russian).
19. A. Z. Grinshpan, *An application of the area principle to Bieberbach-Eilenberg functions*, Mat. Zametki, 11 (1972), 609–618 (in Russian).
20. O. Teichmüller, *Untersuchungen über konforme und quasikonforme Abbildung*, Deutsche Math., Jahrg. **3**, Hf. 6 (1938), 621–678.
21. H. Wittich, *Neuere Untersuchungen über eindeutige analytische Funktionen*, Springer-Verlag, Berlin, 1955.
22. J. A. Jenkins, *On Bieberbach-Eilenberg functions*, Trans. Amer. Math. Soc. **76**, N 3 (1954), 389–396.
23. C. Caratheodory, *Über den Variabilitätsbereich der Fouriershen Konstanten von positiven harmonischen Funktionen*, Rend. Circ. Mat. Palermo **32** (1911), 193–217.

24. P. L. Duren, *Univalent Functions*, Springer-Verlag, New York, 1983.
25. D. Aharohov and H. S. Shapiro, *A minimal-area problem in conformal mapping*, in Proceedings of the Symposium of Complex Analysis (Canterbury 1973), London Math. Soc. Lecture Noto series, 12, Cambridge Univ. Press (1974), pp. 1–5.
26. C. H. Fitzgerald, *Quadratic inequalities and coefficient estimates for schlicht functions*, Arch. Ration. Mech. Anal. **46**, N 5 (1972), 356–368.
27. V. G. Cherednichenko and M. A. Barlukov, *An estimate of coefficients of univalent functions and a potential inverse problem*, preprint N 30, Inst. Mat. SO AN SSSR (1988), 3–24 (in Russian).
28. G. M. Goluzin, *Estimates for analytic functions with bounded mean modulus*, Trudy Mat. Inst. Steklov **18** (1946), 1–87 (in Russian).
29. W. Rogosinski, *On a theorem of Bieberbach-Eilenberg*, Journ. London Math. Soc. **14**, part. I, N 53 (1939), 4–11.
30. N. A. Lebedev, *An application of the area principle to non-overlapping domains*, Trudy Mat. Inst. Steklov **60** (1961), 211–231 (in Russian).
31. A. Z. Grinshpan, *On the coefficients of univalent functions assuming no pair of values w and -w*, Mat. Zametki **11** (1972), 3–11 (in Russian).
32. Z. Nehari, *On the coefficients of Bieberbach-Eilenberg functions*, J. Anal. Math. **23** (1970), 297–303.
33. D. Aharonov, *On Bieberbach-Eilenberg functions*, Bull. Amer. Math. Soc. **76**, N 1 (1970), 101–104.
34. D. Aharonov, *A generalization of a theorem of J. A. Jenkins*, Math. Z. **110** (1969), 218–222.
35. W. Rogosinski, *On the coefficients of subordinate functions*, Proc. London Math. Soc. (2) **48**, pt. 1 (1943), 48–82.

A. Z. Grinshpan
NPO
CNITA
Leningrad 192102
USSR

I. M. Milin
Mineral Processing Research
and Design Institute
Leningrad 199026
USSR

Constantin Carathéodory: An International Tribute (pp. 404–453)
edited by Th. M. Rassias

A NEW CONTRIBUTION TO THE MATHEMATICAL STUDY OF THE CATTLE-PROBLEM OF ARCHIMEDES

Carl C. Grosjean and *Hans E. De Meyer*

ABSTRACT The *Problema Bovinum* of Archimedes is a mathematical question, verbally formulated by way of a Greek epigram, about the composition of an imaginary herd consisting of four kinds of bulls and corresponding cows. The eight unknown numbers are related by seven homogeneous linear equations with very simple coefficients and, in addition, there are two constraints. The complete system of equations admits an infinity of positive integer solutions. In the course of time, the problem has been studied and discussed by at least twenty authors. One of them, named J.F. Wurm, proposed a different interpretation of part of the epigram leading to modification of one of the constraints which simplifies the problem tremendously. Both Wurm's version and the originally accepted formulation of the problem were satisfactorily treated by A. Amthor in an article published in 1880. In the present paper, the cattle-problem is reconsidered, here and there using different techniques, some algebraic, some numerical, giving rise to some complements and minor corrections to Amthor's work.

Solving the subsystem of seven linear equations, the eight unknowns are expressed as integer multiples of an integer parameter. In Wurm's version, taking into account the unmodified constraint yields a smallest solution which at the same time also satisfies the modified one. That solution involves numbers of twelve and thirteen decimal digits.

In the originally accepted formulation of the cattle–problem, the two constraints lead to a Pell equation with a fifteen–digit coefficient. A general method to obtain all the positive integer solutions of a Pell equation based upon the continued fraction development of the arithmetic square root of the coefficient is described, but if it were applied to the special case encountered in the cattle–problem, the calculations would be of an enormous volume. Using a clever artifice, Amthor arrived at another Pell equation of which he determined all the positive integer solutions yielding a solution of the cattle–problem, by the use of seven lemmas. In the present paper, the continued fraction expansion of the square root of the coefficient in Amthor's Pell equation is obtained by computer, the structure of its period is analyzed in detail and the convergents corresponding to the elements of the first period are also calculated by computer. The smallest positive integer solution of Amthor's Pell equation consists of a 45–digit and a 41–digit number. Via a homogeneous linear difference equation of second order, the general representation of all positive integer solutions of that Pellian equation is obtained, expressed in terms of the smallest solution. That representation contains an integer parameter j and it is only for $j = 2329\,n$ with $n = 1, 2, 3, \ldots$ that one finds the numbers which in turn yield the infinite set of solutions of the cattle–problem. The smallest among these solutions results from $j = 2329$. The iterative cycle of a computer program by means of which one can let a machine compute this number is described in detail. Finally, the first twenty or twenty–one decimal digits of the values of the eight unknowns constituting the smallest solution of the cattle–problem, as well as the first twenty–one digits of their sum are calculated, together with the number of digits of which each of these values consists. The number of digits is either 206544 or 206545, and so if a printer filled a page with fifty lines each comprising fifty decimal digits, the eight integers and their sum would occupy a book of 744 pages.

0. FOREWORD

This article has been written in consequence of a talk held in February 1989 by the first author at a meeting of the Permanent Commission for the History of Science, established within the Royal Academy of Science, Literature and Fine Arts of Belgium. In order that it be accessible to everyone interested in the history of science, but perhaps less familiar with mathematical expositions, the paper has

been written more explicitly than is customary for professional articles. Only very few compact mathematical symbols have been used.

1. INTRODUCTION

Around 1773, the known German writer G.E. Lessing discovered an old manuscript in the Wolfenbüttel library, comprising a Greek epigram in twenty-four verses. This epigram verbally states a problem purporting to be one proposed by Archimedes [2] to the mathematicians of Alexandria, by way of a letter to Eratosthenes. Lessing [8] published the epigram, together with a scholium giving a false solution, and also a long mathematical discussion by C. Leiste. The problem consists in finding the composition of an imaginary divine stock of cattle, i.e. the numbers W, X, Y, Z of white, black, piebald and brown bulls, and the numbers w, x, y, z of cows of the corresponding kinds, between which the following relations exist:

$$W = (\frac{1}{2} + \frac{1}{3})X + Z\,, \quad X = (\frac{1}{4} + \frac{1}{5})Y + Z\,, \quad Y = (\frac{1}{6} + \frac{1}{7})W + Z\,,$$
$$w = (\frac{1}{3} + \frac{1}{4})(X + x)\,, \qquad x = (\frac{1}{4} + \frac{1}{5})(Y + y)\,,$$
$$y = (\frac{1}{6} + \frac{1}{6})(Z + z)\,, \qquad z = (\frac{1}{6} + \frac{1}{7})(W + w)\,,$$
$$W + X = \text{a square integer number} = p^2\,, \quad p \in \mathbb{N}_0\,, \text{(*)}$$
$$Y + Z = \text{a triangular integer number} = q(q+1)/2\,, \; q \in \mathbb{N}_0\,, \tag{1}$$

so called on account of

	total	q
1	1	1
1 1	3	2
1 1 1	6	3
1 1 1 1	10	4
1 1 1 1 1	15	5
...	⋮	⋮

(*) The condition reads: p represents a number belonging to the set $\{1, 2, 3, \ldots\}$.

where q labels the successive lines in the triangle of 1's and where the total number of 1's in the first, the second, ... and the q th line equals $q(q+1)/2$.

The first seven relations are homogeneous linear equations with strikingly simple coefficients. As they involve eight unknowns, they do not suffice to define a unique solution. The eight unknowns can be expressed in terms of one degree of freedom, i.e. one parameter which can take on any positive integer value. The last two equations may be regarded as constraints. It turns out, however, that when one joins them to the first seven equations, they do not isolate one or a finite number of solutions. The complete system (1) still admits an infinity of solutions, but these solutions form a tiny portion of those which satisfy only the first seven equations. As will be shown further on, the two constraints are extremely selective; they restrict the solutions tremendously.

In the course of the two centuries following the year 1773, some twenty authors published a variety of comments and opinions concerning the cattle–problem. Since the history of the problem can be found in the existing literature [5–7], we shall not go into details on that matter. Because of its mathematical significance, we solely mention a variant of the cattle–problem originating with J.F. Wurm [10]. This author argued that the part of the epigram formulating the constraint $W + X =$ a square integer, can be interpreted in a different way. Indeed, translating that passage into English, we have: "when the white bulls joined in number with the black, they stood firm with depth and breadth of equal measurement; and the plains of Thrinakia, far–stretching all ways, were filled with their multitude". Hence, the white and the black bulls, packed as a matrix, seem to cover a square *figure*, but the animals not being square individually, Wurm remarked that in this interpretation, $W + X$ is not a square *number*. Rather, that sum should be the product of two integer numbers of the same order of magnitude, preferably with a ratio approximating that of the length and the breadth of an average bull. In this manner, there is a version of the cattle–problem known as Wurm's problem. It was fully treated by A. Amthor [1], a German mathematician, and it is in fact sufficiently simple so that nothing worthwhile can be added to the solution. The

so–called complete problem (Amthor called it "Das Hauptproblem"), corresponding to the system of nine equations as given in (1), was also solved in extenso by Amthor [1]. The treatment of Wurm's problem and an abridged account of the solution of the complete problem can be found in T.L. Heath's book on the works of Archimedes [6]. Amthor's solution of (1) is undoubtedly remarkable, especially because in his time he did not dispose of modern automatic computational facilities and therefore had to rely solely on non–numerical algebraic methods. For instance, in his article [1], he had to establish seven "Hilfssätze" to determine the smallest solution of a Pell equation which satisfies a supplementary condition of divisibility.

With the present article, it is our aim

- to expound how we treated certain parts of the complete cattle–problem by computer;
- to give analytical explanations of certain partial results;
- to make some additions and minor corrections to Amthor's calculations.

2. SOLUTION OF THE CATTLE–PROBLEM

When one eliminates X and Y from the first three equations in (1) by suitable linear combination, one finds

$$297\,W = 742\,Z \qquad \text{or} \qquad 3^3.11\,W = 2.7.53\,Z. \tag{2}$$

In a similar manner, one also obtains

$$99\,X = 178\,Z \qquad \text{or} \qquad 3^2.11\,X = 2.89\,Z \tag{3}$$

and

$$891\,Y = 1580\,Z \qquad \text{or} \qquad 3^4.11\,Y = 2^2.5.79\,Z\,. \tag{4}$$

W, X, Y and Z symbolizing positive integers, Z must be divisible by the smallest common multiple of the coefficients appearing in the left-hand sides of (2)–(4), namely $3^4.11$ or 891. Hence, the result

$$W = 2226\,m\,,\ X = 1602\,m\,,\ Y = 1580\,m\,,\ Z = 891\,m \quad (m = 1,2,3,\ldots)$$

represents all positive integer solutions of the first three equations in (1). This was Leiste's first result. When one inserts (2), (3) and (4) into the next four linear equations of (1) and solves for w, x, y and z, one obtains these unknowns also in terms of Z:

$$1\,383\,129\,w = 2\,402\,120\,Z \quad \text{or} \quad 3^3.11.4657\,w = 2^3.5.7.23.373\,Z \tag{5}$$

$$461\,043\,x = 543\,694\,Z \quad \text{or} \quad 3^2.11.4657\,x = 2.17.15991\,Z \tag{6}$$

$$125\,739\,y = 106\,540\,Z \quad \text{or} \quad 3^3.4657\,y = 2^2.5.7.761\,Z \tag{7}$$

$$461\,043\,z = 604\,357\,Z \quad \text{or} \quad 3^2.11.4657\,z = 13.46489\,Z\,. \tag{8}$$

Again, since the eight unknowns in the problem can only take on positive integer values, Z should now be divisible by $3^4.11.4657$, being the smallest common multiple of $3^4.11$ and the coefficients appearing in the left-hand sides of (5)–(8). Thus, *all* solutions of the subsystem consisting of the first seven homogeneous linear equations of (1) in terms of positive integers are comprised in

$$\left.\begin{aligned}
W &= 2.3.7.53.4657\,n = 10\,366\,482\,n\\
X &= 2.3^2.89.4657\,n = 7\,460\,514\,n\\
Y &= 2^2.5.79.4657\,n = 7\,358\,060\,n\\
Z &= 3^4.11.4657\,n = 4\,149\,387\,n\\
w &= 2^3.3.5.7.23.373\,n = 7\,206\,360\,n\\
x &= 2.3^2.17.15991\,n = 4\,893\,246\,n\\
y &= 2^2.3.5.7.11.761\,n = 3\,515\,820\,n\\
z &= 3^2.13.46489\,n = 5\,439\,213\,n
\end{aligned}\right\} \quad (n = 1,2,3,\ldots) \tag{9}$$

Note how many decimal digits are already required at this stage to write even the smallest solution of the linear subsystem comprised in (1), despite the simplicity

of the coefficients in these equations. The reason lies in the fact that a system of seven linear equations can no longer be called a small system. After all, its solution(s) written à la Cramer require(s) determinants of seven rows and seven columns.

As can be concluded from their decomposition into prime factors, the greatest common divisor of the eight proportionality coefficients in (9) is equal to 1. It is somewhat astonishing that Leiste obtained these coefficients each multiplied by 20. Under this circumstance, when the parameter in his result runs over all positive integers, he only gets one solution out of twenty for the subsystem of seven linear equations. The values appearing in the scholium cited in Lessing's publication correspond to setting the parameter in Leiste's solution equal to 4, which is equivalent to putting $n = 80$ in (9). This, however, does not yield a solution of the entire set of nine equations constituted by (1) because

$$\begin{aligned} W + X &= (10\,366\,482 + 7\,460\,514)\,80 &&= 1\,426\,159\,680 = (37\,764,528\ldots)^2\,, \\ Y + Z &= (7\,358\,060 + 4\,149\,387)\,80 &&= 920\,595\,760 \\ & && = \frac{42\,908,607\ldots \times 42\,909,607\ldots}{2} \end{aligned}$$

which shows that $W + X$ is not a square integer and $Y + Z$ not a triangular number.

Amthor [1] proved that in order to solve Wurm's problem, it is sufficient to find the smallest positive n in (9) for which $Y + Z$ is a triangular number, i.e.

$$Y + Z = \frac{q(q+1)}{2}\,, \qquad\qquad q \in \mathbb{N}_0\,.$$

He easily obtained $q = 1\,643\,921$, which corresponds to $n = 117\,423$ in (9). An average bull measures from muzzle to tail approximately three times its breadth. Hence, when the white bulls mingle with the black to form a rectangular grid packed on a square meadow, the above solution of Wurm's problem yields 861 102 rows and 2 430 954 columns:

$$W + X = (10\,366\,482 + 7\,460\,514)117\,423 = 861\,102 \times 2\,430\,954\,.$$

In the spirit of Wurm's interpretation of the condition on $W + X$, this is a more logical decomposition of that sum into a product of two integer factors than the one given by Amthor, namely, $1\,485\,583 \times 1\,409\,076$. The smallest solution of Wurm's problem demands integers of twelve and thirteen decimal digits, the total number of cattle being $5\,916\,837\,175\,686$. But this is almost nothing compared to the smallest solution of the complete problem which, as we shall see, involves numbers of more than 206 thousand decimal digits.

Continuation of the Solution of the Complete Cattle–Problem

Next, we turn to the eighth equation of (1):

$$W + X = p^2\,, \qquad p \in \mathbb{N}_0\,.$$

From (9), it follows that

$$\begin{aligned} p^2 &= 2.3\,(7.53 + 3.89)\,4657\,n = 17\,826\,996\,n \\ &= 2^2.3.11.29.4657\,n \end{aligned}$$

and the right–hand side is a square integer as soon as

$$n = 3.11.29.4657\,N^2 \tag{10}$$

where N is a new, not yet determined, positive integer. At this stage, all solutions of the first eight equations in (1) are comprised in

$$\left.\begin{aligned} W &= 2.3^2.7.11.29.53.4657^2\,N^2 &&= 46\,200\,808\,287\,018\,N^2 \\ X &= 2.3^3.11.29.89.4657^2\,N^2 &&= 33\,249\,638\,308\,986\,N^2 \\ Y &= 2^2.3.5.11.29.79.4657^2\,N^2 &&= 32\,793\,026\,546\,940\,N^2 \\ Z &= 3^5.11^2.29.4657^2\,N^2 &&= 18\,492\,776\,362\,863\,N^2 \\ w &= 2^3.3^2.5.7.11.23.29.373.4657\,N^2 &&= 32\,116\,937\,723\,640\,N^2 \\ x &= 2.3^3.11.17.29.4657.15991\,N^2 &&= 21\,807\,969\,217\,254\,N^2 \\ y &= 2^2.3^2.5.7.11^2.29.761.4657\,N^2 &&= 15\,669\,127\,269\,180\,N^2 \\ z &= 3^3.11.13.29.4657.46489\,N^2 &&= 24\,241\,207\,098\,537\,N^2 \end{aligned}\right\} \tag{11}$$

$$(N = 1, 2, 3, \ldots),$$

resulting from combining (9) and (10). It remains to determine N so that

$$Y + Z = \frac{q(q+1)}{2}$$

or

$$\frac{q(q+1)}{2} = 3.7.11.29.353.4657^2\, N^2\,. \tag{12}$$

Putting $q = 2s - 1$ (odd) or $2s$ (even) and $N = u.v$, as Amthor did in his way of solving Wurm's problem, would lead to much too voluminous calculations. Hence, preference is to be given to multiplying both sides of (12) by 8 which yields

$$4q^2 + 4q = 2^3.3.7.11.29.353.4657^2\, N^2 \tag{13}$$

and setting $2q + 1 = M$. Then, one obtains a quadratic diophantine equation, known as a Pell equation because of its typical form:

$$M^2 - 410\,286\,423\,278\,424\, N^2 = 1\,. \tag{14}$$

The final part in solving the complete cattle–problem is in this manner reduced to determining the couples of positive integers M, N satisfying this equation. The couple of smallest values yields the N which in turn produces the smallest solution of the problem of Archimedes when inserted into (11).

Positive Integer Solutions of a Pellian Equation

A Pellian equation, say,

$$M^2 - AN^2 = 1\,, \tag{15}$$

whereby A can be any positive integer which is not a square, hence

$$A \in \{2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, \ldots\}\,,$$

is known to have infinitely many positive integer solutions, e.g., when $A = 3$:

$$(M, N) = (2,1), (7,4), (26,15), (97,56), \ldots \tag{16}$$

A method to find the complete set of such solutions in the case of (15) consists in

i) developing $\sqrt{A}$ into a continued fraction of the kind

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \ldots}}}, \quad \text{symbolized by } [a_1, a_2, a_3, a_4, \ldots],$$

where a_1, a_2, a_3, a_4, ... are positive integers. It is well-known that, for any A, the continued fraction is of infinite length and of the mixed-periodic type. More precisely, for the arithmetic square root of any non-square positive integer number A, there comes:

$$\sqrt{A} = [a\,;\, b_1\,,\, b_2\,,\, \ldots\,,\, b_h\,;\, b_1\,,\, b_2\,,\, \ldots\,,\, b_h\,;\, \ldots], \tag{17}$$

where a is the integral part of the decimal representation of $\sqrt{A}$, being the single-element aperiodic part of the expansion, and $b_1\,,\, b_2\,,\, \ldots\,,\, b_h$ is a sequence of elements called the period (of length h), repeating itself an infinite number of times;

ii) calculating the sequence of convergents (or approximants) $\{R_1/S_1\,,\, R_2/S_2\,,\, \ldots,\, R_n/S_n\,, \ldots\}$ of the continued fraction, obtained by truncation after the first, the second, ..., the nth element,.... In the case of (17), this gives:

$$\frac{R_1}{S_1} = \frac{a}{1}\,,\quad \frac{R_2}{S_2} = a + \frac{1}{b_1} = \frac{ab_1 + 1}{b_1}\,,\quad \frac{R_3}{S_3} = a + \frac{1}{b_1 + (1/b_2)} = \frac{ab_1b_2 + b_2 + a}{b_1b_2 + 1}\,,$$

and so on.

It is also well-known that the recurrence relations connecting successive numerators and successive denominators are, in general, for *any* continued fraction $[a_1\,,\, a_2\,,\, a_3\,,\, a_4\,, \ldots]$:

$$\left.\begin{aligned} R_n &= a_nR_{n-1} + R_{n-2} \\ S_n &= a_nS_{n-1} + S_{n-2} \end{aligned}\right\} \qquad (n = 3,4,5,\ldots) \tag{18}$$

starting from $R_1 = a_1$, $S_1 = 1$, $R_2 = a_1 a_2 + 1$, $S_2 = a_2$. Note that the calculation of the numerators is entirely separated from that of the denominators;

iii) If h is the length of the period in the case of the continued fraction development of $\sqrt{A}$, then

- when h is even, i.e., $h \in \{2, 4, 6, \ldots\}$, the infinite sequence of couples

$$(R_h, S_h), (R_{2h}, S_{2h}), \ldots, (R_{jh}, S_{jh}), \ldots \tag{19}$$

constitutes the complete collection of positive integer solutions of eq.(15);

- when h is odd, i.e., $h \in \{1, 3, 5, \ldots\}$, the infinite sequence of couples

$$(R_{2h}, S_{2h}), (R_{4h}, S_{4h}), \ldots, (R_{2jh}, S_{2jh}), \ldots \tag{19'}$$

forms the complete set of positive integer solutions of (15). The reason that the couples bearing an odd multiple of h as subscript are absent in (19') is that they satisfy

$$M^2 - AN^2 = -1 .$$

In this connection, see (31) – (31').

The method may be tested by the reader on the particular case of $M^2 - 3N^2 = 1$ in order to confirm (16).

Therefore, in order to solve (14), one should start by calculating the period in the continued fraction expansion of the square root of the fifteen–(decimal) digit coefficient of N^2. This is an enormous task when it is carried out by hand. In 1867, C.F. Meyer [9] gave it a try but did not succeed as he stopped at the 240th quotient without having found the period. Later, more precisely in 1895, A.H. Bell [3][4] found the smallest solution of the system (1), based on the Pell equation (14)(*),but did nothing more than confirm Amthor's solution obtained fifteen years earlier. Amthor made the problem connected with (14) more tractable by noticing that in this equation the coefficient of N^2 comprises $2^2.4657^2$. He therefore sets

(*) The continued fraction development of the (arithmetic) square root of the fifteen–digit coefficient in eq.(14) has a period of length $h = 203254$.

$$2.4657\,N = u \tag{20}$$

with u a new positive integer. He also puts $4q^2 + 4q = t^2 - 1$ in (13) whereby $t = 2q + 1$ and so arrives at the Pell equation

$$t^2 - 4\,729\,494\,u^2 = 1 \qquad (4\,729\,494 = 2.3.7.11.29.353)\,. \tag{21}$$

This equation also admits infinitely many positive integer solutions, but in contrast to (14), it is not its smallest solution which leads to the smallest solution of (1). Indeed, (20) shows that u should be divisible by 9314. Thus, in replacing (14) by (21) which entails a tremendous simplification as far as the continued fraction development of the square root of the coefficient is concerned, Amthor had to pay the price of obtaining the smallest positive integer solution of (21) in which u is an integer multiple of 9314. This has necessitated the formulation and proof of no less than seven lemmas pertaining to the Pellian equation and its solutions in general.

On the Continued Fraction Development of $\sqrt{4\,729\,494}$

In Amthor's paper [1], one finds in full detail the way of obtaining the continued fraction expansion of V where V means $\sqrt{4\,729\,494}$. The algorithm is very simple:

$$V = 2174,739\ldots = 2174 + r_1 \qquad \text{with} \qquad 0 < r_1 < 1\,,$$

$$r_1 = V - 2174 = \frac{4\,729\,494 - 2174^2}{V + 2174} = \frac{3218}{V + 2174} = \frac{1}{\left(\dfrac{V + 2174}{3218}\right)}$$

$$= \frac{1}{1,351\,379\ldots} = \frac{1}{1 + \dfrac{V - 1044}{3218}} = \frac{1}{1 + r_2} \qquad \text{with} \quad 0 < r_2 < 1\,,$$

$$r_2 = \frac{V - 1044}{3218} = \frac{4\,729\,494 - 1044^2}{3218(V + 1044)} = \frac{1131}{V + 1044} = \frac{1}{\left(\dfrac{V + 1044}{1131}\right)}$$

$$= \frac{1}{2,845\,923\ldots} = \frac{1}{2 + \dfrac{V - 1218}{1131}} = \frac{1}{2 + r_3} \qquad \text{with} \qquad 0 < r_3 < 1\,,$$

$$r_3 = \frac{V-1218}{1131} = \frac{4\,729\,494 - 1218^2}{1131(V+1218)} = \frac{2870}{V+1218} = \frac{1}{\left(\frac{V+1218}{2870}\right)}$$

$$= \frac{1}{1,182\,139\ldots} = \frac{1}{1+\frac{V-1652}{2870}} = \frac{1}{1+r_4} \qquad \text{with} \qquad 0 < r_4 < 1, \quad (22)$$

etc. When the algorithm is repeated over and over again, one arrives at the following mixed-periodic continued fraction expansion of V:

$$\sqrt{4\,729\,494} = [2174; 1,2,1,5,2,25,3,1,1,1,1,1,1,15,1,2,16,1,2,1,1,8,6,$$
$$1,21,1,1,3,1,1,1,1,2,2,6,1,1,5,1,17,1,1,47,3,1,1,\mathbf{6},1,1,3,$$
$$47,1,1,17,1,5,1,1,6,2,2,1,1,1,3,1,1,21,1,6,8,1,1,2,1,16,$$
$$2,1,15,1,1,1,1,1,1,3,25,2,5,1,2,1,4348\,;\,1,2,1,5,2,25,3,$$
$$1,1,1,\ldots]. \qquad (23)$$

Besides the aperiodic element 2174, there is a period of 92 elements repeating itself indefinitely. Amthor found this period, but on account of his slightly confusing way of presenting the result (23), most authors after him, as for instance T.L. Heath and L.E. Dixon, have talked about a period of 91 elements, in contradiction with the true length $h = 92$. Yet, as we have seen, the length of the period is important to select the convergents whose numerator and denominator provide solutions of the Pellian equation (21).

The period in (23) has the following structure :

- its first 91 elements exhibit mirror symmetry with respect to the element in the middle which is a 6 (printed in bold type);
- its last element is equal to twice the aperiodic element.

Such a structure is far from being exceptional among the continued fraction expansions of irrational numbers of the form $\sqrt{A}$ where A is a positive non-square integer(*).For instance, in the case of $A = 14$, we have:

$$\sqrt{14} = [3;1,\mathbf{2},1,6;1,2,\ldots]$$

(*) The condition non-square stems from the fact that for a square A, $\sqrt{A}$ is an integer, and so having no decimal part, it also has no non-vanishing continued fraction expansion. This is in harmony with $M^2 - AN^2 = 1$ then having no non-negative integral solutions except $M = 1, N = 0$.

and other examples are

$$\begin{aligned}
\sqrt{19} &= [4;2,1,\mathbf{3},1,2,\mathbf{8};2,1,3,\ldots], \\
\sqrt{46} &= [6;1,3,1,1,2,\mathbf{6},2,1,1,3,1,12;1,3,1,\ldots], \\
\sqrt{57} &= [7;1,1,\mathbf{4},1,1,14;1,1,4,\ldots], \\
\sqrt{62} &= [7;1,\mathbf{6},1,14;1,6,1,\ldots], \\
\sqrt{67} &= [8;5,2,1,1,\mathbf{7},1,1,2,5,16;5,2,1,\ldots], \\
\sqrt{79} &= [8;1,\mathbf{7},1,16;1,7,1,\ldots], \\
\sqrt{94} &= [9;1,2,3,1,1,5,1,\mathbf{8},1,5,1,1,3,2,1,18;1,2,3,\ldots], \\
\sqrt{151} &= [12;3,2,7,1,3,4,1,1,1,\mathbf{11},1,1,1,4,3,1,7,2,3,24;\ldots], \\
\sqrt{152} &= [12;\mathbf{3},24;3,24;\ldots], \\
\sqrt{244} &= [15;1,1,1,1,1,2,1,5,1,1,9,1,\mathbf{6},1,9,1,1,5,1,2,1,1,1,1,1,30;\ldots].
\end{aligned}$$

In fact, for $2 \leq A \leq 360$, hence on a total of 342 cases, one finds 280 cases involving a period of even length (in which we include

$$\sqrt{k^2 + \frac{2k}{c}} = [k\,;c,2k\,;c,2k\,;\ldots] \qquad (k = 1,2,3,\ldots)$$

whereby $1 \leq c < 2k$ and c a divisor of $2k$, constituting all possible cases with period length 2). The cases with period length $h = 2g$ whereby $g = 2,3,4,\ldots$, all have the above–mentioned characteristics: mirror symmetry with respect to the gth element of the period and the last element of the period equal to twice the aperiodic element. Only occasionally does one encounter a period of odd length. For $2 \leq A \leq 360$, one finds 62 such cases among which eighteen of period length $h = 1$, with the period element equal to the double of the aperiodic element, these cases being of the form $\sqrt{A} = \sqrt{k^2+1}$ $(k = 1,2,3,\ldots)$, and 44 cases of period length $h = 2g+1$ whereby $g = 1,2,3,\ldots$. Examples are:

$$\begin{aligned}
\sqrt{13} &= [3;1,1,1,1,6;1,1,\ldots], \\
\sqrt{29} &= [5;2,1,1,2,10;2,1,\ldots], \\
\sqrt{41} &= [6;2,2,12;2,2,\ldots],
\end{aligned}$$

$$\sqrt{58} = [7;1,1,1,1,1,1,14;1,1,\ldots],$$
$$\sqrt{61} = [7;1,4,3,1,2,2,1,3,4,1,14;1,4,\ldots],$$
$$\sqrt{73} = [8;1,1,5,5,1,1,16;1,1,\ldots],$$
$$\sqrt{74} = [8;1,1,1,1,16;1,1,\ldots],$$
$$\sqrt{130} = [11;2,2,22;2,2,\ldots],$$
$$\sqrt{181} = [13;2,4,1,8,6,1,1,1,1,2,2,1,1,1,1,6,8,1,4,2,26;2,4,\ldots].$$

The periods of odd length $h \ (= 2g+1) \geq 3$ still have the same properties, except that the mirror symmetry is now with respect to the gth comma in the period.

Now, some details about the characteristics of the period. Let A again be any positive non–square integer and let

$$\sqrt{A} = [a;b_1,b_2,b_3,\ldots,b_h;b_1,b_2,\ldots].$$

Inspired by (22), we can write the algorithm to generate the b–elements as follows:

$$\begin{cases} A - c_{n-1}^2 = d_{n-1} \times d_n\,, \\ \left[\dfrac{\sqrt{A}+c_{n-1}}{d_n}\right] = b_n\,, \qquad (n = 1,2,3,\ldots,h) \\ b_n\, d_n - c_{n-1} = c_n\,, \end{cases} \tag{24}$$

with $d_0 = 1$, $c_0 = a =$ integral part of $\sqrt{A}$ and whereby $[x]$ means the largest integer smaller than or equal to the real number x. In this nth cycle, the numbers c_{n-1} and d_{n-1} stem from the preceding cycle when $n \geq 2$. Since $d_0 = 1$, d_1 is obviously a positive integer, but that $d_2, d_3, \ldots$ are integers has to be proved. That $A - c_{n-1}^2$ is divisible by d_{n-1} for $n \geq 2$ can be shown by complete induction:

$$A - c_{n-2}^2 = d_{n-2} \times d_{n-1} \quad \Rightarrow \quad A - c_{n-1}^2 = d_{n-1} \times d_n\,.$$

Indeed,

$$\begin{aligned} A - c_{n-1}^2 &= A - (b_{n-1}\, d_{n-1} - c_{n-2})^2 \\ &= (A - c_{n-2}^2) - b_{n-1}^2\, d_{n-1}^2 + 2c_{n-2}\, b_{n-1}\, d_{n-1} \\ &= (d_{n-2} - b_{n-1}^2\, d_{n-1} + 2c_{n-2}\, b_{n-1})\, d_{n-1}\,. \end{aligned}$$

The nth cycle given by (24) stems from the following step in the continued fraction development of $\sqrt{A}$:

$$r_n = \frac{\sqrt{A}-c_{n-1}}{d_{n-1}} = \frac{d_n}{\sqrt{A}+c_{n-1}} = \frac{1}{\left(\frac{\sqrt{A}+c_{n-1}}{d_n}\right)}$$

$$= \frac{1}{\left[\frac{\sqrt{A}+c_{n-1}}{d_n}\right] + \frac{\sqrt{A}-(b_n d_n - c_{n-1})}{d_n}} = \frac{1}{b_n + r_{n+1}} \qquad (n = 1, 2, \ldots, h) \tag{25}$$

where

$$r_{n+1} = \frac{\sqrt{A}-(b_n d_n - c_{n-1})}{d_n} = \frac{\sqrt{A}-c_n}{d_n}. \tag{25'}$$

Still by complete induction, one can deduce from the equalities in (24), (25)–(25'):

$$1 \le c_n \le a\ (< \sqrt{A}), \quad d_n \ge 1, \quad b_n \ge 1 \qquad (n = 1, 2, 3, \ldots, h)$$

all the symbols in (24) representing integers except $\sqrt{A}$.

In the case of $A = 4\,729\,494$, the computations show that $c_0, c_1, c_2, \ldots$ are all different up to c_{45}, but then something exceptional occurs: it appears that c_{45} is divisible by d_{46}, i.e., $c_{45} = 1827$ and $d_{46} = 609$. In this manner, there comes:

$$r_{46} = \frac{\sqrt{4\,729\,494}-c_{45}}{d_{45}} = \frac{d_{46}}{\sqrt{4\,729\,494}+c_{45}} = \frac{1}{\frac{\sqrt{4\,729\,494}+c_{45}}{d_{46}}}$$

$$= \frac{1}{6,571001\ldots} = \frac{1}{2\frac{c_{45}}{d_{46}} + \frac{\sqrt{4\,729\,494}-c_{45}}{d_{46}}}.$$

Hence, $c_{46} = c_{45}$ and in (24), we find for $n = 46$:

$$4\,729\,494 - 1827^2 = 2285 \times 609 \qquad (\text{being } A - c_{45}^2 = d_{45} \times d_{46}),$$

$$\left[\frac{\sqrt{4\,729\,494}+1827}{609}\right] = 6 \quad (= b_{46}),$$

$$6 \times 609 - 1827 = 1827 \quad (= c_{46}).$$

This, in turn, has as consequence:

$$r_{47} = \frac{\sqrt{4\,729\,494} - c_{46}}{d_{46}} = \frac{\sqrt{4\,729\,494} - c_{45}}{d_{46}} = \frac{d_{45}}{\sqrt{4\,729\,494} + c_{45}}$$
$$= \frac{1}{\left(\dfrac{\sqrt{4\,729\,494} + c_{45}}{d_{45}}\right)} = \frac{1}{b_{45} + r_{48}}$$

because $c_{45} = c_{46}$, $d_{45} = d_{47}$, $b_{45} = b_{47}$, and r_{48} can be written as

$$r_{48} = \frac{\sqrt{4\,729\,494} - c_{44}}{d_{45}},$$

since

$$b_{45}\, d_{45} - c_{44} = c_{45}$$

is an equality which was already obtained in (24) for $n = 45$. In the same way,

$$r_{48} = \frac{\sqrt{4\,729\,494} - c_{44}}{d_{45}} = \frac{d_{44}}{\sqrt{4\,729\,494} + c_{44}} = \frac{1}{\left(\dfrac{\sqrt{4\,729\,494} + c_{44}}{d_{44}}\right)}$$
$$= \frac{1}{b_{44} + r_{49}}$$

with

$$r_{49} = \frac{\sqrt{4\,729\,494} - c_{43}}{d_{44}},$$

etc. Therefore, as the subscript of r increases, those of c and d decrease, and $b_{47} = b_{45}$, $b_{48} = b_{44}, \ldots$, showing the mirror symmetry with respect to b_{46}. The symmetry property ends when $b_{91} = b_1$ is attained. Something similar happens in infinitely many other cases when $\sqrt{A}$ is expanded into a continued fraction.

As for the last element in the period associated with $\sqrt{A}$, on account of the aperiodic part in the development being composed solely of the element a (= the integral part of $\sqrt{A}$) and the mirror symmetry discussed above, the last element is equal to $2a$. Indeed, when

$$\sqrt{A} = [a\,;\, b_1, b_2, \ldots, b_2, b_1, c\,;\, b_1, b_2, \ldots]$$

with period length h, then we have:

$$\sqrt{A} = a + r_1;$$

$$r_1 = \frac{A - a^2}{\sqrt{A} + a} = \frac{1}{\left(\frac{\sqrt{A} + a}{A - a^2}\right)} = \frac{1}{b_1 + r_2}$$

with

$$b_1 = \left[\frac{\sqrt{A} + a}{A - a^2}\right];$$

$$r_2 = \frac{\sqrt{A}}{A - a^2} - \left(b_1 - \frac{a}{A - a^2}\right) = \frac{\frac{A}{(A - a^2)^2} - \left(b_1 - \frac{a}{A - a^2}\right)^2}{\frac{\sqrt{A}}{A - a^2} + \left(b_1 - \frac{a}{A - a^2}\right)}$$

$$= \frac{1}{\left(\frac{\sqrt{A} + ((A - a^2)b_1 - a)}{1 + 2ab_1 - (A - a^2)b_1^2}\right)} = \frac{1}{b_2 + r_3}$$

with

$$b_2 = \left[\frac{\sqrt{A} + ((A - a^2)b_1 - a)}{1 + 2ab_1 - (A - a^2)b_1^2}\right],$$

etc. Consequently, when the period ends as b_2, b_1, c, we have:

$$r_{h-2} = \frac{1}{b_2 (= b_{h-2}) + r_{h-1}}$$

and if r_{h-1} gives rise to the element b_1 by way of

$$r_{h-1} = \frac{1}{b_1 (= b_{h-1}) + r_h},$$

then, necessarily,

$$r_{h-1} = \frac{\sqrt{A} - ((A - a^2)b_1 - a)}{1 + 2ab_1 - (A - a^2)b_1^2}$$

because

$$r_{h-1} = \frac{A - a^2}{\sqrt{A} + ((A - a^2)b_1 - a)} = \frac{1}{\left(\frac{\sqrt{A} - a + (A - a^2)b_1}{A - a^2}\right)}$$

$$= \frac{1}{b_1 + \frac{1}{\sqrt{A} + a}}.$$

In turn, there comes:

$$r_h = \frac{1}{\sqrt{A}+a} = \frac{1}{2a+(\sqrt{A}-a)} = \frac{1}{2a+r_1} = \frac{1}{2a+\dfrac{1}{b_1+r_2}},$$

etc. Hence, $c = 2a$.

Note that b_g, the element in the middle of $b_1, b_2, \ldots, b_2, b_1$ when $h = 2g$, is not necessarily even, as it is the case with $\sqrt{4\,729\,494}$. For instance, in the case of $A = 67$, we find:

$$r_4 = \frac{\sqrt{67}-2}{7} = \frac{1}{\left(\dfrac{\sqrt{67}+2}{9}\right)} = \frac{1}{1+\dfrac{\sqrt{67}-7}{9}},$$

$$r_5 = \frac{\sqrt{67}-7}{9} = \frac{1}{\left(\dfrac{\sqrt{67}+7}{2}\right)} = \frac{1}{7+\dfrac{\sqrt{67}-7}{2}},$$

$$r_6 = \frac{\sqrt{67}-7}{2} = \frac{1}{\left(\dfrac{\sqrt{67}+7}{9}\right)} = \frac{1}{1+\dfrac{\sqrt{67}-2}{9}},$$

etc. In the case of $A = 4\,729\,494$, we had

$$b_{46} = 2\frac{c_{45}}{d_{46}} = 2\frac{1827}{609} = 6,$$

even on account of the divisibility of c_{45} by d_{46}. Here,

$$b_5 = 2\frac{c_4}{d_5} = 2\frac{7}{2} = 7.$$

In general, when the period length is $2g$, the element b_g is

$$b_g = 2\frac{c_{g-1}}{d_g} \begin{cases} \text{even when } c_{g-1} \text{ happens to be divisible by } d_g, \\ \text{odd when } c_{g-1}/d_g \text{ is half-odd integral.} \end{cases}$$

Calculation of the Convergents of the Continued Fraction Expansion of $\sqrt{4\,729\,494}$

It is an easy task for an automatic computer to calculate the convergents one after the other:

$$\frac{R_1}{S_1} = \frac{2174}{1}, \quad \frac{R_2}{S_2} = \frac{2175}{1}, \quad \frac{R_3}{S_3} = \frac{6524}{3}, \quad \frac{R_4}{S_4} = \frac{8699}{4}, \ldots,$$
$$\frac{R_{20}}{S_{20}} = \frac{2R_{19} + R_{18}}{2S_{19} + S_{18}} = \frac{327\,826\,696\,818}{150\,742\,939}, \ldots$$

whereby use is made of the formulae in(18). The numerator and the denominator of each convergent remain of the same order of magnitude when the integral part 2174 is split off, i.e.,

$$\frac{R_n}{S_n} = 2174 + \frac{Q_n}{S_n},$$

with

$$\frac{Q_1}{S_1} = \frac{0}{1}, \quad \frac{Q_2}{S_2} = \frac{1}{1}, \quad \frac{Q_3}{S_3} = \frac{2}{3}, \quad \frac{Q_4}{S_4} = \frac{3}{4}, \ldots,$$

Q_n/S_n being the nth convergent of $[0; 1, 2, 1, 5, 2, 25, \ldots]$, with the numerator and the denominator generated separately by

$$Q_n = b_{n-1}\,Q_{n-1} + Q_{n-2} \qquad S_n = b_{n-1}\,S_{n-1} + S_{n-2} \qquad (n = 3, 4, 5, \ldots, 93). \quad (26)$$

These convergents approximate in an oscillatory manner the irrational number $\sqrt{4\,729\,494} - 2174$ $(= 0,739\,984\,457\ldots)$. It is preferable for checking purposes to let a machine compute the Q's and the S's, and afterwards obtain any R–numerator required by means of $R_n = 2174\,S_n + Q_n$.

In order to find the smallest positive integer solution (t_1, u_1) of (21), one should calculate the ninety–second convergent R_{92}/S_{92}, because

$$t_1 = R_{92} = 2174\,S_{92} + Q_{92}, \qquad u_1 = S_{92}. \quad (27)$$

It turns out that t_1 consists of 45 decimal digits and u_1 comprises 41 decimal digits:

$$t_1 = R_{92} = \quad 109\,931\,986\,732\,829\,734\,979\,866\,232\,821\,433\,543\,901\,088\,049 \qquad (28)$$

$$u_1 = S_{92} = \quad 50\,549\,485\,234\,315\,033\,074\,477\,819\,735\,540\,408\,986\,340\,. \qquad (28')$$

The correctness of this solution has been verified by calculation of t_1^2 and $4\,729\,494\,u_1^2 + 1$, both yielding the same number of 89 decimal digits. More theoretically, that it is the ninety-second convergent which provides a solution of the Pell equation (21), may be checked as follows. If we put

$$\sqrt{4\,729\,494} = 2174 + x\,,$$

then in virtue of the periodicity of the continued fraction expansion of the left-hand side, we have

$$x = \cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{5+\cfrac{1}{2+\dots\;+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{4348+x}}}}}}}}$$

$$= \frac{(4348+x)Q_{92}+Q_{91}}{(4348+x)S_{92}+S_{91}} = \frac{Q_{93}+Q_{92}x}{S_{93}+S_{92}x}\,,$$

so that

$$S_{92}x^2 + (S_{93} - Q_{92})x - Q_{93} = 0\,.$$

This quadratic equation with integer coefficients has a positive and a negative root, the positive one being

$$\sqrt{4\,729\,494} - 2174\,.$$

The negative root is necessarily $-\sqrt{4\,729\,494} - 2174$, and consequently,

$$\frac{S_{93} - Q_{92}}{S_{92}} = 4348\,, \qquad \frac{Q_{93}}{S_{92}} = 4\,729\,494 - 2174^2 = 3218\,. \qquad (29)$$

Now, it is a well-known property of any two successive convergents that their numerators and denominators satisfy

$$Q_n S_{n+1} - Q_{n+1} S_n = (-1)^n \qquad (n = 1, 2, 3, \ldots). \tag{30}$$

Combining (29) with (30) applied to $n = 92$, we get :

$$\left(\frac{S_{93}}{S_{92}} - 4348\right) \frac{S_{93}}{S_{92}} = \frac{Q_{92} S_{93}}{S_{92}^2} = \frac{Q_{93} S_{92} + 1}{S_{92}^2} = 3218 + \frac{1}{S_{92}^2}$$

and so, by addition of 2174^2 on both sides,

$$\left(\frac{S_{93}}{S_{92}} - 2174\right)^2 = 2174^2 + 3218 + \frac{1}{S_{92}^2} = \frac{4\,729\,494\, S_{92}^2 + 1}{S_{92}^2}.$$

But,

$$\frac{S_{93}}{S_{92}} - 2174 = \frac{Q_{92}}{S_{92}} + 2174 = \frac{R_{92}}{S_{92}},$$

and therefore

$$\frac{R_{92}^2}{S_{92}^2} = \frac{4\,729\,494\, S_{92}^2 + 1}{S_{92}^2}. \quad \square$$

Equalities like those in (29) may serve to check the correctness of the calculated integer values of Q_{92}, Q_{93}, S_{92} and S_{93}, but the most remarkable check stems from

$$Q_{92} = S_{91} \quad (= 37\,405\,833\,428\,853\,075\,951\,452\,716\,368\,694\,764\,784\,889)$$

which holds on account of the first equality in (29) and the second equality in (26) applied to $n = 93$:

$$S_{93} = 4348\, S_{92} + S_{91}.$$

The preceding proof may be repeated for the continued fraction development of any irrational $\sqrt{A}$ whereby A is an integer. Let

$$\sqrt{A} = [a; b_1, b_2, \ldots, b_h; b_1, b_2, \ldots, b_h; \ldots]$$

and

R_n/S_n be its nth convergent,

then,

– if h is even : $R_h^2 - AS_h^2 = 1$;

– if h is odd : $R_h^2 - AS_h^2 = -1$.

Actually, along the same lines, one can also prove more generally :

$$R_{jh}^2 - AS_{jh}^2 = (-1)^{jh}, \qquad \forall\, j \in \mathbb{N}_0, \quad \forall\, h \in \mathbb{N}_0,$$

so that

– for even h : $R_{jh}^2 - AS_{jh}^2 = 1, \qquad \forall\, j \in \mathbb{N}_0;$ (31)

– for odd h : $R_{jh}^2 - AS_{jh}^2 = (-1)^j, \qquad \forall\, j \in \mathbb{N}_0.$ (31′)

Construction of a General Expression Comprising all Positive Integer Solutions of Eq.(21)

In accordance with (19), the complete set of positive integer solutions of eq.(21) is given by the infinite sequence of couples

$$(R_{92}, S_{92}), (R_{184}, S_{184}), \ldots, (R_{92j}, S_{92j}), \ldots$$

whereby the R's and the S's are generated separately by the recurrence relations in (18) with

$a_1 = a = 2174$, $a_2 = b_1 = 1$, $a_3 = b_2 = 2$, $a_4 = b_3 = 1, \ldots$

$a_{91} = b_{90} = 2$, $a_{92} = b_{91} = 1$, $a_{93} = b_{92} = 4348$, $a_{94} = b_1 = 1$,
$a_{95} = b_2 = 2$, $a_{96} = b_3 = 1, \ldots$ (with periodic repetition),

starting from $R_1 = 2174$, $R_2 = 2175$, $S_1 = 1$, $S_2 = 1$. As indicated previously, the smallest of these solutions (corresponding to a convergent in the *first* period of the continued fraction development of $\sqrt{4\,729\,494}$) is

$$t_1 = R_{92}\,, \qquad u_1 = S_{92}\,,$$

explicitly given by (28)-(28'), and the jth solution can be represented by

$$t_j = R_{92j}\,, \qquad u_j = S_{92j} \qquad (j = 1, 2, 3, \ldots). \tag{32}$$

It is of considerable importance, within the present context, that all these solutions can be described by one couple of formulae with j as unique variable integer parameter.

By repeated application of the first recurrence relation in (18), it is possible to establish a new recurrence relation involving three consecutive t–values, say t_{j+1}, t_j and t_{j-1}. In extenso, the way in which such a relation is established, proceeds as follows. We have first of all 91 linear relations involving 93 R's :

$$\begin{aligned}
&R_{92(j+1)} - R_{92(j+1)-1} - R_{92(j+1)-2} = 0\,,\\
&R_{92(j+1)-1} - 2R_{92(j+1)-2} - R_{92(j+1)-3} = 0\,,\\
&R_{92(j+1)-2} - R_{92(j+1)-3} - R_{92(j+1)-4} = 0\,,\\
&R_{92(j+1)-3} - 5R_{92(j+1)-4} - R_{92(j+1)-5} = 0\,,\\
&\ldots\\
&R_{92j+3} - 2R_{92j+2} - R_{92j+1} = 0\,,\\
&R_{92j+2} - R_{92j+1} - R_{92j} = 0\,.
\end{aligned} \tag{33}$$

Rewriting the last two equations as

$$\begin{aligned}
&R_{92j+3} - 2R_{92j+2} = R_{92j+1}\\
&R_{92j+2} = R_{92j+1} + R_{92j}\,,
\end{aligned}$$

the entire result is now an inhomogeneous system of 91 linear equations with 91 unknowns in the left-hand sides. The determinant of the coefficients of this Cramer system is equal to 1, because the elements below the main diagonal which is composed of ones, are all equal to zero. Thus, solving for $R_{92(j+1)}$ by means of

Cramer's rule, one gets :

$$R_{92(j+1)} = \begin{vmatrix} 0 & -1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -2 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & -5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ R_{92j+1} & 0 & 0 & 0 & 0 & \dots & 1 & -2 \\ R_{92j+1} + R_{92j} & 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{vmatrix}$$

$$= (R_{92j+1} + R_{92j}) \begin{vmatrix} -1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 1 & -2 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 1 & -5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & -1 \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 \end{vmatrix}$$

$$- R_{92j+1} \begin{vmatrix} -1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 1 & -2 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 1 & -5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{vmatrix}$$

$$= R_{92j+1} \begin{vmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & 5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 3 \end{vmatrix}$$

$$+ R_{92j} \begin{vmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & 5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 \end{vmatrix} . \qquad (34)$$

Along the main diagonal of these two determinants (with 90 rows and 90 columns),

one finds the first 89 elements of the period associated with $\sqrt{4\,729\,494}$, followed by the 90th element plus 1 in the case of the first determinant and by the 90th element itself in the case of the second. Each time, the main diagonal is flanked by elements 1 above and -1 below it.

Similarly, one deduces from the first recurrence relation in (18), ninety-one other linear equations involving $R_{92(j-1)}$, $R_{92(j-1)+1}$, $R_{92(j-1)+2}$, $\ldots$, R_{92j-2}, R_{92j-1}, R_{92j}. That system permits the expression of $R_{92(j-1)}$ in terms of R_{92j-1} and R_{92j} :

$$
\begin{aligned}
R_{92(j-1)} &= \begin{vmatrix}
0 & 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\
0 & 1 & 2 & -1 & 0 & \ldots & 0 & 0 \\
0 & 0 & 1 & 1 & -1 & \ldots & 0 & 0 \\
0 & 0 & 0 & 1 & 5 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
R_{92j-1} & 0 & 0 & 0 & 0 & \ldots & 1 & 2 \\
R_{92j} - R_{92j-1} & 0 & 0 & 0 & 0 & \ldots & 0 & 1
\end{vmatrix} \\
&= -R_{92j-1} \begin{vmatrix}
1 & -1 & 0 & 0 & \ldots & 0 & 0 \\
1 & 2 & -1 & 0 & \ldots & 0 & 0 \\
0 & 1 & 1 & -1 & \ldots & 0 & 0 \\
0 & 0 & 1 & 5 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1 & -1 \\
0 & 0 & 0 & 0 & \ldots & 1 & 3
\end{vmatrix} \\
&\quad + R_{92j} \begin{vmatrix}
1 & -1 & 0 & 0 & \ldots & 0 & 0 \\
1 & 2 & -1 & 0 & \ldots & 0 & 0 \\
0 & 1 & 1 & -1 & \ldots & 0 & 0 \\
0 & 0 & 1 & 5 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1 & -1 \\
0 & 0 & 0 & 0 & \ldots & 1 & 2
\end{vmatrix} .
\end{aligned} \tag{35}
$$

If, in the last two determinants, one multiplies consecutively the first row, the first column, the third row, the third column, ..., the 89th row and the 89th column by -1, one obtains the two determinants in the right-hand side of (34). Hence,

adding (34) and (35) side by side yields :

$$R_{92(j+1)} + R_{92(j-1)} = (R_{92j+1} - R_{92j-1}) \begin{vmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 2 & 1 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & -1 & 3 \end{vmatrix}$$

$$+2R_{92j} \begin{vmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 2 & 1 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{vmatrix} . \qquad (36)$$

Finally, there is the relation between R_{92j+1}, R_{92j} and R_{92j-1} which we did not use so far, since it is located right in between the system (33) and the other one from which we deduced (35). We have

$$R_{92j+1} = 4348\, R_{92j} + R_{92j-1} ,$$

which enables us to eliminate $(R_{92j+1} - R_{92j-1})$ from (36). There comes :

$$R_{92(j+1)} + R_{92(j-1)} = \left\{ 4348 \begin{vmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 2 & 1 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & -1 & 3 \end{vmatrix} \right.$$

$$\left. +2 \begin{vmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 2 & 1 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{vmatrix} \right\} R_{92j} , \qquad (37)$$

which may be rewritten as

$$R_{92(j+1)} + R_{92(j-1)} = 2 \begin{vmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -1 & 2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 5 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & -1 & 2174 \end{vmatrix} R_{92j} \quad (38)$$

whereby the determinant now has 92 rows and 92 columns.

Indeed, when one develops this determinant with respect to its last row, one finds:

$$4348 \begin{vmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & -1 & 5 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{vmatrix}$$

$$+2 \begin{vmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & -1 & 5 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{vmatrix}.$$

It now suffices to add the last column to the preceding column in the first of these determinants in order to effectuate a last row of the form 0 0 0 ... 0 0 1 which enables us to delete that last row and also the last column. These operations give the first determinant in (37) with the right coefficient in front. The similar form of the last column in the second determinant again permits us to delete the last row and the last column in that determinant. This then yields the second determinant in (37) also with the right coefficient in front, and so the proof of the equivalence of (37) and (38) is completed. Note that the main diagonal in the determinant appearing in (38) is composed of the first ninety-two elements of the continued

fraction expansion of $\sqrt{4\,729\,494}$ arranged in reversed order. Once more, that main diagonal is flanked by 1's above and -1's below, with all other elements equal to zero.

On account of $R_{92j} = t_j$, (38) is the linear relation connecting t_{j+1}, t_j and t_{j-1} which we announced earlier. The determinant in (38) could be calculated by computer, but this is not necessary since it can also be evaluated algebraically. From the first recurrence relation in (18) in which we insert successively $n = 92, 91, 90, \ldots, 3$, we obtain :

$$\begin{aligned}
R_{92} - R_{91} - R_{90} &= 0\,,\\
R_{91} - 2R_{90} - R_{89} &= 0\,,\\
R_{90} - R_{89} - R_{88} &= 0\,,\\
R_{89} - 5R_{88} - R_{87} &= 0\,,\\
&\cdots\\
R_3 - 2R_2 - R_1 &= 0\,,
\end{aligned}$$

and this system can be complemented with

$$R_2 - R_1 = 1\,, \qquad R_1 = 2174\,.$$

In this way, we have 92 linear equations forming an inhomogeneous system. More precisely, it is a triangular Cramer system which yields for R_{92} :

$$R_{92} = \begin{vmatrix}
0 & -1 & -1 & 0 & 0 & \ldots & 0 & 0 & 0\\
0 & 1 & -2 & -1 & 0 & \ldots & 0 & 0 & 0\\
0 & 0 & 1 & -1 & -1 & \ldots & 0 & 0 & 0\\
0 & 0 & 0 & 1 & -5 & \ldots & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 0 & 0 & \ldots & 1 & -2 & -1\\
1 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 & -1\\
2174 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1
\end{vmatrix}$$

$$= -\begin{vmatrix} -1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & -2 & -1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -5 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 2174 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -1 & 2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 5 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & -1 & 2174 \end{vmatrix}, \tag{39}$$

the last determinant being precisely that occurring in (38).
In conclusion, we derive from (38) and(39) :

$$t_{j+1} - 2t_1 t_j + t_{j-1} = 0 \tag{40}$$

valid not only for $j = 2, 3, \dots$, but also for $j = 1$ if one agrees to set $t_0 = 1$, being in fact the first number in the trivial solution (1,0) of eq.(21). Indeed, (18) holds for $n = 2$ if one agrees upon $R_0 = 1$, $S_0 = 0$ and also for $n = 1$ if $R_{-1} = 0$, $S_{-1} = 1$. In that convention, $t_0 = 1$ and $u_0 = 0$ hold and extend (32).

Eq.(40) may be regarded as a homogeneous linear difference equation of second order with constant coefficients. Such an equation accepts particular solutions of the form $C\lambda^j$ with C an arbitrary proportionality factor and λ a root of the quadratic equation

$$\lambda^2 - 2t_1\lambda + 1 = 0$$

which one obtains after substitution of the proposed expression into (40) and deletion of $C\lambda^{j-1}$. The roots of this quadratic equation are

$$\lambda_\pm = t_1 \pm \sqrt{t_1^2 - 1}$$

or, in virtue of (t_1, u_1) satisfying eq.(21),

$$\lambda_\pm = t_1 \pm u_1\sqrt{4\,729\,494}\,.$$

The theory of homogenous linear difference equations of second order shows that

$$C_+\lambda_+^j + C_-\lambda_-^j \tag{41}$$

with arbitrary proportionality factors C_+ and C_- comprises all solutions of (40), hence among others the solution t_j. Consequently,

$$t_j = C_1\lambda_+^j + C_2\lambda_-^j \qquad (j = 0, 1, 2, \ldots)\,,$$

this time with well-defined C_1 and C_2, determined by

$$C_1 + C_2 = t_0 = 1\,,$$
$$C_1\lambda_+ + C_2\lambda_- = t_1 \;\; (= R_{92})\,.$$

The (unique) solution is $C_1 = C_2 = 1/2$ and so,

$$t_j = \frac{1}{2}\left\{(t_1 + u_1\sqrt{4\,729\,494})^j + (t_1 - u_1\sqrt{4\,729\,494})^j\right\} \quad (j = 0, 1, 2, \ldots)\,. \tag{42}$$

As for the u–numbers, it is not surprising that the outlined method leads to a difference equation with the same coefficients as in (40) :

$$u_{j+1} - 2t_1u_j + u_{j-1} = 0\,, \tag{40'}$$

since the second recurrence relation in (18) contains the same coefficients as those in the first one. Therefore, u_j is also comprised in (41) and so,

$$u_j = C_3\lambda_+^j + C_4\lambda_-^j \qquad (j = 0, 1, 2, \ldots)\,,$$

with

$$C_3 + C_4 = u_0 = 0$$
$$C_3\lambda_+ + C_4\lambda_- = u_1 \;\; (= S_{92})\,.$$

The solution is $C_3 = -C_4 = 1/(2\sqrt{4\,729\,494})$ and in this way, the formula associated with (42) reads

$$u_j = \frac{1}{2\sqrt{4\,729\,494}} \left\{ (t_1 + u_1\sqrt{4\,729\,494})^j - (t_1 - u_1\sqrt{4\,729\,494})^j \right\} \qquad (j = 0, 1, 2, \ldots). \qquad (42')$$

(42) and (42') represent the trivial solution (1,0) (for $j = 0$) and all positive integer solutions of the Pellian equation (21), expressed in terms of the smallest solution (t_1, u_1) given explicitly by (28)-(28'). That smallest solution clearly corresponds to $j = 1$. It is easy to verify that any (t_j, u_j) satisfies (21) :

$$t_j \pm u_j\sqrt{4\,729\,494} = (t_1 \pm u_1\sqrt{4\,729\,494})^j ,$$

$$t_j^2 - 4\,729\,494\,u_j^2 = (t_j + u_j\sqrt{4\,729\,494}) \times (t_j - u_j\sqrt{4\,729\,494})$$

$$= (t_1 + u_1\sqrt{4\,729\,494})^j \times (t_1 - u_1\sqrt{4\,729\,494})^j = (t_1^2 - 4\,729\,494\,u_1^2)^j = 1 .$$

That (42) and (42') represent integers for any $j \in \mathbb{N}_0$ can be seen in two ways :

1) by the use of Newton's binomium :

$$\left.\begin{aligned} t_j &= t_1^j + 4\,729\,494 \binom{j}{2} t_1^{j-2} u_1^2 + 4\,729\,494^2 \binom{j}{4} t_1^{j-4} u_1^4 + \ldots , \\ u_j &= \binom{j}{1} t_1^{j-1} u_1 + 4\,729\,494 \binom{j}{3} t_1^{j-3} u_1^3 + 4\,729\,494^2 \binom{j}{5} t_1^{j-5} u_1^3 + \ldots , \end{aligned}\right\} \qquad (43)$$

whereby ... does not mean an infinite series but the presence of a finite number of further terms. Indeed, t_j is a homogeneous polynomial in t_1 and u_1 consisting of $[j/2] + 1$ terms whereas u_j is also such a polynomial but involving $[(j+1)/2]$ terms;

2) by their expression in terms of the Chebyshev polynomials of the first and the second kind, respectively. Indeed, in virtue of

$$(t_1 + u_1\sqrt{4\,729\,494})(t_1 - u_1\sqrt{4\,729\,494}) = 1 ,$$

one can set

$$t_1 + u_1\sqrt{4\,729\,494} = e^{\theta}$$

and consequently, one has

$$t_1 - u_1\sqrt{4\,729\,494} = \mathrm{e}^{-\theta}\,, \quad \frac{1}{2}(\mathrm{e}^{\theta} + \mathrm{e}^{-\theta}) = t_1 = \cosh\theta\,,$$
$$\frac{1}{2}(\mathrm{e}^{\theta} - \mathrm{e}^{-\theta}) = u_1\sqrt{4\,729\,494} = \sinh\theta\,.$$

Then,

$$\begin{aligned} t_j &= \frac{1}{2}(\mathrm{e}^{j\theta} + \mathrm{e}^{-j\theta}) = \cosh j\theta = T_j(\cosh\theta) = T_j(t_1) \\ &= T_j\,(109\,931\,986\,\ldots\,088\,049)\,, \qquad (j = 1,2,3,\ldots)\,, \end{aligned} \tag{44}$$

as well as

$$\begin{aligned} u_j &= \frac{1}{2\sqrt{4\,729\,494}}(\mathrm{e}^{j\theta} - \mathrm{e}^{-j\theta}) = \frac{\sinh j\theta}{\sqrt{4\,729\,494}} = \frac{\sinh\theta}{\sqrt{4\,729\,494}}\,U_{j-1}(\cosh\theta) \\ &= u_1\,U_{j-1}(t_1) = 50\,549\ldots340\,U_{j-1}\,(109\,931\ldots049), \quad (j = 1,2,3,\ldots)\,. \end{aligned} \tag{44'}$$

Additional Comments on Eq.(40)

Eq.(40) considered within the context of the continued fraction development of $\sqrt{A}$ where A is a non-square positive integer and the Pell equation $t^2 - Au^2 = 1$, is a general result since it holds for any A whereby t_{j-1}, t_j, t_{j+1} represent the t–part of three successive positive integer solutions of the Pell equation arranged in ascending order and t_1 is the t–part of the smallest positive integer solution. Indeed, if one repeats the proof which precedes (40) from the linear system (33) onward with 4 729 494 replaced by A and 92 replaced by the period length h associated with the expansion of $\sqrt{A}$, one obtains

– for even h :

$$R_{(j+1)h} - 2R_h\,R_{jh} + R_{(j-1)h} = 0 \qquad (j = 1,2,3,\ldots)\,,$$

with $R_0 = 1$, and hence eq.(40) if one sets

$$t_0 = R_0 = 1\,, \qquad t_j = R_{jh} \qquad (j = 1,2,3,\ldots)\,;$$

– for odd h :

$$R_{(j+1)h} - 2R_h R_{jh} - R_{(j-1)h} = 0 \qquad (j = 1, 2, 3, \ldots).$$

In this proof, it appears necessary to make use of the properties

$$b_m = b_{h-m} \qquad (m = 1, 2, \ldots, [\frac{h}{2}]), \qquad b_h = 2a$$

with the symbols stemming from (17). That the result for odd h is different from that for even h and therefore does not yield directly eq.(40), is a consequence of (31′) differing from (31). But, inspired by (19′), let us replace j resp. by $2j + 1$ and $2j - 1$. In this manner, we get

$$R_{(2j+2)h} - 2R_h R_{(2j+1)h} - R_{2jh} = 0,$$
$$R_{2jh} - 2R_h R_{(2j-1)h} - R_{(2j-2)h} = 0,$$

and by subtraction,

$$R_{(2j+2)h} - 2R_h(R_{(2j+1)h} - R_{(2j-1)h}) - 2R_{2jh} + R_{(2j-2)h} = 0$$

or

$$R_{(2j+2)h} - 2(2R_h^2 + 1)R_{2jh} + R_{(2j-2)h} = 0.$$

For $j = 1$, the recurrence formula yields

$$R_{2h} - 2R_h^2 - 1 = 0$$

and so, for odd h,

$$R_{(2j+2)h} - 2R_{2h} R_{2jh} + R_{(2j-2)h} = 0 \qquad (j = 1, 2, 3, \ldots).$$

Putting $t_j = R_{2jh}$ yields eq.(40) and in agreement with (19′) and (31′), the R_{2jh}–values are the t–parts of the positive integer solutions of the Pell equation.

Again, because the second equation in (18) contains the same coefficients as the first, entirely similar results are obtained for the S–denominators in the convergents of the continued fraction development of $\sqrt{A}$. There comes

– for even h :

$$S_{(j+1)h} - 2R_h\, S_{jh} + S_{(j-1)h} = 0 \qquad (j = 1, 2, 3, \ldots)\,,$$

with $S_0 = 0$, thus

$$u_{j+1} - 2t_1\, u_j + u_{j-1} = 0$$

whereby

$$u_0 = S_0 = 0\,, \qquad u_j = S_{jh}\,;$$

– for odd h :

$$S_{(2j+2)h} - 2R_{2h}\, S_{2jh} + S_{(2j-2)h} = 0 \qquad (j = 1, 2, 3, \ldots)$$

or again

$$u_{j+1} - 2t_1\, u_j + u_{j-1} = 0$$

whereby

$$u_0 = S_0 = 0\,, \qquad u_j = S_{2jh}\,.$$

As a consequence, the formulae between (40) and (44′) hold if one replaces 4 729 494 by A. Hence, in general, the positive integer solutions of $t^2 - Au^2 = 1$ may be written as

$$\begin{aligned} t_j &= \frac{1}{2}\left\{(t_1 + u_1\sqrt{A})^j + (t_1 - u_1\sqrt{A})^j\right\} \\ &= t_1^j + \binom{j}{2} A t_1^{j-2} u_1^2 + \binom{j}{4} A^2 t_1^{j-4} u_1^4 + \ldots \\ &= T_j(t_1)\,, \\ u_j &= \frac{1}{2\sqrt{A}}\left\{(t_1 + u_1\sqrt{A})^j - (t_1 - u_1\sqrt{A})^j\right\} \\ &= \binom{j}{1} t_1^{j-1} u_1 + \binom{j}{3} A t_1^{j-3} u_1^3 + \binom{j}{5} A^2 t_1^{j-5} u_1^5 + \ldots \\ &= u_1\, U_{j-1}(t_1) \qquad\qquad (j = 1, 2, 3, \ldots)\,, \end{aligned}$$

in which (t_1, u_1) is the smallest positive integer solution of the Pell equation, explicitly given by

$t_1 = R_h\,, \quad u_1 = S_h$ when the period length h is even,

and

$t_1 = R_{2h}\,, \quad u_1 = S_{2h}$ when the period length h is odd,

in other words, to find (t_1, u_1), one has to look for the convergent belonging to $\sqrt{A}$ whose subscript is equal to the length of the smallest even period in the development of $\sqrt{A}$. Indeed, when h is odd, putting together two consecutive smallest periods yields an entity of length $2h$ which is also a period of the continued fraction expansion of $\sqrt{A}$.

Positive Integer Solutions of Eq.(21) with u Divisible by 9314

According to (20) the positive integer values of N to be inserted into (11) in order to find all solutions of the cattle-problem, stem from the positive integer solutions of the Pellian equation (21) whose u-part is divisible by 2.4657. Hence, the remaining problem consists in selecting from the solutions $(t_1, u_1), (t_2, u_2), (t_3, u_3), \ldots$, precisely those for which u is a 9314-fold, in other words, we have to determine the j-exponents in (42′) for which u_j is divisible by 9314.

First, we point out that in every solution (t_j, u_j) the u-part is even. This is a consequence of u_j being divisible by u_1 according to (44′) and u_1 being even. But, without knowing u_1 and the result (44′), one arrives at the same conclusion as follows, for any $j \in \mathbb{N}_0$:

> $t_j^2 - 1 = 4\,729\,494\,u_j^2$ is even $\Rightarrow t_j^2$ is odd $\Rightarrow t_j$ is odd $\Rightarrow t_j - 1$ and $t_j + 1$ are even with a 4-fold among them $\Rightarrow t_j^2 - 1$ is an 8-fold $\Rightarrow u_j^2$ is a 4-fold $\Rightarrow u_j$ is even.

The j-values for which u_j is divisible by the prime number 4657, are in virtue of Amthor's lemmas all the positive integer multiples of some lowest j-value, ρ say ([1], Lehrsatz 5). This is verified by the consideration that $u_{n\rho}$ is divisible by u_ρ on account of $a^n - b^n$ being divisible by $a - b$.

Amthor determined ρ as a consequence of his seven lemmas formulated and proven in [1]. He obtained

$$\rho = 2329$$

being the upward rounded half of the prime number 4657 by which u had to be divisible. In what follows, this result is rederived by means of a mixture of numerical and algebraic methods.

Since very fast computers with very extended memory capacity exist nowadays, determining ρ could be effectuated by calculating $u_2, u_3, \ldots, u_j, \ldots$ and dividing every u by 4657. But this would involve dealing with rather large integer numbers. Indeed,

$$t_1 + u_1\sqrt{4\,729\,494} \cong 2t_1 = 2,198\,639\,73\ldots \times 10^{44},$$
$$t_1 - u_1\sqrt{4\,729\,494} = 0,454\,826\,68\ldots \times 10^{-44},$$

and so, the number of decimal digits of u_j is roughly equal to the integer part of

$$(44 + \log_{10} 2,198\,639\,73\ldots)j - \log_{10} 2 - \frac{1}{2}\log_{10} 4\,729\,494$$

or $44,342\,154\,j - 3,638\,437$.
Thus, for $j = 2329$, the number of decimal digits of u_{2329} exceeds 100 000. Manipulating such large numbers in examining the divisibility of u_j by 4657 is, however, absolutely unnecessary. The artifice consists in calculating "modulo 4657", which means working with the remainders of the division by 4657 instead of with the integers themselves during a sequence of additions, subtractions and multiplications.

Let us reconsider u_j in the polynomial form as in (43) and let us divide the 45–digit number t_1 and the 41–digit number u_1 (see (28)-(28′)) by 4657. We get :

$$t_1 = 4657\,k + \tau \equiv \tau \pmod{4657} \quad \text{with} \quad \tau = 4406,$$
$$u_1 = 4657\,l + \omega \equiv \omega \pmod{4657} \quad \text{with} \quad \omega = 3051, \tag{45}$$

as well as

$$4\,729\,494 = 1015 \times 4657 + 2639\,.$$

Then,

$$t_1^{j-1} u_1 \equiv \tau^{j-1} \omega \pmod{4657},$$
$$4\,729\,494\, t_1^{j-3} u_1^3 \equiv 2639\, \tau^{j-3} \omega^3 \pmod{4657},$$

etc. Consequently,

$$u_j \equiv \frac{1}{2\sqrt{2639}} [(\tau + \omega\sqrt{2639})^j - (\tau - \omega\sqrt{2639})^j] \pmod{4657} \tag{46}$$

and u_j is divisible by 4657 if and only if the remainder appearing in the right-hand side is an integer multiple of 4657, in other terms,

$$u_j \text{ divisible by } 4657 \iff \frac{1}{2\sqrt{2639}} [(\tau + \omega\sqrt{2639})^j - (\tau - \omega\sqrt{2639})^j] \equiv 0 \pmod{4657}\,.$$

Next, we describe the cycle of an iterative process according to which an automatic computer can be programmed to calculate ρ. Let $\Re(q)$ be the abbreviation of

$\ll$ the non-negative remainder of the division of the non-negative integer q by 4657 $\gg$,

so that for any $q \in \mathbb{N}$,

$$0 \leq \Re(q) < 4657\,.$$

Suppose that the case of $j = 2m-1$ (with $m \in \mathbb{N}_0$) did not yield u_j divisible by 4657 and that just before going over to $j = 2m$, the computer has in memory the following vectors :

$$1,\ 2639 \text{ and } \Re(2639^2), \Re(2639^3), \ldots, \Re(2639^{m-1}); \tag{47}$$

$$1 \left(\text{standing for } \begin{pmatrix} 2m-1 \\ 0 \end{pmatrix}\right) \text{ and } \Re\left(\begin{pmatrix} 2m-1 \\ 1 \end{pmatrix}\right), \Re\left(\begin{pmatrix} 2m-1 \\ 2 \end{pmatrix}\right), \ldots, \Re\left(\begin{pmatrix} 2m-1 \\ 2m-2 \end{pmatrix}\right), \text{ as well as } 1 \left(\text{representing } \begin{pmatrix} 2m-1 \\ 2m-1 \end{pmatrix}\right); \tag{48}$$

$$\Re(\tau^{2m-2}\omega),\ \Re(\tau^{2m-4}\omega^3), \ldots,\ \Re(\omega^{2m-1})\,. \tag{49}$$

Each of these non-negative integers requires two bytes of storage. The case of $j = 2m$ necessitates the consideration of

$$\binom{2m}{1}\tau^{2m-1}\omega + 2639\binom{2m}{3}\tau^{2m-3}\omega^3 + 2639^2\binom{2m}{5}\tau^{2m-5}\omega^5 + \ldots + 2639^{m-1}\binom{2m}{2m-1}\tau\omega^{2m-1}. \tag{50}$$

One applies the well-known rule

$$\binom{2m}{r} = \binom{2m-1}{r-1} + \binom{2m-1}{r}$$

to calculate $\Re\left(\binom{2m}{r}\right)$ for any $r \in \{1, 2, \ldots, 2m-1\}$, i.e.,

$$\Re\left(\binom{2m}{r}\right) = \begin{cases} \Re\left(\binom{2m-1}{r-1}\right) + \Re\left(\binom{2m-1}{r}\right) & \text{if this sum is } < 4657, \\ \Re\left(\binom{2m-1}{r-1}\right) + \Re\left(\binom{2m-1}{r}\right) - 4657 & \text{if the sum of the first two terms is } \geq 4657, \end{cases}$$

and one lets the vector (48) be replaced by

$$1, \Re\left(\binom{2m}{1}\right), \Re\left(\binom{2m}{2}\right), \ldots, \Re\left(\binom{2m}{2m-1}\right), 1 \tag{51}$$

which is one component longer than (48). Next, one calculates

$$\Re\left(\Re\left(\binom{2m}{1}\right) \times \Re\left(\tau^{2m-2}\omega\right)\right),$$

$$\Re\left(2639 \times \Re\left(\binom{2m}{3}\right) \times \Re\left(\tau^{2m-4}\omega^3\right)\right),$$

$$\Re\left(\Re(2639^2) \times \Re\left(\binom{2m}{5}\right) \times \Re(\tau^{2m-6}\omega^5)\right),$$

$$\ldots$$

$$\Re\left(\Re(2639^{m-1}) \times \Re\left(\binom{2m}{2m-1}\right) \times \Re(\omega^{2m-1})\right). \tag{52}$$

These are m integers ≥ 0 and < 4657. According to (46), u_{2m} and (50) differ from each other by an integer multiple of 4657. To examine whether u_{2m} is divisible

by 4657 or not, it suffices to calculate the sum of the m quantities in (52) and to divide it by 4657. Indeed, that sum differs by an integer multiple of 4657 from the sum (50) after the factor τ has been split off and divisibility of (50) by the prime 4657 cannot be caused or helped by the presence of the factor τ (= 4406). So, if the sum of the terms in (52) happens to be an integer multiple of 4657, then $\rho = 2m$ and the machine can stop the computations. Note that for j of the order of magnitude 4000 which means m of the order of 2000, the sum of the terms in (52) still yields a number of at most seven decimal digits.

If, on the contrary, $j = 2m$ does not yield divisibility by 4657, one lets the machine proceed to $j = 2m + 1$. Then, the right-hand side in (46), after leaving out (mod 4657), is

$$\binom{2m+1}{1}\tau^{2m}\omega + 2639\binom{2m+1}{3}\tau^{2m-2}\omega^3 + 2639^2\binom{2m+1}{5}\tau^{2m-4}\omega^5 + \ldots$$
$$+2639^m\binom{2m+1}{2m+1}\omega^{2m+1}. \qquad (53)$$

One applies the same procedure as in the case of $j = 2m$ to generate $\Re\left(\binom{2m+1}{r}\right)$ where $r = 1, 2, \ldots, 2m$ and one lets the vector

$$1, \Re\left(\binom{2m+1}{1}\right), \Re\left(\binom{2m+1}{2}\right), \ldots, \Re\left(\binom{2m+1}{2m}\right), 1$$

replace (51) on the understanding that one integer memory unit is added. After that, one calculates

$$\Re\left(\Re\left(\binom{2m+1}{1}\right) \times \Re(\tau^{2m-2}\omega)\right),$$
$$\Re\left(2639 \times \Re\left(\binom{2m+1}{3}\right) \times \Re(\tau^{2m-4}\omega^3)\right),$$
$$\Re\left(\Re(2639^2) \times \Re\left(\binom{2m+1}{5}\right) \times \Re(\tau^{2m-6}\omega^5)\right),$$
$$\ldots$$
$$\Re\left(\Re(2639^{m-1}) \times \Re\left(\binom{2m+1}{2m-1}\right) \times \Re(\omega^{2m-1})\right),$$

obtains their sum, multiplies that sum by 2460 (which is $\Re(\tau^2)$ because $\tau^2 = 4406^2 = 4168 \times 4657 + 2460$) and determines the remainder of the division of the

product by 4657. Let the result be σ. It differs from the sum of the first m terms in (53) by an integer multiple of 4657. Finally, to take the last term in (53) into account, one lets the machine compute

- $\Re(2639 \times \Re(2639^{m-1}))$
  which yields $\Re(2639^m)$. One lets this value join the vector (47);
- $\Re(3915 \times \Re(\omega^{2m-1}))$ (54)
  which yields $\Re(\omega^{2m+1})$ because $\omega^2 = 3051^2 = 1998 \times 4657 + 3915$;
- $\Re(\Re(2639^m) \times \Re(\omega^{2m+1})) = \nu$
- $\Re(\sigma + \nu)$ which is the remainder of the division of (53) by 4657.

If $\Re(\sigma+\nu) = 0$, divisibility of u_{2m+1} is attained, $\rho = 2m+1$ and the computer can stop the calculations. If, however, $\Re(\sigma+\nu) > 0$, then the machine must proceed to $j = 2m+2$, but first it has to adapt the vector (49) to the following cycle. It has to calculate $\Re(\tau^{2m}\omega)$, $\Re(\tau^{2m-2}\omega^3)$, ..., $\Re(\tau^2\omega^{2m-1})$ by means of the components of the vector still stored in (49). It is clear that for $r = 0, 1, 2, \ldots, m-1$,

$$\Re(\tau^{2m-2r}\omega^{2r+1}) = \Re(2460 \times \Re(\tau^{2m-2r-2}\omega^{2r+1}))$$

since $\Re(\tau^2) = 2460$, and these m values are made to replace those in (49). Complementing this new vector with $\Re(\omega^{2m+1})$ already calculated (see (54)), the iterative cycle which started with $j = 2m-1$ is completed since the vectors (47), (48) and (49) are now in memory with m replaced by $m+1$.

Because u_1 is not an integer multiple of 4657 (see (45)), the described iterative cycle which permits the transition to $j = 2,3$, $j = 4,5$, $j = 6,7$, ..., may be started at $j = 1$, in other terms, $m = 1$. The Siemens 7570 computer of the Central Digital Computing Laboratory of the University of Ghent has needed less than one minute to yield

$$\rho = j_{min} = 2329 \tag{55}$$

in agreement with Amthor's result, after having tested all the smaller j–values $(2 \leq j \leq 2328 \Rightarrow 4657 \nmid u_j)$.

In order to explain why

$$2329 = \frac{1}{2}(4657 + 1) \tag{56}$$

in an elementary way, let us consider

$$\begin{aligned} u_{4658} &= \frac{1}{2\sqrt{4\,729\,494}}\left[(t_1 + u_1\sqrt{4\,729\,494})^{4658} - (t_1 - u_1\sqrt{4\,729\,494})^{4658}\right] \\ &\equiv \frac{1}{2\sqrt{2639}}\left[(\tau + \omega\sqrt{2639})^{4658} - (\tau - \omega\sqrt{2639})^{4658}\right] \pmod{4657} \end{aligned} \tag{57}$$

or

$$\begin{aligned} u_{4658} \equiv \Bigg\{ &\binom{4658}{1}\tau^{4657}\omega + 2639\binom{4658}{3}\tau^{4655}\omega^3 + \dots \\ &+ 2639^{2327}\binom{4658}{4655}\tau^3\omega^{4655} + 2639^{2328}\binom{4658}{4657}\tau\omega^{4657}\Bigg\} \pmod{4657}. \end{aligned} \tag{57'}$$

The binomial coefficients

$$\binom{4658}{3}, \binom{4658}{5}, \dots \binom{4658}{4655}, \tag{58}$$

are integers resulting from the ratio of two products whereby the numerator product contains the prime factor 4657 which cannot be cancelled against one or more factors in the denominator product. Therefore, all terms of the sum between braces in (57') except the first and the last one, are integer multiples of 4657 and so,

$$\begin{aligned} u_{4658} &\equiv (4658\tau^{4657}\omega + 4658(2639)^{2328}\tau\omega^{4657}) \pmod{4657} \\ &\equiv \left\{\tau\omega(\tau^{4656} + 2639^{2328}\omega^{4656})\right\} \pmod{4657}. \end{aligned}$$

Now, we can make use of the well-known minor theorem of Fermat :

$$a^{p-1} - 1 = \text{an integer multiple of } p \ (\text{or} \equiv 0 \pmod{p}) \tag{59}$$

when a is a non-zero integer and p a (positive) prime number which is not a divisor of a. Setting respectively $a = \tau$ and $a = \omega$, with $p = 4657$, we have :

$$\tau^{4656} - 1 = 4406^{4656} - 1 = \text{a multiple of } 4657,$$
$$\omega^{4656} - 1 = 3051^{4656} - 1 = \text{a multiple of } 4657,$$

or equivalently,

$$\tau^{4656} \equiv 1 \pmod{4657}, \tag{60}$$
$$\omega^{4656} \equiv 1 \pmod{4657}. \tag{61}$$

As far as 2639^{2328} is concerned, we deduce from the application of (59) with $a = 2639$ and $p = 4657$:

$$2639^{4656} - 1 = (2639^{2328} - 1)(2639^{2328} + 1) = \text{a multiple of } 4657.$$

Since 4657 is a prime number, divisibility stems a priori either from the first factor or from the second. That in reality it is from the second factor, can be seen as follows :

$$\begin{aligned}
2639^{2328} &= (2639^2)^{1164} \equiv 2106^{1164} \pmod{4657} \\
&= (2106^2)^{582} \pmod{4657} = 1772^{582} \pmod{4657} \\
&= (1772^2)^{291} \pmod{4657} \equiv 1166^{291} \pmod{4657} \\
&= (1166^3)^{97} \pmod{4657} \equiv 4153^{97} \pmod{4657} \\
&\equiv 4153 \times 2538^{48} \pmod{4657} \equiv 4153 \times 813^{24} \pmod{4657} \\
&\equiv 4153 \times 4332^{12} \pmod{4657} \equiv 4153 \times 3171^{6} \pmod{4657} \\
&\equiv 4153 \times 778^{3} \pmod{4657} \equiv 4153 \times 4426 \pmod{4657} \\
&\equiv 4656 \pmod{4657}.
\end{aligned}$$

Thus,

$$2639^{2328} \equiv -1 \pmod{4657} \tag{62}$$

and

$$u_{4658} \equiv \{4406.3051[1+(-1)1]\}(\text{mod } 4657) \equiv 0 \ (\text{mod } 4657). \tag{63}$$

Furthermore, to examine whether or not $j = 4658$ is the smallest exponent in (42′) for which 4657 is a divisor of u_j, we combine (63) with (57) and deduce from this combination :

$$[(\tau+\omega\sqrt{2639})^{2329}+(\tau-\omega\sqrt{2639})^{2329}] \times\frac{1}{2\sqrt{2639}}[(\tau+\omega\sqrt{2639})^{2329}-(\tau-\omega\sqrt{2639})^{2329}] \equiv 0 \ (\text{mod } 4657). \tag{64}$$

It can be shown ad absurdum that the divisibility of the left-hand side by 4657 cannot be ascribed to the first factor. Indeed, if this factor were a multiple of 4657, one would have by squaring

$$(\tau+\omega\sqrt{2639})^{4658}+(\tau-\omega\sqrt{2639})^{4658}+2(\tau^2-2639\omega^2)^{2329} \equiv 0 \ (\text{mod } 4657).$$

But,

$$\begin{aligned}\tau^2-2639\omega^2 &= 4406^2-2639\times 3051^2 \equiv (2460-2639\times 3915)(\text{mod } 4657)\\ &\equiv (2460-2459)(\text{mod } 4657)=1 \ (\text{mod } 4657)\end{aligned} \tag{65}$$

and so,

$$(\tau+\omega\sqrt{2639})^{4658}+(\tau-\omega\sqrt{2639})^{4658} \equiv -2 \ (\text{mod } 4657) \tag{66}$$

or

$$\begin{aligned}&\tau^{4658}+2639\binom{4658}{2}\tau^{4656}\omega^2+2639^2\binom{4658}{4}\tau^{4654}\omega^4+\dots\\ &+2639^{2328}\binom{4658}{4656}\tau^2\omega^{4656}+2639^{2329}\omega^{4658} \equiv -1 \ (\text{mod } 4657).\end{aligned} \tag{66′}$$

The binomial coefficients

$$\binom{4658}{2}, \binom{4658}{4}, \dots, \binom{4658}{4656}$$

being just as those in (58) integers which are divisible by 4657, the preceding congruence reduces to

$$\tau^{4658} + 2639^{2329}\omega^{4658} \equiv -1 \pmod{4657} \tag{67}$$

or, in virtue of (60), (61) and (62),

$$\tau^2(1 + \text{mult. of } 4657) + 2639(-1 + \text{mult. of } 4657)$$
$$\times\omega^2(1 + \text{mult. of } 4657) \equiv -1 \pmod{4657}. \tag{68}$$

Thus,

$$\text{the hypothesis } [(\tau + \omega\sqrt{2639})^{2329} + (\tau - \omega\sqrt{2639})^{2329}] \equiv 0 \pmod{4657}$$
$$\Longrightarrow \tau^2 - 2639\omega^2 \equiv -1 \pmod{4657} \tag{69}$$

and this is in contradiction with (65). The contradiction excludes the hypothesis. If, on the contrary, the second factor in (64) is assumed to be divisible by 4657 and one repeats the previous reasoning, one obtains (66)–(69) but with plus signs replacing the minus signs in the right-hand sides and so, there is agreement with (65), this time. In conclusion,

$$\frac{1}{2\sqrt{2639}}[(\tau + \omega\sqrt{2639})^{2329} - (\tau - \omega\sqrt{2639})^{2329}] \equiv 0 \pmod{4657}$$

which entails

$$u_{2329} \equiv 0 \pmod{4657} \tag{70}$$

on account of (46) applied to $j = 2329$. This j-value being odd, one cannot repeat exactly the same reasoning in order to halve the subscript of u once more. But,

$$2329 = 17 \times 137$$

and therefore, in

$$(\tau + \omega\sqrt{2639})^{2329} - (\tau - \omega\sqrt{2639})^{2329}$$

one could split off, as a factor, either

$$(\tau + \omega\sqrt{2639})^{17} - (\tau - \omega\sqrt{2639})^{17}$$

or

$$(\tau + \omega\sqrt{2639})^{137} - (\tau - \omega\sqrt{2639})^{137}.$$

Such factors, divided by $2\sqrt{2639}$, could a priori be the cause of 4657 being a divisor of u_{17} or u_{137}. That this is not the case stems from the fact that the computer which carried out the iterative process described above, did not stop at $j = 17$ and $j = 137$, but proceeded as far as $j = 2329$ before coming to a halt. That the machine in its search for a u–value which is divisible by 4657 would only need carry out a finite number of cycles was predictable on account of our direct proof that 4657 is a divisor of u_{4658}. Deriving the divisibility of u_{2329} by 4657 from that property constitutes an explanation of (56).

The Positive Integer Solutions of the System of Equations (1)

The complete set of positive integer solutions of the Pell equation (21) whereby the u–part is a multiple of 9314, is described by

$$\begin{aligned} t_{2329n} &= T_{2329n}(t_1) \\ &= \frac{1}{2}\left\{(t_{2329} + u_{2329}\sqrt{4\,729\,494})^n + (t_{2329} - u_{2329}\sqrt{4\,729\,494})^n\right\} \\ &= \frac{1}{2}\left\{(t_1 + u_1\sqrt{4\,729\,494})^{2329n} + (t_1 - u_1\sqrt{4\,729\,494})^{2329n}\right\}, \end{aligned} \tag{71}$$

$$\begin{aligned} u_{2329n} &= u_1 U_{2329n-1}(t_1) \\ &= \frac{1}{2\sqrt{4\,729\,494}}\left\{(t_{2329} + u_{2329}\sqrt{4\,729\,494})^n - (t_{2329} - u_{2329}\sqrt{4\,729\,494})^n\right\} \\ &= \frac{1}{2\sqrt{4\,729\,494}}\left\{(t_1 + u_1\sqrt{4\,729\,494})^{2329n} - (t_1 - u_1\sqrt{4\,729\,494})^{2329n}\right\} \\ &\qquad\qquad (n = 1, 2, 3, \ldots). \end{aligned} \tag{71'}$$

Setting

$$M_n = t_{2329n}, \quad N_n = \frac{u_{2329n}}{9314} (\in \mathbb{N}_0) \qquad (n = 1, 2, 3, \ldots),$$

one finds at the same time the complete set of positive integer solutions of the original Pell equation (14) since

$$M_n^2 - 410\,286\,423\,278\,424\,N_n^2 = t_{2329n}^2 - 4\,729\,494\,u_{2329n}^2 = 1 \qquad (n = 1, 2, 3, \ldots).$$

Inserting the integers

$$\left(\frac{u_{2329n}}{9314}\right)^2 \qquad (n = 1, 2, 3, \ldots)$$

into the place of N^2 in (11), we obtain all positive integer solutions of the system (1). The smallest among them corresponds to choosing $n = 1$:

$$W = 46\,200\,808\,287\,018\,\frac{u_{2329}^2}{9314^2}$$
$$X = \ldots$$
$$\ldots$$

The number of decimal digits in each of the eight integer values composing this smallest solution, as well as some twenty initial digits, may be calculated by the use of decimal logarithms. Knowing that t_1 consists of 45 digits (cfr.(28)), we have

$$\begin{aligned} t_1 + u_1\sqrt{4\,729\,494} &= t_1 + \sqrt{t_1^2 - 1} = 2t_1 - \frac{1}{2t_1} - \frac{1}{8t_1^3} - \ldots \\ &= 2t_1 - O(10^{-44}), \\ t_1 - u_1\sqrt{4\,729\,494} &= \frac{1}{2t_1} + \frac{1}{8t_1^3} + \ldots = O(10^{-44}) \end{aligned}$$

and so,

$$\begin{aligned} u_{2329} &= \frac{1}{2\sqrt{4\,729\,494}}\left[\left(2t_1 - \frac{1}{2t_1} - \ldots\right)^{2329} - \left(\frac{1}{2t_1} + \ldots\right)^{2329}\right] \\ &= \frac{(2t_1)^{2329}}{2\sqrt{4\,729\,494}}\left[1 - \frac{2329}{4t_1^2} + O(t_1^{-4})\right] \\ &= \frac{(2t_1)^{2329}}{2\sqrt{4\,729\,494}}\left[1 - O(0,5 \times 10^{-85})\right]. \end{aligned}$$

Thus, to attain a precision of the order of twenty decimal digits, approximating u_{2329} by the first term is sufficient. There comes :

$$\begin{aligned} \log_{10} u_{2329} \cong\ & 2329(44 + \log_{10} 2,198\,639\,734\,656\ldots) - \log_{10} 2 \\ & -\frac{1}{2}\log_{10} 4\,729\,494 = 103\,269,238\,397\,377\,801\,537\,626\,168\ldots \end{aligned}$$

and

$$u_{2329} \cong 1,731\,399\,858\,951\,771\,056\,429\,417 \times 10^{103269},$$
$$N^2 \cong \left(\frac{u_{2329}}{9314}\right)^2 = 3,455\,590\,635\,455\,937\,050\,630\,380 \times 10^{206530}.$$

Upon insertion into (11), the results are

$$\left.\begin{aligned} W &= 159\,651\,080\,467\,114\,453\,143 \\ X &= 114\,897\,138\,772\,828\,999\,971 \\ Y &= 113\,319\,275\,443\,863\,807\,711 \\ Z &= 63\,903\,464\,823\,090\,286\,500 \\ w &= 110\,982\,989\,237\,331\,903\,972 \\ x &= 75\,359\,414\,205\,454\,263\,981 \\ y &= 54\,146\,089\,457\,145\,667\,802 \\ z &= 83\,767\,688\,241\,852\,443\,869 \end{aligned}\right\} \langle 206\,524 \rangle$$

and the total number of cattle in the herd

$$= 776\,027\,140\,648\,681\,826\,953\ \langle 206\,524 \rangle ,$$

where $\langle 206\,524 \rangle$ means that there are 206 524 more digits to follow (a notation used by Amthor and by Heath).

In his paper [1], Amthor obtained final results for W and the total number of cattle which are in error in the fourth decimal position, i.e.,

$$W = 1598\ \langle 206\,541 \rangle , \qquad \text{total} = 7766\ \langle 206\,541 \rangle ,$$

because he used Briggs logarithms with only four digits following the comma in the last steps of his calculations which is rather hard to understand. In contrast, the total number of digits is correct and agrees with our results :

W, X, Y, w and the sum each involving 206 545 digits, and

Z, x, y, z each consisting of 206 544 digits,

as far as the smallest solution of the system (1) is concerned. It is almost unbelievable that a system of equations of such simplicity as (1) ultimately leads to solutions of colossal magnitude.

Calculating all the digits of $W, X, \ldots, z$ in the smallest solution as well as the sum of these eight numbers, is a task which falls within the capability of some modern computers, but such an undertaking would be nothing more than a stunt. If it were carried out, nonetheless, writing out the nine numbers would require several hundreds of pages. Indeed, the nine integer values involve in total 1 858 901 decimal digits, and if every page were filled with fifty lines of fifty digits each, printing the smallest solution of (1) would demand a book of 744 pages.

ACKNOWLEDGMENTS

The authors wish to thank Ret. Professor Dr. A.J. Rutgers for having drawn their attention on the cattle–problem of Archimedes. They are also indebted to Prof. Dr. C. Vanhelleputte and Dr. C. Coolsaet for helping them find several mathematical and historical papers in the literature dealing with various aspects of the problem. One of the authors (H.D.M.) is indebted to the National Fund for Scientific Research (N.F.W.O. Belgium) for its permanent financial support.

REFERENCES

1. Amthor A., "Das Problema bovinum des Archimedes," *Zeitschrift Math. Phys., hist.-lit.- Abth.*, **25**, 153–171 (1880).

2. *Archimedes opera*, ed. Heiberg J.L., **2**, 450–455 (1881); new ed., **2**, 528–534 (1913).

3. Bell A.H., *Math. Magazine*, Washington, **2**, 163–164 (1895).

4. Bell A.H., *Amer. Math. Monthly*, **2**, 140–141 (1895).

5. Dixon L.E., *Introduction to the Theory of Numbers*, Dover Publ. Inc., New York, **2**, 342–345 (1957).

6. Heath T.L., *The works of Archimedes, with the method of Archimedes*, Cambridge Univ. Press (1897); re–ed. by Dover Publ. Inc., New York, 319–326.

7. Krumbiegel B.,"Das Problema bovinum des Archimedes," *Zeitschrift Math. Phys., hist.-lit. Abth.*, **25**, 121–136 (1880).

8. Lessing G.E., *Zur Geschichte der Literatur*, Braunschweig, **2**, No.13, 421–446 (1773).

9. Meyer C.F., *Ein diophantisches Problem*, Progr., Potsdam, 14 pp. (1867).

10. Wurm J.F., *Jahrbücher für Philologie und Paedagogik*, ed. Jahn J. C., **14**, 194–202 (1830).

Carl C. Grosjean and Hans E. De Meyer
Laboratorium voor Numerieke Wiskunde en Informatica
Rijksuniversiteit-Gent
Krijgslaan 281, S–9
B-9000 Gent, Belgium

Constantin Carathéodory: An International Tribute (pp. 454–460)
edited by Th. M. Rassias

ON NONLINEAR MONOTONE OPERATORS WITH VALUES IN L(X,Y)

N. Hadjisavvas, D. Kravvaritis and G. Pantelidis

1. Introduction

Let X be a locally convex Hausdorff space and X^* its topological dual with $\langle x^*, x\rangle$ written instead of $x^*(x)$.

A multivalued mapping $T : X \to X^*$ is called a monotone operator if

$$\langle x_1^* - x_2^*, x_1 - x_2\rangle \geq 0$$

for all $x_i \in D(T)$ and $x_i^* \in T(x_i), i = 1, 2$.

The properties and the applications of these operators have been discussed in detail in several monographs (cf. [1], [14], [16]). Among other properties, it is known that in case X is a Frechét space, a monotone operator is locally bounded in the interior points of its domain.

In recent years the notion of monotone operator has been extended by replacing the dual $X^* = L(X, \mathbb{R})$ by the space $L(X, Y)$, where Y is an ordered topological vector space (cf. [3], [5], [6], [7]). An important class of such monotone operators consists of the subdifferentials of convex mappings from X into Y ([8], [9], [10], [15], [17]).

In this work we establish various versions of local boundedness of monotone operators when X is a Frechét space and Y a normed lattice. We also discuss the special case of the subdifferential of the indicator function of a convex body.

2. Preliminaries

Let X be a Frechét space and Y a normed lattice. We endow the space $L = L(X, Y)$ of all linear and continuous mappings from X into Y with the topology of simple convergence.

Let T be a nonlinear multivalued operator from X into L. The effective domain and the graph of T are defined as the sets

$$D(T) = \{x \in X : T(x) \neq \emptyset\}$$

and

$$G(T) = \{(x, A) : x \in D(T), A \in T(x)\},$$

respectively. T is said to be *monotone* if

$$(A_1 - A_2)(x_1 - x_2) \geq 0 \quad \text{for all} \quad (x_i, A_i) \in G(T),\ i = 1, 2.$$

A monotone operator is called *maximal* if its graph is not properly contained in the graph of any other monotone operator. An operator $T : X \to L$ is said to be *locally bounded* at $x_0 \in D(T)$ if there exists a neighborhood V of x_0 such that the set

$$T(V) = \cup\{T(x) : x \in V\}$$

is bounded in $L(X, Y)$.

3. Main Results

When $Y = \mathbb{R}$, it is known that a monotone operator is bounded at any interior point of its domain ([2], [11]). The same result is known to hold when X is a Banach space with $D(T) = X$ and Y a normed lattice [6]. The following theorem establishes the local boundedness of monotone operators in our more general setting.

Theorem 1. A monotone operator $T : X \to L$ is locally bounded at all interior points of its domain.

Proof. Suppose T is not locally bounded at $x' \in \operatorname{int} D(T)$. Without loss of generality, we may assume that $x' = 0$. Let d be a metric defining the topology of X. If $U_n = \{x \in X : d(0, x) < \frac{1}{n}\}$, then $T(U_n)$ is not bounded.

Hence, there exists $x_n' \in U_n$ and $A_n \in T(U_n)$ such that $\|A_n x_n'\| > n$. Therefore, $\{A_n\}$ is not equicontinuous. Let $x_n \in U_n$ be such that $A_n \in T(x_n)$. Set $a_n = \max\{1, \|A_n x_n\|\}$ and $B_n = A_n/a_n$. Then $\{B_n\}$ is not equicontinuous, hence not bounded ([13, p. 83]). It follows that there exists $x_0 \in X$ such that $\{\|B_n x_0\|\}$ is not bounded. Now let $\lambda > 0$ be such that $\pm z_0 \in D(T)$, where $z_0 = \lambda x_0$. For $A_0 \in T(z_0), A_0' \in T(-z_0)$ we have

$$(A_n - A_0)(x_n - z_0) \geq 0$$
$$(A_n - A_0')(x_n + z_0) \geq 0\,.$$

so

$$B_n z_0 \leq \frac{A_0}{a_n}(z_0 - x_n) + B_n x_n := u_n$$
$$-B_n z_0 \leq -\frac{A_0'}{a_n}(z_0 + x_n) + B_n x_n := v_n\,.$$

It then follows that

$$\|B_n z_0\| \leq \|u_n\| + \|v_n\|\,.$$

However, it is easy to see that the sequences $\{\|u_n\|\}, \{\|v_n\|\}$ are bounded while $\{\|B_n z_0\|\}$ is not, a contradiction.

Actually, the theorem states that if T is a monotone operator and $x_0 \in \operatorname{int} D(T)$, then there exists a neighborhood U of x_0 such that for all $x \in X$ one has

$$\sup\{\|Ax\|, x' \in U, A \in T(x')\} < \infty\,.$$

When Y is order complete, one can also prove:

Proposition 2. Let Y be order complete and $T : X \to L$ be a monotone operator. Then for any $x_0 \in \operatorname{int} D(T)$ there exists a neighborhood U of x_0 such that

$$\sup_{x,x' \in U} \| \sup_{A \in T(x)} Ax'\| < \infty\,.$$

Proof. Assume that $x_0 = 0$. Let $\mathcal{B}$ be a neighborhood base of 0 in X. By theorem 1, there exists $U_1 \in \mathcal{B}$ such that $T(U_1)$ is bounded, hence it is equicontinuous. So, there exists $U_2 \in \mathcal{B}$ such that $\|Bx\| < 1$ for all $x \in U_2$

and $B \in T(U_1)$. Let $U \in \mathcal{B}$ be circled and such that $U+U \subset U_1 \cap U_2 \cap D(T)$. Given $x, x' \in U$, fix $B_1 \in T(x+x'), B_2 \in T(x-x')$. By the monotonicity of T, one has for all $A \in T(x)$

$$(B_1 - A)(x + x' - x) \geq 0\,, (B_2 - A)(x - x' - x) \geq 0\,,$$

hence

$$B_2 x' \leq Ax' \leq B_1 x'\,.$$

Thus one obtains

$$B_2 x' \leq \sup_{A \in T(x)} Ax' \leq B_1 x'$$

and

$$\| \sup_{A \in T(x)} Ax' \| \leq \|B_1 x'\| + \|B_2 x'\| \leq 2\,.$$

Consequently,

$$\sup_{x, x' \in U} \| \sup_{A \in T(x)} Ax' \| < \infty\,.$$

These boundedness properties may not hold for other points of $D(T)$ as shown by

Proposition 3. ([4]) Let $T : X \to L$ be a maximal monotone operator. If $\operatorname{int}(\overline{\operatorname{co}} D(T)) \neq \emptyset$ and $x_0 \in D(T) \backslash \operatorname{int}(\overline{\operatorname{co}} D(T))$, then $T(x_0)$ contains at least a half line.

An important class of monotone operators consists of the subdifferentials of convex vector valued mappings. A mapping $F : X \to Y \cup \{+\infty\}$ is called *convex* if

$$F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y)$$

for all $x, y \in X$ and $0 \leq \lambda \leq 1$. The effective domain of F is, by definition, the set $D(F) = \{x \in X : F(x) \in Y\}$ which we suppose to be nonempty. The subdifferential of F at $x_0 \in X$ is the set

$$\partial F(x_0) = \{A \in L : A(x - x_0) + F(x_0) \leq F(x)\,, \text{for all}\, x \in X\}\,.$$

The subdifferential operator ∂F is monotone. When $D(\partial F) = X$, it is also a maximal monotone operator [5].

It is not known yet, whether ∂F is a maximal monotone operator whenever $D(\partial F) \neq X$. However, one can prove that if F is the indicator function of a convex body, then ∂F is a maximal monotone operator. This case is a basic tool for the study of monotone operators [12].

Proposition 4. Let M be a convex body in X and $F : X \to Y$ be the mapping defined by

$$F(x) = \begin{cases} 0, & x \in M \\ +\infty, & x \notin M \end{cases}.$$

Then $T := \partial F$ is a maximal monotone operator.

Proof. We may assume that $0 \in \operatorname{int} M$. The gauge p of M is a sublinear function. It is not difficult to see that

$$T(x) = \begin{cases} \{0\}\,, & p(x) < 1 \\ \{A \in L : Ax' \leq Ax\,, \text{ for all } x' \in M\}\,, & p(x) = 1 \\ \emptyset\,, & p(x) > 1\,. \end{cases}$$

In addition, when $p(x) = 1, T(x)$ is a cone containing non-zero elements. Indeed, there exists $x^* \in X^* \backslash \{0\}$ such that $\langle x^*, x' \rangle \leq \langle x^*, x \rangle$ for all $x' \in M$. If $y \in Y^+$, then $A \in T(x)$, where A is defined by

$$Ax' = \langle x^*, x' \rangle y\,, x' \in X\,.$$

In order to prove that T is a maximal monotone operator, let

$$(A_0 - A)(x_0 - x) \geq 0 \quad \text{for all} \quad (x, A) \in G(T)\,. \tag{1}$$

First, we prove that $p(x_0) \leq 1$. If this is not the case, we may take $x = \frac{x_0}{p(x_0)}$ and $A \in T(x) \backslash \{0\}$. Then $A \in T(x)$ for all $\lambda > 0$, hence

$$\lambda A(x_0 - x) \leq A_0(x_0 - x)$$

from which it follows that $A(x_0 - x) \leq 0$, i.e., $Ax_0 \leq 0$. On the other hand, for all $x' \in M$ we have

$$Ax' \leq Ax \leq 0\,,$$

which leads to $A = 0$, a contradiction. Hence $p(x_0) \leq 1$. Now for $x \in M$ we have $0 \in T(x)$. Therefore (1) implies

$$A_0(x_0 - x) \geq 0$$

i.e.,

$$A_0(x - x_0) + F(x_0) \leq F(x)\,, \text{for} \quad x \in X\,.$$

Hence $A_0 \in T(x_0)$ and T is a maximal monotone operator.

References

1. H. Brezis, *Operateurs Maximaux Monotones et Semigroups, de Contractions dans les Espaces de Hilbert*, North Holland, Amsterdam, 1973.
2. P. M. Fitzpatrick, P. Hess and T. Kato, *Local boundedness of monotone-type operators*, Proc. Japan Acad. **48** (1972), 275–277.
3. N. Hadjisavvas, D. Kravvaritis, G. Pantelidis, and I. Polyrakis, *Nonlinear monotone operators with values in $L(X, Y)$*, J. Math. Anal. Appl. **140** (1989), 83–94.
4. N. Hadjisavvas, D. Kravvaritis and G. Pantelidis, *Structural properties of nonlinear monotone operators with values in L* (X,Y), to appear.
5. M. Jouak and L. Thibault, *Monotonie generalisee et sous-differentiels de fonctions convexes vectorielles*, Optimization **16** (1985), 187–199.
6. N. K. Kirov, *Generalized monotone mappings and differentiability of vector-valued convex mappings*, Serdica **9** (1983), 263–274.
7. A. G. Kusraev, *On the subdifferential mapping of convex operators*, Optimisazia (Novosibirsk) **21** (1978), 36–40.
8. A. G. Kusraev and S. S. Kutateladze, *Local convex analysis*, J. Soviet Math. **26** (1984), 2048–2087.
9. N. S. Papageorgiou, *Nonsmooth analysis on partially ordered vector spaces*, Pacific J. Math. **107** (1983), 403–459.
10. N. S. Papageorgiou, *Nonsmooth analysis on partially ordered vector spaces. The subdifferential theory*, Nonlinear Analysis T. M. A. **10** (1986), 615–637.
11. R. T. Rockafellar, *Local boundedness of nonlinear monotone operators*, Michigan Math. J. **16** (1969), 397–407.
12. R. T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc. **149** (1970), 75–88.
13. H. J. Schaefer, *Topological Vector Spaces*, Springer Verlag, 1980.
14. M. M. Vainberg, *Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations*, John Wiley and Sons, New York, 1973.
15. M. Valadier, *Sous-differentiabilité de fonctions convexes à valeurs dans une espace vectoriel ordonné*, Math. Scand. **30** (1972), 65–74.

16. E. Zeidler, *Nonlinear Functional Analysis and Its Applications-Monotone Operators*, Springer Verlag, New York, 1990.
17. J. Zowe, *Subdifferentiability of convex functions with values in an ordered vector space*, Math. Scand. **34** (1974), 69–83.

N. Hadjisavvas
Department of Mathematics
University of the Aegean
83200 Karlovassi
Samos
Greece

D. Kravvaritis
Department of Mathematics
National Technical University of Athens
Zografou Campus
15773 Athens
Greece

G. Pantelidis
Department of Mathematics
National Technical University of Athens
Zografou Campus
15773 Athens
Greece

Constantin Carathéodory: An International Tribute (pp. 461–475)
edited by Th. M. Rassias

FIRST CLASS FUNCTIONS WITH VALUES IN NONSEPARABLE SPACES

Roger W. Hansell

Let $f: T \to X$ be a function with values in a metric space X. By a *function base* for f we mean a collection Γ of subsets of T such that, for any open set U in X, $f^{-1}(U)$ is a union of sets from Γ. Various types of first class functions, such as pointwise limits of sequences of continuous functions and functions whose restriction to any nonempty closed set has a point of continuity, are characterized in terms of the existence of certain kinds of function bases. This yields nonseparable versions of some classical theorems due to R. Baire. In many instances the proofs are more informative and simpler than their classical counterparts. All spaces in this paper are assumed to be at least Hausdorff, and all functions are assumed to take their values in a metric space.

1. Baire Class 1 And F_σ Measurable Functions

A function $f: T \to X$ is of *Baire class* 1 if it is the pointwise limit of a sequence of continuous functions, and is F_σ *measurable* if $f^{-1}(U)$ is an F_σ set in T for each open set U in X. If X is a metric space, then it is easy to see that every Baire class 1 function with values in X is F_σ measurable [16, p. 386]. The converse may fail however, even when X is a two-point discrete space and T = [0, 1] (take f to

be the characteristic function of any point in T). A classical theorem of Baire states that the converse will hold whenever T is a separable metric space and X is the closed unit interval [0, 1] or the real line $\mathbb{R}$ (see, for example, [2, p. 67] or [16, p. 391]). This provides a very useful "internal" criterion for a function to be of Baire class 1. We will show that this result holds more generally when T is assumed only to be a normal space.

In order to obtain a similar result for F_σ measurable functions taking values in an arbitrary Banach space X something more is needed, even when the domain T is a subspace of $\mathbb{R}$. This follows from the fact that Martin's Axiom plus the negation of the Continuum Hypothesis implies the existence of an uncountable set $T \subset \mathbb{R}$ with the property that every subset of T is a relative F_σ set [17]. Thus, if f is any one-to-one function from T onto a discrete subset of a suitably large Banach space, then f is an example of an F_σ measurable function which is not of Baire class 1, since any continuous function defined on T (hence also f) would necessarily have a separable range. We will see that the needed additional property is for the function to have a σ-discrete function base. This concept was recently employed in [11] and [13], although it was introduced previously in [9, §3] where such functions were called "σ-discrete."

Recall that a collection of subsets of a space T is *discrete* if each point of T has a neighborhood meeting at most one member of the collection. The collection is said to be *σ-discrete* if it is a countable union of discrete subcollections. It can be shown that all Borel measurable functions defined on a complete metric space have a σ-discrete function base [9, Th. 3]. Also, Fleissner [6] has shown that this result continues to hold for any metric space provided one assumes the existence of super compact cardinals. Note that *any* function with values in a separable metric space has a countable (and thus σ-discrete) function base. The following lemma shows that the class of all functions having a σ-discrete function base forms a "Baire system" in the sense of Hausdorff [15, p. 191]. In particular, it shows that any Baire class 1 function with values in a metric space has a σ-discrete function base.

Lemma 1.1. *Let* T *be a Hausdorff space and* X *a metric space. Then the class of all functions from* T *to* X *having a σ-discrete function base contains all continuous functions and is close to pointwise limits of convergent sequences.*

Proof. Since X is metrisable, the topology has an open base of the form $\Lambda = \bigcup_{n\geq 1}\Lambda_n$, where each Λ_n is a discrete collection in X.

If $f:T\to X$ is continuous and $\Gamma_n = \{f^{-1}(U) : U\in\Lambda_n\}$, then it easily follows that $\bigcup_{n\ge1}\Gamma_n$ is a σ-discrete function base for f.

Now let $g:T\to X$ be the pointwise limit of the sequence of functions g_n, and suppose each g_n has a function base $\bigcup_{m\ge1}\Gamma_{nm}$ where the collection Γ_{nm} is discrete in T for each $m\ge1$. For each set B in $\bigcup_{n,m\ge1}\Gamma_{nm}$, which we may take to be nonempty, we can enumerate as a sequence $U_B^{(1)}, U_B^{(2)}, \ldots$ the members of

$$\{U\in\Lambda : B\subset g_n^{-1}(U) \text{ for some } n\ge1\},$$

since each point of X can belong to only countably many members of Λ. It follows that each of the families

$$\Gamma_{nmk} = \{B\cap g^{-1}(U_B^{(k)}) : B\in\Gamma_{nm}\} \qquad (n, m, k\ge1)$$

is discrete in T, and it remains only to show that together these form a function base for g.

Let $t\in g^{-1}(W)$ for some open set W in X, and choose $U\in\Lambda$ such that $g(t)\in U$ and $U\subset W$. Since $g_n(t)\to g(t)$, for some n we have $g_n(t)\in U$. By the property of a function base, we can find $m\ge1$ and $B\in\Gamma_{nm}$ such that $t\in B$ and $B\subset g_n^{-1}(U)$. Now for some $k\ge1$ we have $U = U_B^{(k)}$, and so

$$t\in B\cap g^{-1}(U_B^{(k)})\subset g^{-1}(U)\subset g^{-1}(W),$$

and this is what we needed to show.

Recall that a topological space T is *collectionwise normal* if, for each discrete family $\{E_a\}_{a\in A}$ of subsets of T, there is a discrete family of open sets $\{U_a\}_{a\in A}$ such that $cl(E_a)\subset U_a$ for each $a\in A$. All paracompact spaces (hence all metric and compact spaces) are collectionwise normal. Any normal space has the above separation property for any countable discrete collection [5, p. 379].

The following theorem gives the precise relationship between F_σ measurable and Baire class 1 functions.

Theorem 1.2.

(a) *Suppose* T *is a normal space and* X *is either the real line* $\mathbb{R}$ *or the closed unit interval* $[0, 1]$. *Then* $f:T\to X$ *is of Baire class* 1 *if and only if* f *is* F_σ *measurable.*

(b) *Suppose* T *is collectionwise normal and* X *is any closed convex subset of a Banach space. Then* $f:T\to X$ *is of Baire class* 1 *if and only if* f *is* F_σ *measurable and has a* σ*-discrete function base.*

Remark 1.3. In each of the above two cases the key property between the spaces T and X is that any continuous function defined on a closed subspace of T and taking values in X can be continuously extended to all of T. In case (a) this is just the classical theorem of Tietze-Urysohn [5, p. 97]. Now Dugundji [4, Th. 4.1] has shown that any convex subset X of a locally convex linear topological space has the above continuous extension property for any metrisable space T. Further, by a theorem of Dowker [3, Th. 2], whenever a space X has the above extension property for metrisable spaces and X is topologically complete (i.e., X is homeomorphic to a closed subspace of a product of complete metric spaces), then X will also have this extension property for any collectionwise normal space T.

The following lemma isolates the technical part of the proof of Theorem 1.2. In particular, it enables us to eliminate the restrictive assumption that the domain space T be "perfect" in the sense that all open sets are F_σ sets. To my knowledge, this has always been assumed in proofs of part (a) of Theorem 1.2. Note that a similar assumption was made in [10, p. 197] where we gave a nonseparable version of part (a) (see also the correction of [10, Th. 6] given in [12, pp. 389–390]). In addition to being applicable to general paracompact spaces, our present Theorem 1.2 has a much more direct proof.

Lemma 1.4. *Let T be a collectionwise normal space, X a metric space, and suppose $f : T \to X$ is F_σ measurable and has a σ-discrete function base. Then f has a σ-discrete function base consisting of closed G_δ sets. The same result holds with "σ-discrete" replaced by "countable" in the case when T is assumed only to be normal.*

Proof. Let Λ be a σ-discrete open base for the topology of X, so each point of X belongs to only countably many members of Λ. Let $\Gamma = \bigcup_{n\geq 1}\Gamma_n$ be a σ-discrete function base for f. By the collectionwise normality of T, for each $n \geq 1$ there is a discrete collection $\{V_B : B \in \Gamma_n\}$ of open sets in T such that $\mathrm{cl}(B) \subset V_B$ for each $B \in \Gamma_n$. In the case when Γ is countable we may take $T = V_B$ for each $B \in \Gamma$. The remainder of the proof, which requires only that T be a normal space, now applies to either case.

Given $U \in \Lambda$ and $B \in \Gamma$ with $B \subset f^{-1}(U)$ we first show that there is a sequence $\{Z_k\}_{k\geq 1}$ of closed G_δ sets in T such that

$$B \subset \bigcup_{k\geq 1} Z_k \subset V_B \cap f^{-1}[\mathrm{cl}(U)]. \tag{1}$$

Since f is F_σ measurable, we can write

$$\bigcup_{k\geq 1} F_k = f^{-1}(U) \subset f^{-1}[\mathrm{cl}(U)] = \bigcap_{j\geq 1} G_j, \tag{2}$$

where each F_k is closed and each G_j is open in T. Since T is normal, for each j, $k\geq 1$ we can find a closed G_δ set Z_{kj} such that

$$F_k \cap \mathrm{cl}(B) \subset Z_{kj} \subset G_j \cap V_B. \tag{3}$$

It follows that the sets $Z_k = \bigcap_{j\geq 1} Z_{kj}$ are closed G_δ sets in T satisfying (1).

Now, for each nonempty $B\in\Gamma$, let $U_B^{(1)}, U_B^{(2)}, \ldots$ be an enumeration of the sets $U\in\Lambda$ such that $B \subset f^{-1}(U)$, and let $\{Z_{Bk}^{(m)}\}_{k\geq 1}$ be the sequence of closed G_δ sets associated with $B \subset f^{-1}(U_B^{(m)})$ according to the above construction. Let

$$\Gamma_{nmk} = \{Z_{Bk}^{(m)} : B\in\Gamma_n\}$$

and let us show that $\bigcup_{n,m,k\geq 1}\Gamma_{nmk}$ is a σ-discrete function base for f.

Since $Z_{Bk}^{(m)} \subset V_B$ by (1) and $\{V_B : B\in\Gamma_n\}$ is discrete, it follows that each of the collections Γ_{nmk} is discrete in T. Now suppose $t\in f^{-1}(W)$ for some open set W in X and let $U\in\Lambda$ be such that $f(t)\in U$ and $\mathrm{cl}(U)\subset W$. Since G is a function base for f, we have $t\in B \subset f^{-1}(U)$ for some $B\in\Gamma_n$ and $n\geq 1$. Thus $U = U_B^{(m)}$ for some $m\geq 1$, and so $t\in F_{Bk}^{(m)}$ for some $k\geq 1$ by (2). It now follows from (3), the definition of $Z_{Bk}^{(m)}$ and (2) that

$$t\in Z_{Bk}^{(m)} \subset \bigcap_{j\geq 1} G_j = f^{-1}[\mathrm{cl}(U)] \subset f^{-1}(W).$$

That completes the proof.

Proof of Theorem 1.2. It suffices to assume that X is itself a Banach space, or $\mathbb{R}$ in the case of (a), since this easily implies the result for any retract of X (cf. [12, Lemma 7]). In view of Lemma 1.1 it is enough to show that if $f: T \to X$ is F_σ measurable and has a σ-discrete function base, then f is of Baire class 1.

For a given $\varepsilon>0$ we first prove that there is a Baire class 1 function $g: T\to X$ such that $\| f - g \| < \varepsilon$. It will then follow that f is of Baire class 1. To see this, choose Baire class 1 functions g_n such that that $\| f - g_n \| < 2^{-n}$, and let $\{g_{1m}\}_{m\geq 1}$ and $\{g_{nm}\}_{m\geq 1}$ be sequences of continuous functions converging pointwise to the Baire class 1 functions g_1 and $g_{n+1} - g_n$ respectively. Furthermore, we may assume that $\| g_{nm} \| \leq 2^{-n+1}$ for each $m\geq 1$ and $n\geq 2$. It then follows that the continuous functions

$$g_{1n} + g_{2n} + \cdots + g_{nn}$$

will converge pointwise to f (cf. [16, p. 392]).

Given $\varepsilon > 0$ let $\Lambda = \{U_\alpha : \alpha < \lambda\}$ be an open cover of X by sets having diameter less than ε. By Lemma 1.4, f has a function base of the form $\bigcup_{n\geq 1}\Gamma_n$ where each Γ_n is a discrete collection of closed G_δ sets in T. (In case (a) we may further assume that each Γ_n consists of only a single set.) For each $n \geq 1$ and $\alpha < \lambda$ define

$$Z_{\alpha n} = \bigcup\{Z \in \Gamma_n : \alpha \text{ is the least ordinal with } Z \subset f^{-1}(U_\alpha)\} \tag{4}$$

and let $Z_n = \bigcup\{Z_{\alpha n} : \alpha < \lambda\}$. Now it is easy to show that in a normal (resp. collectionwise normal) space the union of a countable (resp. arbitrary) discrete collection of closed G_δ sets (equivalently, zero sets) is again a closed G_δ set (see [5, p. 100]). Hence Z_n is a closed G_δ set and $\{Z_{\alpha n} : \alpha < \lambda\}$ is a discrete collection of closed G_δ sets in T for each $n \geq 1$. Note that

$$\bigcup_{n\geq 1} Z_{\alpha n} \subset f^{-1}(U_\alpha) \quad \text{and} \quad X = \bigcup_{\alpha<\lambda}\bigcup_{n\geq 1} Z_{\alpha n},$$

the latter following from the fact that $\bigcup_{n\geq 1}\Gamma_n$ is a function base for f.

Consider the canonical partition of T associated with the family $\{Z_{\alpha n} : n \geq 1 \text{ and } \alpha < \lambda\}$ by defining

$$D_{\alpha 1} = Z_{\alpha 1} \quad \text{and} \quad D_{\alpha n} = Z_{\alpha n} \setminus \bigcup_{m=1}^{n-1} Z_m \quad (n \geq 2).$$

The important point is that each of the sets $\bigcup_{n\geq 1} D_{\alpha n}$ is an F_σ set in T, since each $D_{\alpha n}$ is of this class. Choosing an F_σ representating for each $D_{\alpha n}$ and then re-indexing, we can find, for each $\alpha < \lambda$, a sequence $\{F_{\alpha m}\}_{m\geq 1}$ of closed sets in T such that

$$\bigcup_{n\geq 1} D_{\alpha n} = \bigcup_{m\geq 1} F_{\alpha m} \subset \bigcup_{n\geq 1} Z_{\alpha n} \subset f^{-1}(U_\alpha), \tag{5}$$

$$\{F_{\alpha m} : \alpha < \lambda\} \text{ is discrete in } T \text{ for each } m \geq 1, \text{ and} \tag{6}$$

$$F_{\alpha m} \cap F_{\beta k} = \emptyset \text{ whenever } \alpha \neq \beta \text{ and for all } m, k \geq 1. \tag{7}$$

It follows that the sets

$$F_n = \bigcup\{F_{\alpha m} : \alpha < \lambda \text{ and } m = 1, \ldots, n\}$$

form an increasing sequence of closed sets covering the space T.

We define a sequence of continuous functions $g_n : F_n \to X$ by first fixing a point $x_\alpha \in U_\alpha$ for each $\alpha < \lambda$ (and independently of n) and defining

$$g_n(t) = x_\alpha \text{ whenever } t \in \bigcup_{m=1}^{n} F_{\alpha m} \subset f^{-1}(U_\alpha).$$

It follows from (6) and (7) that g_n is well defined and continuous. By Remark 1.3, g_n extends to a continuous function on all of T, which we will also denote as g_n. Since $g_n(t) = x_\alpha$ for all $n \geq m$ whenever $t \in F_{\alpha m}$, it follows that the sequence g_n converges pointwise to the Baire class 1 function $g : T \to X$ given by

$$g(t) = x_\alpha \text{ whenever } t \in \bigcup_{m=1}^{\infty} F_{\alpha m} \subset f^{-1}(U_\alpha).$$

Since $x_\alpha \in U_\alpha$ and $\operatorname{diam}(U_\alpha) < \varepsilon$, we have $\| f - g \| < \varepsilon$ as required.

2. Functions With The Point Of Continuity Property

Another well-known theorem of Baire states that, for any complete separable metric space T and real-valued function $f : T \to \mathbb{R}$, f is F_σ measurable (and hence Baire class 1) if and only if $f \mid A$ has a point of continuity for each nonempty closed set $A \subset T$ (see [16, p. 394] or [2, p. 67]). Any function with the latter property will be called a *PC function* (for "point of continuity").

In order to characterize PC functions in terms of a function base we need to introduce a concept which is substantially weaker than a discrete collection of sets. A collection Δ of disjoint subsets of a space T is said to be *scattered* if, for each nonempty subset $A \subset \bigcup \Delta$, some set of the form $A \cap D$, with $D \in \Delta$, is nonempty and open relative to A. Equivalently, Δ is scattered if we can write $\Delta = \{D_\alpha : \alpha < \lambda\}$ and find open sets U_α in T such that

$$D_\alpha \subset U_\alpha \setminus \bigcup_{\beta<\alpha} U_\beta \tag{8}$$

for each $\alpha < \lambda$ (cf. [18, Lemma 2.1] and [14, §2]). Note that a topological space is scattered in the usual sense (each nonempty subset has an isolated point) if and only if the collection of all one-point subsets is scattered. The following lemma gives an important property of scattered collections.

Lemma 2.1. *If Δ is a scattered collection of sets of the first category in T, then $\bigcup \Delta$ is also of the first category in T.*

Proof. Let $\Delta = \{D_\alpha : \alpha < \lambda\}$ and let U_α be open sets in T satisfying (8). By the Banach Category Theorem [16, p. 82], it suffices to show that each point of $\bigcup \Delta$ has a relative open neighborhood of the first category in $\bigcup \Delta$. Proceeding inductively, we may assume that this is true whenever $\Delta = \{D_\alpha : \alpha \le \xi\}$ for some $\xi < \lambda$. But then, if $t \in D_\alpha$ for some $\alpha < \lambda$, then U_α is an open sets containing t such that

$$U_\alpha \cap \bigcup \Delta = \bigcup_{\beta \le \alpha} D_\beta$$

is of the first category in $\bigcup \Delta$.

By a $F\cap G$ set in a space T we mean a set that is the intersection of a closed and an open set (equivalently, the difference of two open or two closed sets). A $(F\cap G)_\sigma$ set is a countable union of $F\cap G$ sets. A topological space T is said to be *hereditarily Baire* if each closed subspace has the Baire category property. All complete metric spaces and all locally compact Hausdorff spaces are hereditarily Baire. Lemma 2.1 enables us to give the following relationship between PC functions and functions having a σ-scattered function base.

Theorem 2.2. *If f is a PC function from a space T to a metric space X, then f has a σ-scattered function base of $F\cap G$ sets in T. Conversely, if T is hereditary Baire, then any function from T to X having a σ-scattered function base of $F\cap G$ sets in T is a PC function.*

Proof. Suppose f is a PC function. For each $\varepsilon>0$ we define inductively an open cover $\Gamma_\varepsilon=\{U_\alpha:\alpha<\lambda\}$ of T such that, if

$$D_\alpha = U_\alpha \setminus \bigcup_{\beta<\alpha} U_\beta,$$

then the diameter of $f(D_\alpha)$ is less than ε for each $\alpha<\lambda$.

Since f has a point of continuity we can find a nonempty open set U_0 in T such that the diameter of $f(U_0)$ is less than ε. Assuming we have defined U_ξ for all $\xi<\alpha$, if

$$A_\alpha = T\setminus\bigcup_{\xi<\alpha} U_\xi \neq \varnothing,$$

then $f|A_\alpha$ has a point of continuity. Hence there is an open set U_α in T such that $U_\alpha\cap A_\alpha$ is nonempty and $f(U_\alpha\cap A_\alpha)$ has diameter less than ε. This defines the desired open cover of T.

Doing this with $\varepsilon=1/n$ for each $n\geq 1$, we obtain a σ-scattered collection $\Gamma=\bigcup_{n\geq 1}\Gamma_n$ such that each D in Γ_n is a $F\cap G$ set in T and the diameter of $f(D)$ is less than $1/n$. It easily follows that Γ is also a function base for f.

Conversely, suppose that f has a function base $\Gamma=\bigcup_{n\geq 1}\Gamma_n$ where each Γ_n a scattered collection of $F\cap G$ sets in T. Since the restriction of f to any closed subspace of T will have a function base of the same type and T is hereditarily Baire, it suffices to show that f has a point of continuity. For each $B\in\Gamma$, let $B=F_B\cap G_B$ where F_B is closed and G_B is open in T. It follows from Lemma 2.1 that

$$M_n=\bigcup\{[F_B\setminus \mathrm{int}(F_B)]\cap G_B : B\in\Gamma_n\}$$

is a set of the first category in T. Since T is a Baire space, there is a point $t\in T$ such that $t\notin M_n$ for each $n\geq 1$, and let us show that f is continuous at t. Let U

be any open subset of X containing $f(t)$, and use the property of a function base to find $B \in \Gamma_n$ such that $t \in B$ and $B \subset f^{-1}(U)$. Since $t \notin M_n$, we have

$$t \in \text{int}(F_B) \cap G_B \subset B \subset f^{-1}(U),$$

proving that f is continuous at t.

Corollary 2.3. *For any space* T *and metric space* X, *if* $f: T \to X$ *has a* σ-*scattered function base of* $(F \cap G)_\sigma$ *sets in* T, *then the set of points of discontinuity of* f *is a set of the first category in* T.

Proof. Note that the assumptions imply that f also has a σ-scattered function base of $F \cap G$ sets in T. From the proof of Theorem 2.2 it follows that the set of points of discontinuity of f is contained $\bigcup_{n \geq 1} M_n$ and this is a set of the first category in T.

Corollary 2.4. (Fort [7, Th. 2]) *For any space* T *and metric space* X, *if* $g: T \to X$ *is of Baire class 1, then the points of discontinuity of* g *is a set of the first category in* T.

Proof. The proof of Lemma 1.1 shows that g has a σ-discrete function base of sets of the form $B \cap g^{-1}(U)$ where B is open in T and $g^{-1}(U)$ is an F_σ set. Since discrete collections are clearly scattered, it follows that g has a function base of the type required in Corollary 2.3.

As noted above, if T is a complete separable metric space, then $f: T \to \mathbb{R}$ is a PC function if and only if it is F_σ measurable. For a general domain space T it is more appropriate to try to relate PC functions with $(F \cap G)_\sigma$ measurable functions. For example, if T is a weakly compactly generated Banach space, then the identity function on T will be weak-to-norm $(F \cap G)_\sigma$ measurable, but it will be F_σ measurable if and only if T is separable (see [14]).

It is easy to give examples of real-valued PC functions which are not Borel measurable, even when the domain space T is a compact Hausdorff space. For example, if T is the ordinal space $[0, \omega_1]$, then any function defined on T is a PC function (since all nonempty subsets contains an isolated point), but not all characteristic functions on T are Borel measurable, since T conains non-Borel sets [8, p. 231]. As we will show, the problem is that T has a scattered partition (in this case the one-

point subsets) which is not $(F\cap G)_\sigma$ additive, i.e., the union of some subcollection is not a $(F\cap G)_\sigma$ set in T.

Theorem 2.5. *For any space* T, *all PC functions defined on* T *and taking values in a metric space will be* $(F\cap G)_\sigma$ *measurable if and only if every scattered partition of* T *is* $(F\cap G)_\sigma$ *additive.*

Proof. Suppose that all PC functions defined on T and taking values in a metric space are $(F\cap G)_\sigma$ measurable, and let $\{D_\alpha : \alpha < \lambda\}$ be some scattered partition of T. Thus $\bigcup_{\alpha<\xi} D_\alpha$ is open in T for each $\xi \le \lambda$. Choose any point $x_\alpha \in D_\alpha$ and let $X = \{x_\alpha : \alpha < \lambda\}$ have the discrete topology. Define $f: T \to X$ so that $f(D_\alpha) = \{x_\alpha\}$ for each $\alpha < \lambda$. Then f is a PC function, since, for any nonempty set $A \subset T$, if α is the least ordinal such that $A \cap D_\alpha \neq \emptyset$, then $f|A$ is continuous at each point of $A \cap D_\alpha$. By assumption, f is $(F\cap G)_\sigma$ measurable. Since each subset of X is open, it follows that the union of any subcollection of $\{D_\alpha: \alpha < \lambda\}$ is a $(F\cap G)_\sigma$ set in T.

Conversely, suppose that every scattered partition of T is $(F\cap G)_\sigma$ additive, and let $f: T \to X$ be a PC function taking values in the metric space X. As was shown in the proof of Theorem 2.2, for each $n \ge 1$, T has a scattered partition Γ_n of $F\cap G$ sets such that the image under f of each set in Γ_n has diameter at most 1/n. It follows from our assumption that the union of each subcollection of $\bigcup_{n\ge1}\Gamma_n$ is a $(F\cap G)_\sigma$ set in T. But $\bigcup_{n\ge1}\Gamma_n$ is also a function base for f, and clearly this implies that f is $(F\cap G)_\sigma$ measurable.

It is well-known and easy to prove that any discrete collection in a space T can be expanded to a discrete collection of closed sets in T (by taking closures), and the union of any discrete family of closed sets will again be a closed set. Consequently, any σ-discete family of F_σ sets in T is F_σ additive. The following lemma, proved in [14, §3], shows that for $F\cap G$ sets this continues to hold for the weaker notion of a *relatively* discrete family. Recall that a collection of sets Γ is *relatively discrete* if it is discrete relative to the subspace $\bigcup\Gamma$; equivalently, each set in Γ has a neighborhood not meeting any other member of Γ.

Lemma 2.6. *The following hold in any topological space* T.

(a) *Each relatively discrete collection in* T *can be expanded to a relatively discrete collection of* $F\cap G$ *sets in* T.

(b) *The union of a relatively discrete collection of* $F\cap G$ *sets in* T *is again of this type.*

(c) *If each collection of open sets in* T *has a* σ*-relatively discrete (resp.* σ*-discrete) refinement, then so does each scattered collection.*

A topological space is said to be *weakly* θ*-refinable* (resp. *subparacompact*) if each open cover has a σ-relatively discrete (resp. a σ-discrete and closed) refinement (see [1, Th. 3.7]). The topological assumption in part (c) of Lemma 2.6 is equivalent to assuming that each subspace of T is weakly θ-refinable (resp. subparacompact for a regular space T). Since metric spaces are hereditarily paracompact, they have the property that each scattered collection has a σ-discrete refinement. It was recently shown in [14] that the weak topology of a Banach space will be hereditarily weakly θ-refinable for a significantly wide class of nonseparable Banach spaces, including those Banach spaces Z which have an equivalent norm $\|\cdot\|$ such that the norm and weak topologies coincide on $\{z : \|z\| = 1\}$. The ordinal space $[0,\omega_1)$ is an example of a space which is not weakly θ-refinable.

THEOREM 2.7. *Let* T *be any Hausdorff space,* X *a metric space and let* $f : T \to X$ *be a PC function.*

(a) *If each open collection in* T *has a* σ*-relatively discrete refinement, then* f *is* $(F\cap G)_\sigma$ *measurable and has a* σ*-relatively discrete function base of* $F\cap G$ *sets in* T.

(b) *If each open collection in* T *have a* σ*-discrete closed refinement, then* f *is* F_σ *measurable and has a* σ*-discrete function base of closed sets in* T.

Proof. Suppose the space T satisfies the assumption in part (a). By Theorem 2.2, f has a function base $\Gamma = \bigcup_{n\geq 1}\Gamma_n$ with each Γ_n a scattered collection of $F\cap G$ sets in T. By (c) of Lemma 2.6, for each $n \geq 1$, Γ_n has a refinement of the form $\bigcup_{n\geq 1}\Gamma_{nm}$ where each collection Γ_{nm} is relatively discrete. By (a) of Lemma 2.6, for each $n, m \geq 1$, Γ_{nm} has a relatively discrete expansion $\{B^* : B \in \Gamma_{nm}\}$ where $B \subset B^*$ and each B^* is a $F\cap G$ sets in T. It follows that, for each n, $m \geq 1$,

$$\Lambda_{nm} = \{B^* \cap C : B \in \Gamma_{nm},\ C \in \Gamma_n \text{ and } B \subset C\}$$

is a relatively discrete collection of $F\cap G$ sets in T (since $B\subset C$ for only one $C\in\Gamma_n$).

To see that $\Lambda=\bigcup_{n,m\geq 1}\Lambda_{nm}$ is a function base for f, suppose $t\in f^{-1}(U)$ for some open set U in X. Then, for some $n\geq 1$ and $C\in\Gamma_n$, we have $t\in C$ and $C\subset f^{-1}(U)$, since Γ is a function base for f. Since $\bigcup_{n\geq 1}\Gamma_{nm}$ is a refinement of Γ_n, there is a $B\in\Gamma_{nm}$ with $B\subset C$ for some $m\geq 1$. It follows that

$$t\in B^*\cap C\in\Lambda_{nm} \quad \text{and} \quad B^*\cap C\subset f^{-1}(U)$$

as required.

Finally, since the union of any subcollection of Λ is a $(F\cap G)_\sigma$ set in T by (b) of Lemma 2.6, it follows that f is $(F\cap G)_\sigma$ measurable.

The proof in the case of (b) is identical using the alternate part of (c) of Lemma 2.6 and the standard properties of discrete collections.

Lastly, we seek to find conditions on T for which all $(F\cap G)_\sigma$ measurable functions defined on T and taking values in a metric space X are PC functions. We conjecture that this is true precisely when T is hereditarily Baire. Here we will show that this holds if, in addition, X is separable or T has countable tightness (i.e., if $t\in \mathrm{cl}(E)$ for some $E\subset T$, then there is a countable set $C\subset E$ such that $t\in \mathrm{cl}(C)$). Clearly all metrisable spaces have countable tightness. The space $C_p(K)$ of all real-valued continuous functions defined on a compact Hausdorff space K with the point-wise topology is known to have countable tightness.

Theorem 2.8. *Let* T *be hereditarily Baire,* X *a metric space and suppose* $f:T\to X$ *is a* $(F\cap G)_\sigma$ *measurable function. If* X *is separable or* T *has countable tightness, then* f *is a PC function.*

Proof. If X is separable, then it has a countable base of open sets $U_1, U_2, \ldots$, and $f^{-1}(U_n)$ is a $(F\cap G)_\sigma$ set in T for each $n\geq 1$. This easily implies that f has a countable function base of $F\cap G$ sets, and thus f is a PC function by Theorem 2.2.

Next we show that f will be continuous at each point of a dense subset of T whenever T is separable. Let D be a countable dense subset of T. For each open set U in X we let

$$f^{-1}(U)=\bigcup_{m\geq 1}F_{Um}\cap G_{Um} \quad \text{and} \quad M_U=\bigcup_{m\geq 1}[F_{Um}\setminus \mathrm{int}(F_{Um})]\cap G_{Um},$$

where F_{Um} is closed and G_{Um} is open in T for each $m \geq 1$. Note that M_U is of the first category in T and $f^{-1}(U) \setminus M_U$ is open in T for each U. For each $n \geq 1$, let Λ_n be a locally finite open cover of X by sets having diameter at most $1/n$, and let

$$M_n = \bigcup\{ M_U : U \in \Lambda_n \text{ and } D \cap f^{-1}(U) \neq \emptyset \}$$
$$W_n = \bigcup\{ U : U \in \Lambda_n \text{ and } D \cap f^{-1}(U) = \emptyset \}.$$

Then M_n is of the first category in T as a countable union of such sets. Further, we must have $f^{-1}(W_n) = M_{W_n}$ since otherwise $f^{-1}(W_n) \setminus M_{W_n}$ would be a nonempty open set in T not meeting D. Since T is a Baire space, $T \neq M_n \cup M_{W_n}$ and so

$$V_n = \bigcup\{ f^{-1}(U) \setminus M_U : U \in \Lambda_n \text{ and } D \cap f^{-1}(U) \neq \emptyset \}$$

is a dense open set in T. It follows that $\bigcap_{n \geq 1} V_n$ is dense in T, and clearly any point of this set is a point of continuity for f (cf. the proof of Theorem 2.2).

Now assume that T has countable tightness. Since this property is inherited by any subspace, and since for any nonempty close subspace F of T, F is a Baire space and $f|F$ is $(F \cap G)_\sigma$ measurable, it is enough to show that f has a point of continuity. Suppose f has no point of continuity, and for each $n \geq 1$ let

$$F_n = \{ t \in T : \operatorname{diam} f(V) \geq 1/n \text{ for each open neighborhood } V \text{ of } t \}.$$

Since the sets F_n are closed and cover the Baire space T, we must have $W = \operatorname{int}(F_m)$ nonempty for some $m \geq 1$. Now each $s \in W$ belongs to the closure of the set of all $t \in T$ such that $\operatorname{dist}[f(t), f(s)] \geq 1/m$, so by countable tightness we can find a countable set $C_s \subset W$ such that

$$s \in \operatorname{cl}(C_s) \quad \text{and} \quad C_s \subset \{ t \in T : \operatorname{dist}[f(t), f(s)] \geq 1/m \}. \tag{10}$$

Since this is true for each point of W, we can iterate (10) to find a countable set $C \subset W$ such that $C_s \subset C$ for each $s \in C$. Since $\operatorname{cl}(C)$ is a closed separable subspace of T, it follows from the above that $g = f|\operatorname{cl}(C)$ has a point of continuity $t \in \operatorname{cl}(C)$. Hence we can find an open set V in T such that

$$t \in V \cap \operatorname{cl}(C) \quad \text{and} \quad \operatorname{diam}[g(V \cap \operatorname{cl}(C))] < 1/m. \tag{11}$$

Let $s \in V \cap C$ and use (10) to get some $s' \in V \cap C_s$. But then $\operatorname{dist}[f(s'), f(s)] \geq 1/m$, by (10), and $s, s' \in V \cap \operatorname{cl}(C)$, and this contradicts (11). That completes the proof.

The latter part of the preceding proof establishes the following corollary.

Corollary 2.9. *Let T be hereditarily Baire and have countable tightness, and let X be a metric space. If $f : T \to X$ is such that $f|S$ has a point of continuity whenever S is a nonempty closed separable subspace of T, then f is a PC function.*

3. First Class Functions Defined On Complete Metric Spaces.

We combine the above results to obtain the following nonseparable extension of the classical theorem of Baire on characterizing Baire class 1 functions defined on a complete separable metric space. (While this paper was being prepared the author obtained from C. Stegall the preprint [19] in which he proves the following theorem using methods quite different from ours.)

Theorem 3.1. *For a complete metric space* T *and a Banach space* X *the following are equivalent for any function* $f : T \to X$.

(i) f *is of Baire class* 1;

(ii) $g \circ f$ *is* F_σ *measurable for each continuous* $g : X \to \mathbb{R}$;

(iii) f *is* F_σ *measurable*;

(iv) f *is a PC function.*

(v) $f|S$ *has a point of continuity for each nonempty closed separable subspace* S *of* T;

Proof. (i) $\to$ (ii) This follows from the fact that f is F_σ measurable [16, p. 386].

(ii) $\to$ (iii) If U is any open set in X, we can find a continuous $g : X \to \mathbb{R}$ such that $x \in U$ if and only if $g(x) \neq 0$. Since $g \circ f$ is F_σ measurable, it follows that $f^{-1}(U)$ is F_σ in T.

(iii) $\to$ (iv) This follows from Theorem 2.8 since any complete metric space is hereditarily Baire and has countable tightness.

(iv) $\to$ (v) This is clear.

(v) $\to$ (iv) This follows from Corollary 2.9.

(iv) $\to$ (i) This follows from part (b) of Theorems 2.7 and 1.2.

REFERENCES

1. D. Burke, *Covering Properties*, in Handbook of set-theoretic topology, North-Holland, Amsterdam, 1984, 347-422.

2. J. Diestel, *Sequences and Series in Banach Spaces*, Springer-Verlag, New York, 1984.

3. C. Dowker, *On a theorem of Hanner*, Ark. för Mat. **2** (1952), 307-313.
4. J. Dugundji, *An extension of Tietze's theorem*, Pacific J. Math. **1** (1951), 353-367.
5. R. Engelking, *General Topology*, PWN, Warsaw, 1977.
6. W. Fleissner, *An axiom for nonseparable Borel theory*, Trans. Amer. Math. Soc. **251** (1979), 309-328.
7. M. Fort, Jr., *Category theorems*, Fund. Math. **42** (1955), 276-288
8. P. Halmos, *Measure Theory*, van Nostrand, New York, 1950.
9. R. Hansell, *Borel measurable mappings for nonseparable metric spaces*, Trans. Amer. Math. Soc. **161** (1971), 145-169.
10. R. Hansell, *On Borel mappings and Baire functions*, Trans. Amer. Math. Soc. **194** (1974), 195-211.
11. R. Hansell, *Extended Bochner measurable selectors*, Math. Ann. **277** (1987), 79-94.
12. R. Hansell, *First class selectors for upper semi-continuous multifunctions*, J. Funct. Anal. **75** (1987), 382-395.
13. R. Hansell, *Sums, products and continuity of Borel maps in nonseparable metric spaces*, Proc. Amer. Math. Soc. **104** (1988), 465-471.
14. R. Hansell, *Descriptive sets and the topology of nonseparable Banach spaces*, preprint (1989), 78 pp.
15. F. Hausdorff, *Set Theorey*, Chelsea, New York, 1957.
16. K. Kuratowski, *Topology*, Vol. 1, Academic Press, New York, 1966.
17. D. Martin and R. Solavay, *Internal Cohen extensions and Souslin's problem*, Ann. Math. **94** (1971), 201-245.
18. E. Michael, *A note on completely metrizable spaces*, Proc. Amer. Math. Soc. **96** (1986), 513-522.
19. C. Stegall, *Functions of the first Baire class with values in Banach spaces*, Proc. Amer. Math Soc. (to appear).

Roger W. Hansell
University of Connecticut
Storrs, Connecticut 06269
U.S.A.

Constantin Carathéodory: An International Tribute (pp. 476–488)
edited by Th. M. Rassias

A NEW QUADRATIC EQUATION

HIROSHI HARUKI

The purpose of this paper is to solve a new quadratic equation on the Gaussian plane and to give its geometric interpretation.

1. Introduction And Statement Of The Result

We consider first the quadratic equation
(see [1], p. 82, [2], pp. 165-200, [4], [6], [8]-[10], [13], [14], [17])

$$f(x+y) + f(x-y) = 2f(x) + 2f(y)\,, \tag{1}$$

where f is an entire function of a complex variable z and x,y are complex variables. We can prove the following theorem:

Theorem A. The only entire solution of (1) is given by $f(z) = \gamma z^2$ where γ is an arbitrary complex constant.

Proof. Since the proof is easy, we omit it.

In this paper we adopt the following definition:

Definition. If the only solution of a given functional equation F whose unknown function is f is $f(z) = \gamma z^2$ where γ is an arbitrary complex constant, i.e., the monomial of degree 2 in z

(including the identically zero function), then the functional equation F is said to be a quadratic equation.

Now we are going to give a new quadratic equation. To this end we give preliminary considerations. We consider the following two Cauchy equations (see [1], pp. 31-42, [2], pp. 11-24)

$$f(x + y) = f(x) + f(y)\,, \tag{2}$$

and

$$f(x + y) = f(x)\, f(y)\,, \tag{3}$$

where f is an entire function of a complex variable and x,y are complex variables. If we replace x and y by s and it in (2), (3) respectively, where s,t are real variables and we take the absolute values of the resulting equalities, then we obtain the following two functional equations:

$$\left|f(s + it)\right| = \left|f(s) + f(it)\right| \quad \text{(Robinson's functional equation; see [5], [15])}$$

and

$$\left|f(s + it)\right| = \left|f(s)\, f(it)\right| \quad \text{(Hille's functional equation; see [5], [11], [12], [16])},$$

where f is an unknown entire function of a complex variable and s,t are real variables.

In a similar way, replacing x and y by s and it in (1) where s,t are real variables and taking the absolute values of the resulting equality yields the following functional equation:

$$\left|f(s + it) + f(s - it)\right| = 2\left|f(s) + f(it)\right|, \tag{4}$$

where f is an unknown entire function of a complex variable z and s,t are real variables.

The purpose of this paper is to solve (4), i.e., to prove the following theorem:

Theorem 1.1. The only entire solution of (4) is given by $f(z) = \gamma z^2$ where γ is an arbitrary complex constant.

To prove the above theorem we shall use a special trick (see Lemma 2.1 in the next section).

2. Lemmas

We shall apply the following three lemmas to prove the theorem in Section 1.

Preceding to state Lemma 2.1 we shall explain some notations.

Let f be an entire function of a complex variable z. Since f is an entire function, we can expand f in a power series at any point. Let its power series expansion at $z = 0$ be

$$f(z) = \sum_{n=0}^{+\infty} c_n z^n , \tag{5}$$

where each of $c_n (n = 0, 1, 2, \cdots)$ is a complex constant.

If we set

$$a_n = \mathrm{Re}(c_n), \ b_n = \mathrm{Im}(c_n) \ (n = 0, 1, 2, \ldots) , \tag{6}$$

then, by (5) we obtain

$$f(z) = \sum_{n=0}^{+\infty} \left(a_n + ib_n\right) z^n \tag{7}$$

for all complex z, where a_n, b_n $(n = 0, 1, 2, \ldots)$ are all real constants by (6).

We may now state Lemma 2.1.

Lemma 2.1. We use the same notations as above. If we set

$$\phi(z) = \sum_{n=0}^{+\infty} a_n z^n \quad \text{and} \quad \psi(z) = \sum_{n=0}^{+\infty} b_n z^n , \tag{8}$$

then we obtain

(i)
$$\phi(z) = \frac{1}{2}\left(f(z) + \overline{f(\bar{z})}\right) \quad \text{and} \quad \psi(z) = \frac{1}{2i}\left(f(z) - \overline{f(\bar{z})}\right) . \tag{9}$$

for all complex z;

(ii) $\phi(z)$ and $\psi(z)$ are entire functions of a complex variable z;

(iii) $\overline{\phi(z)} = \phi(\bar{z})$ and $\overline{\psi(z)} = \psi(\bar{z})$ for all complex z;

(iv) if z is real, then $\phi(z)$ and $\psi(z)$ are also real.

Proof. (i) By (7) we have

$$\overline{f(\bar{z})} = \overline{\sum_{n=0}^{+\infty} (a_n + ib_n)\bar{z}^n} = \sum_{n=0}^{+\infty} \overline{(a_n + ib_n)z^n} \tag{10}$$
$$= \sum_{n=0}^{+\infty} (a_n - ib_n)z^n$$

for all complex z.

Adding (7), (10) side by side yields

$$\phi(z) = \frac{1}{2}\left(f(z) + \overline{f(\bar{z})}\right)$$

for all complex z.

Subtracting (10) from (7) side by side yields

$$\psi(z) = \frac{1}{2i}\left(f(z) - \overline{f(\bar{z})}\right).$$

Q.E.D.

for all complex z.

(ii) Since, by hypothesis, f is an entire function of z, so is $\overline{f(\bar{z})}$. Hence, by (i) of Lemma 2.1 Q.E.D.

$\phi(z)$ and $\psi(z)$ are entire functions of z.

(iii) By (8) we obtain

$$\overline{\phi(z)} = \overline{\sum_{n=0}^{+\infty} a_n z^n} = \sum_{n=0}^{+\infty} \overline{a_n z^n} = \sum_{n=0}^{+\infty} a_n \bar{z}^n = \phi(\bar{z})$$

for all complex z.

Similarly we can prove that $\overline{\psi(z)} = \psi(\bar{z})$ holds for all complex z.

Q.E.D.

(iv) By (iii) of Lemma 2.1 $\overline{\phi(z)} = \phi(\bar{z})$ and $\overline{\psi(z)} = \psi(\bar{z})$ holds for all complex z. If z is real, then $\bar{z} = z$. So, if z is real, $\overline{\phi(z)} = \phi(z)$ and $\overline{\psi(z)} = \psi(z)$ hold.

Q.E.D.

Lemma 2.2. Let g be an entire function of a complex variable z. Then the only entire solution of the functional equation

$$g(2z) = 16g(z) \tag{11}$$

is given by $g(z) = d_4 z^4$ where d_4 is an arbitrary complex constant.

Proof. Since g is an entire function of a complex variable z, we can expand g in a power series at any point. Let its power series expansion at $z = 0$ be

$$g(z) = \sum_{n=0}^{+\infty} d_n z^n, \tag{12}$$

where each of $d_n (n = 0, 1, 2, \ldots)$ is a complex constant.

Substituting (12) into (11) and equating the coefficients of $z^n (n = 0, 1, 2, ...)$ yields

$$2^n d_n = 16 d_n \quad (n = 0, 1, 2, ...),$$

and so

$$d_n = 0 \quad (n = 0, 1, 2, 3, 5, 6, ...)$$

if $n \neq 4$.

Hence, by (12) we obtain

$$g(z) = d_4 z^4 . \tag{13}$$

Direct substitution shows that (13) satisfies our original equation (11).

Q.E.D.

Remark. About Lemma 2.2 see [3].

Lemma 2.3. Let h be an entire function of two complex variables x, y. Then the only entire solution of the functional equation

$$h(2x, 2y) = 16 h(x, y) \tag{14}$$

is given by $h(x,y) = d_{40}x^4 + d_{31}x^3y + d_{22}x^2y^2 + d_{13}xy^3 + d_{04}y^4$ where $d_{40}, d_{31}, d_{22}, d_{13}, d_{04}$ are arbitrary complex constants.

Proof. Since h is an entire function of two complex variables x, y, we can expand h in a double power series at any point. Let its power series expansion at $(x,y) = (0,0)$ be

$$h(x,y) = \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} d_{mn} x^m y^n , \tag{15}$$

where each of $d_{mn} (m = 0, 1, 2, \cdots; n = 0, 1, 2, \cdots)$ is a complex constant.

Substituting (15) into (14) and equating the coefficients of $x^m y^n (m = 0, 1, 2, \cdots; n = 0, 1, 2, \cdots)$ yields

$$2^{m+n} d_{mn} = 16 d_{mn} \ (m = 0, 1, 2, \cdots; n = 0, 1, 2, \cdots) ,$$

and so

$$d_{mn} = 0$$

if $m + n \neq 4$.

Hence we obtain $d_{mn} = 0$ except $d_{40}, d_{31}, d_{22}, d_{13}, d_{04}$.

Therefore, by (15) we obtain

$$h(x,y) = d_{40}\,x^4 + d_{31}\,x^3y + d_{22}\,x^2\,y^2 + d_{13}\,xy^3 + d_{04}\,y^4\,. \tag{16}$$

Direct substitution shows that (16) satisfies our original equation (14). Q.E.D.

Remark. About Lemma 2.3 see [3].

3. Proof Of The Theorem

We may now prove the theorem in Section 1.

We may assume that $f(z) \not\equiv 0$.

If we set $s = 0$, $t = 0$ in (4), then we obtain

$$f(0) = 0\,, \tag{17}$$

and so, by (9),

$$\phi(0) = 0 \quad \text{and} \quad \psi(0) = 0\,. \tag{18}$$

We show first that f is an even function of z. Setting $s = 0$ in (4) and using (17) yields

$$\left|f(it) + f(-it)\right| = 2\left|f(it)\right| \tag{19}$$

for all real t.

Replacing t by $-t$ in (19) and using (19) again yields

$$\left|f(it)\right| = \left|f(-it)\right| \tag{20}$$

for all real t.

By the triangle inequality for complex numbers we have

$$\left|f(it)\right| + \left|f(-it)\right| \geq \left|f(it) + f(-it)\right| \tag{21}$$

for all real t.

Combining (21) with (19), (20) yields

$$2\left|f(it)\right| = \left|f(it)\right| + \left|f(-it)\right| \geq \left|f(it) + f(-it)\right| = 2\left|f(it)\right| \tag{22}$$

for all real t.

Consequently, by (22) the equality

$$\left|f(it)\right| + \left|f(-it)\right| = \left|f(it) + f(-it)\right| \tag{23}$$

occurs for all real t.

By (23) and by a well-known fact there exists a real number $A(t)$ for each real t such that

$$A(t) > 0 \tag{24}$$

and

$$f(-it) = A(t)\, f(it) \tag{25}$$

hold in R.

(If $f(it_0) = f(-it_0) = 0$ $(t_0 \in R)$ (see (20)), then we adopt the convention that $A(t_0) = 1$.)

By (20), (25) and by the above convention we have

$$\left|A(t)\right| = 1 \tag{26}$$

for all real t.

By (24), (26) we obtain

$$A(t) = 1$$

for all real t, and so, by (25),

$$f(-it) = f(it)$$

for all real t.

Therefore, by the Identity Theorem we obtain

$$f(-z) = f(z)$$

for all complex z.

So f is an even function of a complex variable z.

We use the same notations as those in Lemma 2.1. By (7), (8), (10) we have

$$f(z) = \phi(z) + i\psi(z) \quad \text{and} \quad \overline{f(\overline{z})} = \phi(z) - i\psi(z) \tag{27}$$

for all complex z.

By (4) we obtain

$$\left|f(z) + f(\bar{z})\right|^2 = 4\left|f(s) + f(it)\right|^2,$$

and so

$$\left(f(z) + f(\bar{z})\right)\left(\overline{f(z)} + \overline{f(\bar{z})}\right) = 4\left(f(s) + f(it)\right)\left(\overline{f(s)} + \overline{f(it)}\right), \tag{28}$$

where $z = s + it (s, t \in R)$.

Substituting (27) into (28) and observing $\bar{i} = -i$ yields

$$\begin{aligned}&\left(\phi(z) + \phi(\bar{z}) + i(\psi(z) + \psi(\bar{z}))\right)\left(\overline{\phi(z)} + \overline{\phi(\bar{z})} - i\left(\overline{\psi(z)} + \overline{\psi(\bar{z})}\right)\right) \\ &= 4\left(\phi(s) + \phi(it) + i(\psi(s) + \psi(it))\right)\left(\overline{\phi(s)} + \overline{\phi(it)} - i(\overline{\psi(s)} + \overline{\psi(it)})\right).\end{aligned} \tag{29}$$

Since, by Lemma 2.1, we have
$\overline{\phi(z)} = \phi(\bar{z})$, $\overline{\phi(\bar{z})} = \phi(z)$, $\overline{\psi(z)} = \psi(\bar{z})$, $\overline{\psi(\bar{z})} = \psi(z)$, $\overline{\phi(s)} = \phi(s)$, $\overline{\phi(it)} = \phi(-it)$, $\overline{\psi(s)} = \psi(s)$ and $\overline{\psi(it)} = \psi(-it)$, by (29) we obtain

$$\begin{aligned}&\left(\phi(z) + \phi(\bar{z})\right)^2 + \left(\psi(z) + \psi(\bar{z})\right)^2 = 4\left(\phi(s) + \phi(it)\right. \\ &+ i\left(\psi(s) + \psi(it)\right)\left(\phi(s) + \phi(-it) - i(\psi(s) + \psi(-it))\right),\end{aligned}$$

and so

$$\begin{aligned}&\left(\phi(s + it) + \phi(s - it)\right)^2 + \left(\psi(s + it) + \psi(s - it)\right)^2 \\ &= 4\left(\phi(s) + \phi(it) + i(\psi(s) + \psi(it))\right)\left(\phi(s) + \phi(-it) - i\left(\psi(s) + \psi(-it)\right)\right)\end{aligned} \tag{30}$$

for all real s, t.

By (30) and by the Identity Theorem we obtain

$$\begin{aligned}&\left(\phi(x+y) + \phi(x-y)\right)^2 + \left(\psi(x+y) + \psi(x-y)\right)^2 \\ &= 4\left(\phi(x) + \phi(y) + i\left(\psi(x) + \psi(y)\right)\right)\left(\phi(x) + \phi(-y) - i(\psi(x) + \psi(-y))\right).\end{aligned} \tag{31}$$

for all complex x, y.

Since, as already proved, f is even, by Lemma 2.1 (i) ϕ and ψ are also even.

Hence we have

$$\phi(-y) = \phi(y) \quad \text{and} \quad \psi(-y) = \psi(y) \tag{32}$$

for all complex y.

By (31), (32) we have the functional equation

$$\Big(\phi(x+y) + \phi(x-y)\Big)^2 + \Big(\psi(x+y) + \psi(x-y)\Big)^2$$
$$= 4\Big(\phi(x) + \phi(y)\Big)^2 + 4\Big(\psi(x) + \psi(y)\Big)^2 . \tag{33}$$

If we set $y = x$ in (33), then we have

$$\Big(\phi(2x) + \phi(0)\Big)^2 + \Big(\psi(2x) + \psi(0)\Big)^2 = 16\phi(x)^2 + 16\psi(x)^2 \tag{34}$$

for all complex x.

By (18), (34) we obtain

$$\phi(2x)^2 + \psi(2x)^2 = 16(\phi(x)^2 + \psi(x)^2) \tag{35}$$

for all complex x.

Setting

$$g(x) = \phi(x)^2 + \psi(x)^2 \tag{36}$$

for all complex x and using (35) yields

$$g(2x) = 16g(x) \tag{37}$$

for all complex x.

By (37) and by Lemma 2.2 we obtain

$$g(x) = Ax^4 , \tag{38}$$

where A is a complex constant.

By (36), (38) we have

$$\phi(x)^2 + \psi(x)^2 = Ax^4 . \tag{39}$$

By (33) we obtain

$$\Big(\phi(x+y)^2 + \psi(x+y)^2\Big) + \Big(\phi(x-y)^2 + \psi(x-y)^2\Big) +$$
$$2\Big(\phi(x+y)\,\phi(x-y) + \psi(x+y)\,\psi(x-y)\Big)$$

$$= 4\left(\phi(x)^2 + \psi(x)^2\right) + 4\left(\phi(y)^2 + \psi(y)^2\right) + 8\left(\phi(x)\,\phi(y) + \psi(x)\,\psi(y)\right). \tag{40}$$

By (39), (40) we have

$$2\left(\phi(x+y)\,\phi(x-y) + \psi(x+y)\,\psi(x-y)\right) = 4Ax^4 + 4Ay^4 - A(x+y)^4 - A(x-y)^4 + 8\left(\phi(x)\,\phi(y) + \psi(x)\,\psi(y)\right). \tag{41}$$

If we set

$$h(x,y) = \phi(x)\,\phi(y) + \psi(x)\,\psi(y) \tag{42}$$

and

$$R(x,y) = 4Ax^4 + 4Ay^4 - A(x+y)^4 - A(x-y)^4, \tag{43}$$

then, by (41), we have

$$2h(x+y,\,x-y) = R(x,y) + 8h(x,y). \tag{44}$$

Replacing x,y by $x+y,\,x-y$ in (44) yields

$$2h(2x,2y) = R(x+y,\,x-y) + 8h(x+y,\,x-y). \tag{45}$$

By some simple calculations and by (43) we have

$$R(x+y,\,x-y) = -4R(x,y). \tag{46}$$

Substituting (44), (46) into (45) yields

$$2h(2x,2y) = -4R(x,y) + 4\left(R(x,y) + 8h(x,y)\right)$$

or

$$h(2x,2y) = 16h(x,y) \tag{47}$$

for all complex x,y.

By (47) and by Lemma 2.3 we obtain

$$h(x,y) = A_1 x^4 + A_2 x^3 y + A_3 x^2 y^2 + A_4 x y^3 + A_5 y^4, \tag{48}$$

where A_1, A_2, A_3, A_4, A_5 are complex constants.

Applying $\dfrac{\partial^5}{\partial x^3 \partial y^2}$ to both sides of (48) yields

$$\frac{\partial^5}{\partial x^3 \partial y^2} h(x,y) = 0 \tag{49}$$

for all complex x,y.

By (42), (49) we obtain

$$\phi'''(x)\,\phi''(y) + \psi'''(x)\,\psi''(y) = 0 \tag{50}$$

for all complex x,y.

Applying $\frac{\partial}{\partial y}$ to both sides of (50) yields

$$\phi'''(x)\,\phi'''(y) + \psi'''(x)\,\psi'''(y) = 0 \tag{51}$$

for all complex x,y.

Setting $y = x$ in (51) yields

$$\phi'''(x)^2 + \psi'''(x)^2 = 0 \tag{52}$$

for all complex x.

When x is real, by Lemma 1.1 (iv) $\phi(x)$ and $\psi(x)$ are also real, and so, so are $\phi'''(x)$ and $\psi'''(x)$. So by (52), when x is real

$$\phi'''(x) \equiv 0 \quad \text{and} \quad \psi'''(x) \equiv 0\,. \tag{53}$$

By (53) we obtain

$$\phi(x) = ax^2 + a_1x + a_2 \tag{54}$$

and

$$\psi(x) = bx^2 + b_1x + b_2\,, \tag{55}$$

where a, a_1, a_2, b, b_1, b_2 are real constants.

As already proved, $f(z)$ is an even function of a complex variable z. Consequently, by Lemma 1.1 (i) each of $\phi(z)$ and $\psi(z)$ is an even function of z. Hence we obtain

$$\phi'(0) = 0 \quad \text{and} \quad \psi'(0) = 0\,. \tag{56}$$

By (18), (54), (55), (56) we have

$$a_1 = a_2 = b_1 = b_2 = 0\,. \tag{57}$$

By (54), (55), (57) we have

$$\phi(x) = ax^2 \quad \text{and} \quad \psi(x) = bx^2 \tag{58}$$

when x is real.

By (58) and by the Identity Theorem we obtain

$$\phi(z) = az^2 \quad \text{and} \quad \psi(z) = bz^2 \tag{59}$$

for all complex z, where a,b are real constants. If we set $\gamma = a + ib$, then, by (27), (59) we obtain

$$f(z) = \gamma z^2,$$

Q.E.D.

where γ is a complex constant.

4. A Geometric Interpretation Of The Functional Equation (4) From The Standpoint Of Conformal Mapping (see [5], [7])

In this section we shall state a geometric interpretation of (4).

To this end we shall apply the following mapping property of $w = f(z) = z^2$.

Horizontal and vertical lines on the z-plane are mapped into an orthogonal family of confocal parabolas with common focus at $w = 0$ and with common principal axis on the real axis of the w-plane under the mapping function $w = f(z) = z^2$. Consider an arbitrary point $z = s + it$ on the z-plane where s,t are nonzero real numbers. Then, under the mapping function $w = f(z) = z^2$, the horizontal and vertical lines passing through the point $z = s + it$ on the z-plane are mapped on two parabolas with common focus F and with common principal axis on the real axis of the w-plane. Let the vertices of the above two parabolas be A,B and let the point of intersection of the common chord and the common principal axis AB be C. Then

$$\overline{AF} = \overline{BC}$$

holds.

Since the proof is easy, we omit its details

$$\left(\left| \frac{f(s+it) + f(s-it)}{2} \right| = \overline{CF},\ f(s) = s^2 \quad \text{and} \quad f(it) = -t^2 \right).$$

Acknowledgement

The author wishes to thank Professor J. Aczél for his helpful suggestions. This work was supported by the Natural Sciences and Engineering Research Council of Canada No. A-4012.

REFERENCES

1. Aczél, J., Lectures on functional equations and their applications. Academic Press, New York and London, 1966.
2. Aczél, J. and Dhombres, J., Functional equations in several variables. Cambridge University Press, Cambridge, New York, New Rochelle, Melbourne, Sydney, 1989.
3. Aczél, J. and Kiesewetter, H., Über die Reduktion der Stufe bei einer Klasse von Funktionalgleichungen. Publ. Math. Debrecen. 5 (1957), 348-363.
4. Aczél, J., The general solution of two functional equations by reduction to functions additive in two variables and with aid of Hamel bases. Glasnik Mat. 20 (1965), 65-73.
5. Aczél, J. and Haruki, H., Commentary to Einar Hille's collected works. (Edited by R.R. Kallman) The MIT Press, Cambridge, Mass., and London, England, 1975, pp. 651-658.
6. Baker, J.A., On quadratic functionals continuous along rays. Glasnik Mat. 3 (23) (1968), 215-229.
7. Carathéodory, C., Conformal representation. Cambridge University Press, 1963.
8. Davison, T.M.K., Röhmel's equation for quadratic forms. Aequationes Math. 34 (1987), 78-81.
9. Fischer, P. and Mokanski, J.P., A class of symmetric biadditive functionals. Aequationes Math. 23 (1981), 169-174.
10. Heuvers, K.J., A family of symmetric biadditive nonbilinear functions. Aequationes Math. 29 (1985), 14-18.
11. Hille, E., A Pythagorean functional equation. Ann. of Math. 24 (1923), 175-180.
12. Hille, E., A class of functional equations. Ann. of Math. 29 (1928), 215-222.
13. Kurepa, S., On quadratic forms. Aequationes Math. 34 (1987), 125-138.
14. Rätz, J., Quadratic functionals satisfying a subsidiary inequality. In General Inequalities 1 (Proc. First Internat. Conf. on General Inequalities, Oberwolfach. 1976). Birkhäuser, Basel-Stuttgart, 1978, pp. 261-270.
15. Robinson, R.M., A curious trigonometric identity. Am. Math. Monthly 64 (1957), 83-85.
16. Sato, T., On the functional equality $|f(x+iy)| = |f(x)|\,|f(iy)|$. J. College Arts Sci. Chiba Univ. 4, No. 2 (1963), 9-10.
17. Volkmann, P., Eine Charakterisierung der positiv definiten quadratischen Formen. Aequationes Math. 11 (1974), 174-182.

Hiroshi Haruki
Department of Pure Mathematics
University of Waterloo
Waterloo, Ontario, Canada

Constantin Carathéodory: An International Tribute (pp. 489–494)
edited by Th. M. Rassias

The Characterization of Determinant and Permanent Functions by the Binet-Cauchy Theorem

Konrad J. Heuvers and Daniel S. Moak

1 INTRODUCTION

Throughout the paper K will denote a field of characteristic 0. A function $\mu : \mathrm{K} \to \mathrm{K}$ satisfying the multiplicative form of Cauchy's functional equation,

$$\mu(xy) = \mu(x)\mu(y) \tag{1}$$

for all $x, y \in \mathrm{K}$, is called a multiplicative function on K.

Let $M_n(\mathrm{K})$ denote the set of all square $n \times n$ matrices over K and let $M_{n\times n}(\mathrm{K})$ denote the set of all rectangular $m \times n$ matrices over K. For any square matrix $A = (a_{ij}) \in M_n(\mathrm{K})$ two matrix functions from $M_n(\mathrm{K})$ to K are defined in terms of the two linear characters of the symmetric group S_n. The permanent of A is defined in terms of the identically one character via

$$\operatorname{per} A = \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} \tag{2}$$

and the determinant function is defined in terms of the alternating character $\zeta(\sigma) = \pm 1$ via

(3) $$\det A = \sum_{\sigma \in S_n} \zeta(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \quad [10,11].$$

If $f : M_n(\mathrm{K}) \to \mathrm{K}$ satisfies the Cauchy equation

(4) $$f(AB) = f(A)f(B)$$

for all $A, B \in M_n(\mathrm{K})$ it is well known that $f(A) = \mu(\det A)$ for all $A \in M_n(\mathrm{K})$ where μ is an arbitrary multiplicative function on K [1,2,3,8,9].

2 NOTATION

In order to simplify our notation we have adopted a formal "product" notation for repeated terms inside $n-$tuples. Accordingly,

$$(\underbrace{x_1, \ldots, x_1}_{s_1}, \ldots, \underbrace{x_n, \ldots, x_n}_{s_n})$$

will be denoted by $(x_1^{s_1}, \cdots, x_n^{s_n})$ or (x^s) where s_i is the number of times that x_i appears together inside the $n-$tuple. If $s_i = 0$ then x_i does not appear. Thus, each s_i is a non-negative integer and for the $n-$tuple $s = (s_1, \ldots, s_n)$ of non-negative integers we will let $|s| = s_1 + \cdots + s_n$. Let $A = [a_1, \ldots, a_m]$ be an $n \times m$ matrices with $n \times 1$ columns a_j, $j = 1, \ldots, m$, and let $B = \langle b_{(1)}, \ldots, b_{(m)} \rangle$ be an $m \times n$ matrix with $1 \times n$ rows $b(i)$, $i = 1, \ldots, m$, where $n \leq m$. Let $A^s = [a_1^{s_1}, \ldots, a_m^{s_m}]$ and $B_s = \langle b_{(1)}^{s_1}, \ldots, b_{(m)}^{s_m} \rangle$ for $|s| = n$ denote the square $n \times n$ matrices corresponding to s which are formed from A and B.

Let $Z_+ = \{0, 1, 2, \ldots\}$ be the set of non-negative integers. If $\alpha = (\alpha_1, \ldots, \alpha_q) \in Z_+^q$ is a $q-$tuple of non-negative integers we let $|\alpha| = \alpha_1 + \cdots + \alpha_q$. For an $m-$tuple $s = (s_1, \ldots, s_m) \in Z_+^m$ let $s! = s_1! \ldots s_m!$ and if $|s| = n$ then $\binom{n}{s}$ denotes the multinomial coefficient

$$\frac{n!}{s!} = \frac{n!}{s_1! \ldots s_m!} = \binom{n}{s_1 \ldots s_m} .$$

If J denotes the square $n \times n$ matrix with all entries one then $E = (\frac{1}{n})J$ is the square $n \times n$ matrix with all entries $\frac{1}{n}$. If $\sigma \in S_n$ and $A \in M_n(\mathrm{K})$ is the square matrix $A = [a_1, \ldots, a_n] = \langle a_{(1)}, \ldots, a_{(n)} \rangle$ with columns a_j and rows $a_{(i)}$ then $A^\sigma = [a_{\sigma(1)}, \ldots, a_{\sigma(n)}]$ and $A_\sigma = \langle a_{\sigma(1)}, \ldots, a_{\sigma(n)} \rangle$. Let e_i, $i = 1, \ldots, n$, denote the columns and $\mathrm{e}_{(j)}$, $j = 1, \ldots, n$, the rows of the $n \times n$ identity matrix. Then $D = [d_1\mathrm{e}_1, \ldots, d_n\mathrm{e}_n] = \langle d_1\mathrm{e}_{(1)}, \ldots, d_n\mathrm{e}_{(n)} \rangle$ is the diagonal matrix with diagonal entries $d_1, \ldots, d_n$.

3 BINET-CAUCHY THEOREMS

Both the determinant and permanent functions satisfy Binet-Cauchy Theorems. Let $A \in M_{n\times m}(\mathrm{K})$ and $B \in M_{m\times n}(\mathrm{K})$ for $n \leq m$. Then for the determinant function its Binet-Cauchy Theorem is given by

$$\det(AB) = \sum_{|s|=n} \det A^s \det B_s, \tag{5}$$

where each $s_i = 0$ or 1 [10,9]. For the permanent function its Binet-Cauchy Theorem is given by

$$\operatorname{per}(AB) = \frac{1}{n!} \sum_{|s|=n} \binom{n}{s} \operatorname{per} A^s \operatorname{per} B_s, \tag{6}$$

where

$$\binom{n}{s} = \frac{n!}{s_1! \dots s_m!} \quad [10,11].$$

For $f : M_n(\mathrm{K}) \to \mathrm{K}$, $n \leq m$, $A \in M_{n\times m}(\mathrm{K})$, and $B \in M_{m\times n}(\mathrm{K})$

$$f(AB) = \frac{1}{n!} \sum_{|s|=n} \binom{n}{s} f(A^s) f(B_s) \tag{7}$$

will be called the Binet-Cauchy functional equation. It is the intention of this article to summarize the solutions of this equation.

4 SOLUTIONS OF THE BINET-CAUCHY FUNCTIONAL EQUATION FOR SQUARE MATRICES

In 1988 Heuvers, Cummings, and K. P. S. Bhaskara Rao [4] established the following result for square matrices A and B.

Theorem 1. *If $f : M_n(\mathrm{K}) \to \mathrm{K}$ is non-constant and satisfies the Binet-Cauchy functional equation (7) for $A, B \in M_n(\mathrm{K})$ and if $f(E) \neq 0$ where $E = (\frac{1}{n})J$ then the general solution of (7) is given by $f(A) = \phi(\operatorname{per} A)$ for $A \in M_n(\mathrm{K})$ where ϕ is an isomorphism from K into K.*

In 1989 Heuvers and Moak [6] solved (7) for square matrices A and B when $f(E) = 0$ and f is non-constant.

Theorem 2. *If $f : M_n(\mathrm{K}) \to \mathrm{K}$ is non-zero and satisfies the Binet-Cauchy functional equation (7) for $A, B \in M_n(\mathrm{K})$ and if $f(E) = 0$, then the general solution to (7) is given by $f(A) = \mu(\det A)$ for $A \in M_n(\mathrm{K})$ where μ is a non-constant multiplicative function on* K.

Thus for $f(E) = 0$ and $f \neq 0$ the solution of (7) for square matrices A and B is the same as the solution of (4).

5 SOLUTIONS OF THE FUNCTIONAL EQUATION FOR SQUARE AND RECTANGULAR PAIRS OF MATRICES.

In 1964 S. Kurepa [9] showed the following for matched pairs A and B of square or rectangular matrices one dimension away from being square.

Theorem 3. *If a non-zero $f : M_n(\mathrm{K}) \to \mathrm{K}$ satisfies*

$$f(AB) = \sum_{|s|=n} f(A^s) f(B_s) \tag{8}$$

where each $s_i = 0$ or 1, $n \leq m \leq n+1$, $A \in M_{n\times m}(\mathrm{K})$ and $B \in M_{m\times n}(\mathrm{K})$ then $f(A) = \phi(\det A)$ for $A \in M_n(\mathrm{K})$ where ϕ is an isomorphism of K *into* K.

It was this pioneering work which initiated the investigation which led to Theorem 1 and the next result.

In 1989 Heuvers and Moak [5] showed the following for matched pairs A and B of square matrices or rectangular matrices one dimension away from being square.

Theorem 4. *If $f : M_n(\mathrm{K}) \to \mathrm{K}$ is non-zero and satisfies (7) for $A, B \in M_n(\mathrm{K})$ and for $A \in M_{n\times(n+1)}(\mathrm{K})$ and $B \in M_{(n+1)\times n}(\mathrm{K})$, and if $f(E) = 0$, then $f(A) = \phi(\det A)$ for $A \in M_n(\mathrm{K})$ where ϕ is an isomorphism of* K *into* K.

Thus, the solution of (7) under the same condition as in Theorem 3 leads to the same solution.

Remark.

In equation (7) if each $s_i = 0$ or 1 then $\binom{n}{s} = 1$ so (7) and (8) are the same. If furthermore A and B are square matrices then (7) reduces to (4).

6 SOLUTION OF THE FUNCTIONAL EQUATION FOR RECTANGULAR MATRICES.

In 1989 Heuvers and Moak [7] obtained the following result for $A \in M_{n\times(n+r)}(\mathbf{K})$ and $B \in M_{(n+r)\times n}(\mathbf{K})$ for a fixed $r \geq 0$.

Theorem 5. *Let $f : M_n(\mathbf{K}) \to \mathbf{K}$ be a non-constant function such that for a fixed $r \geq 0$ f satisfies (7) for $A \in M_{n\times(n+r)}(\mathbf{K})$ and $B \in M_{(n+r)\times n}(\mathbf{K})$. Then the general solution of (7) for $A \in M_n(\mathbf{K})$ is given by $f(A) = \phi(\operatorname{per} A)$ if $f(E) \neq 0$, $f(A) = \phi(\det A)$ if $f(E) = 0$ and $r \geq 1$, and $f(A) = \mu(\det A)$ if $f(E) = 0$ and $r = 0$ where ϕ is an isomorphism of $\mathbf{K}$ into $\mathbf{K}$ and μ is a non-constant multiplicative function.*

On of the major tools used to prove Theorem 1, Theorem 2, Theorem 4, and Theorem 5 was the following theorem of multinomial type which was proved in 1988 by Heuvers, Cummings, and K. P. S. Bhaskara Rao [4, Theorem 3]

Theorem 6. *Let X be a non-empty set and let V be a vector space over $\mathbf{K}$. Let $\Phi, \Psi : X^n \to V$ be functions satisfying*

$$\Psi(x_1, \ldots, x_n) = \sum_{|s|=n} \binom{n}{s} \Phi(x_1^{s_1}, \ldots, x_n^{s_n}) \tag{9}$$

for all $(x_1, \ldots, x_n) \in X^n$. Then

$$\Phi(x_1, \ldots, x_n) = \sum_{|s|=n} c_s \Psi\left(x_1^{s_1}, \ldots, x_n^{s_n}\right) \tag{10}$$

for fixed constant c_s depending only on the $\binom{n}{s}$.

In particular if

$$\sum_{|s|=n} \binom{n}{s} \Phi_1\left(x_1^{s_1}, \ldots, x_n^{s_n}\right) = \sum_{|s|=n} \binom{n}{s} \Phi_2\left(x_1^{s_1}, \ldots, x_n^{s_n}\right) \tag{11}$$

then

$$0 = \sum_{|s|=n} \binom{n}{s} \left(\Phi_1\left(x_1^{s_1}, \ldots, x_n^{s_n}\right) - \Phi_2\left(x_1^{s_1}, \ldots, x_n^{s_n}\right)\right) . \tag{12}$$

By Theorem 6 it then follows that $\Phi_1 - \Phi_2 = 0$ or $\Phi_1 = \Phi_2$.

From these results we see that (7) the Binet-Cauchy functional equation is the source of the common properties of $\det A$ and $\operatorname{per} A$. The value of $f(E)$ is sufficient to distinguish between the two functions. Thus, equation (7) characterizes these important functions.

REFERENCES

[1] J. Aczèl, **Lectures on Functional Equations and Their Applications**, Academic Press, New York, 1966.

[2] J. Aczèl and J. Dhombres, **Functional Equations in Several Variables**, Cambridge University Press, New York, 1989.

[3] D. Ž. Djoković, On the homomorphisms of the general linear group, **Aequationes Math.** 4: 99-102 (1970).

[4] K. J. Heuvers, L. J. Cummings, and K. P. S. Bhaskara Rao, A characterization of the permanent function by the Binet-Cauchy Theorem, **Linear Algebra Appl.** 101: 49-72 (1988).

[5] K. J. Heuvers and D. S. Moak, The Binet-Cauchy functional equation and non-singular multi-indexed matrices, to appear in **Linear Algebra Appl.**

[6] K. J. Heuvers and D. S. Moak, The solution of the Binet- Cauchy functional equation for square matrices, to appear in **Discrete Math.**

[7] K. Heuvers and D. S. Moak, The Binet-Pexider functional equation for rectangular matrices, to appear in **Aequationes Math.**

[8] M. Hosszú, Megjeqyzések mátrixok skalár értékü multiplikativ függrényéröl (Hungarian), Nehézipari Müszaki Egyetem Közl. 5(1960), 173-177.

[9] S. Kurepa, On a characterization of the determinant, **Glas. Mat.-Fiz. Astr.** 14:97-113 (1959).

[10] M. Marcus, **Finite Dimensional Multilinear Algebra**, Part I, Marcel Dekker, New York, 1972.

[11] H. Minc, **Permanents**, Addison-Wesley, Reading, Mass., 1978.

Konrad J. Heuvers and Daniel S. Moak
Michigan Technological University
Houghton, Michigan 49931
USA

Constantin Carathéodory: An International Tribute (pp. 495–500)
edited by Th. M. Rassias

PROBLEMS IN THE THEORY OF UNIVALENT FUNCTIONS

Liubomir Iliev

To my teacher on conformal mappings, Constantin Carathéodory

1. Let us denote by $S_k, k = 1, 2, \ldots (S_1 = S)$ the class of functions

$$(S_k) \qquad \begin{aligned} f_k(z) &= z + c_1^{(k)} z^{k+1} + c_2^{(k)} z^{2k+1} + \ldots \\ &= z + c_{k+1} z^{k+1} + c_{2k+1} z^{2k+1} + \ldots \end{aligned}$$

which are regular, univalent and k-symmetric in the disc $D : |z| < 1$.

In 1928 Szegö showed that if $f(z) \in S, |z_1| < 1, |z_2| < 1, z_1 \neq z_2$, then (see [1]):

$$(S_z) \qquad \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| \geq \left(\frac{1 - |z_2|}{1 + |z_2|} \right)^2 \frac{|1 - \bar{z}_2 z_1|}{(|z_1 - z_2| + |1 - \bar{z}_2 z_1|)^2} .$$

With the help of this inequality he proved the following

Theorem (S_z). The partial sums of a function $f(z) \in S$:

$$\sigma_n(z) = z + c_2 z^2 + \ldots + c_n z^n \, , \, n = 1, 2, \ldots$$

are univalent in the disc $|z| < 1/4$. The constant $1/4$ cannot be substituted by a greater one.

Later on, Iliev [2] (1949) proved

Theorem (I). If $f_k(z) \in S_k, k = 1, 2, \ldots$ and $|z_1| \leqq r < 1, |z_2| \leqq r, z_1 \neq z_2$, then

$$\left| \frac{f_k(z_1) - f_k(z_2)}{z_1 - z_2} \right| \geqq \frac{1 - r^2}{(1 + r^k)^{4/k}} . \tag{I}$$

Inequality (I) is exact for $k = 1$ and $k = 2$.

Using this inequality Iliev proved in [2], [3] and in [4], [5] respectively the following theorems.

Theorem (I_1). If

$$f_2(z) = z + c_3 z^3 + \ldots \in S_2 ,$$

then its partial sums

$$\sigma_n^{(2)}(z) = z + c_3 z^3 + \ldots + c_{2n+1} z^{2n+1} , n = 1, 2, \ldots$$

are univalent in the disc $|z| < 1/\sqrt{3}$. The constant $1/\sqrt{3}$ cannot be substituted by a greater one.

Theorem (I_2). If

$$f_3(z) = z + c_1^{(3)} z^4 + \ldots \in S_3 ,$$

then its partial sums

$$\sigma_n^{(3)}(z) = z + c_1^{(3)} z^4 + \ldots + c_n^{(3)} z^{3n+1} , n = 1, 2, \ldots$$

are uinvalent in the disc $|z| < \sqrt[3]{3}/2$. This constant cannot be replaced by a greater one.

Using the Szegö's inequality, in 1939 V. Levin [6] proved that in the case $n > 16$ the partial sums $\sigma_n^{(1)}(z)$ are functions univalent in the disc $|z| < 1 - 6 \ln n/n$.

By means of inequality (I) Iliev [5], [7] proved:

(A) For $n > 14$ the partial sums $\sigma_n^{(1)}(z)$ are univalent in the disc $|z| < 1 - 4\ln n/n$;

(B) For $n > 11$ the partial sums $\sigma_n^{(2)}(z)$ are univalent in the disc $|z| < (1 - 3\ln n/n)$

(C) The partial sums $\sigma_n^{(3)}(z)$ (cf. Th. (I_2)) are univalent in the disc $|z| < \left\{1 - \frac{8}{3}\frac{\ln\theta(n+1)}{n+1}\right\}^{1/3}$, $\theta = 7,96^{3/8}.3^{1/4}.2^{7/8}$.

Further, Iliev found discs in which the polynomials $\sigma_n^{(1)}(z)/z, \sigma_m^{(2)}(z)/z$ and $\sigma_n^{(3)}(z)/z$ do not vanish.

These applications show the importance of the results of the kind of Theorems (Sz), (I).

2. Let $L(z_1, z_2)$ be a curve $z = z(s), 0 \leq s \leq \bar{s}, z_1 = z(0), z_2 = z(\bar{s}), |z_1| < |z_2|$ for which $z'(s)$ and $r'(s) = |z(s)|'$ exist and are continuous functions except for a finite number of values of s. Here the parameter s denotes the length of the arc.

Denote by $\mathcal{L}(z_1, z_2, f)$ the image of $L(z_1, z_2)$ by means of $f(z) \in S$. The denotations $\overline{L}(z_1, z_2)$ and $\overline{\mathcal{L}}(z_1, z_2, f)$ are used for the lengths of $L(z_1, z_2)$ and $\mathcal{L}(z_1, z_2, f)$, respectively.

In [8], [9] Iliev established the following theorems.

Theorem I. If $f(z) \in S$ and $|z_1| < |z_2| < 1$, then

$$\frac{1 - |z_1||z_2|}{(1 + |z_1|)^2(1 + |z_2|)^2} \leq \frac{\overline{\mathcal{L}}(z_1, z_2, f)}{\overline{L}(z_1, z_2)} \leq \frac{1 - |z_1||z_2|}{(1 - |z_1|)^2(1 - |z_2|)^2},$$

where the upper estimate is true provided $r'(s) \geq 0$.

For $|z| \leq r < 1$ we obtain

Theorem I*. If $f(z) \in S$ and $|z_1| < |z_2| \leq r < 1$, then

$$\frac{1 - r}{(1 + r)^3} = \frac{\overline{\mathcal{L}}(z_1, z_2, f)}{\overline{L}(z_1, z_2)} = \frac{1 + r}{(1 - r)^3},$$

where the upper estimate is true provided $r'(s) \geqq 0$.

As a corollary we obtain the following

Theorem Ī. If $f(z) \in S$ and $|z_1| < |z_2| \leqq r < 1$, then

$$\frac{1-|z_1||z_2|}{(1+|z_1|)^2(1+|z_1|)^2} \leqq \left|\frac{f(z_1)-f(z_2)}{z_1-z_2}\right| \leqq \frac{1-|z_1||z_2|}{(1-|z_1|)^2(1-|z_2|)^2},$$

where the left inequality holds if the segment joining the points $f(z_1)$ and $f(z_2)$ lies entirely in the image $f(D)$ of the unit disc by means of $f(z)$, while the right inequality is true if, on the segment joining z_1 and z_2, $|z|$ only increases or only decreases.

If $f(z) \in S$ is a convex function, then under the same conditions the following inequalities hold:

$$\frac{1}{(1+|z_1|)(1+|z_2|)} \leqq \left|\frac{f(z_1)-f(z_2)}{z_1-z_2}\right| \leqq \frac{1}{(1-|z_1|)(1-|z_2|)}.$$

If $f_k(z) \in S_k$, then under the conditions of Theorem I we have the inequalities:

$$\frac{1}{|z_1|-|z_2|}\int_{|z_1|}^{|z_2|}\left(\frac{1-r^k}{1+r^k}\right)^3 \frac{dr}{(1-r^k)^{2/k}} \leqq \left|\frac{f_k(z_1)-f_k(z_2)}{z_1-z_2}\right|$$

$$\leqq \frac{1}{|z_2|-|z_1|}\int_{|z_1|}^{|z_2|}\left(\frac{1+r^k}{1-r^k}\right)^3 \frac{dr}{(1+r^k)2/k},$$

and if $f_k(z) \in S_k$ is, in addition, a convex function, then under the same conditions we get:

$$\frac{1}{|z_2|-|z_1|}\int_{|z_1|}^{|z_2|}\frac{dr}{(1+r^k)^{2/k}} \leqq \left|\frac{f_k(z_1)-f_k(z_2)}{z_1-z_2}\right|$$

$$\leqq \frac{1}{|z_2|-|z_1|}\int_{|z_1|}^{|z_2|}\frac{dr}{(1-r^k)^{2/k}}.$$

For these theorems to have corresponding applications, it is necessary that some of the conditions in their statements could be discharged. Thus, the following problem arises.

Problem I. Is the condition $r'(s) \geqq 0$ necessary in Theorems I, I*?

For $k > 3$ the following problem remains open.

Szegö supposed that

$$\left|c_n^{(k)}\right| = o(n^{-1+2/k}) \text{ as } n \to \infty .$$

This estimate was suggested by the Littlewood's inequality

$$\left|c_{n-1}^{(1)}\right| = |c_n| < en .$$

For $k = 2$ Littlewood and Paley [10] established that $\left|c_n^{(2)}\right| = o(1)$. V. Levin [11] proved this estimate for $k = 3$. On the contrary, as it was shown by Littlewood [12], the assumption is not true for $k > 3$, even if $f_k(z)$ is bounded in the unit disc.

Therefore, we can calculate that there exist three positive constants A_1, A_2, A_3, not depending on n, such that for any n the following inequality

$$\left|c_n^{(k)}\right| \leq A_k n^{-1+2/k}, k = 1, 2, 3$$

holds. By A_1, A_2, A_3 we denote the smallest of these constants.

On proving the Bieberbach conjecture, it was established that $A_1 = 1$. Further, according to V. Levin, $A_2 < 3,4$. K. Joh found that $A_3 < 7,96$. In this manner, the Bieberbach conjecture concerning the coefficients of univalent functions is extended as follows: Which are the exact values of A_1, A_2, A_3? After the conjecture has been established (i.e., $A_1 = 1$) the following problem still remains open.

Problem II. Find the exact values of A_2 and A_3.

As for the applications of Theorems I, I* and $\bar{\text{I}}$, it is necessary to solve the following

Problem III. Which is the order of $\left|c_n^{(k)}\right|$ for $k > 3$?

References

1. G. Szegö, Math. Ann. **100** (1928), 188–201.
2. Soviet Math. Dokl. (4) **69** (1949), 491–494.

3. L. Iliev. Comptes Rendus Acad. Bulg. Sci. (1) **2** (1949), 21–24.
4. Soviet Math. Dokl. (1) **84** (1952), 9–12.
5. Soviet Math. Dokl. (4) **100** (1955), 621–622.
6. V. Levin. Jahresbericht der Deutschen Mathematiker Vereinigung **42** (1939), 68–70.
7. Soviet Math. Dokl. (1) **70** (1950), 9–11.
8. L. Iliev. Banach Center Publications, vol. **11**, Warsaw (1983), 89–100.
9. L. Iliev, Pliska, Studia Math. Bulg. **4** (1981), 137–141.
10. J. E. Littlewood and R. E. A. C. Paley, Journal London Math. Soc. **7** (1932), 167–169.
11. V. Levin, Math. Z. **38** (1933), 306–311.
12. J. E. Littlewood, Quart. J. Math. (Oxford Ser.) **9** (1938), 14–20.

Liubomir Iliev
Institute of Mathematics
Bulgarian Academy of Sciences
ul. Acad. G. Bonchev, Block 8
P. O. Box 373, 1113 Sofia
Bulgaria

Constantin Carathéodory: An International Tribute (pp. 501–526)
edited by Th. M. Rassias

SYSTEMS DEVELOPMENT SIMULATION PROBLEMS AND C. CARATHEODORY'S CONCEPTS

V. V. Ivanov

Recently, the necessity of mathematical simulation multiple precision was one of the reasons for integro-functional models creation and their extensive propagation. The problem arose of transference and development of the given results in the differential equations theory and the optimization theory on the basis of differential models for the case of the more general models. The present article has been devoted to the generalization of several well-known results, including that of C. Caratheodory, for the case of simulation of the developing (evolutionary) system (DS) with the prehistory. In addition the appropriate software and some applications were described in brief.

1. On Mathematical Models of Systems Development

According to [3], any DS to which one can practically refer any natural DS and any artificial DS being created or already created by human beings and functioning with their participation, contains the following main specific features:

(i) the subsystem for realization of the internal functions of the system perfection as a whole including itself, that is self-perfection of the subsystem perfection;

(ii) the subsystem for realization of the external functions of interaction with the environment;

(iii) the inflow of deficient resources from the outside;

(iv) allocation of the system resources among their internal and external functions;

(v) the out of date products and the prehistory of DS.

The formalized representation of the above-mentioned features results

in the base mathematical model (m.m.) of DS:

$$m(t) = \int_0^t \alpha(t,\tau)\lambda(t,\tau)y(\tau)m(\tau)d\tau\,, 0 \le y(\tau), \lambda(t,\tau) \le 1\,,$$

$$c(t) = \int_{t_0}^t \beta(t,\tau)\mu(t,\tau)[1-y(\tau)]m(\tau)d\tau\,, 0 \le \mu(t,\tau) \le 1\,,$$

$$P(t) = \int_0^t \{\lambda(t,\tau)y(\tau) + \mu(t,\tau)[1-y(\tau)]\}m(\tau)d\tau,$$

$$M(t) = \int_0^t m(\tau)d\tau\,,$$

$$G(t) = M(t) - P(t), \hat{f}(t) \ge f(t) = m(t) + c(t)\,, t \ge t_0 > 0\,, \qquad (1)$$

where $m(t)$ is the rate of creation of the first kind of new products (resources) number at the time instant t which provides the fulfilment of the internal functions of DS, that is restoration of itself and creation of the second kind products; $y(\tau)m(\tau)$ is a share of $m(\tau)$ for fulfilment of internal functions in the subsystem A of restoration and perfection of the system as a whole; $\lambda(t,\tau)$ is a relative share of the intensity of $y(\tau)m(\tau)$ products utilization at the instant t; $\alpha(t,\tau)$ is the efficiency index for functioning of the subsystem A along the channel $\lambda(t,\tau)y(\tau)m(\tau) - m(t)$, i.e., the number of units of $m(t)$ created in the unit of time starting from the instant t per one unit of $\lambda(t,\tau)y(\tau)m(\tau)$; $c(t)$ is the rate of creation of the second kind of new products number at the instant t which provides the realization of the external functions of DS; $\mu(t,\tau)$ and $\beta(t,\tau)$ are similar to λ and α, respectively but for the subsystem B of creation of the second kind products; $P(t)$ is the total number of the first kind products functioning at the instant t; $M(t)$ is the total number of the first kind products to be created during the time t; $G(t)$ is the total number of the out-of-date products at the time t; $\hat{f}(t)$ is the rate of the resources inflow from the outside ($m(t)$ and $c(t)$ are measured in units of $\hat{f}(t)$; t_0 is the starting point of modelling; $[0, t_0]$ is the prehistory of DS for which all the functions are given (their values will be noted by the same symbols but with index "0", e.g. $m(t) \equiv m_0(t), t \in [0, t_0]$).

It is obvious that all relations (1) are faithful representations by definition. And here, the functions α and β can depend on $m, c, \lambda, \mu, y, P, f$, in the general case. Thus in the general case, m.m. (1) is a system of nonlinear integro-functional relations which consists of 5 equalities and 9 inequalities connecting 14 values, namely: $m, c, \lambda, \mu, y, P, t, \alpha, t_0, \beta, M, G, \hat{f}, 0$.

The typical suitable examples used to interpret all the above-mentioned values can be as follows:

1. The economics as a whole. Then (see [3],[6]) A is a subsystem (a group) of the capital goods industry and B is a subsystem of the consumer goods industry; $m(t)$ is the rate of production of the new work places (WP) number in A and B; $\alpha(t,\tau)$ is labour productivity in the group A, i.e., the quantity of WP created for the unit of time starting from the time instant t by one worker from the group A at WP created at the time instant τ; $\beta(t,\tau)$ is labour productivity at WP created in the group B at the time τ; $P(t)$ is the total number of functioning WP at the time t, which can be equal to the quantity of labour resources; $G(t)$ is the total number of obsolete (or lying in reserve) WP at the time t; $\hat{f}(t)$ is the rate of the inflow into economics from the outside, e.g., from the biosphere or the cosmos [6].

2. The biosphere. Then [6] A is a subsystem of re-creation of the living substance of the planet, mainly of the phytomass by way of photosynthesis; B is a subsystem of creation of the so-called bioboned substance, mainly of the oxygen; $m(t)$ is the rate of creation of the new living substance quantity; $P(t)$ is the total quantity of the functioning living substance at the time t; $G(t)$ is the total quantity of the dead substance for the time t which is mainly equal to the humus by mass; α is the specific rate of reproduction of $m(t)$; β is the specific rate of production of $c(t)$ and so on.

We should note three special cases of m.m. (1) when λ and μ have the form

$$\lambda(t,\tau) = \lambda(t-\tau), \mu(t,\tau) = \mu(t-\tau); \tag{2}$$

$$\lambda(t,\tau), \mu(t,\tau) = \begin{cases} 0, & 0 \le \tau < a(t), \\ 1, & t \le \tau \ge a(t); \end{cases} \tag{3}$$

$$\lambda(t,\tau) = \begin{cases} 0, 0 \le \tau < a_1(t), \\ \lambda(t-\tau), t \ge \tau \ge a_1(t); \end{cases} \quad \mu(t,\tau) = \begin{cases} 0, 0 \le \tau < a_2(t), \\ \mu(t-\tau), t \ge \tau \ge a_2(t) \end{cases} \tag{4}$$

The case (2) corresponds to the stationary process of the intensity; the case (3) means that the products created previous to a certain temporal threshold $a(t), a(t) < t$, at the instant t, are never used but those created after $a(t)$ are used entirely; the case (4) extends the previous ones.

Instead of the relations (1), in particular, we shall have

$$m(t) = \int_{a(t)}^{t} \alpha(t,\tau) y(\tau) m(\tau) d\tau, \tag{5}$$

$$P(t) = \int_{a(t)}^{t} m(\tau)d\tau\,, \tag{6}$$

$$c(t) = \int_{a(t)}^{t} \beta(t,\tau)[1 - y(\tau)]m(\tau)d\tau\,, \tag{7}$$

$$G(t) = \int_{0}^{a(t)} m(\tau)d\tau\,, \tag{8}$$

$$0 \le y(\tau) \le 1\,, a(t) < t, t \ge t_0 > a(t_0) = 0\,, \tag{9}$$

and maybe

$$\frac{da}{dt} \ge 0\,, t \ge 0\,. \tag{10}$$

Denoting $P'(t)$ by $p(t)$ we convert (5),(6) to the form (assuming a necessary smoothness of m and a and $m(a) \neq 0$):

$$m(t) = \int_{a(t)}^{t} \alpha(t,\tau)y(\tau)m(\tau)d\tau\,, \frac{da}{dt} = \frac{m(t) - p(t)}{m(a)}\,. \tag{11}$$

As one can see, even in the simplified formulation, the m.m. (11) is reduced to the relations in which along with nonlinear integral equations of the unusual form (where a variable lower bound $a(t)$ can be the desired unknown function) there appears the nonlinear differential-difference (functional) equation.

It is not hard to introduce different additions and generalizations of the m.m. (1) [3], [6], [15], [17]. Really, n-products mathematical model of DS, $n > 2$, can be formally written almost in the same form (1) under condition that $m(t), c(t), p(t)$ and $G(t)$ are the vector-functions and $\alpha, \beta, \lambda, \mu, y$ are the appropriate functional matrices (1 is the identity matrix and the inequalities for vectors and matrices signify inequalities of the same name for all of their appropriate components). However, the relation $\hat{f}(t) \ge f(t) = m(t) + c(t)$ constitutes the exception now. It should be replaced by

$$\hat{f}(t) \ge f(t) = \sum_{i=1}^{r} m_i(t) + \sum_{k=1}^{s} c_k(t)\,, r + s = n\,.$$

It is also not hard to describe a continuous m.m. of DS in the simular form considering t and τ as many-dimensional independent variables and examining the appropriate integrals as multivariate ones. A stochastic similarity of (1) can be obtained by considering α and β as functions of a

random factor ω. A discrete similarity of (1) can also be represented in the form (1) if the integrals of (1) are understood in the sense of Stieltjes.

It is easy to show that we can obtain [3] a greate many of the well-known m.m. as the special cases of (1) by means of selection of the functions α and β. We shall dwell on the connection of m.m. in question with the classical models, among them the models [1], Bd 1. Everybody is familiar with the approach of the so-called "black-box" when only the input $X = (x_1, \ldots, x_n)$ and the output $Y = (y_1, \ldots, y_p)$ of a dynamic system are given. We have in the linear approximation

$$Y(t) = \int_{t-T}^{t} K(t,\tau)X(\tau)d\tau, \tag{12}$$

where T is the upper bound for all transients termination time, K the matrix of the pulse transition functions $K_{ji}(t,\tau)$ which are the response functions of a system when $x_i(\tau) = \delta(\tau - t)$ along the channel $i - j$ at the input i, where δ is Dirac's δ-function, and at the rest of inputs all $x_K(\tau) \equiv 0$. Nonlinear dynamic system can also be represented as (12) but K will depend on X. The m.m. (1) deals with the so-called "grey box" when the structure of a dynamic system is partly revealed. Indeed it is possible to say that in the case (1) the matrix K has been factored into three factors α, λ and y or β, μ and $1 - y$ such that each of them has its applied sense. In addition, several of the outputs of the system in the case (1) have served as its inputs. At last, due to functions $P(t), G(t)$ and $\bar{f}(t)$ we deal with the so-called open dynamic system in the case (1). The essential difference consists more in that all the values in (1) are non-negative by definition and a diminution of the output values is regulated not by a sign but for example, by the rate of $a(t)$ growth in the relations (7)–(11).

In the case of adiabatic quasi-static processes the variation of a system's entropy [1], Bd. 1; [2], [14] is as follows:

$$\eta - \eta_0 = \int_{X_0}^{X} \frac{\tilde{\alpha}(x)dx}{C} = \int_{t_0}^{t} \frac{\tilde{\alpha}(X(t))}{C}\frac{dx}{dt}dt = 0, \tag{13}$$

where C is constant. Therefore, assuming in (1) $\lambda, \mu \equiv 1, \alpha(t,\tau) \equiv \alpha(\tau)$, $\beta(t,\tau) \equiv \beta(\tau)$ we obtain for $t \geq t_0$

$$m(t) = m_0 + \int_0^t \alpha(\tau)y(\tau)m(\tau)d\tau, c(t) = \int_{t_0}^{t} \beta(\tau)[1 - y(\tau)]m(\tau)d\tau,$$

$$df = dm + dc, dm = \alpha(t)y(t)dt = \tilde{\alpha}(x)dx, dc = \beta(t)[1 - y(t)]dt = \tilde{\beta}(y)dy.$$

Hence on the strength of (13):

$$m(t) = m_0 e^{\int_{t_0}^{t} \alpha(t)y(t)dt} = m_0 e^{\int_{x_0}^{x} \tilde{\alpha}(X)dx} = m_0\,, \qquad df = dc = \beta(t)[1 - y(t)]dt\,. \tag{14}$$

The last relation means that in the case of adiabatic processes the whole external work is perhaps the result of the internal energy of a system. From (14) it follows that for any $t > t_0$ there existed a state $m(t) \neq m_0$ which cannot be accessible according to the Caratheodory's concept of adiabatic inaccessibility.

The very detailed comparison between the m.m. under consideration and various similar contemporary ones by other authors can be found in [12].

Concluding the section we shall dwell on the question of completion of the given class of m.m. or determination of the so-called "light box", i.e., the construction of additional set of relations so that it would be possible to determine all the elements of DS in future knowing its prehistory and predicting only its separate elements or parameters of exogenous nature.

Believing that new "technologies" $\alpha(t,\tau)$ and $\beta(t,\tau)$ are also the "products" of DS of $m(t)$-type and admitting for the sake of simplicity [3] that

$$\alpha(t,\tau) = \beta(t,\tau) = \alpha(\tau)e^{-c_\alpha(t-\tau)}\,, \tag{15}$$

where c_α defines the rate of deterioration of previously created technologies we have instead of (5)–(7):

$$\begin{aligned}
&\alpha(t) = k_m \int_{a(t)}^{t} \alpha(\tau)e^{-c_\alpha(t-\tau)}x(\tau)m(\tau)d\tau\,,\\
&m(t) = \int_{a(t)}^{t} \alpha(\tau)e^{-c_\alpha(t-\tau)}z(\tau)m(\tau)d\tau\,,\\
&P(t) = \int_{a(t)}^{t} m(\tau)d\tau\,, x + z = y\,,\\
&c(t) = \int_{a(t)}^{t} \alpha(\tau)e^{-c_\alpha(t-\tau)}[1 - y(\tau)]m(\tau)d\tau\,,\\
&a(t) < t, 0 \leq x, y, z \leq 1, T \geq t \geq t_0 > a(t_0) = 0\,,
\end{aligned} \tag{16}$$

where k_m is coefficient with dimensionality α/m.

Other approaches to the completion of m.m. (1) can be based on (see [3] p. 329–333) the system disaggregation and aggregation methods as well as on application of several extremal concepts.

2. Existence and Uniqueness Theorems

Considering $\alpha(t,\tau)y(\tau)$ and $P(t)$ as assigned (y is usually found from solution of some optimization problem) we shall examine explicitly the question about the existence and the uniqueness of solutions for the system (5), (6) relative to $m(t)$ and $a(t)$ on any given segment $[t_0,T]$. The other cases of solutions of equations for the proposed mathematical model shall be considered briefly later on.

Theorem 1. Let $m_0, \alpha y, P$ be positive functions and here

$$m_0 \in C_{[0,t_0]}, \alpha y \in C^{(1)}_{[0,T]\times[0,T]}, P \in C^{(1)}_{[t_0,T]},$$

where C and $C^{(1)}$ are continuous and continuously differentiable spaces of functions, respectively. Then over $[t_0,T]$ the system (5), (6) has the unique positive solution $m(t)$ and $a(t)$ with m and $a \in C^{(1)}_{[t_0,T]}$ and $a(t) < t$.

Proof. Let us introduce formally the relations

$$m(t) = \Psi(m)(t) \equiv \Phi(t_0) - \Phi\left\{M_0^{-1}[M(t) - P(t)]\right\} + \int_{t_0}^{t} \alpha(t,\tau)y(\tau)m(\tau)d\tau; \tag{17}$$

$$a(t) = M_0^{-1}[M(t) - P(t)], M(t) = \int_0^t m(\tau)d\tau, \tag{18}$$

where

$$\Phi(t) = \int_0^t \alpha(t,\tau)y_0(\tau)m_0(\tau)d\tau, M_0(t) = \int_0^t m_0(\tau)d\tau, t \in [0,t_0]. \tag{19}$$

Since Φ and M_0 and hence M_0^{-1} are continuously differentiable monotone increasing functions, the Volterra-type operator Ψ effects contracted mapping of $C_{[t_0,t_1]}$ into itself, where $t_1 > t_0$, and has the property: $a(t) \leq$

$t_0, t \in [t_0, t_1]$. By virtue of (18), (19) and $M(t_0) - P(t_0) = 0$ such t_1 will always be found. It follows that the unique positive solution of the equation $m = \Psi(m)$ on the segment $[t_0, t_1]$ can be found by the method of simple iteration

$$m_{k+1} = \Psi(m_k), k = 0, 1, \ldots; m_0 \equiv 0;$$
$$m = \lim_{k \to \infty} m_k = \sum_{k=0}^{\infty} (m_{k+1} - m_k). \tag{20}$$

On the strength of the well-known Weierstrass theorem, we have from (20) that m is the continuous function. But then it follows from (17) that $m \in C^{(1)}_{[t_0, t_1]}$. It also follows from (18) that $a(t)$ defined after $m(t)$ also pertains to $C^{(1)}_{[t_0, t_1]}$. And now it is easy to see that $m(t)$ and $a(t)$ obtained by virtue of (17)–(20) is the desired solution of the initial system (5), (6) on the segment $[t_0, t_1]$. The first instant t_1 for which $a(t_1) = t_0$, is hitherto unknown and found in the process of solution. If $a(t) \leq t_0$ for $t_0 \leq t \leq T$ it follows that the constructed solution is valid over the whole given segment and the theorem is proved. In the similar way it is proved that for $T > t_1$ the problem has the unique positive solution over $[t_1, t_2], a(t_2) = t_1$, then over $[t_2, t_3], a(t_3) = t_2$ and so on until the solution over the whole segment $[t_0, T]$ has been obtained. The latter is possible by virtue of the conditions of the theorem

$$P(t) \leq \max_{t \in [0,T]} m(t)[t - a(t)] \leq \left\{ \max_{t \in [0, t_0]} m_0(t) + \max_{0 \leq \tau \leq t \leq T} [\alpha(t, \tau) y(\tau)] \times \right.$$
$$\left. \max_{t \in [t_0, T]} P(t) \right\} \times [t - a(t)], t \in [t_0, T],$$

whence

$$t - a(t) \geq \min_{t \in [t_0, T]} P(t) / \left\{ \max_{t \in [0, t_0]} m_0(t) \right.$$
$$\left. + \max_{0 \leq \tau \leq t \leq T} [\alpha(t, \tau) y(\tau)] \max_{t \in [t_0, T]} P(t) \right\} > 0. \tag{21}$$

As it was already noted above, in the general case the function α can depend on the unknown function m. Therefore we shall consider the system

$$m(t) = \int_{a(t)}^{t} \alpha(t, \tau, m(\tau)) d\tau, a(t) < t,$$
$$P(t) = \int_{a(t)}^{t} m(\tau) d\tau, T \geq t > t_0 > a(t_0) = 0, \tag{22}$$

where $\alpha(t,\tau,m)$, for $T \geq t \geq \tau \geq 0$ and $m \in R^+$, i.e. $m \geq 0$, and $P(t) \geq 0$ are the given functions and m and a are the unknown ones. The appropriate theorems of solutions existence and uniqueness for the system (22), moreover for the very general assumptions extending the conditions of the theorem (1), will be obtained on the base of the widely propagated concepts and results of C. Caratheodory [1], [16] holding to the notations and succession of presentation just as it has been done by J. Warga in [16].

The set $\{\alpha(t,\cdot,\cdot)\}$ will be called the collection of the Caratheodory functions $B_t, B_t \equiv B(t,\bar{t},V;R)$ where t is a paramater of the collection if for any fixed $t, 0 \leq t \leq T$

(i) for any $\tau \in \bar{t} \equiv [0,t]$ the function $\alpha(t,\tau,\cdot) \in C(V,R)$ where $V \subset R$;

(ii) for any $m \in V$ the function $\alpha(t,\cdot,m)$ is measurable;

(iii) there exists the integrable function $\psi_\alpha : \overline{T} \to R$

such that

$$\sup_{t\in\overline{T}} |\alpha(t,\tau,\cdot)|_{\sup} = \sum_{t\in\overline{T}} \sup_{m\in V} |\alpha(t,\tau,m)| \leq \psi_\alpha(\tau).$$

The set B_t will be the normed space if we assume

$$|\alpha|_{B_t} = \int_0^t |\alpha(t,\tau,\cdot)|_{\sup} d\tau : B_t \to R.$$

The function $\alpha(\cdot,\tau,m)$ will be considered continuous over $\overline{T} = [0,T]$ for any τ and $m, 0 \leq \tau \leq \cdot$ and $m \in V$. In the notations [16] one can also write

$$\alpha \in C(\overline{T}, B(\overline{T},V;R)),\ \alpha(t,\tau,m) = 0, \tau > t.$$

Theorem 2. The existence of local solutions. Let $\overline{T} = [0,T], T > t_0$ and $\alpha \in B_t$, and let α be continuous by t for any $t \in \overline{T}, P$ be absolutely continuous over $[t_0,T], m_0$ be integrable over $\bar{t}_0 = [0,t_0]$ and $m_0 \in V$, all the functions α, P and m_0 be positive and also

$$0 < m_0^- \leq \inf_{t\in\bar{t}_0} m_0(t) \leq \sup_{t\in\bar{t}_0} m_0(t) \leq m_0^+ < \infty,$$

$$\sup_{\substack{0\leq\tau\leq t\leq T \\ m\in V}} \alpha(t,\tau,m) \leq \alpha^+ < \infty.$$

Let us also assume $\tilde{T} = [t_0, \tilde{t}]$, where

$$\tilde{t} = \min(t_1, t_2, t_3),$$
$$t_1 = \sup\{t \in [t_0, T] | [m(t_0) + b](t - t_0) + P(t) - P(t_0) \leq t_0 m_0^-\},$$
$$t_2 = \sup\{t \in [t_0, T] | \int_{t_0}^{t} \alpha(t, \tau, \cdot)_{\sup} d\tau \leq b_1\},$$
$$t_3 = t_0 + \frac{m_0^-}{2\alpha^+}$$

and introduce the number $b > 0$ such that

$$\frac{\alpha^+}{m_0^-}\{[m(t_0) + b](t - t_0) + P(t) - P(t_0)\} + b_1 \leq b, t \in \tilde{T}.$$

If a closed sphere $S^F(m(t_0), b) \subset V$, then there exist the positive functions $\tilde{m} : \tilde{T} \to V$ and $\tilde{a} : \tilde{T} \to \tilde{t}_0$ such that they are the solutions of the system (22) over $\tilde{T}$. And here, $\tilde{m}$ will be continuous and $\tilde{a}$ is absolutely continuous over $\tilde{T}$.

Proof. Let us introduce formally the equation

$$m(t) = m(t_0) - \Phi[M_0^{-1}(M(t) - P(t))] + \int_{t_0}^{t} \alpha(t, \tau, m(\tau)) d\tau \equiv \Psi(m)(t), \quad (23)$$

where

$$M_0(t) = \int_0^t m_0(\tau) d\tau, \Phi(t) = \int_0^t \alpha(t, \tau, m_0(\tau)) d\tau, t \in \tilde{t}_0 \quad (24)$$

and assume

$$K = \{m \in C(\tilde{T}, R) | |m(t) - m(t_0)| \leq b, t \in \tilde{T}\}.$$

Since the restriction $\alpha|t, \tilde{t} \times V$ pertains to $B(t, \tilde{t}, V; R)$ for any $t \in \tilde{T}$ the function $\tau \to \alpha(t, \tau, m(\tau)) : \tilde{t} \to R$ is integrable for any $t \in \tilde{T}$ and $m \in K$. Consequently, the function $F(m)$,

$$F(m)(t) = \int_{t_0}^{t} \alpha(t, \tau, m(\tau)) d\tau \quad (25)$$

is defined for any $t \in \bar{T}$ and $m \in K$, and $F(m) \in C(\bar{T}, R)$. Further, on the strength of the condition of the theorem

$$|\Psi(m)(t) - m(t_0)| \leq \frac{\alpha^+}{m_0^-}[(m(t_0)+b)(t-t_0) + P(t) - P(t_0)] + b_1 \leq b, t \in \bar{T}.$$

This means that $\Psi(K) \subset K$. And what is more

$$|\Psi(m)(t) - \Psi(m)(t')| \leq \int_{t'}^{t} \alpha(t, \tau, \cdot)_{\sup} d\tau + (t' - t_0)\omega_\alpha(t - t')$$
$$+ \frac{\alpha^+}{m_0^-}[(m(t_0)+b)(t-t') + P(t) - P(t')] + \frac{m_0^+ t_0}{m_0^-}$$
$$\times \omega_\alpha \left[\frac{(m(t_0)+b)(t-t') + P(t) - P(t')}{m_0^-} \right] (t_0 < t' \leq t \leq \bar{t}, m \in K),$$

where ω_α is a continuous module for the function α over the first variable t. Consequently, $\Psi(K)$ is the bounded and uniformly continuous subset from $C(\bar{T}, R)$. It follows [16] that $\overline{\Psi(K)}$ is a compactum. Now let $\lim\limits_j m_j = m$ in K. Then it is obvious that for the operator Ψ_1,

$$\Psi_1(m)(t) = \Phi\left[M_0^{-1}\left(\int_0^t m(\tau)d\tau - P(t)\right)\right], t \in \bar{T} \qquad (26)$$

the relation $\lim\limits_j \Psi_1(m_j)(t) = \Psi_1(m)(t)$ in $K, t \in \bar{T}$ is valid and therefore [16] for the operator Ψ it will also be $\lim\limits_j \Psi(m_j)(t) = \Psi(m)(t), t \in \bar{T}$. Thus, the mapping $\Psi : K \to K$ is continuous. At last it is easy to see that K is a closed and convex set. By Schauder's fixed point theorem it follows that the mapping $\Psi : K \to K$ has the fixed point $\tilde{m}$ on $\bar{T}$. But then on the strength of (23), (24) and (18) (with replacement of m and a by $\tilde{m}$ and $\tilde{a}$, respectively) $\tilde{m}$ and $\tilde{a}$ will be the solutions of (22) which have automatically the desired properties by virtue of the conditions assumed and the structure of (22) itself.

Theorem 3. The extension of local solutions. Let conditions of the Theorem 2 take place and there exists an integrable function $\psi : \overline{T} \to R$ and positive increasing and continuous function $\varphi : (0, \infty) \to (0, \infty)$ such that

$$\sup_{t \in \overline{T}} \alpha(t, \tau, m) \leq \varphi(m)\psi(\tau) \quad (\tau \in \overline{T}, m \in R), \lim_{r \to \infty} \int_0^r \frac{ds}{\varphi(s)} = \infty. \qquad (27)$$

Then there exist the positive functions $\overline{m} : \overline{T} \to R$ and $\overline{a} : \overline{T} \to \overline{T}_a, T_a < T$ which are the solutions of (22) on $\overline{T}$. In addition $\overline{m}$ will be continuous and $\overline{a}$ absolutely continuous on $\overline{T}$.

Proof. Let

$$0 < P^- \leq \inf_{t \in \overline{T}} P(t) \leq \sup_{t \in \overline{T}} P(t) \leq P^+ < \infty .$$

Then similar to the case (21) it is not difficult to obtain

$$t - a(t) \geq P^- / (m_0^+ + \alpha^+ P^+) > 0 \tag{28}$$

for any $t \in [t_0, t']$ for which the solution of Eq. (23) exists. The relation (28) means that the behaviour of the function a cannot serve as an obstruction to the extension of the solution $m(t)$ and $a(t)$ over the whole $\overline{T}$. But as shown in [16] the property (27) can be used as the sufficient condition for extension of the solution on any segment $\overline{T}$ in the case of equations of the form $m(t) = m(t_0) + F(m)(t)$, where the operator F is given by the relation (25). The analysis of the corresponding proof in [16] shows that the operator Ψ_1 given by the relation (26) and the operator $\Psi = F - \Psi_1$ possess just the same properties as F. This fact provides the extension of the solution of (23) over the whole segment $\overline{T}$.

Theorem 4. The uniqueness of solution. Under the conditions of Theorems 2 and 3, it follows that if there exist integrable functions ψ_1 and ψ_2 on $\overline{T}$ such that

$$|\alpha(t_1, \tau, m_0(\tau)) - \alpha(t_2, \tau, m_0(\tau)| \leq \psi_1(\tau)|t_1 - t_2| (\tau \in \overline{T}; t_1, t_2 \in \overline{T}) ; \tag{29}$$

$$|\alpha(t, \tau, m_1) - \alpha(t, \tau, m_2)| \leq \psi_2(\tau)|m_1 - m_2| (t, \tau \in \overline{T}; m_1, m_2 \in V \subset R) \tag{30}$$

then $\overline{m}(t)$ and $\overline{a}(t)$ (on the strength of Theorem 2) are the unique solutions of the system (22).

Proof. If $m_1(t)$ and $m_2(t)$ are two solutions of the Eq. (23), we have

$$m_1(t) - m_2(t) = \int_{t_0}^{t} [\alpha(t, \tau, m_1(\tau) - \alpha(t, \tau, m_2(\tau))] d\tau$$
$$- \{\Phi[M_0^{-1}(M_1(t) - P(t))] - \Phi[M_0^{-1}(M_2(t) - P(t))]\} ,$$

whence on the strength of (24), (29) and (30)

$$|m_1(t) - m_2(t)| \le \int_{t_0}^{t} \psi_2(\tau)|m_1(\tau) - m_2(\tau)|d\tau + \left[\alpha^+ + \int_0^T \psi_1(\tau)d\tau\right] \times \frac{1}{m_0^-} \int_{t_0}^{t} |m_1(\tau) - m_2(\tau)|d\tau .$$

But from Granuola's inequality [16] there follows $|m_1(t) - m_2(t)| \le 0$ and hence $m_1(t) = m_2(t)$.

If the functions α, a and P or more generally $\lambda, \mu, \alpha, \beta$ and P are assigned, for determination of $m_1 = ym$ and $m_2 = (1-y)m$ in (1) we have the system of linear integral equations of Volterra-type:

$$m_1(t) + m_2(t) = \int_0^t \alpha(t,\tau)\lambda(t,\tau)m_1(\tau)d\tau ,$$
$$P(t) = \int_0^t \lambda(t,\tau)m_1(\tau)d\tau + \int_0^t \mu(t,\tau)m_2(\tau)d\tau, T \ge t \ge t_0 > 0 . \tag{31}$$

The problems of solutions existence and uniqueness for the system (31) and several similar linear systems are examined in [3], [15] and [17]. Many results obtained in these works can be strengthened by the way of introduction of the appropriate classes of the Caratheodory functions.

Let us dwell in brief on the nonlinear system (16). We shall assume that functions P, x and z, and parameters k_m and c_α, and also all the elements of the system (16) on the prehistory are given. Then the first three equations in (16) are Volterra-type system relative to three unknown functions α, m and a on the segment $\bar{T} = [t_0, T]$. It is obvious that in this case the existence of a local solution (see Theorems 1 and 2) will also take place but the similarity of Theorem 3 will not be applicable since the condition of the type (27) will be violated. However, it is not hard to find the sufficient condition such that any solution (16), provided that they exist on $[t_0, T]$, will be bounded on this segment.

Theorem 5. Let $c_\alpha > 0, x(t)$ and $z(t)$ be integrable, and $P(t), z(t)$ be positive for $t \in \bar{T}$, and also

$$[1 - e^{-c_\alpha(t-t_0)}]m(t_0) \le \frac{c_\alpha}{c}, c = k_m \max_{t\in\bar{T}} \left(\frac{x(t)}{z(t)}\right) \max_{t\in\bar{T}} z(t) . \tag{32}$$

Then any non-negative solutions (16) $m(t)$, $\alpha(t)$ and $a(t)$ possess the properties

$$\begin{aligned} m(t) &\le m(t_0)\frac{e^{-c_\alpha(t-t_0)}}{1-\frac{cm(t_0)}{c_\alpha}\left[1-e^{-c_\alpha(t-t_0)}\right]}, \\ \alpha(t) &\le k_m \max_{t\in\bar{T}}\left(\frac{x(t)}{z(t)}\right) m(t), \\ t-a(t) &\ge \frac{P(t)}{\max\limits_{0\le t\le T} m(t)}. \end{aligned} \tag{33}$$

Proof. It is easy to see that

$$\begin{aligned} \alpha_1(t) = \alpha(t)e^{c_\alpha(t)} &= k_m \int_{a(t)}^{t} \alpha_1(\tau)\times(\tau)m(\tau)d\tau\,, m(t)e^{c_\alpha t} \\ &= \int_{a(t)}^{t} \alpha_1(\tau)z(\tau)m(\tau)d\tau\,, \end{aligned}$$

from which

$$\begin{aligned} \alpha_1(t) &\le k_m \max\left(\frac{x}{z}\right) m(t)e^{c_\alpha t}\,, \\ m(t)e^{c_\alpha t} &\le k_m \max\left(\frac{x}{z}\right) \max z \int_{t_0}^{t} e^{c_\alpha \tau} m^2(\tau)d\tau + m(t_0)e^{c_\alpha t_0} \\ &= c\int_{t_0}^{t} [e^{\frac{c_\alpha}{2}\tau} m(\tau)]^2 d\tau + c_0, \\ c_0 &= m(t_0)e^{c_\alpha t_0}\,, \end{aligned}$$

and hence

$$[m(t)e^{\frac{c_\alpha t}{2}}]^2 \le e^{-c_\alpha t}\left\{c\int_{t_0}^{t}[e^{\frac{c_\alpha \tau}{2}} m(\tau)]^2 d\tau + c_0\right\}^2.$$

Assuming $v = \int_{t_0}^{t}[e^{\frac{c_\alpha \tau}{2}} m(\tau)]^2 d\tau$ we find

$$dv \le e^{-c_\alpha t}(cv + c_0)^2 dt\,, v(t_0) = 0\,,$$

hence

$$\int_0^v \frac{dv}{(cv+c_0)^2} \le \int_{t_0}^{t} e^{-c_\alpha t} dt = \frac{e^{-c_\alpha t} - e^{-c_\alpha t_0}}{c_\alpha}$$

and

$$cv + c_0 \le \frac{c_0}{1 - \frac{cm(t_0)}{c_\alpha}[1 - e^{-c_\alpha(t-t_0)}]},$$

$$m(t) \le m(t_0)\frac{e^{-c_\alpha(t-t_0)}}{1 - \frac{cm(t_0)}{c_\alpha}[1 - e^{-c_\alpha(t-t_0)}]}$$

so the first inequallity in (33) is proved. After that, the second and third inequalities in (33) become obvious.

As a consequence, from Theorem 5 on the basis of similarities of the Theorems 1–4, it is not difficult to prove the existence and uniqueness of solutions for the system (16) provided the conditions (32) are realized.

3. Examples of Optimization Problems

3.1. The external DS function maximization problem:

$$I_1 = \int_{t_0}^{T} c(t)dt \equiv \int_{t_0}^{T} \left\{ \int_{a(t)}^{t} \beta(t,\tau)[1 - y(\tau)]m(\tau)d\tau \right\} dt = \max_{y} \qquad (34)$$

provided that the relations (5)–(10) or (15) and (16) are observed.

3.2. The average internal DS expenditures minimization problem:

$$I_2 = \int_{t_0}^{T} P(t)dt \equiv \int_{t_0}^{T} \left[\int_{a(t)}^{t} m(\tau)d\tau \right] dt = \min_{y} \qquad (35)$$

provided that the restrictions (5)–(10) and maybe the restrictions on $c(t)$: $c(t) \ge c^-(t)$, where $c^-(t)$ is given, are observed.

3.3. The problem of DS's high speed of operation:

$$I_3 = T - t_0 = \min_{y} \qquad (36)$$

provided that (5)–(10) and the conditions $\int_{t_0}^{T} c(t)dt \in C^*$ and $P \in P^*$, where sets C^*, P^* are given, are observed.

3.4. The DS viability maximization problem:

$$I_4 = \min[m^+(t) - m(t), m(t) - m^-(t), c^+(t) - c(t), c(t) - c^-(t)] = \max_{y} \qquad (37)$$

provided that (5)–(10) and the inequalities $m^-(t) \leq m(t) \leq m^+(t)$ and $c^-(t) \leq c(t) \leq c^+(t)$ with the assigned functions $m^\pm(t)$ and $c^\pm(t)$, are observed.

3.5. The problem of out-of-date DS products minimization:

$$I_5 = \int_{t_0}^{T} G(t)dt \equiv \int_{t_0}^{T} \left[\int_0^{a(t)} m(\tau)d\tau \right] dt = \min_y \tag{38}$$

provided that (5)–(9) are observed.

3.6. The external DS function minimization problem:

$$I_1 = \int_{t_0}^{T} c(t)dt = \min_y \tag{39}$$

provided that (5)–(10) or (15) and (16) are observed.

3.7. The active DS's "life" maximization problem:

$$I_3 = T - t_0 = \max_y \tag{40}$$

provided that (5)–(10) or (15) and (16) and the restriction $c(t) \geq c^0(t)$ where $c^0(t)$ is given, are observed.

In the case of the example 1 (Sec. 2) when DS is economy as a whole, 3.1 is the maximization problem for the number of consumer goods during the design period $T - t_0$; 3.2 is the minimization problem for the average labour inputs; 3.3 is the minimization problem for the time of consumption preassigned level attainment provided that the preassigned labour inputs are given. In the case of the natural DS as opposed to the artificial one the problems 3.6 and 3.7 can have more sense than 3.1 and 3.3. If, for example, a DS is a population of viruses in the human organism, in particular, the population of HIV [8] and $c(t)$ is the aggressive factor of HIV, then just the problem 3.6 makes sense. If a DS is a human being himself, the attractiveness namely the problem 3.7 is obvious [6].

4. Qualitative Investigation of Optimization Problem

Let us investigate the problem (34) in detail. The results of this investigation are transferred both on the problems (35)–(40) and some other optimization problems [17].

4.1. The existence of solutions

Proof of solution existence for problem 3.1 is based on the results of the monograph by J. Warga [16]. However, as is shown brilliantly by L. C. Young [18], one of the essential sources for creation of contemporary means of the optimal control theory based on the so-called generalized nonsmooth classes of an admissible control circuit is furnished by the earlier mentioned notions and other C. Caratheodory's concepts and results.

Let us introduce the details of the appropriate proof following Ju. P. Jacenko's presentation [17].

1°. In order to adapt theorems of [16] to problem 3.1 let us represent it in the form adopted in [16], namely to determine

$$\inf g_0(x, y, a) \tag{41}$$

on the set

$$H(Y) = \{(x, y, a) \in X \times Y \times B | x = F(x, y, a)\} \tag{42}$$

under the restriction

$$g_1(x, y, a) = 0\,, \tag{43}$$

where $x = (m, c)$ are the state variables, y is a control function, a is a control parameter,

$$\begin{aligned} g_0(x, y, a) &= \int_{t_0}^{t} c(t)dt\,, \\ g_1(x, y, a) &= \int_{a(t)}^{t} m(\tau)d\tau - P(t) \end{aligned} \tag{44}$$

and a state equation $\dot{x} = F\dot{x}$ has the form

$$x(t) = \int_0^t f[t, \tau, x(\tau), y(\tau), a(t)]d\tau \tag{45}$$

or

$$m(t) = \int_0^t f_1[t, \tau, m(\tau), y(\tau), a(t)]d\tau\,,$$

$$c(t) = \int_0^t f_2[t, \tau, m(\tau), y(\tau), a(t)]d\tau\,, \tag{46}$$

$$f_1 = \begin{cases} \alpha(t,\tau)y_0(\tau)m_0(\tau), a(t) \le \tau \le t_0\,, \\ \alpha(t,\tau)y(\tau)m(\tau), \hat{a}(t) \le \tau \le T\,, \\ 0 \quad , 0 \le \tau \le a(t)\,, \end{cases}$$

$$f_2 = \begin{cases} \beta(t,\tau)(1 - y_0(\tau))m_0(\tau), a(t) \le \tau \le t_0 \\ \beta(t,\tau)(1 - y(\tau))m(\tau), \hat{a}(t) \le \tau \le T\,, \\ 0, \quad 0 \le \tau < a(t)\,, \end{cases}$$

$$\hat{a}(t) = \max[t_0, a(t)]\,. \tag{47}$$

Let us introduce also the set

$$A(Y) = \{(x, y, a) \in X \times Y \times B | x = F(x, y, a), g_1(x, y, a) = 0\}$$

and accept $X = C_{[t_0,T]} \times C_{[t_0,T]}, B = \{a(t) \in C_{[t_0,T]} | 0 \le a(t) \le T\}$.

Following [16], as a control space Y we shall consider the space U of ordinary (Lebesgue measurable) controlling y functions: $[t_0, T] \to R$ and $\mathcal{P}$ that of generalized controlling functions σ representing functions of the time with the values of the set of measures. Respectively let us determine the admissible sets of control values

$$R^*(t) = [y_{\min}(t), 1] \subset R\,,$$

the control functions

$$U^* = \{y \in U | y(t) \in R^*(t)\}$$

and the generalized control functions

$$\mathcal{P}^* = \{\sigma \in \mathcal{P} | \sigma(\overline{R}^*(t)) = 1\}\,.$$

Following [16] we shall call:

(i) the point $(x, y, a) \in H(U^*)$ as a minimizing U-solution of the problem if it minimizes the function g_0 on the set $A(U)$;

(ii) the sequence (x_j, y_j, a_j) in $H(U)$ as an approximate minimizing U-solution if

$$\begin{aligned} &\lim\ g_1(x_j, y_j, a_j) = 0\,, \\ &\lim\ g_0(x_j, y_j, a_j) \le \liminf\ g_0(x_j, y_j, a_j)\,; \end{aligned} \tag{48}$$

(iii) the point $(x, \sigma, a) \in H(\mathcal{P}^*)$ as a minimizing generalized solution if it minimizes the function g_0 on the set $A(\mathcal{P}^*)$.

The state equation (45) is substituted by [16]

$$x = F(x, \sigma, a) \equiv \int_0^t d\tau \int_{R^*(\tau)} f[t, \tau, x(\tau), r, a(t)]\sigma(\tau)(dr) \tag{49}$$

under transition from ordinary controls $y \in U$ to the generalized controls $\sigma \in \mathcal{P}$.

The physical sense of the substitution consists in that for every t, τ, x and a we calculate not a function itself but its mean value by $r \in R^*(\tau)$ and here $\sigma(\tau)$ determines what kinds of values r and of what weight are involved in averaging.

2°. Let us prove the existence of a generalized solution of the problem (41)–(47). Let $V \subset R \times R$ and let V be the set of values for a state variable x under $y \in R^*$ and $a \in B$. There are the following facts:

(i) on the strength of continuity and monotonicity of the operator in a state equation (45) the set V is bounded and closed;

(ii) the set B is compact;

(iii) the function $g = (g_0, g_1) : X \times \mathcal{P}^* \times B \to R \times R$ is continuous;

(iv) the set $A(\mathcal{P}) \neq \emptyset$(by virtue of $y_{\min}(t) \le 1,\ t \in [t_0, T]$).

Furthermore, the following properties of the function $f(f_1, f_2)$ are true:

$$f : [t_0, T] \times [0, T] \times V \times R^* \times B \to R \times R\,,$$

(i) f is continuous over t under $(t, x, y, a) \in [0, T] \times V \times R^* \times B$;

(ii) f is continuous on (x, y, a) under $(t, \tau) \in [t_0, T] \times [0, T]$;

(iii) f has discontinuity under $\tau = a(t)$ but it is measurable by τ under $(t, x, y, a) \in [t_0, T] \times V \times R^* \times B$;

(iv) there exists $\varphi : [0, T] \to R \times R$ such that

$$\sup_{[t_0, T] \times V \times R^* \times B} |f(\cdot, \tau, \cdot, \cdot, \cdot)| \le \varphi(\tau), \tau \in [0, T]\,.$$

Thus all the conditions of Theorem VII.I.1 from [16] have been realized and hence the problem (41)–(47) has the minimizing generalized solution $(x, \sigma, a) \in V \times \mathcal{P}^* \times B$.

3°. Let us prove the existence of the ordinary solution y on the set of measurable functions for the problem (41)–(47). Indeed there are the following properties of the problem:

(i) $g(x, \sigma, a) \equiv \overline{g}(x, a), \overline{g}(x, a) : X \times B \to R \times R$ for all $(x, \sigma, a) \in X \times \mathcal{P}^* \times B$;

(ii) the set $R^*(\tau)$ is closed for all $\tau \in [0, T]$;

(iii) the set of functions $\psi(r) = f(\cdot, \tau, x, r, a), \psi : R \to L^1_{[t_0, T]}$

makes up a convex subset $\{\psi(r) | r \in R^*(\tau)\}$ in the space of measurable functions $[t_0, T] \to R \times R$ (it follows from the formula (47)) under any fixed $(\tau, x, a) \in [0, T] \times V \times B$. Thus for the problem (41)–(47) the theorem VII.I.4 is valid on the strength of which the problem under consideration has the minimizing U-solution (x, y, a).

4.2. A problem solution structure investigation

The first essential result on the properties of solutions of the problem 3.1 has been obtained by V. M. Glushkov and V. V. Ivanov (see [3]) and consisted qualitatively in that for "small" $T - t_0$ the desired $y(t)$ is minimally possible ($y(t) = y_{\min}(t)$ by virtue of the restriction $\frac{da}{dt} \geq 0$) but for "large" $T - t_0$ the desired $y(t)$ may differ from the minimally possible on the larger initial part of the segment $[t_0, T]$ and only on the smaller final part of $[t_0, T]$ the desired $y(t)$ is minimally possible. The notions "small" and "large" depend on the values of functions α and β, namely, the greater the functions in question, the nearer to t_0 is the boundary between "small" and "large" segments. The obtained result was in essense an implication of the fact that for "large" $T - t_0$ a growth of $m(\tau)[1 - y(\tau)]$ overtakes a diminution of $1 - y(\tau)$. The result has obtained, in the sequel, the important qualitative very general interpretation: the record of an external function for any DS can be obtained only under conditions of its sufficiently comfortable guarantee, that is, under very significant fraction of resources sent to internal needs of DS.

The solution structure of the problem 3.1 and other optimization problems has been investigated in detail mainly by Ju. P. Jacenko. Let us refer to the appropriate results below without proofs (detailed proofs can be found in [17]).

Lemma [3], [17]. The variation of Lagrange function for the optimal control problem 3.1 has the form

$$\delta L = \int_{t_0}^{T} \left\{ - \int_{t}^{\underline{a}^{-1}(t)} [\alpha(t,\tau)\psi_1(\tau) + \beta(t,\tau)]d\tau\, m(t)\delta y(t) + [\psi_1(t) \right.$$
$$\left. \times\, \alpha(t,a(t))y(a(t)) - \beta(t,a(t))[1 - y(a(t))] + \psi_2(t)]m(a(t))\delta a(t) \right\} dt\,, \tag{50}$$

where δy and δa are variations of independent variables y and a, ψ_1 and ψ_2 satisfy the equation

$$\psi_1(t) = \int_{t}^{\underline{a}^{-1}(t)} \{\alpha(t,\tau)y(t)\psi_1(\tau) + \psi_2(\tau) - \beta(t,\tau)[1 - y(t)]\}\, d\tau\,,$$
$$\underline{a}^{-1}(t) = \begin{cases} a^{-1}(t), & t_0 \le t \le a(T)\,, \\ T, & a(T) \le t \le T\,. \end{cases} \tag{51}$$

The following results are obtained, in particular, in [3], [17] on the basis of (50) and (51).

Theorem 6. There exists the instant $\theta, t_0 \le \theta < T$ such that $I'_{1y}(y_{\min}, t) < 0$ under $t \in (\theta, T)$ and $y^*(t) \equiv y_{\min}(t)$. If $\theta > t_0$ then the segment $[t_0, \theta]$ consists of subintervals, on the every one of them either $I'_{1y}(y^*, t) > 0$ and $y^*(t) \equiv 1$ or $I'_{1y}(y^*, t) < 0$ and $y^*(t) \equiv y_{\min}(t)$ or $I'_{1y}(y^*, t) \equiv 0$ and $y_{\min}(t) \le y^*(t) \le 1$, and here, all of these cases are possible depending on the given functions of the problem.

The more refined results can be obtained in the special cases [17]

A. $$\beta(t,\tau) = K(t)\alpha(t,\tau) \qquad (\text{then } \psi_1(t) \equiv -K(t)) \tag{52}$$

B. It is known *a priori* that

$$a^*(T) \le t_0 \ (\text{then } a^{-1}(t) \equiv T \text{ and } y(a(\tau)) \equiv y_0(a(\tau)))\,. \tag{53}$$

Theorem 7. For the problem 3.1 in the cases (52) and (53) there exists the "best" function $\tilde{a}(t), t \in [t_0, T]$ such that $I'_{1y}(\tilde{a}, t) \equiv 0, t \in [t_0, \theta], t_0 \le \theta < T$ and $I'_{1y}(\tilde{a}, t) < 0, t \in (\theta, T)$. Depending on the values of a disagreement

$$\Delta_a = |\tilde{a}(t_0) - a(t_0)|\,,\ \Delta_{a'} = |\tilde{a}'(t_0) - a'(t_0)|$$

and a length of the planning interval $T-t_0$, the problem solutions $a(t)$ and $y(t)$ can have the following behaviour variants:

(i) $y(t) \equiv y_{\min}(t), t \in [t_0, T]$;

$$\text{(ii)}\quad y(t) = \begin{cases} 1, & t \in [t_0, \tau_1], a(t) \text{ does not intersect } \bar{a}(t) \\ y_{\min}(t), & t \in [t_1, T]; \end{cases}$$

$$\text{(iii)}\quad y(t) = \begin{cases} 1, & t \in [\tau_{i-1}, \tau_i), \\ y_{\min}(t), & t \in (\tau_i, \tau_{i+1}), i = \overline{1, N}, \tau_0 = t_0, \tau_N = T, \end{cases} \qquad (54)$$

$a(t)$ intersects $\bar{a}(t)$ N times, $N \geq 1$.

The case of unlimited quantity of switchings is possible.

4.3. On the uniqueness of solutions

The above mentioned investigation of the possible solutions structure for the problem 3.1 allows us to prove the uniqueness of its solution (in cases A and B). Since on the strength of previous theorem the problem solution a^*, y^*, and m^* is determined by the function $\bar{a}$, the uniqueness of solutions follows from that of $\bar{a}$.

Theorem 8 [17]. If the conditions A and B are fulfilled, $y_{\min}(t) \leq 1, t \in [t_0, T]$, a function β is monotonous by τ and α and P are slowly varying functions (i.e., α'_t, α'_τ and $P'(t) \ll 1$), the problem 3.1 has the unique solution a^*, y^* and m^*.

Theorem 9 [17]. If $y_{\min}(t) \leq 1, t \in [t_0, T]$, and it is known *a priori* that $a^*(T) \leq t_0$ and $\beta(t,\tau)[1 - y_0(\tau)]$ and $\alpha(t,\tau)y_0(\tau)$ are monotone increasing functions of τ, the problem 3.1 has the unique solution.

In the case when $\beta(t,\tau)$ is not a monotone function of τ in [3], [17] the instance of the nonuniqueness of solution for the problem 3.1 is constructed.

Theorem 10 (on a highway) [17]. Let the following conditions

(i) $\beta(\tau) = b^\tau, b > 1$ or $\beta(\tau) = \tau^s, s > 0$;

(ii) $\left(\frac{\alpha(t)}{\beta(t)}\right)'$ is a rapidly enough decreasing function of t

be fulfilled. Then for the problem 3.1 there exists the "best" function $\bar{a}(t)$ ("highway") and also for any $\theta > t_0$ there exists $\overline{T}(\theta)$ such that for all $T \geq \overline{T}(\theta)$ the behaviour of $\bar{a}(t)$ does not depend on the values $T - t_0$ on the segment $[t_0, \theta]$ and is determined only by functions β and α. In

addition the problem 3.1 solution $a^*(t) \to \tilde{a}(t)$ under $T, t, T-t \to \infty$ if $0 \le y_{\min}(t) \le y^*(t) \le 1$.

5. On Numerical Methods DS Simulation and Appropriate Software and Applications

One can draw information on numerical methods for simulation of DS in [4], [5], [9], [11]. Taking into account the great "stiffness" of the corresponding systems of Volterra-type equations in the sense of the great magnitude

$$\max[\alpha(T-t_0), \beta(T-t_0)]$$

it has been necessary to construct and apply the so-called optimal, in accuracy and the number of the necessary basic computer operations, algorithms for their solution. On the strength of significant complexity of the desired optimal controls structure it has been necessary to develop the so-called adaptive algorithms of optimization [11].

The appropriate software is contained in [7], [13]. The peculiarity of this software consists in that the so-called estimating subroutines are frequently contained in it, side by side with the ordinary solving subroutines. The former subroutines estimate the number of the necessary basic operations, the required computer memory capacity and different kinds of errors accompanying a process of the applied problems solution on a computer, namely, the errors due to input data inaccuracy and incompleteness, inaccuracy of approximate methods and round-of-errors during realization of the corresponding algorithms on a computer.

Among various possible numerous applications we dwell briefly on those stated in [6], [8] and [10].

Along with m.m. of the economy and biosphere two other bonds are introduced as well in [6]: what a human being takes from the nature and what he gives it in return. As a result, the interconnected m.m. of DS, namely, the mathematical model of human activities and biosphere which was defined and investigated by academician V. I. Vernadsky as noosphere, was obtained. A qualitative and numerical investigation of this m.m. resulted in the following conclusions:

1. The volume of consumption $c(T)$ cannot be preassigned if natural resources remain limited.

2. There are critical fractions of a living, bio-boned substance and humus of biosphere consumed by human beings, the exceeding of which results in irreproducible losses during a process of bio-geo-chemical circulation of substances in a biosphere.

3. The similar result will take place if a fraction of a solar energy utilized by human beings exceeds also a certain threshold.

4. Even insignificant harmful anthropogenic effects on biosphere (in the form of out-of-date products of human activity $G(t)$) can serve as a motive of significant disasters in ecology.

5. On the other hand, on the strength of the laws of nature, just a component $G(t)$, which turns into a certain fraction of natural resources, can serve, due to human intellect, as a source for a noosphere prosperity with unlimited growth of its resources.

The analysis of mathematical model [8] for the immune network of an AIDS patient leads to the new probable immunological methods of the struggle with HIV consisting in a creation of conditions for the extreme possible tolerance to the component of $m(t)$-type and simultaneously for the maximum possible aggressiveness to the component of $c(t)$-type for the whole population of HIV as DS in the human organism.

The mathematical model [8] is based on the very complicated and perfect m.m. of an immune network which have been developed earlier in [10]. One of the implications of a tendency to the construction of the more precise m.m. of an immune network was the fact that the desired m.m. turns out to be not differential but integro-differential ones. A qualitative and a numerical investigation of m.m. [10] resulted in its subsequent refinement in the article [8].

References

1. C. Caratheodory, *Gesammelte mathematische Schriften*, Münhen, 1954–1957, – Rd. 1–5.
2. Ya. M. Gel'fer, *History and Methodology of Thermodynamics and Statistical Physics*, Moscow: Vysshaja shkola, 1981, 536 p. (in Russian).
3. V. M. Glushkov, V. V. Ivanov and V. M. Janenko, *Developing System Modelling*, Moscow: Nauka, 1983, 352 p. (in Russian).
4. V. V. Ivanov, *The theory of approximate methods and their application to the numerical solution of singular integral equations*, Leyden Noordhoff intern. publ., 1976, 330 p.

5. V. V. Ivanov, *Methods of Computation*, Reference guidebook, Kiev: Naukova dumka, 1986, 584 pp. (in Russian).

6. V. V. Ivanov, *Mathematical model of a noosphere* (in Russian), Acad. of Sci., Ukr. SSR, V. M. Glushkov Institute of Cybernetics. Kiev, 1989, 17 pp. (manuscript deposited in VINITI 13.03.89, N 1627-B-89).

7. V. V. Ivanov, M. D. Babich, A. I. Berezovskij and P. N. Bessarab et al., *Complex of routines* POM-1 (in Russian), Acad. of Sci., Ukr. SSR, V. M. Glushkov Institute of Cybernetics, Kiev: 1985, 1250 pp. (dep. in GOS FAP SSSR, N 5086000156).

8. V. V. Ivanov, V. N. Korzhova, *Mathematical model of an immune network for patients by AIDS* (in Russian), Acad. of Sci., Ukr. SSR, V. M. Glushkov Inst. of Cyber, Kiev: 1989, 18 pp. (dep. in VINITI 9.01.90, N133-B-90).

9. V. V. Ivanov, Ju. G. Tvalodze, A. Sh. Zhuzhunashvili, *On solution of Volterra-type integral equations with a preassigned accuracy* (in Russian), Acad. of Sci., GSSR, N. I. Mushelishvili Institute of Computing Mathematics, Tbilisi: 1989, 31 pp. (dep. in VINITI 23.03.89, N 1886-B-89).

10. V. V. Ivanov, V. M. Janenko, L. N. Fontalin, V. G. Nesterenko, *Modelling of idiotype-antiidiotypical interactions of immune network with regard to distinction of lymphocytes into subpopulation*, in: Mathematical Modelling, North-Holland, 1983, pp. 141–149.

11. V. V. Ivanov, Ju. P. Jacenko, *Adaptive algorithms of optimization in the integral macro-economical model*, DAN SSSR **290** (5) (1986) 1053–1058 (in Russian).

12. V. V. Ivanov, Ju. P. Jacenko, U. E. Galiev, *Comparison of some integral dynamical macro-economic models*, Avtomatika **4** (1986) 47–53 (in Russian).

13. V. S. Mikhalevich, V. V. Ivanov, L. D. Zdorenko, Ju. P. Jacenko et al., *Complex of routines MDS-1* (in Russian) Acad. of Sci., Ukr. SSR, V. M. Glushkov Inst. of Cyber., Kiev: 1986, 711 pp. (dep. in GOS FAP SSSR, N 50880000105).

14. O. Perron, *Constantin Caratheodory*, Jahrb, Ber. Dtsch. Math. Ver. **55** (1952) 39–51.

15. A. E. Vuginshtein, V. V. Ivanov, *Analytic investigation of continuous models of developing systems*, Soviet Math. Dokl. **273** (2) (1984) 600–604.

16. J. Warga, *Optimal Control of the Differential and Functional Equations*, New York and London: Academic Press (translated into Russian, Moscow: Nauka, 1977. – 624pp.)

17. Ju. P. Jacenko, *Integral Dynamic Models and Optimal Control Problems* (in Russian), Acad. of Sci., Ukr. SSR, V. M. Glushkov Inst. of Cyber., Kiev: 1985, 246 pp. (dep. in VINITI, N 8436-B-85).

18. L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory*, Philad. London Toronto: W. B. Saunders Comp., 1969, 488 pp. (translated into Russian, Moscow: Mir, 1974, 488 pp.).

V. V. Ivanov
V. M. Glushkov Institute of Cybernetics
Academy of Sciences of the Ukrainian SSR
252207, Kiev 207, USSR

Constantin Carathéodory: An International Tribute (pp. 527–542)
edited by Th. M. Rassias

ON CONTINUOUS SOLUTIONS OF THE EQUATION OF INVARIANT CURVES

Witold Jarczyk

Given a transform

$$T(x,y) = (f(x,y), g(x,y))$$

of the real plane one can ask about curves which are invariant under T, that is curves C satisfying the condition $T(C) \subset C$. If C is the graph of a function, say φ, the fact that C is invariant under T means analytically that φ is a solution of the equation

$$\text{(E)} \qquad \varphi(f(x,\varphi(x))) = g(x,\varphi(x))\,.$$

Equation (E) has been extensively studied by many authors (cf., for instance, Hadamard [2], Lattès [4] and Montel [5]). It is also the main subject of Chapter XIV of the monograph [3] by M. Kuczma, where the reader can find many further references concerning Eq. (E).

The approach to Eq. (E) presented here is an extension of some ideas used by J. Dhombres in the study of the equation

$$\text{(D)} \qquad \varphi(x+\varphi(x)) = c\varphi(x)$$

being a special case of (E) (see [1, Ch. 6, Sc. 2]). Theorem 1 below is a general result from which we shall derive two theorems concerning the equation

$$\text{(A)} \qquad \varphi(x+\varphi(x)) = p(\varphi(x))\,.$$

As a corollary we shall obtain the result of Dhombres [1, Theorem 6.4].

To begin with we are going to prove the following fact.

Lemma. Let I be a real interval and let $h : I \to I$ be a continuous function satisfying the condition

$$\text{if } x \in I,\ n \in \mathbb{N} \text{ and } h^{n+1}(x) = h^n(x) \text{ then } h(x) = x.$$

Assume that for every $x \in I$ the sequence $(h^n(x) : n \in \mathbb{N})$ converges in I. If the function $H : I \to I$, given by

$$H(x) = \lim_{n\to\infty} h^n(x),$$

is continuous then there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ such that $a \leq b$ and

$$H(x) = \begin{cases} a, & x \in I \cap (-\infty, a), \\ x, & x \in I \cap [a, b], \\ b, & x \in I \cap (b, +\infty). \end{cases}$$

Proof. Since

$$h^n \circ h = h^{n+1}, \quad n \in \mathbb{N},$$

we have

$$H \circ h = H. \tag{1}$$

Thus

$$H \circ h^n = H, \quad n \in \mathbb{N},$$

whence, by the continuity of H,

$$H \circ H = H.$$

Therefore $X = H(I)$ is an interval which is a closed subset of I and

$$H(x) = x, \quad x \in X.$$

Put $a = \inf X$ and $b = \sup X$. Assume that $b < \sup I$ and suppose that $H(x) \neq b$ for an $x \in I \cap (b, +\infty)$. Then $H(x) < b$ and, in particular,

$a < b$. Thus, by the equality $H(b) = b$ and the continuity of H, we can find an $x_0 \in (b, +\infty)$ with the property

$$H(x_0) \in (a, b).$$

Choose a number $n \in \mathbb{N}$ in such a manner that $h^n(x_0), h^{n+1}(x_0) \in (a, b)$. Then, on account of (1), (E) and again (1), we have

$$h^{n+1}(x_0) = H(h^{n+1}(x_0)) = H(h^n(x_0)) = h^n(x_0).$$

By the assumptions this means that $h(x_0) = x_0$, i.e., $H(x_0) = x_0$, whence $x_0 \in X$ which is impossible. Consequently,

$$H(x) = b, \quad x \in I \cap (b, +\infty).$$

Similarly, if $a > \inf I$ then

$$H(x) = a, \quad x \in I \cap (-\infty, a),$$

which completes the proof.

Let us consider the following hypothesis.

(H) I and J are real intervals, $0 \in J$. The functions $f : I \times J \to I$ and $g : I \times J \to J$ are continuous and satisfy the conditions

$$f(x, y) = x \text{ iff } y = 0, \ x \in I, y \in J, \tag{2}$$

$$g(x, y) = 0 \text{ iff } y = 0, \ x \in I, y \in J. \tag{3}$$

Given functions f and g such that hypothesis (H) holds, put

$$F_1(x, y) = f(x, y) \quad \text{and } G_1(x, y) = g(x, y),$$

$$F_{n+1}(x, y) = f(F_n(x, y), G_n(x, y))$$

and

$$G_{n+1}(x, y) = g(F_n(x, y), G_n(x, y))$$

for every $n \in \mathbb{N}$ and $x \in I, y \in J$.

Our main result reads as follows.

Theorem 1. Let hypothesis (H) hold and assume that for every $(x, y) \in I \times J$ the sequence $(F_n(x, y) : n \in \mathbb{N})$ converges in I and the function $F : I \times J \to I$, given by

$$F(x, y) = \lim_{n \to \infty} F_n(x, y),$$

is invertible with respect to the second variable and continuous.

If $\varphi : I \to J$ is a continuous solution of Eq. (E) then there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying the conditions

(i) $a \leq b$,

(ii) if $a \in \mathbb{R}$ then for every $x \in I \cap (-\infty, a)$ there is a $y \in J$ such that $F(x,y) = a$, if $b \in \mathbb{R}$ then for every $x \in I \cap (b, +\infty)$ there is a $y \in J$ such that $F(x,y) = b$,

(iii) if $a < b$ then

$$f(x, F(x,\cdot)^{-1}(a)) \leq a\,, \quad x \in I \cap (-\infty, a)\,,$$
$$f(x, F(x,\cdot)^{-1}(b)) \geq b\,, \quad x \in I \cap (b, +\infty)\,;$$

and

$$\varphi(x) = \begin{cases} F(x,\cdot)^{-1}(a)\,, & x \in I \cap (-\infty, a)\,, \\ 0\,, & x \in I \cap [a, b]\,, \\ F(x,\cdot)^{-1}(b)\,, & x \in I \cap (b, +\infty)\,. \end{cases} \tag{4}$$

Conversely, for every $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying conditions (i)–(iii) the function $\varphi : I \to J$ given by (4) is a continuous solution of Eq. (E).

Proof. Using (3) and a simple induction one can easily show that for every $n \in \mathbb{N}$

$$G_n(x,y) = 0 \quad \text{iff} \quad y = 0\,, \quad x \in I, y \in J$$

whence, on account of (2),

$$F_{n+1}(x,y) = F_n(x,y) \quad \text{iff} \quad y = 0\,,\ x \in I, y \in J\,. \tag{5}$$

Since $f(x,0) = x$ for every $x \in I$, it follows from (5) that

$$F_n(x,0) = x\,, \quad x \in I\,,$$

for every $n \in \mathbb{N}$. Therefore $F(x,0) = x$ for every $x \in I$, whence

$$F(x,y) = x \quad \text{iff} \quad y = 0\,, \quad x \in I, y \in J\,, \tag{6}$$

because of the invertibility of F with respect to the second variable. Another simple induction yields the condition

$$F_{n+1}(x,y) = F_n(f(x,y), g(x,y))\,, \quad x \in I, y \in J\,,$$

for every $n \in \mathbb{N}$, so we have

$$F(x,y) = F(f(x,y), g(x,y)), \quad x \in I, y \in J. \tag{7}$$

Let $\varphi : I \to J$ be a continuous solution of Eq. (E). The function $h : I \to I$, defined by

$$h(x) = f(x, \varphi(x)),$$

is continuous. We shall show that for every $n \in \mathbb{N}$ and $x \in I$

$$h^n(x) = F_n(x, \varphi(x)) \quad \text{and} \quad \varphi(h^n(x)) = G_n(x, \varphi(x)). \tag{8}$$

In the case $n = 1$ the first relation is clear and the second follows immediately from (E). Fix a positive integer n and assume (8) for all $x \in I$. Then

$$\begin{aligned} h^{n+1}(x) &= h(h^n(x)) = f(h^n(x), \varphi(h^n(x))) \\ &= f(F_n(x, \varphi(x)), G_n(x, \varphi(x))) \\ &= F_{n+1}(x, \varphi(x)) \end{aligned}$$

and, in view of (E),

$$\begin{aligned} \varphi(h^{n+1}(x)) &= \varphi(h(h^n(x))) = g(h^n(x), \varphi(h^n(x))) \\ &= g(F_n(x, \varphi(x)), G_n(x, \varphi(x))) \\ &= G_{n+1}(x, \varphi(x)) \end{aligned}$$

for every $x \in I$, which proves (8) for $n + 1$.

If $x \in I, n \in \mathbb{N}$ and $h^{n+1}(x) = h^n(x)$ then, according to the first of conditions (8) and relation (5), we obtain $\varphi(x) = 0$ which, due to (2), means that

$$h(x) = f(x, \varphi(x)) = f(x, 0) = x.$$

Consequently, the function h fulfils the assumptions of the Lemma.

By virtue of Lemma there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ such that $a \leq b$ and

$$F(x, \varphi(x)) = \begin{cases} a, & x \in I \cap (-\infty, a), \\ x, & x \in I \cap [a, b], \\ b, & x \in I \cap (b, +\infty). \end{cases} \tag{9}$$

If $x \in I \cap [a,b]$ then, in view of (6), we have $\varphi(x) = 0$. Thus statements (i) and (ii) hold true and φ has form (4). To prove (iii) assume that $a < b$ and fix a point $x \in I \cap (-\infty, a)$. Making use of relations (E), (7) and (9) we have

$$\begin{aligned} F(f(x,\varphi(x)),\varphi(f(x,\varphi(x)))) &= F(f(x,\varphi(x)),g(x,\varphi(x))) \\ &= F(x,\varphi(x)) = a\,, \end{aligned}$$

whence, by (4), (9) and the inequality $a < b$,

$$f(x,F(x,\cdot)^{-1}(a)) = f(x,\varphi(x)) \leq a\,.$$

Similarly, if $x \in I \cap (b,+\infty)$ then

$$f(x,F(x,\cdot)^{-1}(b)) \geq b\,.$$

Now fix any $a \in [-\infty,+\infty)$ and $b \in (-\infty,+\infty]$ satisfying conditions (i)–(iii) and define the function $\varphi : I \to J$ by formula (4). Then (cf. (6)) equality (9) holds true. Clearly φ is continuous in the set $I \cap (a,b)$.

Assume that $a \in \mathbb{R}$ and fix a point $x_0 \in I \cap (-\infty, a]$. Suppose that there exists a sequence $(x_n : n \in \mathbb{N})$ of points of I such that $\lim\limits_{n\to\infty} x_n = x_0$, $\lim\limits_{n\to\infty} \varphi(x_n) = y_0 \in \operatorname{cl} I$ and $y_0 \neq \varphi(x_0)$. According to (9) we have

$$\lim_{n\to\infty} F(x_n,\varphi(x_n)) = a\,. \tag{10}$$

The functions $F(x,\cdot), x \in I$, are strictly monotonic. Assume, for instance, that $F(x_0,\cdot)$ is strictly increasing (due to the continuity of F this means that all functions $F(x,\cdot)$ are strictly increasing) and $y_0 > \varphi(x_0)$. In the remaining cases the argument is quite similar. Fix a point $y^* \in I \cap (\varphi(x_0), y_0)$. The function $F(\cdot,y^*)$ is continuous and (cf. (9))

$$F(x_0,y^*) > F(x_0,\varphi(x_0)) = a\,.$$

Thus there exist a number $c > a$ and a positive number ε such that

$$F(x,y^*) > c, \quad x \in I \cap (x_0 - \varepsilon, x_0 + \varepsilon)\,.$$

Therefore, since the functions $F(x,\cdot)$, $x \in I$, are increasing,

$$F(x,y) > c, \quad x \in I \cap (x_0 - \varepsilon, x_0 + \varepsilon), y \in J \cap (y^*, +\infty).$$

So, choosing a number $n_0 \in \mathbb{N}$ in such a way that $x_n \in (x_0 - \varepsilon, x_0 + \varepsilon)$ and $\varphi(x_n) > y^*$ for $n \geq n_0$, we have

$$F(x_n, \varphi(x_n)) > c > a, \quad n \geq n_0,$$

which contradicts property (10). This proves that the function φ is continuous in the set $I \cap (-\infty, a]$. Analogously one can show the continuity of φ in the set $I \cap [b, +\infty)$.

Finally we shall verify that φ is a solution of Eq. (E). At first let us consider the case $a = b$. Then necessarily $a \in \mathbb{R}$. Moreover, due to (9),

$$F(x, \varphi(x)) = a \tag{11}$$

for every $x \in I$. Fix an $x \in I$. Then, by virtue of (11) (used for x and then for $f(x, \varphi(x))$) and (7), we get

$$\begin{aligned} F(f(x,\varphi(x)), \varphi(f(x,\varphi(x)))) &= a \\ &= F(x,\varphi(x)) = F(f(x,\varphi(x)), g(x,\varphi(x))) \end{aligned}$$

which, since F is invertible with respect to the second variable, means that equality (E) is fulfilled.

Now assume that $a < b$. If $x \in I \cap (-\infty, a)$ then, according to (iii) and (4), $f(x, \varphi(x)) \leq a$. Thus, by (9), relation (11) is satisfied by x as well as $f(x, \varphi(x))$. Hence and by (7)

$$\begin{aligned} &F(f(x,\varphi(x)), \varphi(f(x,\varphi(x)))) \\ &\quad = F(x,\varphi(x)) = F(f(x,\varphi(x)), g(x,\varphi(x))) \end{aligned}$$

and equality (E) holds true. If $x \in I \cap (b, +\infty)$ we proceed similarly. In the case $x \in I \cap [a, b]$ it is enough to observe that (cf. (4)) $\varphi(x) = 0$, so, due to conditions (2) and (3),

$$f(x, \varphi(x)) = x \quad \text{and} \quad g(x, \varphi(x)) = 0 = \varphi(x)$$

and (E) follows immediately. This completes the proof.

Now we shall apply Theorem 1 to Eq. (A).

Theorem 2. Let $p : \mathbb{R} \to \mathbb{R}$ be a continuous function such that

$$p(x) = 0 \quad \text{iff} \quad x = 0, \quad x \in \mathbb{R}.$$

Assume that for every $x \in \mathbb{R}$ the series $\sum_{n=0}^{\infty} p^n(x)$ converges and the function $P_+ : \mathbb{R} \to \mathbb{R}$, given by

$$P_+(x) = \sum_{n=0}^{\infty} p^n(x),$$

is invertible and continuous.

If $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous solution of Eq. (A) then there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying the conditions

(iv) $a \leq b$,

(v) if $a \in \mathbb{R}$ then $(0, +\infty) \subset P_+(\mathbb{R})$, if $b \in \mathbb{R}$ then $(-\infty, 0) \subset P_+(\mathbb{R})$,

(vi) if $-\infty < a < b$ then

$$P_+^{-1}(x) \leq x, \quad x \in (0, +\infty),$$

if $a < b < +\infty$ then

$$P_+^{-1}(x) \geq x, \quad x \in (-\infty, 0);$$

and

$$\varphi(x) = \begin{cases} P_+^{-1}(a - x), & x \in (-\infty, a), \\ 0, & x \in \mathbb{R} \cap [a, b], \\ P_+^{-1}(b - x), & x \in (b, +\infty). \end{cases} \tag{12}$$

Conversely, for every $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying conditions (iv)–(vi) the function $\varphi : \mathbb{R} \to \mathbb{R}$ given by (12) is a continuous solution of Eq. (A).

Proof. Put $I = J = \mathbb{R}$,

$$f(x, y) = x + y \quad \text{and} \quad g(x, y) = p(y), \quad x, y \in \mathbb{R}.$$

Then hypothesis (H) is fulfilled. Moreover,

$$F_n(x,y) = x + \sum_{k=0}^{n-1} p^k(y) \quad \text{and} \quad G_n(x,y) = p^n(y),$$

for every $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$, whence

$$F(x,y) = x + P_+(y), \quad x, y \in \mathbb{R}. \tag{13}$$

In particular, F is invertible with respect to the second variable and continuous.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous solution of Eq. (A). Then, by virtue of Theorem 1, there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ such that conditions (i)–(iii) and equality (4) hold. If $a \in \mathbb{R}$ and $x \in (0, +\infty)$ then, by (ii) and (13), there is a $y \in \mathbb{R}$ for which

$$(a - x) + P_+(y) = a,$$

i.e., $x = P_+(y)$. This means that $(0, +\infty) \subset P_+(\mathbb{R})$. Analogously one can check that $(-\infty, 0) \subset P_+(\mathbb{R})$ provided $b \in \mathbb{R}$. Now assume that $-\infty < a < b$ and fix an $x \in (0, +\infty)$. Then, using (13) and (iii), we get

$$\begin{aligned} P_+^{-1}(x) &= (a - x) + P_+^{-1}(a - (a - x)) + (x - a) \\ &= (a - x) + F(a - x, \cdot)^{-1}(a) + (x - a) \\ &= f(a - x, F(a - x, \cdot)^{-1}(a)) + (x - a) \\ &\leq a + (x - a) = x. \end{aligned}$$

Similarly we can verify that if $a < b < +\infty$ and $x \in (-\infty, 0)$ then $P_+^{-1}(x) \geq x$. Equality (12) follows directly from (4) and (13).

The converse also can be proved by reduction to Theorem 1. Nevertheless we present here some immediate argument.

Fix $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying conditions (iv)–(vi) and define the function $\varphi : \mathbb{R} \to \mathbb{R}$ by (12). Clearly φ is continuous. We shall check that φ is a solution of Eq. (A). It follows from the definition of P_+ that

$$P_+(p(x)) = P_+(x) - x, \quad x \in \mathbb{R}. \tag{14}$$

If $a = b$ then

$$\varphi(x) = P_+^{-1}(a - x), \quad x \in \mathbb{R},$$

whence, on account of (14),

$$P_+(\varphi(x+\varphi(x))) = a - x - \varphi(x)$$
$$= P_+(\varphi(x)) - \varphi(x) = P_+(p(\varphi(x)))$$

for every $x \in \mathbb{R}$, i.e., (A) holds. Now assume that $a < b$. If $x < a$ then $a - x \in (0, +\infty)$, so, by (12) and (vi),

$$x + \varphi(x) = x + P_+^{-1}(a - x) \leq x + (a - x) = a$$

whence, using (12) and (14), again we get (A). A similar argument can be used if $x > b$.

Also the next result is a consequence of Theorem 1 although the assumptions considered here are different from those imposed in Theorem 2.

Theorem 3. Let p be a homeomorphism mapping $\mathbb{R}$ onto itself. Assume that for every $x \in \mathbb{R}$ the series $\sum_{n=1}^{\infty} p^{-n}(x)$ converges and the function $P_- : \mathbb{R} \to \mathbb{R}$, given by

$$P_-(x) = \sum_{n=1}^{\infty} p^{-n}(x),$$

is invertible and continuous.

If $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous solution of Eq. (A) then there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying the conditions

(vii) $a \leq b$,

(viii) if $a \in \mathbb{R}$ then $(-\infty, 0) \subset P_-(\mathbb{R})$, if $b \in \mathbb{R}$ then $(0, +\infty) \subset P_-(\mathbb{R})$,

(ix) if $-\infty < a < b$ then

$$p^{-1}(P_-^{-1}(x)) \geq x, \quad x \in (-\infty, 0),$$

if $a < b < +\infty$ then

$$p^{-1}(P_-^{-1}(x)) \leq x, \quad x \in (0, +\infty);$$

and

$$\varphi(x) = \begin{cases} P_-^{-1}(x-a), & x \in (-\infty, a), \\ 0, & x \in \mathbb{R} \cap [a,b], \\ P_-^{-1}(x-b), & x \in (b, +\infty). \end{cases} \tag{15}$$

Conversely, for every $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying conditions (vii)–(ix) the function $\varphi : \mathbb{R} \to \mathbb{R}$ given by (15) is a continuous solution of Eq. (A).

Proof. It follows from the assumption that

$$\lim_{n\to\infty} p^{-n}(x) = 0, \quad x \in \mathbb{R},$$

whence, by the continuity of p^{-1}, we get $p^{-1}(0) = 0$. Therefore

$$p(x) = 0 \quad \text{iff} \quad x = 0, \quad x \in \mathbb{R}.$$

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous solution of Eq. (A). The functions $h : \mathbb{R} \to \mathbb{R}$ and $h^* : \mathbb{R} \to \mathbb{R}$, defined by

$$h(x) = x + \varphi(x) \quad \text{and} \quad h^*(x) = x - p^{-1}(\varphi(x)), \quad x \in \mathbb{R}, \tag{16}$$

are continuous and, in view of (A),

$$\varphi(h(x)) = p(\varphi(x)), \quad x \in \mathbb{R}. \tag{17}$$

According to (17) we have

$$h(x) - p^{-1}(\varphi(h(x))) = h(x) - \varphi(x) = x$$

for every $x \in \mathbb{R}$, i.e.,

$$h^*(h(x)) = x, \quad x \in \mathbb{R}, \tag{18}$$

which implies the invertibility of h. So the function h is strictly monotonic and, consequently, there exist limits

$$u = \lim_{x\to-\infty} h(x) \quad \text{and} \quad v = \lim_{x\to+\infty} h(x).$$

If $u \in \mathbb{R}$ then by (16) $\lim_{x\to-\infty} \varphi(x) = +\infty$ whence, by virtue of (17),

$$|\varphi(u)| = \lim_{x\to-\infty} |\varphi(h(x))| = \lim_{x\to-\infty} |p(\varphi(x))| = +\infty$$

which is impossible. Therefore $u \in \{-\infty, +\infty\}$ and, similarly, $v \in \{-\infty, +\infty\}$. Moreover, due to the monotonicity of h, we have $u \neq v$. Thus, since the function h is continuous,

$$h(\mathbb{R}) = \mathbb{R}.$$

Hence and from (18) we deduce that $h^* = h^{-1}$. Therefore, by (17), we obtain

$$\varphi(h^*(x)) = p^{-1}(\varphi(h(h^*(x)))) = p^{-1}(\varphi(x)), \quad x \in \mathbb{R},$$

which means that the function φ satisfies the equation

$$\varphi(x - p^{-1}(\varphi(x))) = p^{-1}(\varphi(x)). \tag{A*}$$

Putting $I = J = \mathbb{R}$,

$$f(x,y) = x - p^{-1}(y) \quad \text{and} \quad g(x,y) = p^{-1}(y), \quad x,y \in \mathbb{R},$$

we see that hypothesis (H) holds. Moreover,

$$F_n(x,y) = x - \sum_{k=1}^{n} p^{-k}(y) \quad \text{and} \quad G_n(x,y) = p^{-n}(y)$$

for every $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$ and

$$F(x,y) = x - P_-(y), \quad x,y \in \mathbb{R}.$$

Applying Theorem 1 in a similar way as in the proof of Theorem 2 one can show that the function φ has form (15) with some $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying (vii), (viii) and (ix).

Now fix any $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying conditions (vii)–(ix) and define φ by formula (15). Making use of Theorem 1 or proceeding as in the proof of Theorem 2 we infer that φ is a continuous solution of Eq. (A*). Define $h : \mathbb{R} \to \mathbb{R}$ and $h^* : \mathbb{R} \to \mathbb{R}$ by (16). Then we have

$$\varphi(h^*(x)) = p^{-1}(\varphi(x)), \quad x \in \mathbb{R}, \tag{19}$$

and

$$h^*(x) + \varphi(h^*(x)) = x - p^{-1}(\varphi(x)) + p^{-1}(\varphi(x)) = x$$

for every $x \in \mathbb{R}$, whence

$$h(h^*(x)) = x\,, \quad x \in \mathbb{R}\,. \tag{20}$$

In particular, the function h^* is invertible. Repeating the argument used in the first part of the proof one can show that

$$h^*(\mathbb{R}) = \mathbb{R}\,.$$

Therefore, due to condition (20), $h^* = h^{-1}$ whence, by virtue of (19), we obtain

$$\varphi(h(x)) = p(\varphi(h^*(h(x)))) = p(\varphi(x))\,, \quad x \in \mathbb{R}\,,$$

and, consequently, φ satisfies Eq. (A).

Remark. In connection with Theorem 2 (as well as Theorem 3) it is desirable to know some simple conditions ensuring the convergence of the series $\sum_{n=0}^{\infty} p^n(x)$ and the continuity of its sum. One of the possible answers to this problem is the following.

Let $p : \mathbb{R} \to \mathbb{R}$ be a continuous function such that

$$0 < p(x)/x < 1\,, \quad x \in \mathbb{R}\backslash\{0\}\,. \tag{21}$$

If p has the continuous derivative in an interval $[-\delta, \delta]$(δ is positive), $0 < p'(0) < 1$ and there exist positive numbers M and μ such that

$$|p'(x) - p'(0)| \leq M|x|^{\mu}\,, \quad x \in [-\delta, \delta]\,,$$

(the latter condition is certainly fulfilled if p has the second derivative at zero) then the series $\sum_{n=0}^{\infty} p^n(x)$ converges for every $x \in \mathbb{R}$ and its sum is a continuous function.

To see this fix $u, v \in \mathbb{R}$ such that $-\infty < u < 0 < v < +\infty$ and put $s = p'(0)$. Since $p'(0) > 0$ we can assume that δ is so small that the function p increases in $[-\delta, \delta]$. It follows from (21) that

$$\lim_{n\to\infty} p^n(x) = 0$$

uniformly on the interval $[u, v]$. Choose an $n_0 \in \mathbb{N}$ in such a way that

$$p^{n_0}(x) \in [-\delta, \delta], \quad x \in [u, v]. \tag{22}$$

According to a result of G. Szekeres [6] (see also [3, Theorems 6.3 and 6.2]) the sequence $(p^n(x)/s^n : n \in \mathbb{N})$ converges for every $x \in [-\delta, \delta]$ and its limit $q : [-\delta, \delta] \to \mathbb{R}$ is a continuous (even of class C^1) function. Since the functions $p^n/s^n, n \in \mathbb{N}$, increase in $[-\delta, \delta]$ it follows that

$$\lim_{n\to\infty} p^n(x)/s^n = q(x)$$

uniformly in the set $[-\delta, \delta]$. Put

$$K = \sup q([-\delta, \delta])$$

and let $n_1 \in \mathbb{N}$ be such that

$$|p^n(x)/s^n - q(x)| \leq 1, \quad x \in [-\delta, \delta], \ n \geq n_1.$$

Fix $n \geq n_0 + n_1$ and $x \in [u, v]$. Then, by (22), we have $p^{n_0}(x) \in [-\delta, \delta]$ whence

$$\begin{aligned} 0 < p^n(x) &= p^{n-n_0}(p^{n_0}(x)) \\ &\leq s^{n-n_0}(q(p^{n_0}(x)) + 1) \leq (K+1)s^{n-n_0}. \end{aligned}$$

Since $s \in (0, 1)$ this means that the series $\sum_{n=0}^{\infty} p^n(x)$ converges uniformly in $[u, v]$ and completes the proof.

Finally we shall find the form of all continuous solutions $\varphi : \mathbb{R} \to \mathbb{R}$ of Eq. (D) where $c \in \mathbb{R}\backslash\{-1, 0, 1\}$. This is the main part of Theorem 6.4 from the book [1] by J. Dhombres. The special cases where $c = -1, 0, 1$ (considered also by Dhombres) are classical. If $c = -1$ then putting $h(x) = x + \varphi(x)$ for every $x \in \mathbb{R}$, we see that h is an involution, i.e., it satisfies the equation

$$h^2(x) = x.$$

The form of all such continuous functions is well knwon (cf. for instance [3, Theorems 15.3, 15.2 and Lemma 15.2]). In the case $c = 0$ the function φ satisfies Eq. (D) if and only if h is a solution of the equation of idempotence

$$h^2(x) = h(x).$$

The reader can easily observe that a continuous function $h : \mathbb{R} \to \mathbb{R}$ satisfies this equation if and only if there exists an interval X (maybe a singleton) such that $h_{|X}$ is the identity function and $h(\mathbb{R}) = X$. If $c = 1$ Eq. (D) becomes the well-known Euler's equation

$$\varphi(x + \varphi(x)) = \varphi(x)$$

whose only continuous solutions defined in $\mathbb{R}$ are constant functions (see [7]).

The remaining cases are described in the following result.

Corollary. Assume that $c \in (0, +\infty)\backslash\{1\}$. A function $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous solution of Eq. (D) if and only if there exist $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ such that $a \leq b$ and

$$\varphi(x) = \begin{cases} (c-1)(x-a), & x \in (-\infty, a), \\ 0, & x \in \mathbb{R} \cap [a, b], \\ (c-1)(x-b), & x \in (b, +\infty). \end{cases} \tag{23}$$

Assume that $c \in (-\infty, 0)\backslash\{-1\}$. A function $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous solution of Eq. (D) if and only if either

$$\varphi(x) = 0, \quad x \in \mathbb{R},$$

or there exists an $a \in \mathbb{R}$ such that

$$\varphi(x) = (c-1)(x-a), \quad x \in \mathbb{R}. \tag{24}$$

Proof. Assume that $0 < |c| < 1$ and put

$$p(x) = cx, \quad x \in \mathbb{R}.$$

Then the series $\sum_{n=0}^{\infty} p^n(x)$ converges for every $x \in \mathbb{R}$ and

$$P_+(x) = \sum_{n=0}^{\infty} p^n(x) = \frac{x}{1-c}, \quad x \in \mathbb{R}.$$

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous solution of Eq. (D). By Theorem 2 the function φ has form (23) with some $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$ satisfying conditions (iv)-(vi). If, in addition, c is negative then, by (vi), either $a = -\infty$ and $b = +\infty$ or $a = b \in \mathbb{R}$, i.e., either φ is the zero function or φ is of form (24).

To see that all functions of form (23) satisfy (D) in the case $c \in (0, 1)$ and all functions of form (24) are solutions of Eq. (D) provided $c \in (-1, 0)$ one can apply Theorem 2 or simply a direct calculation.

If $|c| > 1$ then we can proceed in a similar way recalling Theorem 3 instead of Theorem 2.

References

1. J. Dhombres, *Some Aspects of Functional Equations*, Chulalongkorn University Press, Bangkok, 1979.
2. J. Hadamard, *Sur l'tération et solutions asymptotiques des équations différentielles*, Bull. Soc. Math. France **29** (1901), 224–228.
3. M. Kuczma, *Functional equations in a single variable*, Monografie Mat. **46**, PWN – Polish Scientific Publishers, Warszawa, 1968.
4. S. Lattés, *Sur une classes d'équations fonctionnelles*, C. R. Acad. Sci. Paris **137** (1903), 905–908.
5. P. Montel, Leçons sur les Récurrences et leurs Applications, Paris, 1957.
6. G. Szekeres, *Regular iteration of real and complex functions*, Acta Math. **100** (1958), 203–258.
7. R. Wagner, *Eindeutige Lösungen der Funktionalgleichung* $f[x+f(x)]=f(x)$, Elem. Math. **14** (1959), 73–78.

Witold Jarczyk
Institute of Mathematics
Silesian University
40-007 Katowice
Poland

Constantin Carathéodory: An International Tribute (pp. 543–547)
edited by Th. M. Rassias

BOUNDS FOR AN OPTIMAL SEARCH

Richard D. Järvinen

ABSTRACT

An object is expected to be located at a certain point in the plane. Its position coordinates are assumed to be independently normally distributed. In this setting a summary of an earlier analysis that gives an optimal searching procedure to maximize the probability of detecting the object at all times during a search using an imperfect detecting apparatus is succinctly provided. New to this paper are bounds for the existence of feasible solutions (searching procedures) to detect the object and their impact on an analysis to produce the optimal solution. The mathematical context of this paper is the minimization of a nonlinear function of several variables subject to linear constraints where both the number of variables to use as well as their values must be determined.

1. Introduction

The mathematics in this paper deals with optimization and is explained in the context of an optimal searching procedure. We assume that an object (particle, vehicle) in the plane that is to be found follows a bivariate normal distribution in the independent variables x and y. As usual, their means and standard deviations are denoted μ_x, μ_y and σ_x, σ_y, respectively. Of all regions in the plane with a fixed probability of the object being within it, an ellipse centered at the origin and coaxial with the coordinate axes has the smallest area [5,p. 328]. In this paper to optimally search for the object in question, assuming that one searches only small regions sequentially, means to maximize the probability of locating the object at all times during the searching procedure. The actual object being pursued might be a satellite and the apparatus used to find it might be a radar with a well focused beam pulsing and scanning the sky, or the object might be

a small tumor and the searching apparatus that of a laser gun which destroys with each firing a small amount of tissue as it is moved from position to position in the hunt for the tumor.

We assume that the apparatus used to carry out the search is not a perfect detector. In the case of the radar this means that the radar will not always say that it is seeing the satellite when it actually is, or in the case of the laser, that the laser will not always destroy the tumor even when it has it as its target. We do not assume, then, that the probability of detection on one pulse of the radar, $P(D_1)$, is 1. To maximize the probability of finding the object, certain concentric, coaxial elliptical regions will be searched and some of them will be searched multiple times. A context involving the use of a radar to carry out the search and an accompanying analysis is more fully described in [3]. Some highlights of that mathematical analysis are presented in the remainder of this introduction.

The mathematical circumstance with which we deal becomes this. We wish to minimize the nonlinear function

$$(1) \qquad A_T = -2\pi\sigma_x\sigma_y \sum_{k=1}^{N} \ln [1 - P(E_k)]$$

subject to the linear constraints

$$(2) \qquad P(A) = P(D_1)[P(E_1) + q\,P(E_2) + \ldots + q^{k-1}P(E_k) + \ldots + q^{N-1}P(E_N)], \text{ where } q = 1 - P(D_1)$$

$$P(E_k) > 0, \quad k = 1, 2, \ldots, N$$

where $P(A)$ is the probability of acquisition, i.e., the probability of actually finding the object upon completion of the search--$P(A)$ is a user input value. $P(E_k)$ is the probability that the vehicle exists in the kth innermost ellipse in the family of concentric, coaxial ellipses to be searched; again see [3] for details. N denotes the total number of ellipses in the family of ellipses to be searched and also represents the maximum number of times any one ellipse is searched. Finally, A_T denotes the total area to be searched, where if a given region (always an ellipse) is searched twice, then its area is added in twice.

Mathematically interesting in this setting is the fact that both

the number of ellipses to be used in the searching procedure as well as their sizes must be determined. Saying the latter in mathematical terms is to say that we wish to minimize a real valued function of several real variables by not only specifying the values of the variables but also the number of them that must be used.

In the earlier work [3] it is shown that the area of the kth ellipse, an arbitrary member of the family of elliptical regions to be searched to obtain a feasible solution, is

$$A_k = 2\pi\sigma_x\sigma_y \ln [N P(D_1) q^{k-1}/(1 - P(A) - q^N)] \tag{3}$$

where the Lagrange multiplier method and/or bordered Hessian can be employed to establish this result. By a feasible solution we mean a searching procedure which, when completed, yields the probability of acquisition, P(A), the input value.

2. Bounds for Feasible Solutions

In this section we establish four results. These are:

1. There is a smallest N for which there is a feasible solution.
2. There is a largest N for which there is a feasible solution.
3. The values of A_T decrease with increasing N if $q < 1/4$.
4. The optimal solution occurs for the feasible solution with the largest N when $q < 1/4$.

Lemma 2.1. There is a smallest N for which there is a feasible solution, and it is the smallest N for which

$$1 - P(A) - q^N > 0 \tag{4}$$

Proof: To establish this claim all we need show is that

$$N P(D_1) q^{k-1}/(1 - P(A) - q^N) > 1, \quad k = 1, 2, \ldots, N \tag{5}$$

The numerator is smallest when k = N. Thus (5) holds when

$$[1 - P(A)]/q^{N-1} < 1 - P(D_1) + N P(D_1) \tag{6}$$

We observe that the left side is less than or equal to 1 because

(7) $$1 - P(A) - q^{N-1} \leq 0$$

since N is the smallest N for which

(8) $$1 - P(A) - q^{N} > 0$$

The right side is greater than or equal to 1; it equals 1 if N = 1. Note that (5) remains true if N = 1. QED

Lemma 2.2. There is a largest N for which there is a feasible solution.

Proof. A_N becomes negative as N gets large since the numerator in the quotient in the logarithm in (3), when k = N, goes to zero by the nth term test applied to the convergent series

(9) $$\Sigma\, N\, q^{N-1}$$

Feasible solutions exist only when A_k, k = 1, 2, ... , N, are positive. Here, A_k is the area in the kth ellipse of the family of ellipses to be searched. QED

Lemma 2.3. For feasible solutions and $q < 1/4$, the values of A_T are strictly monotonically decreasing as N increases.

Proof. In [3] we find that A_T (given by (1) above) is also expressed

(10) $$A_T = 2\pi\sigma_x\sigma_y \ln [N\, P(D_1)\, q^{(N-1)/2}/(1 - P(A) - q^{N})]^{N}$$

Letting b_N be the quotient in the logarithm that is raised to the Nth power, we find that

(11) $$b_{N+1}/b_N \leq [(N+1)/N][q^{1/2}][1 - P(A) - q^{N}][1 - P(A) - q^{N+1}]^{-1}$$
$$\leq [(N+1)/N][q^{1/2}]$$
$$< 1 \text{ if } q < 1/4$$

Hence, A_T is strictly monotonically decreasing under the given conditions. QED

We note that $P(D_1) > 0.75$ is a practical range of probability of detection values.

Lemma 2.4. If $1 - P(D_1) < 1/4$, then the optimal solution occurs for the largest N for which $A_T > 0$.

Proof. Lemma 2.4 follows at once from Lemma 2.3 since Lemma 2.3 tells us the area to be searched is smallest when N is largest. QED

There are other ways to formulate conditions for determining the value of N for which the optimal solution occurs. One can always compute (by computer methods) the total area to be searched for each feasible solution--we know there are finitely many feasible solutions as a result of the propositions proved above--and take the value of N to be the one that corresponds to the feasible solution for which the total search area A_T is least.

REFERENCES

[1] Buck, R.C., Advanced Calculus (McGraw-Hill, New York, 1956).

[2] Gnedenko, B.V., The Theory of Probability (Chelsea, New York, 1962).

[3] Järvinen, R.D., Conditions for an Optimal Search. In: Selected Studies: Physics-Astrophysics, Mathematics, History of Science, T.M. Rassias and G.M. Rassias (editors) (North-Holland, Amsterdam, 1982).

[4] Stone, L.D., Theory of Optimal Search (Academic Press, New York, 1975).

[5] Uspensky, J.V., Introduction to Mathematical Probability (McGraw-Hill, New York, 1937).

Richard D. Järvinen
Winona State University
Winona, Minnesota 55987
U.S.A.

Constantin Carathéodory: An International Tribute (pp. 548–553)
edited by Th. M. Rassias

ON ANALYTIC PATHS

James A Jenkins

1. In many contexts a useful technical role is played by the intersection and composition properties of analytic curves. These questions seem to have been treated first in a complete and precise manner by Minda [2] and, as he has said, the statements and purported proofs of other authors are frequently vague or actually erroneous. Our present purpose is to show how, by reordering the material, a very significant technical simplification of the treatment of the primary results can be obtained.

2. Since our terminology is rather different from Minda's we will first give a summary of a number of definitions. There is no significant difference in working on a general Riemann surface or in the plane so to avoid unnecessarily pedantic phraseology we will assume that all entities lie in the plane.

Definition 1. A path is a continuous function $z(t)$, $0 \leq t \leq 1$. An open path is a continuous function $z(t)$, $0 < t < 1$. A closed path is a path for which $z(0) = z(1)$. An arc is the homeomorphic image of a closed

segment. An open arc is the (1,1) continuous image of an open segment. Two paths $z_1(t), z_2(t)$ are said to be equivalent if there is a homeomorphism $\lambda(t)$ of the interval [0,1] onto itself (sensed) so that $z_1(\lambda(t)) = z_2(t), 0 \leq t \leq 1$, (This is evidently an equivalence relation.) A path $z_2(t)$ is said to be a subpath of $z_1(t)$ if it equivalent to a path $z_1^{(a,b)}(t)$,

$$z_1^{(a,b)}(t) = z_1(a + (b-a)\,t), 0 \leq a \neq b \leq 1.$$

The set of points given by $z(t)$, $t \in [0,1]$, is denoted by $\{z(t)\}$.

Definition 2. An analytic path $z(t)$ is one for which there exists a domain Δ symmetric under reflection in the real axis, containing the segment [0, 1] and a function $f(z)$ regular in Δ with $f'(t) \neq 0, 0 \leq t \leq 1$, with $z(t) = f(t), 0 \leq t \leq 1$. An analytic closed path $z(t)$ is one for which there exists a domain Δ containing $|z| = 1$, symmetric under reflection in this circumference and a function $f(z)$ regular in Δ with $f'(z) \neq 0, |z| = 1$, with $z(t) = f(e^{2\pi i t}), 0 \leq t \leq 1$.

Definition 3. A locally analytic path $z(t)$ is one for which for every $\hat{t}$, $0 \leq \hat{t} \leq 1$, there is a neighborhood Δ of $\hat{t}$ symmetric in the real axis containing an open interval (t', t'') with $t' < \hat{t} < t''$ and a conformal mapping $\varphi(z)$ of Δ onto a neighborhood $\equiv$ of $z(\hat{t})$ such that $z(t) = \varphi(t)$, $t \in (t', t'') \cap [0,1]$. A locally analytic closed path $z(t)$ is one for which for every $\hat{t}, 0 \leq \hat{t} \leq 1$, there is a neighborhood Δ of $e^{2\pi i \hat{t}}$ symmetric in $|z| = 1$ containing an open arc of that circumference with end points $e^{2\pi i t'}$, $e^{2\pi i t''}$ and a conformal mapping $\varphi(z)$ of Δ onto a neighborhood $\equiv$ of $z(\hat{t})$ such that $z(t) = \varphi(e^{2\pi i t}), t \equiv t^* \pmod 1, t^* \in (t', t'')$.

3. It is immediate that an analytic path is locally analytic. The converse is

essentially true.

Theorem 1. *A locally analytic path is equivalent to an analytic path.*

Let $z(t)$ be a locally analytic path. Let T be the subset of $(0,1]$ such that, for $\tau \in T$, $z^{(0,\tau)}(t)$ is equivalent to an analytic path. We have at once that T is non–void and open relative to $(0,1]$. To see that it is closed relative to $(0, 1]$ we take $\hat{t} \in (0,1]$ which is the limit of a sequence of values $\{t_n\}$ in T. (Evidently we may assume $t_n < \hat{t}$.) Let $\equiv$, φ be the neighborhood and function of Definition 3. For ρ sufficiently small the image C_ρ of $|z-\hat{t}| = \rho$ under φ will lie in $\equiv$ meeting $\{z(t)\}$ in points $z(t'_\rho)$, $z(t''_\rho)$ with $t'_\rho < \hat{t} < t''_\rho$. We choose t_n so large that $t'_\rho < t_n < \hat{t}$. There exists λ as in Definition 1 with $z^{(0,\, t_n)}(\lambda(t)) = \tilde{z}(t)$, an analytic path with corresponding domain and function Δ, f as in Definition 2. Let $\lambda^{-1}(t'_\rho) = \tilde{t}$. Let $\Delta_\epsilon(a,b)$, $0 \le a < b \le 1$, denote the subset (defined for $\epsilon > 0$ sufficiently small) of Δ consisting of points within ϵ of $[a, b]$. We choose t^*, $\tilde{t} < t^* < 1$, and ϵ sufficiently small that f is univalent on $\Delta_\epsilon(t^*,1)$ and $f(\Delta_\epsilon(t^*,1))$ lies inside C_ρ. For suitable σ, $0 < \sigma < \rho$, an arc on C_σ will provide a crosscut β of $f(\Delta_\epsilon(t^*,1))$ dividing this domain into subdomains A' and B' so that $\tilde{z}(t^*)$, $\tilde{z}(1)$ lie in A' and B' is bounded by β and a crosscut γ of the inside of C_σ dividing the latter domain into subdomains B' and C'. The image b of β under f^{-1} divides $\Delta_\epsilon(0,1)$ into subdomains A, B with $B = f^{-1}(B')$. The image c of γ under φ^{-1} divides $|z-\hat{t}| < \sigma$ into subdomains $\tilde{B}$, $\tilde{C}$ with $\tilde{B} = \varphi^{-1}(B')$, $\tilde{C} = \varphi^{-1}(C')$. Then there exists a domain $\mathcal{D}$ symmetric in the real axis and divided by crosscuts λ, μ (also symmetric in the real axis) into

subdomains $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ for which there exist conformal mappings F of $\Delta_\epsilon(0,1)$ onto $\mathscr{A} \cup \lambda \cup \mathscr{B}$ and Φ of $|z-\hat{t}| < \sigma$ onto $\mathscr{B} \cup \mu \cup \mathscr{C}$ (both possessing reflectional symmetry in the real axis) such that $fF^{-1} = \varphi\Phi^{-1}$ on $\mathscr{B}$. This follows of course from the General Uniformization Theorem but from an expository point of view it is worth noting that is requires only the simplest step of blending of domains, see for example [1, pp. 98,99]. We can further assume that [0,1] lies in $\mathscr{D}$ with $fF^{-1}(0) = z(0)$, $\varphi\Phi^{-1}(1) = z(\hat{t})$. Setting

$$\tilde{f} = fF^{-1} \text{ on } \mathscr{A} \cup \lambda \cup \mathscr{B}$$

$$\tilde{f} = \varphi\Phi^{-1} \text{ on } \mathscr{B} \cup \mu \cup \mathscr{C}$$

we obtain a regular function in $\mathscr{D}$ which provides an equivalent analytic parametrization of $z^{(0,\hat{t})}(t)$. Thus T is closed relative to (0,1] and $T \equiv (0,1]$.

The preceding proof can readily be modified to prove the corresponding result for closed paths.

Theorem 1′. *A locally analytic closed path is equivalent to an analytic closed path.*

4. Theorem 2. *Let $z_1(t)$, $z_2(t)$ be analytic paths. Then there are two possibilities.*

I. *There exist only finitely many values t such that $z_1(t) = z_2(\hat{t})$ for $0 \le \hat{t} \le 1$. In this case the sets $\{z_1(t)\}$, $\{z_2(t)\}$ have only a finite number of common points.*

II. *The paths $z_1(t)$, $z_2(t)$ have a common subpath. In this case $z_1(t)$, $z_2(t)$ are subpaths of an analytic path.*

Suppose there is a sequence of distinct values $\{t_j\}$ such that $z_1(t_j) = z_2(\hat{t}_j)$, $j = 1,2,\ldots$ We can assume that the values t_j converge to t^* and that also $\hat{t}_j$ converge to $\hat{t}^*_j$. There exist neighborhoods of $z_1(t^*)$, $z_2(\hat{t}^*)$ as in Definition 3 images of $\Delta, \hat{\Delta}$ under conformal mappings $\varphi, \hat{\varphi}$. In a neighborhood of t^* $\hat{\varphi}^{-1}\varphi$ is a conformal mapping ψ. In particular from a certain stage on the $\hat{t}_j$ are distinct. Thus $\psi'(t^*)$ is real and from the symmetry properties of $\varphi, \hat{\varphi}$ is follows that ψ maps a segment on the real axis onto another such. Thus $z_1(t), z_2(t)$ have a common subpath.

If we choose this subpath to be maximal it is seen at once that at least one of the corresponding intervals on [0, 1] for $z_1(t)$ or $z_2(t)$ has an end point at 0 or 1. Thus there are one, two, three, or four subpaths of $z_1(t), z_2(t)$ which can be parametrized consecutively so as to obtain a path $z(t)$ which has $z_1(t), z_2(t)$ as subpaths. Clearly this can be done so that $z(t)$ is locally analytic and so analytic by Theorem 1.

Analogues to Theorem 2 with one or both of $z_1(t), z_2(t)$ closed analytic paths are readily formulated.

REFERENCES

1. Carathéodory, C., Conformal Representation, second edition, Cambridge Tracts in Mathematics and Mathematical Physics, No. 28, Cambridge 1952.

2. Minda, D., "Regular Analytic Arcs and Curves", Colloq., Math. 38, 73–82 (1977).

James A. Jenkins
Washington University
St Louis Missouri 63130
U.S.A.

Constantin Carathéodory: An International Tribute (pp. 554–573)
edited by Th. M. Rassias

STABILITY OF REACTION-DIFFUSION SYSTEM WITH SELF- AND CROSS-DISPERSION IN MATHEMATICAL ECOLOGY*

Xinhua Ji

In this paper, the effects of self and cross-dispersion on the global stability of interacting and dispersing species systems have been studied in homogeneous habitat and also in heterogeneous habitat which arises due to ecological and environmental factors. First, two and three species are to be studied, then a model of several species. It has been shown that for n-number species the stability of the positive equilibrium state requires the dominance of self-dispersal of the species over the cross-dispersal in the way as $d_{ii}\, d_{jj} > (n!-1)\, d_{ij}\, d_{ji}, i \neq j$; where d_{ii}, d_{jj} are self-dispersion and d_{ij}, d_{ji} are cross-dispersion.

Outline

1. Introduction

We consider the dynamics of interacting and dispersing species, first in a one dimensional homogeneous habitat, then in heterogeneous habitat in which heterogeneity is due to ecological and environmental characteristics. The effects of environmental and ecological factors can be studied by dividing the habitat into p-number of patches $(l_{k-1} \leq x \leq l_k; k = 1, \ldots, p)$

*Research supported by the National Natural Science Foundation of China

such that the growth rate of the species, their interaction and dispersion coefficients are constant but different in different patches. In such a case the system governing the dynamics of several interacting and dispersing species in the k-th patch can be written as,

$$\frac{\partial u_i^{(k)}}{\partial t} = f_i^{(k)}(u_1^{(k)}, \ldots, u_n^{(k)}) + \frac{\partial}{\partial x}\sum_{j=1}^{n} d_{ij}^{(k)} \frac{\partial u_j^{(k)}}{x}, \qquad (1.1)$$
$$i = 1, \ldots, n; \; k = 1, \ldots, p,$$

where $u_i^{(k)}; i = 1, \ldots, n$; denote the density of the i-th species in the k-th patch. $d_{ii}^{(k)}$ and $d_{ij}^{(k)} (i \neq j)$ are self and cross-dispersion of the i-th species in the k-th patch respectively. In general, $d_{ij}^{(k)}, i \neq j$, may be positive, negative or zero depending upon the interaction between the species. If $d_{ii}^{(k)}$ are thought of as the sum of two terms, then because of usual natural dispersal due to environmental/ecological factors, d_{ii} may also be taken to be positive, negative or zero.

We first consider the cases of two and three species, then consider Lotka-Volterra mdoel of several species. By employing Liapunov function we have proved that the stability of the positive equilibrium state requires the dominance of self-dispersal of the species over the cross-dispersal in a way as $d_{ii}d_{jj} > (n! - 1)d_{ij}d_{ji}, i \neq j$, where d_{ii}, d_{jj} are self-dispersion and d_{ij}, d_{ji} are cross-dispersion. A simple example shows that the equilibrium state may become unstable in the case when cross-dispersion coefficients dominate over self-dispersion.

2. Mutualistic Model of Two Species in Homogeneous Habitat

In paper [4], J. B. Shukle, V. N. Pal and S. Gakkhar investigated the stability of competition model and prey-predator model of two species with self and cross-dispersion in a one dimensional heterogeneous habitat. Carrying on we consider first the following mutualistic model as

$$\frac{\partial u_1}{\partial t} = u_1(b_1 - a_{11}u_1 + a_{12}u_2) + \frac{\partial}{\partial x}\left(D_1\frac{\partial u_1}{\partial x} + d_1\frac{\partial u_2}{\partial x}\right), \qquad (2.1)$$

$$\frac{\partial u_2}{\partial t} = u_2(b_2 + a_{21}u_1 - a_{22}u_2) + \frac{\partial}{\partial x}\left(d_2\frac{\partial u_1}{\partial x} + D_2\frac{\partial u_2}{\partial x}\right), \qquad (2.2)$$

where $b_1, b_2, a_{11}, a_{12}, a_{21}, a_{22}$ are positive constants. D_1, D_2 are self-dispersion, d_1, d_2 are cross-dispersion.

The non-trivial positive equilibrium state $(\bar{u}_1, \bar{u}_2)$ for the systems (2.1) and (2.2) can be obtained by solving

$$b_1 - a_{11}\bar{u}_1 + a_{12}\bar{u}_2 = 0\,,$$
$$b_2 + a_{21}\bar{u}_1 - a_{22}\bar{u}_2 = 0\,.$$

Using the transformations

$$u_i = \bar{u}_i + v_i\,, i = 1, 2\,,$$

the systems (2.1) and (2.2) can be written as

$$\frac{\partial v_1}{\partial t} = u_1(-a_{11}v_1 + a_{12}v_2) + \frac{\partial}{\partial x}\left(D_1\frac{\partial v_1}{\partial x} + d_1\frac{\partial v_2}{\partial x}\right)\,, \qquad (2.1)'$$

$$\frac{\partial v_2}{\partial t} = u_2(a_{21}v_1 - a_{22}v_v) + \frac{\partial}{\partial x}\left(d_2\frac{\partial v_1}{\partial x} + D_2\frac{\partial v_2}{\partial x}\right)\,, \qquad (2.2)'$$

which may be associted with the initial and boundary conditions as follows:

$$v_i(x, 0) = F_i(x)\,,\ i = 1, 2\,, \qquad (2.3)$$

$$v_i(L, t) = v_i(0, t) = 0\,,\ i = 1, 2\,. \qquad (2.4)$$

The condition (2.3) shows the initial distributions of two species. By condition (2.4), we mean that the species densities are at equilibrium at the boundaries of the habitat.

We discuss the global stability of the systems (2.1)′ and (2.2)′ with the conditions (2.3) and (2.4). To this end we consider the following Liapunov function E, as

$$E = \int_0^L \left\{\bar{u}_1\left[\frac{v_1}{\bar{u}_l} - \log\left(1 + \frac{v_1}{\bar{u}_1}\right)\right] + cu_2\left[\frac{v_2}{\bar{u}_2} - \log\left(1 + \frac{v_2}{\bar{u}_2}\right)\right]\right\} dx\,, \qquad (2.5)$$

where positive constant c is to be determined. Differentiating the function E with respect to t along the solution of system (2.1)′ and (2.2)′ with conditions (2.3), (2.4) we get dE/dt, as

$$\frac{dE}{dt} = -\int_0^L (P + Q)dx \qquad (2.6)$$

where

$$P = a_{11}v_1^2 - (a_{12} + ca_{21})v_1v_2 + ca_{22}v_2^2 \tag{2.7}$$

and

$$\begin{aligned} Q = &\left(\frac{u_1}{u_1^2}\right) D_1 \left(\frac{\partial v_1}{\partial x}\right)^2 + \left(\frac{\bar{u}_1}{u_1^2}d_1 + \frac{\bar{u}_2}{u_2^2}cd_2\right) \frac{\partial v_1}{\partial x}\frac{\partial v_2}{\partial x} \\ &+ \left(\frac{\bar{u}_2}{u_2^2}\right) cD_2 \left(\frac{\partial v_2}{\partial x}\right)^2 . \end{aligned} \tag{2.8}$$

By using Sylvester criterion for positive definiteness we have the following statement

(i) Set

$$R_2 = \begin{pmatrix} a_{11} & -\frac{1}{2}(a_{12} - ca_{21}) \\ -\frac{1}{2}(a_{12} - ca_{21}) & ca_{22} \end{pmatrix} .$$

Then P is positive definite if and only if

$$\det R_2 > 0 .$$

(ii) Set

$$S_2 = \begin{pmatrix} D_1\bar{u}_1/u_1^2 & \frac{1}{2}(d_1\bar{u}_1/u_1^2 + cd_2\bar{u}_2/u_2^2) \\ \frac{1}{2}(d_1\bar{u}_1/u_1^2 + cd_2\bar{u}_2/u_2^2) & cD_2\bar{u}_2/u_2^2 \end{pmatrix} .$$

Then Q is positive definite if and only if

$$D_1 > 0 \text{ and } \det S_2 > 0 \text{ (with } u_1 > 0 , u_2 > 0) .$$

Theorem 1. Suppose $D_1 > 0$ and $D_2 > 0$. If conditions

$$a_{11}a_{22} > a_{12}a_{21} , \tag{2.9}$$

$$D_1D_2 > d_1d_2 \tag{2.10}$$

are satisfied, then the equilibrium state $(\bar{u}_1, \bar{u}_2)$ is global stable in bounded region

$$A = \{(u_1, u_2) : 0 < u_i^0 \le u_i \le U_i^0 , i = 1, 2\}$$

(where u_i^0, U_i^0 are positive constants).

Proof. (i) It can be seen that $\det R_2 > 0$, iff

$$ca_{11}a_{22} - \frac{1}{4}(a_{12} + ca_{21})^2 > 0\,. \tag{2.11}$$

And (2.11) is equivalent to

$$g(c) \equiv c^2 a_{21}^2 - 2(2a_{11}a_{22} - a_{12}a_{21}) + a_{12}^2 < 0\,.$$

Let

$$x_1 = \left[2a_{11}a_{22} - a_{12}a_{21} - 2\sqrt{a_{11}a_{22}(a_{11}a_{22} - a_{12}a_{21})}\right] \Big/ a_{21}^2$$

and

$$x_2 = \left[2a_{11}a_{22} - a_{12}a_{21} + 2\sqrt{a_{11}a_{22}(a_{11}a_{22} - a_{12}a_{21})}\right] \Big/ a_{21}^2\,.$$

Under the condition (2.9) we see that

$$x_2 > x_1 > 0\,.$$

Thus we have that if (2.9) is satisfied, then

$$g(c) < 0 \text{ for } x_1 < c < x_2\,.$$

(ii) If $d_1^2 + d_2^2 \neq 0$, e.g. $d_2 \neq 0$, proceeding in a same manner as for (i), we get

$$y_1 = \left(\frac{\bar{u}_1}{u_1^2}\right) \frac{(2D_1D_2 - d_1d_2) - 2\sqrt{D_1D_2(D_1D_2 - d_1d_2)}}{d_2^2\left(\frac{\bar{u}_2}{u_2^2}\right)}$$

$$y_2 = \left(\frac{\bar{u}_1}{u_1^2}\right) \frac{(2D_1D_2 - d_1d_2) + 2\sqrt{D_1D_2(D_1D_2 - d_1d_2)}}{d_2^2\left(\frac{\bar{u}_2}{u_2^2}\right)}\,.$$

Under the condition (2.10) we see that

$$y_2 > y_1 > 0$$

which implies that if (2.10) is satisfied, then we have

$$\det S_2 > 0 \text{ for } y_1 < c < y_2$$

in the region

$$A = \{(u_1, u_2) : u_i^0 \le u_i \le U_i^0, i = 1, 2\} \tag{2.12}$$

provided

$$(x_1, x_2) \cap (y_1, y_2) \neq \emptyset, \tag{2.13}$$

where u_i^0 are the lower bounds and U_i^0 are the upper bounds for u_i, and are positive constants.

In case of $d_1^2 + d_2^2 = 0$, it is obvious that $\det S_2 > 0$ for any $c > 0$ in region A. Thus, (2.13) is satisfied naturally in this case.

Therefore we can conclude that $dE/dt < 0$. The equilibrium state $(\bar{u}_1, \bar{u}_2)$ is global stable in region A. The proof of Theorem 1 is complete.

The condition (2.10) implies that the stability of the equilibrium state requires the dominance of self-dispersal of the species over the cross-dispersal.

The assumption (2.13) expresses a relationship between the growth rate of the species, their interaction coefficients and their dispersion coefficients.

Therefore from Theorem 1 we have the following stable cases: Suppose $D_1 > 0, D_2 > 0$, and $a_{11}a_{22} > a_{12}a_{21}$, under the hypothesis (2.13),

(1) $d_1 = 0, d_2 = 0$,
(2) $d_1 = 0, d_2 \neq 0$,
(3) $d_1 \neq 0, d_2 = 0$,
(4) $d_1 < 0, d_2 > 0$,
(5) $d_1 > 0, d_2 < 0$,
(6) $d_1 > 0, d_2 > 0, D_1 D_2 > d_1 d_2$,
(7) $d_1 < 0, d_2 < 0, D_1 D_2 > d_1 d_2$.

In any one of the above seven cases the equilibrium state $(\bar{u}_1, \bar{u}_2)$ is global stable in region A given by (2.12).

Next we consider unstability of this model. On linearising the system (2.1)′ and (2.2)′, we get

$$\frac{\partial v_1}{\partial t} = -a_{11}u_1 v_1 + a_{12}u_1 v_2 + \frac{\partial}{\partial x}\left(D_1 \frac{\partial v_1}{\partial x} + d_1 \frac{\partial v_2}{\partial x}\right), \tag{2.14}$$

$$\frac{\partial v_2}{\partial t} = a_{21}u_2 v_1 - a_{22}u_2 v_2 + \frac{\partial}{\partial x}\left(d_2 \frac{\partial v_1}{\partial x} + D_2 \frac{\partial v_2}{\partial x}\right). \tag{2.15}$$

We employ the following scalar function

$$H = \int_0^L v_1 v_2 dx\,.$$

Differentiating the function H with respect to t along the solutions of systems (2.14) and (2.15) with (2.3) and (2.4) we get dH/dt as

$$\frac{dH}{dt} = \int_0^L (\bar{P} + \bar{Q}) dx$$

where

$$\bar{P} = a_{21}\bar{u}_2 v_1^2 - (a_{11}\bar{u}_1 + a_{22}\bar{u}_2) v_1 v_2 + a_{12}\bar{u}_1 v_2^2$$

and

$$\bar{Q} = -d_2 \left(\frac{\partial v_1}{\partial x}\right)^2 - (D_1 + D_2)\frac{\partial v_1}{\partial x}\frac{\partial v_2}{\partial x} - d_1 \left(\frac{\partial v_2}{\partial x}\right)^2 .$$

By means of the same method as above we can draw the conclusion that in case of $a_{12}a_{21} > a_{11}a_{22}$, if $d_1 d_2 > D_1 D_2$, then under certain conditions we have $\bar{P} > 0$ and $\bar{Q} > 0$ which implies $dH/dt > 0$. That is to say, if the cross-dispersion dominates over the self-dispersion, the equilibrium state may become unstable and the species may not survive.

3. Stability of Three Species in Homogeneous Habitat

3.1. *Mutualistic model*

For mutualistic model of three species a model of interaction function is

$$\begin{aligned} f_1(u_1, u_2, u_3) &= u_1(b_1 - a_{11}u_1 + a_{12}u_2 + a_{13}u_3)\,, \\ f_2(u_1, u_2, u_3) &= u_2(b_2 + a_{21}u_1 - a_{22}u_2 + a_{23}u_3)\,, \\ f_3(u_1, u_2, u_3) &= u_3(b_3 + a_{31}u_1 + a_{32}u_2 - a_{33}u_3)\,. \end{aligned}$$

The considered corresponding system in this section is

$$\frac{\partial v_1}{\partial t} = u_1(-a_{11}v_1 + a_{12}v_2 + a_{13}v_3) + \frac{\partial}{\partial x}\left(D_1\frac{\partial v_1}{\partial x} + d_{12}\frac{\partial v_2}{\partial x} + d_{13}\frac{\partial v_3}{\partial x}\right), \tag{3.1}$$

$$\frac{\partial v_2}{\partial t} = u_2(a_{21}v_1 - a_{22}v_2 + a_{23}v_3) + \frac{\partial}{\partial x}\left(d_{21}\frac{\partial v_1}{\partial x} + D_2\frac{\partial v_2}{\partial x} + d_{23}\frac{\partial v_3}{\partial x}\right), \tag{3.2}$$

$$\frac{\partial v_3}{\partial t} = u_3(a_{31}v_1 + a_{32}v_2 - a_{33}v_3) + \frac{\partial}{\partial x}\left(d_{31}\frac{\partial v_1}{\partial x} + d_{32}\frac{\partial v_2}{\partial x} + D_3\frac{\partial v_3}{\partial x}\right), \tag{3.3}$$

associated with Initial conditions

$$v_i(x,0) = F_i(x)\,,\ i = 1,2,3\,, \tag{3.4}$$

and Boundary conditions

$$v_i(L,t) = v_i(0,t) = 0\,,\ i = 1,2,3\,. \tag{3.5}$$

To discuss the global stability of systems (3.1), (3.2), (3.3) with conditions (3.4), (3.5) we consider the Liapunov function E as follows

$$E = \int_0^L \left\{\sum_{i=1}^{3} c_i\left[v_i - \bar{u}_i \log\left(1 + \frac{v_i}{\bar{u}_i}\right)\right]\right\} dx\,. \tag{3.6}$$

From (3.6) we get

$$\frac{dE}{dt} = -\int_0^L \left[P(v_1,v_2,v_3) + Q\left(\frac{\partial v_1}{\partial x}, \frac{\partial v_2}{\partial x}, \frac{\partial v_3}{\partial x}\right)\right] dx$$

where

$$P(v_1,v_2,v_3) = \sum_{j=1}^{3} c_j a_{jj} v_j^2 - \sum_{\substack{i,j=1 \\ i\neq j}}^{3} (c_i a_{ij} + c_j a_{ji}) v_i v_j$$

and

$$Q\left(\frac{\partial v_1}{\partial x}, \frac{\partial v_2}{\partial x}, \frac{\partial v_3}{\partial x}\right) = \sum_{j=1}^{3} c_j \frac{\bar{u}_j}{u_j^2} D_j \left(\frac{\partial v_j}{\partial x}\right)^2 + \sum_{\substack{i,j=1 \\ i\neq j}}^{3} \left(c_i \frac{\bar{u}_i}{u_i^2} d_{ij} + c_j \frac{\bar{u}_j}{u_j^2} d_{ji}\right) \frac{\partial v_i}{\partial x}\frac{\partial v_j}{\partial x}\,.$$

In the same way as in Sec. 2 we get

$$R_3 = \begin{pmatrix} c_1a_{11} & -\frac{1}{2}(c_2a_{21} - c_1a_{12} & -\frac{1}{2}(c_3a_{31} - c_1a_{13}) \\ -\frac{1}{2}(c_1a_{12} - c_2a_{21}) & c_2a_{22} & -\frac{1}{2}(c_3a_{32} - c_2a_{23}) \\ -\frac{1}{2}(c_1a_{13} - c_3a_{31}) & -\frac{1}{2}(c_2a_{23} - c_3a_{32}) & c_3a_{33} \end{pmatrix}$$

and

$$S_3 = \begin{pmatrix} \frac{a_1}{u_1^2}c_1D_1 & \frac{1}{2}\left(\frac{a_2}{u_2^2}c_2d_{21} + \frac{a_1}{u_1^2}c_1d_{12}\right) & \frac{1}{2}\left(\frac{a_3}{u_3^2}c_3d_{31} + \frac{a_1}{u_1^2}c_1d_{13}\right) \\ \frac{1}{2}\left(\frac{a_1}{u_1^2}c_1d_{12} + \frac{a_2}{u_2^2}c_2d_{21}\right) & \frac{a_2}{u_2^2}c_2D_2 & \frac{1}{2}\left(\frac{a_3}{u_3^2}c_3d_{32} + \frac{a_2}{u_2^2}c_2d_{23}\right) \\ \frac{1}{2}\left(\frac{a_1}{u_1^2}c_1d_{13} + \frac{a_3}{u_3^2}c_3d_{31}\right) & \frac{1}{2}\left(\frac{a_2}{u_2^2}c_2d_{23} + \frac{a_3}{u_3^2}c_3d_{32}\right) & \frac{a_3}{u_3^2}c_3D_3 \end{pmatrix}.$$

Corresponding to (i), (ii) in Sec. 2, we have the conditions (I) and (II) under which

$$P(v_1, v_2, v_3) \text{ and } Q\left(\frac{\partial v_1}{\partial x}, \frac{\partial v_2}{\partial x}, \frac{\partial v_3}{\partial x}\right)$$

are positive definite respectively. They are:

(I) $P(v_1, v_2, v_3)$ is positive definite if and only if

(1) $\det R_2 > 0$ and

(2) $\det R_3 > 0$

are valid simultaneously.

(II) $Q(\partial v_1/\partial x, \partial v_2/\partial x, \partial v_3/\partial x)$ is positive definite if and only if

(1) $D_1 > 0$,

(2) $\det S_2 > 0$, and

(3) $\det S_3 > 0$

are valid simultaneously.

Theorem 2. Suppose $D_1 > 0, D_2 > 0$ and $D_3 > 0$. If the conditions

$$a_{ii}a_{jj} > 4a_{ij}a_{ji}, i \neq j; i, j = 1, 2, 3 \tag{3.7}$$

and

$$D_iD_j > 4d_{ij}d_{ji}, i \neq j; i, j = 1, 2, 3 \tag{3.8}$$

are satisfied, then the equilibrium state $(\bar{u}_1, \bar{u}_2, \bar{u}_3)$ is global stable in a bounded region, under certain hypothesis expressing a relationship between interaction coefficients and dispersion coefficients.

Proof. We first calculate det R_3,

$$\begin{aligned}\det R_3 =& c_1c_2c_3a_{11}a_{22}a_{33} - \frac{1}{4}c_2a_{22}(c_1a_{13} + c_3a_{31})^2 \\ &- \frac{1}{4}c_3a_{33}(c_1a_{12} + c_2a_{21})^2 \\ &- \frac{1}{4}c_1a_{11}(c_2a_{23} + c_3a_{32})^2 \\ &- \frac{1}{4}(c_1a_{12} + c_2a_{21})(c_2a_{23} + c_3a_{32})(c_3a_{31} + c_1a_{13})\,.\end{aligned}$$

By means of the same method as in Sec. 2, from condition (3.7) we have the following estimate

$$c_1c_2a_{11}a_{22} > (c_1a_{12} + c_2a_{21})^2\,, \tag{3.9}$$

$$c_2c_3a_{22}a_{33} > (c_2a_{23} + c_3a_{32})^2\,, \tag{3.10}$$

$$c_3c_1a_{33}a_{11} > (c_3a_{31} + c_1a_{13})^2\,, \tag{3.11}$$

where c_1, c_2, c_3 are positive constants. Combining (3.9), (3.10) and (3.11) we get

$$c_1c_2c_3a_{11}a_{22}a_{33} > (c_1a_{12} + c_2a_{21})(c_2a_{23} + c_3a_{32})(c_3a_{31} + c_1a_{13})\,. \tag{3.12}$$

Then from (3.9), (3.10), (3.11) and (3.12) we immediately obtain that there exist positive constants c_1, c_2, c_3 such that

$$\det\ R_3 > 0\,,$$

provided (3.7). From (3.7) we also have

$$a_{11}a_{22} > 4a_{12}a_{21} > a_{12}a_{21}$$

which means that

$$\det\ R_2 > 0$$

with $c = c_1/c_2$. Therefore we have the positive definiteness of $P(v_1, v_2, v_3)$. Thus (I) has been proved.

Proceeding in a similar way we can also obtain (II). That is, if $D_1 > 0$, then there exist positive constants c_1, c_2, c_3 such that $\det S_3 > 0$ and $\det S_2 > 0$ with $c = c_1/c_2$ under the condition (3.8).

There are certain intervals which the positive constants c_1/c_2 and c_2/c_3 belong to. They are

$$c_1/c_2 \in (x_1^{12}, x_2^{12}) \text{ and } c_2/c_3 \in (x_1^{23}, x_2^{23}) \tag{3.13}$$

where

$$x_1^{12}, x_2^{12} = \frac{a_{11}a_{22} - 2a_{12}a_{21} \mp \sqrt{a_{11}a_{22}(a_{11}a_{22} - 4a_{12}a_{21})}}{a_{12}^2},$$
$$x_1^{23}, x_2^{23} = \frac{a_{22}a_{33} - 2a_{23}a_{32} \mp \sqrt{a_{22}a_{33}(a_{22}a_{33} - 4a_{23}a_{32})}}{a_{23}^2}.$$

(3.13) expresses a condition under which (I) is true. On the other hand from the omitted details of the proof of (II) we obtain

$$c_1/c_2 \in (y_1^{12}, y_2^{12}) \text{ and } c_2/c_3 \in (y_2^{23}, y_3^{23}), \tag{3.13'}$$

where

$$y_1^{12}, y_2^{12} = \left(\frac{\bar{u}_2}{\bar{u}_1}\right)\left(\frac{u_1}{u_2}\right)^2 \left(\frac{D_1D_2 - 2d_{12}d_{21} \mp \sqrt{D_1D_2(D_1D_2 - 4d_{12}d_{21})}}{d_{12}^2}\right),$$
$$y_1^{23}, y_2^{23} = \left(\frac{\bar{u}_3}{\bar{u}_2}\right)\left(\frac{u_2}{u_3}\right)^2 \left(\frac{D_2D_3 - 2d_{23}d_{32} \mp \sqrt{D_2D_3(D_2D_3 - 4d_{23}d_{32})}}{d_{23}^2}\right)$$

in the bounded region

$$A = \{(u_1, u_2, u_3) : u_i^0 \le u_i \le U_i^0, i\, 1, 2, 3\} \tag{3.14}$$

where u_i^0 and $U_i^0, i = 1, 2, 3$, are bounded positive constants. So, the assumption corresponding to (2.13) is that

$$(x_1^{12}, x_2^{12}) \cap (y_1^{12}, y_2^{12}) \neq \emptyset \tag{3.15}$$

and

$$(x_1^{23}, x_2^{23}) \cap (y_1^{23}, y_2^{23}) \neq \emptyset. \tag{3.16}$$

We thus have $dE/dt < 0$ in region A given by (3.14). Theorem 2 is therefore proved.

From Theorem 2 we have following twenty stable cases: Suppose $D_1 > 0, D_2 > 0, D_3 > 0$ and $a_{ii}a_{jj} > 4a_{ij}a_{ij}, i \neq j; i,j = 1,2,3$, also under the hypothesis (3.15)–(3.16),

(1) $d_{ij} = 0, i \neq j\,; i,j = 1,2,3\,;$

(2) $d_{kj} = d_{jk} = d_{ki} = d_{ik} = 0\,, d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0$, $i \neq j \neq k; i,j,k = 1,2,3;$

(3) $d_{kj} = d_{jk} = 0, d_{ki}d_{ik} = 0$ with $d_{ki}^2 + d_{ik}^2 \neq 0$, $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0, i \neq j \neq k; i,j,k = 1,2,3;$

(4) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0, i \neq j; i,j = 1,2,3.$

(The cases (1)–(4) may be summarized as the case that $d_{ij}d_{ji} = 0, i \neq j; i,j = 1,2,3$, i.e., if we say d_{ij} and d_{ji} to be a pair, then the product of every pair of cross-dispersion coefficients of species vanishes.)

(5) $d_{ij} = d_{ji} = d_{jk} = d_{kj} = 0, d_{ki}d_{ik} < 0,$ $i \neq j \neq k; i,j,k = 1,2,3;$

(6) $d_{ij} = d_{ji} = d_{jk} = d_{kj} = 0, d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ki}d_{ik},$ $i \neq j \neq k; i,j,k = 1,2,3;$

(7) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} = 0$ with $d_{kj}^2 + d_{jk}^2 \neq 0,$ $d_{ki}d_{ik} < 0, i \neq j \neq k; i,j,k = 1,2,3;$

(8) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} = 0$ with $d_{jk}^2 + d_{kj}^2 \neq 0,$ $d_{ki}d_{ik} > 0$ with $D_iD_k > 4d_{ik}d_{ki}, i \neq j \neq k; i,j,k = 1,2,3;$

(9) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0,$ $d_{jk}d_{kj} = 0$ with $d_{jk}^2 + d_{kj}^2 \neq 0,$ $d_{ik}d_{ki} < 0, i \neq j \neq k; i,j,k = 1,2,3.$

(10) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0,$ $d_{jk}d_{kj} = 0$ with $d_{jk}^2 + d_{kj}^2 \neq 0,$ $d_{ik}d_{ki} > 0$ with $D_iD_k > 4d_{ik}d_{ki},$ $i \neq j \neq k; i,j,k = 1,2,3.$

(The cases (5)–(10) can be summarised as the case that $d_{ij}d_{ji} = 0, d_{jk}d_{kj} = 0, d_{ki}d_{ik} \neq 0$ with $D_kD_i > 4d_{ik}d_{ki}$, i.e., the product of only one pair of cross-dispersion coefficients of species does not vanish.)

(11) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} < 0, d_{ki}d_{ik} < 0,$ $i \neq j \neq k; i,j,k = 1,2,3;$

(12) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} < 0,$ $d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ik}d_{ki},$ $i \neq j \neq k; i,j,k = 1,2,3;$

Proceeding in a similar way we can also obtain (II). That is, if $D_1 > 0$, then there exist positive constants c_1, c_2, c_3 such that $\det S_3 > 0$ and $\det S_2 > 0$ with $c = c_1/c_2$ under the condition (3.8).

There are certain intervals which the positive constants c_1/c_2 and c_2/c_3 belong to. They are

$$c_1/c_2 \in (x_1^{12}, x_2^{12}) \text{ and } c_2/c_3 \in (x_1^{23}, x_2^{23}) \tag{3.13}$$

where

$$x_1^{12}, x_2^{12} = \frac{a_{11}a_{22} - 2a_{12}a_{21} \mp \sqrt{a_{11}a_{22}(a_{11}a_{22} - 4a_{12}a_{21})}}{a_{12}^2},$$
$$x_1^{23}, x_2^{23} = \frac{a_{22}a_{33} - 2a_{23}a_{32} \mp \sqrt{a_{22}a_{33}(a_{22}a_{33} - 4a_{23}a_{32})}}{a_{23}^2}.$$

(3.13) expresses a condition under which (I) is true. On the other hand from the omitted details of the proof of (II) we obtain

$$c_1/c_2 \in (y_1^{12}, y_2^{12}) \text{ and } c_2/c_3 \in (y_2^{23}, y_3^{23}) , \tag{3.13'}$$

where

$$y_1^{12}, y_2^{12} = \left(\frac{\bar{u}_2}{\bar{u}_1}\right)\left(\frac{u_1}{u_2}\right)^2 \left(\frac{D_1D_2 - 2d_{12}d_{21} \mp \sqrt{D_1D_2(D_1D_2 - 4d_{12}d_{21})}}{d_{12}^2}\right),$$
$$y_1^{23}, y_2^{23} = \left(\frac{\bar{u}_3}{\bar{u}_2}\right)\left(\frac{u_2}{u_3}\right)^2 \left(\frac{D_2D_3 - 2d_{23}d_{32} \mp \sqrt{D_2D_3(D_2D_3 - 4d_{23}d_{32})}}{d_{23}^2}\right)$$

in the bounded region

$$A = \{(u_1, u_2, u_3) : u_i^0 \le u_i \le U_i^0 , i\, 1, 2, 3\} \tag{3.14}$$

where u_i^0 and $U_i^0, i = 1, 2, 3$, are bounded positive constants. So, the assumption corresponding to (2.13) is that

$$(x_1^{12}, x_2^{12}) \cap (y_1^{12}, y_2^{12}) \neq \emptyset \tag{3.15}$$

and

$$(x_1^{23}, x_2^{23}) \cap (y_1^{23}, y_2^{23}) \neq \emptyset . \tag{3.16}$$

We thus have $dE/dt < 0$ in region A given by (3.14). Theorem 2 is therefore proved.

From Theorem 2 we have following twenty stable cases: Suppose $D_1 > 0, D_2 > 0, D_3 > 0$ and $a_{ii}a_{jj} > 4a_{ij}a_{ij}, i \neq j; i,j = 1,2,3$, also under the hypothesis (3.15)–(3.16),

(1) $d_{ij} = 0, i \neq j\,; i,j = 1,2,3\,;$

(2) $d_{kj} = d_{jk} = d_{ki} = d_{ik} = 0, d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0$, $i \neq j \neq k; i,j,k = 1,2,3;$

(3) $d_{kj} = d_{jk} = 0, d_{ki}d_{ik} = 0$ with $d_{ki}^2 + d_{ik}^2 \neq 0$, $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0, i \neq j \neq k; i,j,k = 1,2,3;$

(4) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0, i \neq j; i,j = 1,2,3.$

(The cases (1)–(4) may be summarized as the case that $d_{ij}d_{ji} = 0, i \neq j; i,j = 1,2,3$, i.e., if we say d_{ij} and d_{ji} to be a pair, then the product of every pair of cross-dispersion coefficients of species vanishes.)

(5) $d_{ij} = d_{ji} = d_{jk} = d_{kj} = 0, d_{ki}d_{ik} < 0,$ $i \neq j \neq k; i,j,k = 1,2,3;$

(6) $d_{ij} = d_{ji} = d_{jk} = d_{kj} = 0, d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ki}d_{ik},$ $i \neq j \neq k; i,j,k = 1,2,3;$

(7) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} = 0$ with $d_{kj}^2 + d_{jk}^2 \neq 0,$ $d_{ki}d_{ik} < 0, i \neq j \neq k; i,j,k = 1,2,3;$

(8) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} = 0$ with $d_{jk}^2 + d_{kj}^2 \neq 0,$ $d_{ki}d_{ik} > 0$ with $D_iD_k > 4d_{ik}d_{ki}, i \neq j \neq k; i,j,k = 1,2,3;$

(9) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0,$ $d_{jk}d_{kj} = 0$ with $d_{jk}^2 + d_{kj}^2 \neq 0,$ $d_{ik}d_{ki} < 0, i \neq j \neq k; i,j,k = 1,2,3.$

(10) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0,$ $d_{jk}d_{kj} = 0$ with $d_{jk}^2 + d_{kj}^2 \neq 0,$ $d_{ik}d_{ki} > 0$ with $D_iD_k > 4d_{ik}d_{ki},$ $i \neq j \neq k; i,j,k = 1,2,3.$

(The cases (5)–(10) can be summarised as the case that $d_{ij}d_{ji} = 0, d_{jk}d_{kj} = 0, d_{ki}d_{ik} \neq 0$ with $D_kD_i > 4d_{ik}d_{ki}$, i.e., the product of only one pair of cross-dispersion coefficients of species does not vanish.)

(11) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} < 0, d_{ki}d_{ik} < 0,$ $i \neq j \neq k; i,j,k = 1,2,3;$

(12) $d_{ij} = d_{ji} = 0, d_{jk}d_{kj} < 0,$ $d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ik}d_{ki},$ $i \neq j \neq k; i,j,k = 1,2,3;$

(13) $d_{ij} = d_{ji} = 0$,
$d_{jk}d_{kj} > 0$ with $D_jD_k > 4d_{jk}d_{kj}$,
$d_{ki}d_{ik} > 0$ with $D_iD_k > 4d_{ik}d_{ki}$,
$i \neq j \neq k; i, j, k = 1, 2, 3$;

(14) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0$,
$d_{jk}d_{kj} < 0, d_{ki}d_{ik} < 0, i \neq j \neq k; i, j, k = 1, 2, 3$;

(15) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0$,
$d_{jk}d_{kj} < 0$,
$d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ki}d_{ik}$,
$i \neq j \neq k; i, j, k = 1, 2, 3$;

(16) $d_{ij}d_{ji} = 0$ with $d_{ij}^2 + d_{ji}^2 \neq 0$,
$d_{jk}d_{kj} > 0$ with $D_jD_k > 4d_{jk}d_{kj}$,
$d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ki}d_{ik}$,
$i \neq j \neq k; i, j, k = 1, 2, 3$.

(The case (11)–(16) can be summarized as the case that there is only one pair of species, the product of whose cross-dispersion coefficients vanishes.)

(17) $d_{ij}d_{ji} < 0, i \neq j; i, j, k = 1, 2, 3$;

(18) $d_{ij}d_{ji} < 0, d_{jk}d_{kj} < 0$,
$d_{ki}d_{ik} > 0$ with $D_kD_i > 4d_{ik}d_{ki}$,
$i \neq j \neq k; i, j, k = 1, 2, 3$;

(19) $d_{ij}d_{ji} < 0$,
$d_{jk}d_{kj} > 0$ with $D_kD_j > 4d_{jk}d_{kj}$,
$d_{ik}d_{ki} > 0$ with $D_iD_k > 4d_{ik}d_{ki}$,
$i \neq j \neq k, i, j, k = 1, 2, 3$;

(20) $d_{kj}d_{jk} > 0$ with $D_kD_j > 4d_{kj}d_{jk}, k \neq j; k, j = 1, 2, 3$.

(The cases (17)–(20) can be summarized as the case that there is only one pair of species, the product of whose cross-dispersion coefficients vanishes.)

In any one of these twenty cases the equilibrium state $(\bar{u}_1, \bar{u}_2, \bar{u}_3)$ is global stable in the bounded region A given by (3.14).

3.2. *Competition model*

For competition model of three species, a model of interacting function is

$$f_1(u_1, u_2, u_3) = u_1(b_1 - a_{11}u_1 - a_{12}u_2 - a_{13}u_3),$$
$$f_2(u_1, u_2, u_3) = u_2(b_2 - a_{21}u_1 - a_{22}u_2 - a_{23}u_3),$$
$$f_3(u_1, u_2, u_3) = u_3(b_3 - a_{31}u_1 - a_{32}u_2 - a_{33}u_3).$$

Processing and calculating in the same way as in Sec. 3.1, we get in this case

$$R_3 = \begin{pmatrix} c_1a_{11} & \frac{1}{2}(c_1a_{12}+c_2a_{21}) & \frac{1}{2}(c_1a_{13}+c_3a_{31}) \\ \frac{1}{2}(c_1a_{12}+c_2a_{21}) & c_2a_{22} & \frac{1}{2}(c_2a_{23}+c_3a_{32}) \\ \frac{1}{2}(c_1a_{13}+c_3a_{31}) & \frac{1}{2}(c_2a_{23}+c_3a_{32}) & c_3a_{33} \end{pmatrix}$$

and S_3 is the same as for mutualistic model of three species. Omitting the proof we give a theorem on global stability of competition model of three species with self and cross-dispersion as follows

Theorem 3. Suppose $D_i > 0, (i = 1, 2, 3)$. If

$$a_{ii}a_{jj} \geq 3a_{ij}a_{ji}$$

and

$$D_iD_j > 4d_{ij}d_{ji} \tag{3.17}$$

for $i \neq j; i, j = 1, 2, 3$, then the positive equilibrium state $(\bar{u}_1, \bar{u}_2, \bar{u}_3)$ is global stable in the bounded region A, under certain hypothesis expressing a relationship between interaction coefficients and dispersion coefficients similar to (3.15) and (3.16).

3.3. *Prey-Predator model*

For prey-predator model there are many different cases. Here in this section we only choose a model that is easy to deal with. Other different models can be dealt with in the same way.

The interacting function of the model is

$$\begin{cases} f_1 = u_1(b_1 - a_{11}u_1 - a_{12}u_2), \\ f_2 = u_2(-b_2 + a_{21}u_1 - a_{22}u_2 - a_{23}u_3), \\ f_3 = u_3(-b_3 + a_{32}u_2 - a_{33}u_3), \end{cases} \tag{3.18}$$

whose matrix R_3 is

$$R_3 = \begin{pmatrix} c_1a_{11} & \frac{1}{2}(c_1a_{12}-c_2a_{21}) & 0 \\ \frac{1}{2}(c_1a_{12}-c_2a_{21}) & c_2a_{22} & \frac{1}{2}(c_2a_{23}-c_3a_{32}) \\ 0 & \frac{1}{2}(c_2a_{23}-c_3a_{32}) & c_3a_{33} \end{pmatrix}.$$

By calculating we know that in this case there always exist positive constants c_1, c_2, c_3 such that $\det R_3 > 0$. Therefore we have the following theorem.

Theorem 4. The positive equilibrium state of prey-predator model (3.18) with self and cross-dispersion is global stable if $D_i > 0, i = 1, 2, 3$, and (3.17) is satisfied in the bounded region A under (3.13)'.

4. Stability of Several Species in Homogeneous Habitat

In this section we will show that the skill used to analyse the stability of two and three species in the preceding sections is effective to deal with the stability of system of n-number species.

For the Lotka-Volterra model of n species the interaction function is

$$f_i = u_i \left(b_i + \sum_{j=1}^{n} a_{ij} u_j \right), i = 1, \ldots, n.$$

The considered system with self and cross-dispersion is

$$\frac{\partial v_i}{\partial t} = u_i \sum_{j=1}^{n} a_{ij} v_j + \frac{\partial}{\partial x} \sum_{j=1}^{n} d_{ij} \frac{\partial v_j}{\partial x}, \quad i = 1, \ldots, n, \tag{4.1}$$

associated with initial conditions

$$v_i(x, 0) = F_i(x), i = 1, \ldots, n, \tag{4.2}$$

and boundary conditions

$$v_i(0, t) = v_i(L, t) = 0, \; i = 1, \ldots, n. \tag{4.3}$$

To analyse the global stability of system (4.1) with conditions (4.2) and (4.3) we employ the Liapunov function E as follows,

$$E = \int_0^L \left\{ \sum_{i=1}^{n} c_i \left[v_i - \bar{u}_i \log \left(1 + \frac{v_i}{\bar{u}_i} \right) \right] \right\} dx, \tag{4.4}$$

where $c_1, \ldots, c_n$ are positive constants. Differentiating E with respect to t along the solutions of system (4.1) with conditions (4.2) and (4.3) we get

$$\frac{dE}{dt} = -\int_0^L \left[W(v) + \overline{W}\left(\frac{\partial v}{\partial x}\right) \right] dx ,$$

where

$$W(v) = -v(CA + A^T C)v^T ,$$

$$\overline{W}\left(\frac{\partial v}{\partial x}\right) = \left(\frac{\partial v}{\partial x}\right)(BD + D^T B)\left(\frac{\partial v}{\partial x}\right)^T ,$$

$$v = (v_1, \ldots, v_n) ,$$

$$\frac{\partial v}{\partial x} = \left(\frac{\partial v_1}{\partial x}, \ldots, \frac{\partial v_n}{\partial x}\right) .$$

$$C = \operatorname{diag}(c_1, \ldots, c_n) ,$$

$$B = \operatorname{diag}\left(c_1 \frac{\bar{u}_1}{u_1^2}, \ldots, c_n \frac{\bar{u}_n}{u_n^2}\right) ,$$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \cdots & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} , \quad D = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ & \cdots & \\ d_{n1} & \cdots & d_{nn} \end{pmatrix} .$$

Omitting the proof we have the following Theorem which furnishes a sufficient condition for global stability of the positive equilibrium state $(\bar{u}_1, \ldots, \bar{u}_n)$ of system (4.1).

Theorem 5. If there exists a positive diagonal matrix $C = \operatorname{diag}(c_1, \ldots, c_n)$ such that matrix $CA + A^T C$ is negative definite and matrix $BD + D^T B$ is positive definite simultaneously. The functions $W(v)$ and $\overline{W}\left(\frac{\partial v}{\partial x}\right)$ do not vanish along the solutions of the system (4.1). Then the positive equilibrium state $(\bar{u}_1, \ldots, \bar{u}_n)$ of system (4.1) is global stable in bounded region

$$A = \{(u_1, \ldots, u_n) : u_i^0 \le u_i \le U_i^0, i = 1, \ldots, n\} \tag{4.5}$$

(where u_i^0 and U_i^0 are positive constants).

Based upon Theorem 5 we have the following Theorem which will also provide a sufficient condition for global stability to system (4.1).

Theorem 6. Suppose $d_{ii} > 0, a_{ii} < 0, i = 1, \ldots, n$. If the conditions

$$a_{ii}a_{jj} > (n! - 1)a_{ij}a_{ji} , \; i \neq j ; i, j = 1, \ldots, n , \tag{4.6}$$

and

$$d_{ii}d_{jj} > (n!-1)d_{ij}d_{ji}\,,\ i \neq j\,;\ i,j = 1,\ldots,n\,, \tag{4.7}$$

are satisfied, then the positive equilibrium state $\bar{u}_1 = (\bar{u}_1,\ldots,\bar{u}_n)$ is global stable in region A given by (4.5) under the hypothesis

$$\left(x_1^{ij}, x_2^{ij}\right) \bigcap \left(y_1^{ij}, y_2^{ij}\right) \neq \emptyset\,, i < j\,;\ i,j = 1,\ldots,n\,, \tag{4.8}$$

where

$$x_1^{ij}, x_2^{ij} = \frac{2a_{ii}a_{jj} - \lambda a_{ij}a_{ji} \mp 2\sqrt{a_{ii}a_{jj}(a_{ii}a_{jj} - \lambda a_{ij}a_{ji})}}{\lambda a_{ij}^2}\,,$$

$$y_1^{ij}, y_2^{ij} = \left(\frac{u_i}{u_j}\right)^2 \left(\frac{\bar{u}_j}{\bar{u}_i}\right) \frac{(2d_{ii}d_{jj} - \lambda d_{ij}d_{ji}) \mp 2\sqrt{d_{ii}d_{jj}(d_{ii}d_{jj} - \lambda d_{ij}d_{ji})}}{d_{ij}^2}$$

with $\lambda = n! - 1$.

A brief proof (for the details see the reference [5]):

$$\det\left[-\frac{1}{2}(CA + A^TC)\right] = \sum_{\sigma} \operatorname{sgn}\sigma\, r_{1\sigma(1)}\cdots r_{n\sigma(n)}\,,$$

where

$$r_{ii} = -c_i a_{ii}\,,\ r_{ij} = -\frac{1}{2}(c_i a_{ij} + c_j a_{ji})\,, i \neq j\,,$$

and the sum runs over all $n!$ permutations σ of the n items $\{1,\ldots,n\}$ and the "sign" of a permutation σ, sgn σ, is $+1$ or -1, according to whether the minimum number of transpositions necessary to achieve it starting from $\{1,2,\ldots,n\}$ is even or odd. Thus,

$$\begin{aligned}
&\det\left[-\frac{1}{2}(CA + A^TC)\right] \\
&= \sum_{\sigma} \operatorname{sgn}\sigma \sum_{m=0}^{n} \prod_{\substack{i=1 \\ i\neq j_k \neq j_h}}^{n} c_i(-a_{ii}) \prod_{\substack{k,h=1 \\ k\neq h}}^{m} \left(-\frac{1}{2}\right)(c_{j_k}a_{j_kj_h} + c_{j_h}a_{j_hj_k}) \\
&> \prod_{i=1}^{n} c_i|-a_{ii}| - \sum_{\sigma}\sum_{m=2}^{n} \frac{1}{2^m} \prod_{\substack{i=1 \\ i\neq j_k\neq j_h}}^{n} c_i|a_{ii}| \prod_{\substack{k,h=1 \\ k\neq h}}^{m} |c_{j_k}a_{j_kj_h} + c_{j_h}a_{j_hj_k}|\,.
\end{aligned}$$

From (4.6) we have

$$c_{j_k}a_{j_kj_k}c_{j_h}a_{j_hj_h} > \frac{n!-1}{4}(c_{j_k}a_{j_kj_h} + c_{j_h}a_{j_hj_k})^2\,, k \neq h$$

which implies

$$\frac{1}{n!-1}\prod_{i=1}^{n} c_i\,|a_{ii}| > \frac{1}{2^m}\prod_{\substack{i=1\\ i\neq j_k\neq j_h}}^{n} c_i\,|a_{ii}| \prod_{\substack{k,h=1\\ k\neq h}}^{m} |c_{j_k}a_{j_kj_h} + c_{j_h}a_{j_hj_k}|$$

which leads to

$$\det\left[-\frac{1}{2}(CA - A^TC)\right] > 0\,.$$

Therefore we have proved that the matrix $CA + A^TC$ is negative definite provided (4.6) and $a_{ii} < 0, i = 1, \dots, n$. In the same way we can prove that $BD - D^TB$ is positive definite if $d_{ii} > 0, i = 1, \dots, n$, and (4.7) valid under the hypothesis (4.8) in the region A. The proof of Theorem 6 is finished.

5. Stability of Several Species in Heterogeneous Habitat

As mentioned in the beginning of this paper, in this case we divide the habitat into p-number of patches ($l_{k-1} \le x \le l_k, k = 1, 2, \dots, p$) such that the growth rate of the species, their interaction and dispersion coefficients are constant but different in different patches. In such a case the system of n species in the k-th patch can be written as

$$\frac{\partial v_i^{(k)}}{\partial t} = u_i^{(k)}\sum_{j=1}^{n} a_{ij}^{(k)} v_j^{(k)} + \frac{\partial}{\partial x}\sum_{j=1}^{n} d_{ij}^{(k)}\frac{\partial v_j^{(k)}}{\partial x}\,, \qquad (5.1)$$
$$i = 1, \dots, n; k = 1, \dots, p\,.$$

The system (5.1) may be associated with the following initial conditions

$$\begin{aligned} v_i^{(k)}(x, 0) &= F_i^{(k)}(x)\,,\\ F_i^{(k)}(l_k) &= F_i^{(k+1)}(l_k)\,, \end{aligned} \qquad (5.2)$$

boundary conditions

$$v_i^{(1)}(0, t) = v_i^{(n)}(l_p, t) = 0\,, \qquad (5.3)$$

matching conditions

$$\begin{aligned} v_i^{(k)}(l_k, t) &= v_i^{(k+1)}(l_k, t)\,,\\ \left[\sum_{j=1}^{n} d_{ij}^{(k)}\,\frac{\partial v_j^{(k)}}{\partial x}\right]_{x=l_k} &= \left[\sum_{j=1}^{n} d_{ij}^{(k+1)}\frac{\partial v_j^{(k+1)}}{\partial x}\right]_{x=l_k}\,. \end{aligned} \qquad (5.4)$$

The condition (5.2) implies that the initial distributions of the species in different patches are continuous on the common boundaries, i.e., at the interface of the two patches. The conditions (5.4) prescribe the matching of species densities and fluxes across interfaces of the patches.

The employed Liapunov function is

$$E = \sum_{k=1}^{p} \int_{l_{k-1}}^{l_k} \left\{ \sum_{i=1}^{n} c_i \left[v_i^{(k)} - \bar{u}_i^{(k)} \log \left(1 + \frac{v_i^{(k)}}{u_i^{(k)}} \right) \right] \right\} dx . \tag{5.5}$$

Differentiating E with respect to t along system (5.1) with conditions (5.2),(5.3),(5.4), we get

$$\frac{dE}{dt} = -\sum_{k=1}^{p} \int_{l_{k-1}}^{l_k} \left[W(v^{(k)}) + \overline{W} \left(\frac{\partial v^{(k)}}{\partial x} \right) \right] dx ,$$

where

$$-W(v^{(k)}) = v^{(k)}(CA^{(k)} + (A^{(k)})^T C)(v^{(k)})^T , \quad \text{with } v^{(k)} \neq 0 ,$$

$$\overline{W} \left(\frac{\partial v^{(k)}}{\partial x} \right) = \left(\frac{\partial v^{(k)}}{\partial x} \right) (B^{(k)} D^{(k)} + (D^{(k)})^T B^{(k)}) \left(\frac{\partial v^{(k)}}{\partial x} \right)^T ,$$

$$\text{with } \frac{\partial v^{(k)}}{\partial x} \neq 0,$$

$$A^{(k)} = \left(a_{ij}^{(k)} \right)_{n \times n} , D^{(k)} = \left(d_{ij}^{(k)} \right)_{n \times n} ,$$

$$C = \operatorname{diag}(c_1, \ldots, c_n) ,$$

$$B^{(k)} = \operatorname{diag} \left(c_1 \frac{\bar{u}_1^{(k)}}{(u_1^{(k)})^2}, \ldots, c_n \frac{\bar{u}_n^{(k)}}{(u_n^{(k)})^2} \right) ,$$

$$v^{(k)} = \left(v_1^{(k)}, \ldots, v_n^{(k)} \right) ,$$

$$\frac{\partial v^{(k)}}{\partial x} = \left(\frac{\partial v_1^{(k)}}{\partial x}, \ldots, \frac{\partial v_n^{(k)}}{\partial x} \right) .$$

From the results obtained in Theorem 6 we have

Theorem 7. If the following conditions are satisfied for the positive integer $k, k = 1, \ldots, p$,

(1) $d_{ii}^{(k)} > 0 , d_{jj}^{(k)} d_{ss}^{(k)} > (n! - 1) a_{js}^{(k)} a_{sj}^{(k)} , j \neq s ,$

(2) $a_{ii}^{(k)} < 0 , a_{jj}^{(k)} a_{ss}^{(k)} > (n! - 1) a_{js}^{(k)} a_{sj}^{(k)} , j \neq s ,$

where $i, j, s = 1, \ldots, n$. Then the positive equilibrium state is global stable in a bounded region under a hypothesis similar to (4.8).

References

1. Lansun Chen, *Models of Mathematical Ecology and Research Methods*, (Chinese), Science Press, Beijing, 1988.
2. Thomas G. Hallam and Simon A. Levin, Editors, *Mathematical Ecology, An Introduction*, Biomathematics Volume 17, Springer-Verlag Berlin Heidelberg, 1986.
3. Loo-Keng Hua, Ziqian Wu and Wei Lin, *Second Order Linear Partial Differential Equation Systems of Two Unknown Functions with Constant Coefficients in Two Independent Variables*, Sciences Press, Beijing, 1979.
4. J. B. Shukla, V. N. Pal and Sunita Gakkhar, *Stability of Interacting Model with Self and Cross-Dispersion in a Patchy Habitat.*
5. Xinhua Ji, *A Stability Theorem of the Lotka-Volterra System of n-species.*

Xinhua Ji
Institute of Mathematics
Academia Sinica
Beijing 100080
China

Constantin Carathéodory: An International Tribute (pp. 574–580)
edited by Th. M. Rassias

PARAMETRICALLY ADDITIVE SUM FORM INFORMATION MEASURES

PL. Kannappan and P.K. Sahoo

ABSTRACT

In this paper we seek the representation of a class of information measures that possess the sum form and satisfy parametric (2,3)-additivity. The measures we obtain contain some well known information measures such as the Shannon's entropy $H_n(P) = -\sum_{i=1}^n p_i \log p_i$ and the entropy of degree β $H_n^\beta(P) = (2^{1-\beta}-1)^{-1}(\sum_{i=1}^n p_i^\beta - 1)$. This paper fills some gaps left out in [8,9].

1. INTRODUCTION

Let $\mathbf{R}$ be the set of all real numbers and let I_o be the unit open interval $]0,1[$. Let $\Gamma_n^o = \{P = (p_1, p_2, ..., p_n) \mid 0 < p_k < 1, \sum_{k=1}^n p_k = 1\}$ and let $\Gamma_n = \{P = (p_1, p_2, ..., p_n) \mid 0 \leq p_k \leq 1, \sum_{k=1}^n p_k = 1\}$ denote the closure of Γ_n^o. An *information measure* is a sequence of mapping $I_n : \Gamma_n^o \to \mathbf{R}\,(n = 2, 3, ...)$. If there exists a *generating function* $f : I_o \to \mathbf{R}$ such that

$$I_n(P) = \sum_{i=1}^n f(p_i) \qquad P \in \Gamma_n^o, \tag{1.1}$$

then $\{I_n\}$ is said to have the *sum form*. A *fundamental information measure* is an information measure which possess the sum form. A brief survey of results related to fundamental information measures can be found in [1].

In this sequel we study certain parametrically additive fundamental information measures, I_n, that is those $\{I_n\}$ satisfying

$$I_{lm}(P * Q) = I_l(P) + I_m(Q) + \lambda I_l(P) I_m(Q) \tag{1.2}$$

for all $P \in \Gamma_l^o$, $Q \in \Gamma_m^o$, $P * Q := (p_1q_1, p_2q_1, ..., p_iq_j, ..., p_lq_m) \in \Gamma_{lm}^o$ and $\lambda \in \mathbf{R}$. Moreover, if the sequence $\{I_n\}$ satisfies (1.1), we arrive at the functional equation

$$\sum_{i=1}^{l}\sum_{j=1}^{m} f(p_iq_j) = \sum_{i=1}^{l} f(p_i) + \sum_{j=1}^{m} f(q_j) + \lambda \sum_{i=1}^{l} f(p_i) \sum_{j=1}^{m} f(q_j) \tag{1.3}$$

where $P \in \Gamma_l^o$, $Q \in \Gamma_m^o$ and $\lambda \in \mathbf{R}$.

The functional equation (1.3) was solved in [3,6,7,11] under various regularity conditions and in [13] without any regularity condition. In all these cited papers [3,6,7,11,13] the functional equation was solved with the use of 0-probability and 1-probability, that is allowing P and Q to be in the closure of Γ_l^o and Γ_m^o, respectively. The use of these extreme values of the probabilities makes the functional equation (1.3) easily solvable. Also, the use of 0-probability and 1-probability requires awkward definitions like $0^\beta = 0$, $0 \log 0 = 0$. It is also *apriori* quite possible that there may exist solutions other than those on [0,1] restricted to]0,1[as shown in [2] for a fundamental equation of information. If the 0-probability and 1-probability are excluded from the domain of (1.3), then the corresponding domain is referred to as open domain.

On open domain, the (Lebesgue) measurable solution of (1.3) with $l, m \geq 3$ was given in [7,8] for $\lambda \neq 0$ and in [4,15] for $\lambda = 0$. The case when $l = 2$ and $m = 3$ was left out for $\lambda \neq 0$. We would like to point out that the general solution of (1.3) when $l, m \geq 3$ and $\lambda \neq 0$ is known [12]. When $\lambda = 0$, the general solution of (1.3) with no regularity assumptions on f is still an *open problem*. Here we find the measurable solution of (1.3) when $l = 2$, $m = 3$ and $\lambda \in \mathbf{R}$, that is of the equation

$$\sum_{i=1}^{2}\sum_{j=1}^{3} f(p_iq_j) = \sum_{i=1}^{2} f(p_i) + \sum_{j=1}^{3} f(q_j) + \lambda \sum_{i=1}^{2} f(p_i) \sum_{j=1}^{3} f(q_j) \tag{1.4}$$

for all $P \in \Gamma_2^o$, $Q \in \Gamma_3^o$.

2. SOLUTION OF (1.3) ON I_o

The case $\lambda = 0$ is covered in [4] and the measurable solution is given by

$$f(p) = ap\log p + bp + b \tag{2.1}$$

where a and b are arbitrary constants.

Now we consider the case when $\lambda \neq 0$. Then (1.4) reduces to

$$\sum_{i=1}^{2}\sum_{j=1}^{3} g(p_iq_j) = \sum_{i=1}^{2} g(p_i)\sum_{j=1}^{3} g(q_j) \tag{2.2}$$

where

$$g(p) := \lambda f(p) + p. \tag{2.3}$$

Lemma 1. Let $g : I_o \to \mathrm{R}$ be measurable and satisfy (2.2) for all $P \in \Gamma_2^o$ and $Q \in \Gamma_3^o$. Then g is given by either

$$g(p) = p^\alpha \tag{2.4}$$

or

$$g(p) = ap + b \tag{2.5}$$

where α is an arbitrary constant and a and b are constants satisfying

$$(a + 6b) = (a + 2b)(a + 3b). \tag{2.6}$$

Proof: Letting $p_1 = p$, where $p \in I_o$, in (2.2), we obtain $\sum_{j=1}^{3} h(q_j) = 0$, where

$$h(q) := \{g(pq) + g((1-p)q) - g(q)\,(g(p) + g(1-p))\}. \tag{2.7}$$

Then from [15], we obtain

$$g(pq) + g((1-p)q) - g(q)\{g(p) + g(1-p)\} = \phi(p)[3q - 1] \tag{2.8}$$

where $\phi : I_o \to \mathrm{R}$. Now for $x \in]0, 1]$ we replace q by $xq \in I_o$ in (2.8) to obtain

$$g(xpq) + g(x(1-p)q) - g(xq)\{g(p) + g(1-p)\} = \phi(p)[3xq - 1]. \tag{2.9}$$

Similarly replacing q by $x(1-q) \in I_o$ in (2.8) we get

$$g(xp(1-q))+g(x(1-p)(1-q))-g(x(1-q))\{g(p)+g(1-p)\} = \phi(p)[3x(1-q)-1]. \tag{2.10}$$

Adding (2.9) to (2.10), we obtain

$$g(xpq) + g(x(1-p)q) + g(xp(1-q)) + g(x(1-p)(1-q))-$$

$$-\{g(xq) + g(x(1-q))\}\{g(p) + g(1-p)\} = \phi(p)\,[3x-2] \tag{2.11}$$

for $x \in]0,1]$ and $p,\ q \in I_o$. Putting $x = 1$ in (2.11) and using the symmetry of the left side of (3.12), we get

$$\phi(p) = \phi(q) = a_o \quad \text{(constant)}. \tag{2.12}$$

By (2.8) and (2.12), (2.11) becomes

$$g(xpq) + g(x(1-p)q) + g(xp(1-q)) + g(x(1-p)(1-q))-$$

$$-g(x)\{g(p)+g(1-p)\}\{g(q)+g(1-q)\} = a_o[3x-2] + \{g(p)+g(1-p)\}\,a_o[3x-1]. \tag{2.13}$$

Again using the symmetry of the left side in p and q, we have

$$a_o[3x-1]\{g(p)+g(1-p)\} = a_o[3x-1]\{g(q)+g(1-q)\} \tag{2.14}$$

for all $p,\ q \in I_o$. If $a_o \neq 0$, then

$$g(p)+g(1-p) = c_o \quad \text{(constant)}. \tag{2.15}$$

Now (2.15) and (2.12) in (2.8) yields,

$$g(u)+g(v) = c_0 g(u+v) + a_o\{3(u+v)-1\} \tag{2.16}$$

with $u = pq$ and $v = (1-p)q$, which is a Pexider equation. Thus $g(p) = ap + b$, where a and b are constants. Letting g into (2.2) we get (2.6). Next suppose $a_o = 0$. Then (2.8) with (2.13) becomes

$$g(pq) + g((1-p)q) = g(q)\{g(p)+g(1-p)\}. \tag{2.17}$$

The measurable solution of (2.17) can be obtained from [10,14] as

$$g(p) = p^{\alpha} \tag{2.18}$$

where α is an arbitrary constant.

This completes the proof of Lemma 1.

Theorem 2. Let $f : I_o \to \mathrm{R}$ be (Lebesgue) measurable and satisfy the functional equation (1.4) for all $P \in \Gamma_2^o$, $Q \in \Gamma_3^o$ and $\lambda \neq 0$. Then f is given by either

$$f(p) = \frac{p^{\alpha} - p}{\lambda} \tag{2.19}$$

or

$$f(p) = \frac{(a-1)p + b}{\lambda} \tag{2.20}$$

where a and b are constants satisfying (2.6). The constant α in (2.19) is an arbitrary real constant.

Proof: Follows from (2.3) and Lemma 1.

3. ADDITIVE FUNDAMENTAL INFORMATION MEASURES

In this section we display the form of all measurable sum form information measures that satisfy parametric (2,3)-additivity.

Theorem 3. Let $I_n : \Gamma_n^o \to \mathrm{R}\,(n = 2, 3, ...)$ be an information measure possessing the (Lebesgue) measurable sum form, that is

$$I_n(P) = \sum_{k=1}^{n} f(p_k), \qquad P \in \Gamma_n^o \tag{3.1}$$

and satisfying parametric (2,3)-additivity, that is

$$I_8(P * Q) = I_2(P) + I_3(Q) + \lambda\, I_2(P) I_3(Q) \qquad P \in \Gamma_2^o,\ Q \in \Gamma_3^o.$$

Then I_n is of the form

$$I_n(P) = \begin{cases} aH_n(P) & \text{if } \lambda = 0 \\ bH_n^{\beta}(P) & \text{if } \lambda \neq 0, \end{cases}$$

where a and b are arbitrary constants, and $H_n(P) = -\sum_{i=1}^{n} p_i \log p_i$ and $H_n^{\beta}(P) = (2^{1-\beta} - 1)^{-1}(\sum_{i=1}^{n} p_i^{\beta} - 1)$.

ACKNOWLEDGEMENTS

This research was supported in parts by grants from College of Arts and Sciences and the Graduate Programs and Research.

REFERENCES

[1] Aczel, J.,"Characterizing information measures: Approching the end of an era", *Lecture Notes in Computer Science*, No 286, Springer Verlag, NY, pp. 359- 383.

[2] Aczel, J. and Ng, C.T., "Determination of all symmetric, recursive information measures of multiplicative type of n positive discrete probability distributions", *Linear Algebra Appl.*, 52/53 (1983), 1-30.

[3] Behara, M. and Nath, P., "Additive and nonadditive entropies of finite measurable partitions", Probability and Information Theory, Vol. 2, Lecture Notes in Math., Vol. 296, Springer, Berlin 1973, 102-138.

[4] Ebanks., B.R., "Measurable solutions of functional equations connected with information measures on open domain", *Utilitas Math.*, 27 (1985), 217-223.

[5] Havrda, J. and Charvat, F., "Quantification method of classification processes, concept of structural a-entropy", *Kybernetika* (Prague), 3 (1967), 30-35.

[6] Kannappan, PL., "On a generalization of some measures in information theory", *Glasnik Mat.*, Ser III, 9(29) (1974), 81-93.

[7] Kannappan, PL., "On some functional equations from additive and nonadditive measures I", *Proc. Edinburgh Math. Soc.*, (2) 23 (1980), 145-150.

[8] Kannappan, PL. and Sahoo, P.K., "On a functional equation connected to sum form nonadditive information measures on an open domain", *C.R. Math. Rep. Acad. Sci.* Canada, 7 (1985), 45-50.

[9] Kannappan, PL. and Sahoo, P.K., "On a functional equation connected to sum form nonadditive information measures on an open domain - I", *Kybernetika*, 22 (1986), 268-275.

[10] Kannappan, PL and Sahoo, P.K., Parametrically Additive Sum Form Weighted Information Measures. To appear.

[11] Losonczi, L., "A characterization of entropies of degree α", *Metrika*, 28 (1981), 237-244.

[12] Losonczi, L., "Sum form equations on an open domain I", *C.R. Math. Rep. Acad. Sci.* Canada, 7 (1985), 85-90.

[13] Losonczi, L. and Maksa, Gy., "On some functional equations of the information theory", *Acta Math. Acad. Sci. Hungar*, 39 (1982), 73-82.

[14] Maksa, Gy., "The general solution of a functional equation arising in information theory", *Acta Math. Acad. Sci. Hungar*, 49 (1987), 213-217.

[15] Sahoo, P.K., "On some functional equations connected to sum form information measures on open domains", *Utilitas Math.*, 23 (1983), 161-175.

Pl. Kannappan
Department of Pure Mathematics
University of Waterloo
Waterloo, Ontario, N2L 3G1
CANADA

P. K. Sahoo
Department of Mathematics
University of Louisville
Louisville, KY 40292
USA

Constantin Carathéodory: An International Tribute (pp. 581–591)
edited by Th. M. Rassias

NEAREST AND FARTHEST POINTS OF CLOSED SETS IN HYPERBOLIC SPACES

W. A. Kirk

ABSTRACT. It is shown that for a wide class of uniformly convex hyperbolic metric spaces, the set of points of the space which have a nearest point in a given closed set S is dense in the space. If S is also bounded the same is true of the set of points of the space which have a farthest point in S. The algorithm involved is essentially one devised by Edelstein for uniformly convex Banach spaces.

1. INTRODUCTION. In [11] the writer observed that Krasnoselskii's iteration process for nonexpansive mappings extends from a Banach space setting to a much wider class of spaces — the so–called metric spaces of 'hyperbolic type'. This class of spaces includes all normed linear spaces as well as the Hilbert ball endowed with the hyperbolic metric ([10], also see [9]) and the cartesian product of such Hilbert balls ([13]). Other examples are discussed by Reich and Shafrir in [16], who propose this class of spaces as an appropriate background for the study of nonlinear operator theory. We note in particular that investigations of hyperbolic metrics in spaces of more than one dimension originate with Caratheodory [5].

Our purpose here is to show that results of Edelstein [7, 8] on nearest and farthest points of closed sets in uniformly convex Banach spaces also extend to this wider setting. In doing so we show that these results are essentially 'geometric' in nature, requiring no additional underlying topological structure and only the linear structure associated with the hyperbolic nature of the metric. (Corresponding results in Banach spaces (cf., [1], [2], [14], [15]) are formulated either in reflexive spaces or involve weak compactness assumptions.)

2. HYPERBOLIC METRIC SPACES. We suppose (X,ρ) is a metric space containing a family $\mathscr{L}$ of metric lines such that distinct points $x,y \in X$ lie on exactly one member $\ell(x,y)$ of $\mathscr{L}$. We shall use the symbol $s[x,y]$ to denote the metric segment of $\ell(x,y)$ which joins x and y. For each $t \in [0,1]$ there is a unique point z in $s[x,y]$ for which

$$\rho(x,z) = t\rho(x,y) \text{ and } \rho(z,y) = (1-t)\rho(x,y).$$

Adopting the notation of [9], we shall denote this point $(1-t)x \oplus ty$.

We shall say that ρ is a metric of *hyperbolic type* if the following condition holds:

$$\rho(\tfrac{1}{2}x \oplus \tfrac{1}{2}y, \tfrac{1}{2}x \oplus \tfrac{1}{2}z) \leq \tfrac{1}{2}\rho(y,z)$$

for all x,y and z in X. Since this condition with strict inequality is an axiom of hyperbolic geometry (cf., [18]), if ρ is a metric of hyperbolic type we shall refer to (X,ρ) as a *hyperbolic metric space.*

The *modulus of convexity* $\delta_X: X \times (0,\infty) \times (0,2] \longrightarrow [0,1]$ of a hyperbolic metric space (X,ρ) is defined by setting

$$\delta_X(a,r,\epsilon) = \inf\{1 - \rho(a,\tfrac{1}{2}x \oplus \tfrac{1}{2}y)/r\}$$

where the infimum is taken over all points x and y satisfying $\rho(a,x) \leq r$, $\rho(a,y) \leq r$ and $\rho(x,y) \geq \epsilon r$. If δ is always positive, X is said to be *uniformly convex.*

Several examples of uniformly convex hyperbolic metric spaces are given in [16]. In particular, the (infinite dimensional) Hilbert ball H is a uniformly convex hyperbolic metric space. (A precise formula for the modulus δ_H is given in [9, p. 107].) The Hilbert ball, as well as all of the hyperbolic metric spaces alluded to in the Introduction, satisfy another assumption we shall need. Specifically, we shall assume the hyperbolic inequality is uniform in the following rather weak sense:

(*) Given any bounded set S in X, for each $\epsilon > 0$ there exists $\epsilon' > 0$ such that if $x,y,z \in S$ and $\rho(y,z) \geq \epsilon$, then $\rho(\frac{1}{2}x \oplus \frac{1}{2}y, \frac{1}{2}x \oplus \frac{1}{2}z) \geq \epsilon'$.

We shall also assume the that motions (surjective isometries) of X are *transitive* in the sense that given any two points $x,y \in X$ there exists a motion of X which maps x to y. The Hilbert ball satisfies this property as well ([9, p. 98]). This assumption effectively means that the modulus δ no longer depends on the point a.

We shall use standard notation. In particular, the symbol B(a;r) will denote a closed ball centered at $a \in X$ with radius r. For a given space X we set $\delta = \delta_X$.

3. **MAIN THEOREMS.** The following theorems were originally formulated in a uniformly convex Banach space setting. The first is due to Edelstein [7] and the second, independently, to Edelstein [8] and Steckin [17]. (Steckin's version of Theorem 2 asserts further that the complement of C_2 is of the first category.)

Theorem 1. *Suppose* X *is a uniformly convex hyperbolic metric space which is complete, has transitive motions, and satisfies* (*). *Let* S *be a nonempty closed and bounded subset of* X *and let*

$$C_1 = \{c \in X: \exists\, s \in S \text{ such that } \rho(c,s) = \sup\{\rho(c,x): x \in S\}\}.$$

Then C_1 *is dense in* X.

Theorem 2. *Suppose* X *is as above. Let* S *be a nonempty closed subset of* M *and let*

$$C_2 = \{c \in X: \exists\, s \in S \text{ such that } \rho(c,s) = \inf\{\rho(c,x): x \in S\}\}.$$

Then C_2 *is dense in* X.

We should mention that each of these theorems has been extended in Banach space settings. Asplund [1] has shown that if S is a bounded and closed

subset of a reflexive locally uniformly convex Banach space, then the set C_1 of points of X which have a farthest point in S contains a dense G_δ set. Subsequently, Lau [14] showed that if S is assumed to be weakly compact then the result follows if X is an arbitrary Banach space. Further extensions of Theorem 1 may be found in Zizler [19] and Deville–Zizler [6]. In [15, Theorem 10] Lau shows that in any reflexive space with Kadec–Klee norm, if K is a nonempty closed set then the set of points of $X\backslash K$ with a nearest point in K is of the second category. Since Konjagin [12, Theorem 9] has shown that if X is a Banach space which does not belong to this class then there exists a closed nonempty set K for which the set of points of $X\backslash K$ with nearest points in K is not dense, this class characterizes the density property for closed sets. (For related results, see Borwein [2] and Borwein and Giles [3].)

Our proofs will require the following:

Proposition 1. *Let* X *be a uniformly convex hyperbolic space which transitive motions and satisfies* (*), and *let* r *and* d *be fixed positive numbers. Suppose* $c, c' \in X$ *satisfy* $\rho(c,c') = r$. *Then*

$$(1) \qquad \lim_{\xi \to 0} \operatorname{diam} [B(c;d) \cap (X \backslash B(c';d+r-\xi))] = 0;$$

$$(2) \qquad \lim_{\xi \to 0} \operatorname{diam} [B(c;d-r+\xi) \cap (X \backslash B(c';d))] = 0.$$

Moreover, the convergence in (1) *and* (2) *is uniform for all such* c, c' *lying in a bounded set.*

Proof. Since (1) and (2) are equivalent (replace d in (1) with $d+r-\xi$ and reverse the roles of c and c') we shall only prove (1). The proof is by contradiction.

Suppose there exist $\epsilon > 0$, a sequence $\{\xi_i\}$ of positive numbers with $\xi_i \to 0$, and sequences $\{c_i\}$ and $\{c_i'\}$ lying in a ball $B(a;R) \subset X$ satisfying $\rho(c_i,c_i') \equiv r$ for which $\operatorname{diam}(S_i) \geq \epsilon$, where

$$S_i = B(c_i;d) \cap (X \setminus B(c_i';d+r-\xi_i)).$$

Moreover, since X has transitive motions, we may assume $c_i' \equiv a \in X$. Next, for each i, let $u_i \in X$ satisfy

$$c_i \in \text{seg}[a,u_i] \text{ and } \rho(c_i,u_i) = d.$$

Then $u_i \in S_i$, so by assumption there exists $w_i \in S_i$ for which $\rho(u_i,w_i) \geq \epsilon/2$. Select
$h_i \in \text{seg}[a,w_i]$ satisfying $\rho(a,h_i) = d$. Since $\rho(u_i,w_i) \geq \epsilon/2$ for all i the condition (*) implies that $\{h_i\}$ is bounded away from $\{c_i\}$; thus there exists $\epsilon' > 0$ such that
$\rho(h_i,c_i) \geq \epsilon' r$.

Now observe that

$$\lim_{i \to \infty} [r + \rho(h_i,w_i)] = \lim_{i \to \infty} \rho(a,w_i) \leq r + d;$$

hence $\lim_{i \to \infty} \rho(h_i,w_i) \leq d$. Since $\lim_{i \to \infty} \rho(c_i,w_i) \leq d$, it follows that

$$\lim_{i \to \infty} \rho(w_i, \tfrac{1}{2}h_i \oplus \tfrac{1}{2}c_i) \leq d.$$

Since $\delta = \delta(a,r,\epsilon') > 0$ it is possible to choose $\eta > 0$ so that

$$(1 - \delta)r + d + \eta \leq k < r + d.$$

Also, since $\rho(a,c_i) = \rho(a,h_i) \equiv r$, we have

$$\rho(a,\tfrac{1}{2}h_i \oplus \tfrac{1}{2}c_i) \leq (1-\delta)r.$$

Thus for i sufficiently large

$$\rho(a,w_i) \leq \rho(a,\tfrac{1}{2}h_i \oplus \tfrac{1}{2}c_i) + \rho(w_i,\tfrac{1}{2}h_i \oplus \tfrac{1}{2}c_i) \leq (1-\delta)r + d + \eta \leq k.$$

This clearly contradicts $\lim_{i\to\infty} \rho(a,w_i) = r + d$.

We are now ready to prove the theorems. A reformulation of Proposition 1 will facilitate the proofs.

Proposition 2. *Let* X *be as in Proposition 1 and let* S *be a bounded subset of* X. *Then if* ϵ, d, *and* r *are fixed positive numbers there exist* $\xi = \xi(\epsilon,d,r) > 0$ *and* $\xi' = \xi'(\epsilon,d,r) > 0$ *such that if* c, c' $\in$ S *satisfy* $\rho(c,c') = r$, *then*

(1') $$\operatorname{diam}[B(c;d)\cap(X\setminus B(c';d+r-\xi))] < \epsilon;$$

(2') $$\operatorname{diam}[B(c;d-r+\xi)\cap(X\setminus B(c';d))] < \epsilon.$$

Proof of Theorem 1. Fix $c_0 \in X$ and let $d_0 = \sup\{\rho(c_0,x): x \in S\}$. Clearly we may suppose $d_0 > 0$. Let $r \in (0,d_0)$ be arbitrary and let $\{r_i\}$ and $\{\epsilon_i\}$ be sequences of positive numbers satisfying $\Sigma r_i < r$ and $\epsilon_i \to 0$. Select $\xi_1 = \xi(\epsilon_1,d_0,r_1)$ as in condition (1') and choose $x_0 \in S$ so that $\rho(x_0,c_0) > d_0 - \xi_1$. Now let $c_1 \in X$ be chosen so that

$$c_0 \in \operatorname{seg}[c_1,x_0] \text{ and } \rho(c_1,c_0) = r_1,$$

set $d_1 = \sup\{\rho(c_1,x): x \in S\}$, and let $\xi_2 = \xi(\epsilon_2,d_1,r_2)$. We proceed by induction. Having defined c_n, x_n, and d_n, with

$$c_{n-1} \in \text{seg}[c_n, x_{n-1}] \text{ and } \rho(c_n, c_{n-1}) = r_n,$$

$d_n = \sup\{\rho(c_n, x): x \in S\}$, and $\rho(x_n, c_n) \geq d_n - \xi_{n+1}$ where $\xi_{n+1} = \xi(\epsilon_{n+1}, d_n, r_{n+1})$, set $d_{n+1} = \sup\{\rho(c_{n+1}, x): x \in S)$ where $c_{n+1} \in X$ is chosen so that

$$c_n \in \text{seg}[c_{n+1}, x_n] \text{ and } \rho(c_{n+1}, c_n) = r_{n+1}.$$

Since

$$d_{n+1} \geq \rho(c_{n+1}, x_n) = r_{n+1} + \rho(c_n, x_n),$$

and since the argument terminates if equality holds, we may suppose it is possible to choose $x_{n+1} \in S$ so that both the following hold:

(3) $\rho(c_{n+1}, x_{n+1}) > d_{n+1} - \xi_{n+2}$ and $\rho(c_{n+1}, x_{n+1}) > r_{n+1} + \rho(c_n, x_n)$,

where $\xi_{n+2} = \xi(\epsilon_{n+2}, d_{n+1}, r_{n+2})$. In particular, for $n > k$ the triangle inequality and the second inequality in (3) imply

$$\rho(c_{k+1}, x_{n+1}) \geq \rho(c_{n+1}, x_{n+1}) - \sum_{i=k+2}^{n+1} r_i > \rho(c_{k+1}, x_{k+1}).$$

Hence

$$\rho(c_{k+1}, x_{n+1}) > r_{k+1} + \rho(c_k, r_k) > r_{k+1} + d_k - \xi_{k+1}.$$

Since clearly $x_{n+1} \in B(c_k; d_k)$ for all k, we have for all $n > k$:

$$x_{n+1} \in B(c_k;d_k) \cap (X \setminus B(c_{k+1};d_k + r_{k+1} - \xi_{k+1})).$$

Therefore, if $n, m > k$, $\rho(x_n,x_m) \leq \epsilon_{k+1}$ proving that $\{x_n\}$ is a Cauchy sequence. Thus $\{x_n\}$ converges to a point $s \in S$ and, since $\rho(c_{n+1},c_n) = r_n$ with $\Sigma r_i < r$, $\{c_n\}$ is also Cauchy with limit $c \in B(c_0;r)$. Clearly $\rho(c,s) = \sup\{\rho(c,x): x \in S\}$, completing the proof.

Proof of Theorem 2. This proof is dual to that of of Theorem 1, using (2') instead of (1'). Fix $c_0 \in X$ and let $d_0 = \inf\{\rho(c_0,x): x \in S\}$. Assume $d_0 > 0$ and let $r \in (0,d_0)$ be arbitrary. As before, let $\{r_i\}$ and $\{\epsilon_i\}$ be sequences of positive numbers satisfying $\Sigma r_i < r$ and $\epsilon_i \longrightarrow 0$. Select $\xi_1' = \xi'(\epsilon_1,d_0,r_1)$ as in (2') and choose $x_0 \in S$ so that $\rho(c_0,x_0) < d_0 + \xi_1'$. Now choose $c_1 \in \text{seg}[c_0,x_0]$ satisfying $\rho(c_0,c_1) = r_1$, let $d_1 = \inf\{\rho(c_1,x): x \in S\}$, and set $\xi_2' = \xi(\epsilon_2,d_1,r_2)$. Now suppose c_n, x_n, and d_n have been defined with $c_n \in \text{seg}[c_{n-1},x_{n-1}]$, $d_n = \inf\{\rho(c_n,x): x \in S\}$, and so that $\rho(x_n,c_n) < d_n + \xi_{n+1}'$ where $\xi_{n+1}' = \xi'(\epsilon_{n+1},d_n,r_{n+1})$. Let $c_{n+1} \in \text{seg}[c_n,x_n]$ satisfy $\rho(c_n,c_{n+1}) = r_{n+1}$ and then set $d_{n+1} = \inf\{\rho(c_{n+1},x): x \in S\}$. Since

$$d_{n+1} \leq \rho(c_{n+1},x_n) = \rho(c_n,x_n) - r_{n+1},$$

with the argument terminating if equality holds, it is possible to choose $x_{n+1} \in S$ so that both the following hold:

$$\rho(x_{n+1},c_{n+1}) < r_{n+1} + \xi_{n+2}' \text{ and } \rho(x_{n+1},c_{n+1}) < \rho(c_n,x_n),$$

where $\xi'_{n+2} = \xi'(\epsilon_{n+2}, d_{n+1}, r_{n+2})$. The proof may now be completed exactly as in Theorem 1.

Acknowledgment. Part of this work was done while the author was visiting the University of Milan. He wishes to thank Peter Kenderov for bringing much of the literature, and in particular [14], to his attention.

REFERENCES

1. Asplund, E., Farthest points in reflexive locally uniformly rotund Banach spaces, Israel J. Math 4, 213–216 (1966).

2. Borwein, J. M., Weak local supportability and applications to approximation, Pacific J. Math. 82, 323–338 (1979).

3. Borwein, J. M., and Giles, J. R., The proximal normal formula in Banach space,Trans. Amer. Math. Soc. 302, 371–381 (1987).

4. Busemann, H., Spaces with non–positive curvature, Acta Math. 80, 259–310 (1948).

5. Caratheodory, C., Uber das Schwarzsche Lemma bei analytischen Funktionen von zwei komplexen Veranderlichen, Math. Ann. 97, 76–98 (1926).

6. Deville, R., and Zizler, V., Farthest points in W^*–compact sets, Bull. Austral. Math. Soc.38, 433–439 (1988).

7. Edelstein, M., Farthest points of sets in uniformly convex Banach spaces, Israel J. Math. 4, 171–176 (1966).

8. Edelstein, M., On nearest points of sets in uniformly convex Banach spaces, Israel J. London Math. Soc. 43, 375–377 (1968).

9. Goebel, K., and Reich, S., *Uniform Convexity, Hyperbolic Geometry and Nonexpansive Mappings*, Marcel Dekker, New York and Basel, 1984.

10. Geobel, K., Sekowski, T., and Stachura, A, Uniform convexity of the hyperbolic metric and fixed points of holomorphic mappings in the Hilbert ball, Nonlinear Anal., Theory Methods & Applications 4, 1011–1021 (1980).

11. Kirk, W. A., Krasnoselskii's iteration process in hyperbolic space, Numer. Funct. Anal. and Optimiz. 4(4), 371–381 (1981–82).

12. Konjagin, S. V., On approximation properties of closed sets in Banach spaces and the characterization of strongly convex spaces, Soviet Math. Dokl. 21, 418–422 (1980).

13. Kuczumow, T., and Stachura, A., Fixed points of holomorphic mappings in the cartesian product of n unit Hilbert balls, Canad. Math. Bull. 29, 281–286 (1986).

14. Lau, K–S., Farthest points in weakly compact sets, Israel J. Math. 22, 168–174 (1975).

15. Lau, K–S., Almost Chebyshev subsets in reflexive Banach spaces, Indiana Univ. Math. J. 27, 791–795 (1978).

16. Reich, S., and Shafrir, I., Nonexpansive iterations in hyperbolic spaces, preprint.

17. Steckin, S. B., Approximation properties of sets in normed linear spaces, Rev. Roum. Math. Pures et Appl. 8, 5–18 (1963) (Russian).

18. Young, W. H., On the analytical basis of non–euclidean geometry, Amer. J. Math. 33, 240–286 (1911).

19. Zizler, V., On some extremal problems in Banach spaces, Math. Scand. 32, 214–224 (1973).

W. A. Kirk
Department of Mathematics
University of Iowa
Iowa City, Iowa 52242
U. S. A.

Constantin Carathéodory: An International Tribute (pp. 592–628)
edited by Th. M. Rassias

ON CARATHÉODORY'S THEORY OF DISCONTINUOUS EXTREMALS AND GENERALIZATIONS

Manfred Kracht and Erwin Kreyszig

ABSTRACT

This paper concerns Carathéodory's fundamental work in the calculus of variations, its origins (Secs. 1, 2), its basic ideas (Sec. 3) and their impact on modern work in optimal control (Sec. 4), minimal surfaces (Sec. 5), and functional analysis (Secs. 6, 7).

0. INTRODUCTION

Constantin Carathéodory (1873–1950) was born in Berlin, descending from an old and highly respected Greek family. He wrote his doctoral thesis while he studied at the University of Göttingen and had it published in 1904. This thesis "Über die diskontinuierlichen Lösungen in der Variationsrechnung" (On discontinuous solutions in the calculus of variations)[8],[1],3–79 opened a long series of fundamental contributions to the calculus of variations, one of Carathéodory's main fields of work.

In this paper we show that, whereas activity in the calculus of variations proper, as measured in terms of numbers of publications, has been decreasing for some time, the impact of the main ideas in the field has infiltrated, transformed, and fertilized various other areas. Selecting some important ones of the latter, we shall observe particularly ideas initially resulting from Carathéodory's work, whose profound effects deserve to become known in more detail. A main theme to be considered is that of the reduction of differentiability assumptions, typical of a central trend in functional analysis and its application to partial differential

equations. Another theme is the interrelation between the calculus of variations and functional analysis in general.

Other fields of Carathéodory's work include complex analysis and measure and integration. Since we shall not deal with these, be it permitted that we add at least a few lines from a highly important, little known document[5],226–229, the letter of application of October 1917 from Berlin University to the Minister of Education to appoint Carathéodory to a professorship, signed by E. Schmidt, H.A. Schwarz, Schottky, Planck, and others:

> "... Carathéodory succeeded in proving the Landau-Picard theorem in a surprisingly simple way, in shedding bright light on the mysterious character of this theorem [and] in making it substantially more precise ... [His investigations give] a complete solution of the extremely difficult problem of the behavior of the boundary under conformal mapping of general regions ... All his papers in Analysis are penetrated by the spirit of Geometry. [He] uses his extraordinary spatial imagination ["*Raumanschauung*"] as a most powerful tool ..."

It is this "spirit of Geometry" that we shall sense as the background of much of his work to be investigated here.

1. ON THE CLASSICAL BASIS OF CARATHÉODORY'S WORK

In this section we characterize the development of those main ideas of the so-called classical theory of the calculus of variations that were most relevant as a basis of Carathéodory's work. This theory was created stepwise by Euler, Lagrange, and Jacobi, and completed by Weierstrass. We recall that in the simplest case, one is concerned with the extremization (minimization or maximization) of a functional (integral)

$$J[y] = \int_{x_0}^{x_1} F(x, y, y')\, dx, \quad y(x_0) = y_0, \quad y(x_1) = y_1, \quad x_0 < x_1 \tag{1.1}$$

in a given class of functions, subject to the indicated boundary conditions. For any variational problem, the functions y in the domain of the functional satisfying the additional conditions are called the *admissible functions* of the problem. Presently we assume that $y \in C^1([x_0, x_1])$.

Early theory concerned necessary conditions for an *extremal* (solution $y = y(x)$, solution curve of the problem). Then came sufficient conditions and finally existence questions, the latter with some delay because problems arose from physics or geometry, where the existence of an extremal was "obvious". Existence was first treated by theorems on differential equations and later by "direct methods" (Sec. 2).

Johann Bernoulli gave the earliest impetus to the calculus of variations, in 1696, by posing the famous problem of the *brachistochrone* (curve of fastest descent): Find the curve

$$C : y = y(x) \quad \text{from} \quad P_0 : (x_0, y_0) \quad \text{to} \quad P_1 : (x_1, y_1), \quad y_1 > y_0$$

in a vertical plane (with horizontal x-axis and downward y-axis) such that a mass particle, being initially at rest, slides down C without friction in the shortest possible time; thus, minimize (1.1) with

$$F(x, y, y') = k[(1 + y'^2)/(y - y_0)]^{\frac{1}{2}}, \qquad k \text{ constant.}$$

Carathéodory wrote two little known interesting critical articles [8],*II*,93–107,108–128 on Bernoulli's work in which he convincingly demonstrated that the work already contained rudiments of ideas of Weierstrass on *fields*. He also discovered and developed an elegant extension of Bernoulli's method to other problems ("Carathéodory's method", cf. Sec. 3).

The earliest general necessary condition for an extremal appeared in a paper by Euler of 1736 and again in Euler's book of 1744, the first systematic treatment of the calculus of variations (from a modern viewpoint authoritatively commented on by Carathéodory[8],*V*,111–174): If $y = y(x)$ is an extremal of (1.1) [of class $C^2([x_0, x_1])$], it must satisfy the *Euler-Lagrange equation*

$$F_y - \frac{d}{dx} F_{y'} = 0, \tag{1.2}$$

written out,

$$F_{y'y'} y'' + F_{y'y} y' + F_{y'x} - F_y = 0.$$

This suggests to call (1.1) a *regular problem* when $F_{y'y'}$ is never zero, and then to assume that $F_{y'y'} > 0$. Equation (1.2) is obtained from

$$\bar{y} = y + \epsilon\eta, \quad \eta \in C^2([x_0, x_1]), \quad \eta(x_0) = \eta(x_1) = 0 \tag{1.3}$$

and $\partial J[\bar{y}]/\partial\epsilon|_{\epsilon=0} = 0$.

A remarkable claim of Carathéodory [l.c.,165] states that the famous Maupertuis's *principle of least action* (i.e., minimize the action integral $\int mv\, ds$ over the path; m = mass, v = speed, s = arc length) was the driving force ("der treibende Faktor") of Euler's work on the calculus of variations.

With the publication of his famous Memoir in 1760-1761, at the age of only 24, Lagrange became the main initiator of the *theory* of the calculus of variations. He is also the founder of analytical mechanics, which he related to the calculus of variations by deriving his equations of motion from the minimization of the action integral. A great advantage of his method over Euler's is its extendibility to double integrals (to which he turned in 1760-61), triple integrals, etc., for instance, to the extremization of

$$J[z] = \iint_{\Omega} F(x, y, z, z_x, z_y)\, dx\, dy \tag{1.4}$$

over a domain Ω in the xy-plane subject to given boundary conditions. The corresponding Euler-Lagrange equation is

$$F_z - \frac{\partial}{\partial x}F_{z_x} - \frac{\partial}{\partial y}F_{z_y} = 0. \tag{1.5}$$

Lagrange also was the first to express the need for denoting the *first variation* by a special symbol δ, for which he gave rules of operation (but no definition). In modern terms, the first variation of y in (1.3) is

$$\delta y = \epsilon\eta(x), \tag{1.6}$$

and the first variation of the functional (1.1) is

$$\delta J = \epsilon\frac{\partial J[\bar{y}]}{\partial\epsilon}\Big|_{\epsilon=0} = \epsilon\int_{x_0}^{x_1}(F_y\eta + F_{y'}\eta')\, dx. \tag{1.7}$$

The *second variation* of (1.1) is

$$\delta^2 J = \frac{\epsilon^2}{2}\frac{\partial^2 J[\bar{y}]}{\partial\epsilon^2}\Big|_{\epsilon=0} = \frac{\epsilon^2}{2}\int_{x_0}^{x_1}(F_{yy}\eta^2 + 2F_{yy'}\eta\eta' + F_{y'y'}\eta'^2)\, dx. \tag{1.8}$$

It was introduced by Legendre in 1786, formally motivated by Taylor's theorem

$$J[y+\epsilon\eta] = J[y] + \delta J + \delta^2 \bar{J}, \tag{1.9}$$

and conceptually by a beginning search for sufficient conditions for an extremal, by Legendre, Jacobi, and others, an evolution brought to completion by Weierstrass, whose work provided a basis for Carathéodory's (see[9]Preface). (In (1.9), the tilda means that the arguments are $y+\bar{\epsilon}\eta, y'+\bar{\epsilon}\eta'$ with $\bar{\epsilon} \in (0,\epsilon]$.)

In 1836, in a letter to Encke, Jacobi[16],IV,39–55 communicated a sufficient condition for a *weak minimum* of (1.1) (A. Kneser's term[17],54), that is, a minimum when a family of admissible functions y satisfying

$$|y-\bar{y}| < \rho \qquad (\rho > 0) \tag{1.10}$$

as well as

$$|y'-\bar{y}'| < \rho \tag{1.11}$$

is considered a neighborhood of $\bar{y}$. Sufficient for an extremal to give a weak minimum is [cf. (1.2)]

$$F_{y'y'} > 0 \tag{1.12}$$

together with the so-called

Jacobi condition. The *conjugate point* of x_0, defined as the smallest zero of a solution of

$$\frac{d}{dx}\left(F_{y'y'}\frac{dw}{dx}\right) - \left(F_{yy} - \frac{d}{dx}F_{yy'}\right)w = 0, \quad w(x_0) = 0, \quad (x \geq x_0) \tag{1.13}$$

is greater than x_1. Here $w(x) = \partial y/\partial\alpha|_{\alpha=0}$ and $\alpha = 0$ corresponds to $\bar{y}$ in the family of extremals $y = y(x,\alpha)$.

Proceeding with proverbial "Weierstrassian rigor", Weierstrass became convinced that it would be essential to extend the domain of (1.1) by considering also a *strong minimum* (A. Kneser's term, l.c.), that is, a minimum when a family of admissible functions satisfying only (1.10) is considered a neighborhood of $\bar{y}$. "In fact, if one realizes the question with which the calculus of variations is concerned, one recognizes that the problem must be solved in this sense"[23],187,

(unpublished). A sufficient condition for a strong minimum needed new concepts, that of a field and the excess function (E-function), the latter of which Weierstrass introduced in 1879, "a turning point in the history of the calculus of variations" (Bolza[6],84b).

Weierstrass defined a *field of extremals* of (1.1) to be a domain Ω in the xy-plane such that through every point of Ω there passes precisely one extremal of a one-parameter family of extremals of (1.1) depending continuously on the parameter. To define the E-function, he started from the *slope function* $p = p(x, y)$, the slope at (x, y) of the extremal of a field of extremals $y = h(x, \alpha)$; thus

$$p(x, y) = h'(x, \alpha)|_{\alpha=\alpha(x,y)}. \tag{1.14}$$

With this he defined the *E-function* by

$$E(x, y, p, y') = F(x,\ y,\ y') - F(x, y, p) - (y' - p)F_{y'}(x, y, p) \tag{1.15}$$

where $y = y(x)$ is any C^1-curve in the field of extremals. He was then able to prove that if for an extremal $y = \bar{y}(x)$ of the field the above sufficient conditions for a weak minimum are satisfied and if $E \geq 0$ at every point in the field and for every y', then $\bar{y}(x)$ gives a strong minimum of (1.1).

"It is perhaps a unique case that the ideas of a great master which revolutionized a whole science [mathematics] became [generally known] only slowly and through underground channels", wrote Carathéodory[8],V,343 in 1927 when Weierstrass's "Vorlesungen über Variationsrechnung"[31],VII were finally officially published.

2. CALCULUS OF VARIATIONS IN GÖTTINGEN AT HILBERT'S TIME

In 1902, when Carathéodory came to Göttingen from Berlin, where he had been H.A. Schwarz's student, he found a stimulating atmosphere for the calculus of variations. Indeed, in Problem 23 of his famous Paris talk of 1900 on unsolved problems, Hilbert had drawn attention to Weierstrass's work and to A. Kneser's book[17], which Carathéodory[8],V,337 called "the first presentation on the modern calculus of variations, [which was] enormously successful" in greatly increasing

research activities in the field between 1900 and 1910. Five of the over twenty-five doctoral theses supervised by Hilbert between 1900 and 1907 were on the calculus of variations. Hilbert's interest in this area, so far remote from his famous "Zahlbericht", was most likely motivated by the state of Analysis at that time: with complex analysis securely founded and impressively developed, activity had shifted to boundary value problems, with the *Dirichlet problem* for the Laplace equation

$$\Delta u = 0 \quad \text{in } \Omega; \quad u|_{\partial\Omega} = f, \qquad u \in C^2(\Omega) \cap C^0(\overline{\Omega}), \quad \Omega \subset \mathbf{R}^2 \text{ or } \mathbf{R}^3 \tag{2.1}$$

and eigenvalue problems for the wave equation in the center.

Now an existence "proof" for (2.1) in general domains had been based on the so-called *Dirichlet principle*, which may be stated as follows. There exists a unique function u that minimizes the functional *(Dirichlet integral)*

$$J[u] = \int_\Omega |\text{grad } u|^2 \, dx, \quad u|_{\partial\Omega} = f \in C^0(\partial\Omega), \quad \Omega \subset \mathbf{R}^2 \text{ or } \mathbf{R}^3 \tag{2.2}$$

among all functions $u \in C^1(\Omega) \cap C^0(\overline{\Omega})$ which take on given values f on the boundary $\partial\Omega$ of Ω and that function u satisfies (2.1). – Note that (2.1) is the Euler-Lagrange equation of (2.2). – But in 1870 Weierstrass[31],II,49–54 pointed out that the Dirichlet principle in its general form is invalid, the faulty conclusion of existence being a consequence of a conceptual confusion of "minimum" and "greatest lower bound." His counterexample is (l.c., 53)

$$J[\phi] = \int_{-1}^{1} [x\phi'(x)]^2 dx, \quad \phi(-1) = a, \ \phi(1) = b \neq a \tag{2.3}$$

where $\phi \in C^1([-1,1])$. Clearly, $J[\phi] \geq 0$, but $J[\phi] = 0$ for no such ϕ. Although this merely shows that there are variational problems without solution, rather than directly implying that the Dirichlet principle is faulty – because the latter concerns a different integral – (a fact that is often overlooked), it is clear that the situation for the latter integral is basically the same.

Now after this criticism there was no more *general* principle for handling corresponding problems, but each had to be attacked by a different method,

ingenious in nature (C. Neumann, Schwarz, Poincaré), but confusingly heterogeneous, even after Poincaré's great effort in unification. It seems that in this lamentable situation, Hilbert set his hope in the calculus of variations because it had produced general principles in the past (cf. Sec. 1). Not intimidated by Weierstrass, Hilbert[15],*III*,10–37 was able to re-establish the Dirichlet principle within proper limits as a valid method of proof, in two papers of 1900 and 1901 (reprinted 1905). In his first paper [l.c., 11] he proposed the following more general formulation of the Dirichlet principle.

> "Every regular problem of the calculus of variations [Sec. 1] has a solution as soon as with respect to the nature of the given boundary conditions suitable restrictive assumptions are satisfied and, if necessary, the concept of a solution is suitably generalized. – In what way this principle can serve as a guiding star for finding rigorous and simple existence proofs will be shown by the following examples."

These are (a) shortest arcs on a surface, and (b) the Dirichlet problem for the Laplace equation in a plane domain with continuously curved boundary and class $C^1(s)$ boundary values. This initiated the *direct methods* (methods without the use of the Euler-Lagrange equations), which became of basic importance in the existence theory of the calculus of variations. (A forerunner of these methods was Euler's almost forgotten "direct difference method"[32],289–291.) Another solution of the Dirichlet problem by direct methods was given later, in 1907, by Lebesgue[20],*IV*,91–122.

Apart from this splendid beginning of his intentions to uniformize Analysis by methods of the calculus of variations, Hilbert made no further effort to employ the calculus of variations for that goal, but soon turned to integral equations as the seemingly more promising tool for a uniform approach to the central problems of Analysis mentioned above, and was soon able to progress far beyond the landmark set by Fredholm in 1900-1903. Nevertheless, the Göttingen atmosphere remained receptive to ideas in the calculus of variations, as subsequent doctoral theses document, Carathéodory's of 1904 being the most outstanding of them. We shall discuss Carathéodory's thesis and work resulting from it in the next section, and then turn to applications related to Carathédory's theory.

3. ON CARATHÉODORY'S WORK

In 1903, Hans Hahn, having just completed his doctoral work under von Escherich in Vienna, came to Göttingen and gave a talk on von Escherich's theory of the second variation. Carathéodory[8],V,405 reported that "all were very surprised that according to that theory there are exceptional cases in which no solutions of the variational problem seem to exist." And he tried to find a simple example: project an open hemisphere S from its center into its tangent plane at the South Pole and find a curve of given length L on S with given endpoints A and B, of spherical distance $d_S(A,B) < L$, whose image is as long as possible or as short as possible. He conjectured that the image must consist of two segments that make a corner, and he was able to calculate the E-function and solve the problem. We emphasize that this example is typical because in an isoperimetric problem, in which an integral (1.1) is to be extremized, whereas another integral

$$\int_{\bar{x}_0}^{\bar{x}_1} G(x,\ y,\ y')\ dx$$

is to have a given value, the above exceptional case occurs if the Euler equation (1.2) has the same solutions as that for G, and this case occurs in the example because geodesic arcs on the sphere are mapped onto geodesic arcs (straight segments) in the plane.

A few weeks later, Carathéodory had constructed the framework of his doctoral thesis "On discontinuous solutions in the calculus of variations"[8],I,3−79, where the unfortunate term "*discontinuous solution*" means extremal with corners. Not much on such solutions had been done before. The earliest publication on them appeared in 1871, written by Todhunter. Serious shortcomings of it were corrected in 1876 by Erdmann, in a paper that also contained the Weierstrass-Erdmann corner conditions[6],37−38, that had been used first by Weierstrass eleven years earlier in his lectures. A reason for the lack of further work may have been that the situation in the case of the catenoid[8],I,3 had been regarded as typical – which it is not. Hence when Carathéodory started his work, the corner conditions were known, but a further theory of discontinuous solutions was missing. Carathéodory created such a theory, resulting from his own ideas under the influence of the stimulating Göttingen atmosphere, but without having a supervisor

of his thesis in the usual sense. Being too shy to approach Hilbert or Klein, in 1904 he presented his thesis to Minkowski, with whom he had closer contact.

In his thesis, Carathéodory treated (1.1), but in parametric form, writing

$$J = \int_{t_0}^{t_1} F(x, y, x', y')\, dt, \qquad (t_0 < t_1) \tag{3.1}$$

a form which also Weierstrass had found very advantageous in most of his work, and he was able to develop his theory in a detailed and lucid way. He obtained necessary and sufficient conditions for the occurrence of broken extremals, proved the existence of a field of broken extremals in the neighborhood of a cusp, and developed the theory of conjugate points. He also considered isoperimetric problems; here the above example appeared in full[8],I,57-69.

After the Heidelberg International Congress of Mathematicians (1904), Carathéodory came into closer contact with Klein and Hilbert, who insisted that he should immediately write his "Habilitationsschrift". This work entitled (in German) "On strong maxima and minima for simple integrals" was completed already in 1905 (and published in 1906 [l.c., 80-142]).

It concerned *strong extremals* (functions that give a strong extremum of single integrals) and generalizations of Hilbert's results, which had already been extended in 1904 from Problem (a) [above] to more general problems, by Bolza, who recognized that the success of Hilbert's method depends on the two facts that

(1) the problem is "*definite*", that is, the integrand must have the same sign on each curve element of the domain Ω, and

(2) each point P in Ω must be "*regular*", that is, P has a neighborhood which can be covered simply and without gaps by *strong* extremals beginning at P.

In Chap. I, Carathéodory showed that (2) is essentially a consequence of (1) if one also admits discontinuous solutions; here, "essentially" means that there may be exceptional points (along certain curves only) that are not regular. He then showed that Hilbert's method can be readily extended to positive definite problems, whereas the regularity of all points in Ω alone is not sufficient for such

an extension, as is proved by a simple discussion of the "isoperimetric integral" [l.c., 141-142]

$$J = \int_{t_0}^{t_1} (x'y + [x'^2 + y'^2]^{\frac{1}{2}})\, dt. \tag{3.2}$$

Beginning with the thesis, Carathéodory's work had profound influence on the general long-term trend that concerns efforts of achieving increasing generality by weakening differentiability assumptions. In an appendix to his thesis, Carathéodory proposed a generalized Bernoulli method for positive definite problems and published his new approach, later known as *Carathéodory's method*, in 1908[8],*I*,170−187. This method has the advantage of reduced differentiability assumptions on F in (3.1). He assumed (1.)–(3.):

(1.) F is positive homogeneous in x', y' of first degree,

$$F(x, y, kx', ky') = kF(x, y, x', y'), \qquad k > 0. \tag{3.3}$$

(2.) For fixed x, y the curve $F(x, y, \xi, \eta) = 1$ is a strictly convex oval in the $\xi\eta$-plane containing the origin in its interior, so that F is a positive definite function of its last two variables.

(3.) $F_{x'}, F_{y'}$ exist and are continuous.

Hence the Euler equations do not make sense in this case. He called a family of curves $\phi(x, y) = t$ *geodesically parallel* if the equation

$$F_{x'} = \phi_x, \qquad F_{y'} = \phi_y \tag{3.4}$$

give the same angle $\theta = \theta(x, y)$. Note that by (1.), the left-hand sides $F_{x'}$ and $F_{y'}$ are homogeneous in x', y' of degree zero; thus they depend only on θ determined by

$$x' = (x'^2 + y'^2)^{\frac{1}{2}} \cos\theta, \qquad y' = (x'^2 + y'^2)^{\frac{1}{2}} \sin\theta.$$

Then for a curve C intersecting such a family and at each point (x, y) making the angle $\theta(x, y)$ with the positive x-axis, we have by differentiating (3.3) with respect to k, setting $k = 1$, and using (3.4),

$$F_{x'}x' + F_{y'}y' = F(x, y, x', y') = \phi_x x' + \phi_y y' = \frac{d\phi}{dt} = 1,$$

whereas $F(x, y, x', y') > 1$ along every other curve, as can be shown. Hence C minimizes the integral J in (3.1).

The nature of this method becomes most perspicuous in dynamical (or optical) applications, where it is related to Hamilton's ideas. Then F in (3.1) is the Lagrange function L of the dynamical system, and Carathéodory introduced an arbitrary family of surfaces $S(t,\ x,\ y) = const$, imposing conditions that led to the Hamilton-Jacobi equation

$$\frac{\partial S}{\partial t} + H(t,\ x,\ y,\ \text{grad}\ S) = 0 \tag{3.5}$$

(H the Hamiltonian of the system), and defining S as a solution of (3.5). Then for any path,

$$S(t_1,\ x_1,\ y_1) - S(t_0,\ x_0,\ y_0) \geq \int_{t_0}^{t_1} L(t,\ x,\ y,\ x',\ y')\ dt$$

with equality for trajectories intersecting those surfaces (surfaces of constant least action, wave fronts in optics) in geodesic descent. The direction of the latter may not be unique, so that discontinuous solutions arise quite naturally.

It seems surprising that Hamilton, Jacobi, and other classical masters working also in partial differential equations never fully developed and exploited the relations between the calculus of variations and partial differential equations. The idea of making this relationship the basis of a new approach to the calculus of variations became the theme of Carathéodory's classic, which appeared in 1935 under the title "Variationsrechnung und partielle Differentialgleichungen erster Ordnung"[9], and in 1965, 1967, and 1982 in translation under the title "Calculus of Variations and Partial Differential Equations of the First Order", a masterpiece which, over fifty years after its first appearance, still provides "une présentation extrêment lucide, et, à beaucoup d'égards, étonnamment moderne." (René Thom,[29],580.) Carathéodory's approach entails substantial gain in simplicity and depth of insight as well as novelty of presentation, even in the discussion of the standard parts of the theory. The tx-space $\mathbf{R}^{n+1}, x = (x_1, \ldots, x_n)$, is used from the very beginning and throughout. *Extremals* are defined to be curves in $\mathbf{R}^n$ along which the integral (3.1) (with $x_1, \ldots, x_n$ instead of $x,\ y$) locally has

at least a weak extremum. There is no need for going into further details, but we can refer to a profound (English) Zentralblatt review [11 (1935), 356-357] by Tibor Radó, whose advice helped to improve the book (see[8],I,402). In the Preface, Carathéodory remarks that the book also emphasizes physical and differential geometric applications and incorporates Weierstrass's ideas, and as two other approaches he mentions "the variational calculus of Lagrange, which now forms a part of the tensor calculus [and], second, the theory of Tonelli, in which the more subtle relations of the minimum problem to set theory are developed."

Whereas Carathéodory's results were many-sided, his approach was uniform and based on his theory of geodesic fields. H. Rund (in[29],496–536) has shown that this theory is less restrictive than H. Weyl's field theory, which is more popular in physics, probably because of its greater calculational accessibility. Carathéodory's theory was generalized by Boerner [Math. Z. 46 (1940), 720-742] and by Le Page [Bull. Acad. Roy. Belg. (5) 27 (1941), 27-46], who used exterior differential forms as a natural tool, together with the generalized Stokes's theorem. Boerner later called this very elegant treatment a "Königsweg" to the calculus of variations. But there remains a substantial discrepancy between the necessary and the sufficient conditions obtained so far in this "generalized Carathéodory theory", for m-fold integrals depending on n functions to be varied independently. This theory thus presents various open questions for further study.

Other trends that result directly or indirectly from Carathéodory's ideas will be investigated in Secs. 4-7 in terms of important ongoing developments. To the latter also belongs *Morse theory (Calculus of variations in the large)*, which we have not included here because the leading ideas of the earlier development in the field, as created by Poincaré[30], G.D. Birkhoff, Morse, Lusternik, and Schnirelmann, can be seen from[21], [25], [26], [27] and recent activity from the excellent survey by R. Bott[7].

4. OPTIMAL CONTROL

Whereas the number of research articles in the calculus of variations proper has been decreasing over the years, the influence of the calculus of variations in other fields – optimal control, minimal surfaces, functional analysis, partial

differential equations – has been steadily increasing. In many of these applications, the ideas in the work by Carathéodory and extensions to *Lagrange problems* (problems with side conditions that result from differential equations) by Carathéodory, Bliss, Hestenes, and others had profound impacts during the initial and later stages of evolutions.

This is true for optimal control, which initiated from engineering problems short before 1960, has developed into a large field of its own, and is presently expanding into various directions, making it virtually impossible to survey even special portions of it. In view of this, as well as of the availability of comprehensive monographs (for instance[2], [33]), we shall attempt to provide a general understanding of this new area, which abunds of unsolved problems, by explaining some general concepts and leading ideas in the center of the development. It is interesting to note that, beginning in 1968, Mathematical Reviews officially recognized the close interrelation between optimal control and the calculus of variations by extending the headline of subject #49 to "Calculus of Variations and Optimal Control."

Many control problems arise in connection with nonlinear dynamical systems governed by a system of differential equations

$$y' = F(y, u), \qquad y(t_0) = y_0 \tag{4.1}$$

where $' = d/dt$, t is time, $y = [y_j]$ is an n-vector whose components are functions of t, and $u = [u_k]$ is an r-vector whose components are functions of t. For instance, when $n = 2$, then $y = y(t)$ may be the displacement of a linearly damped system [with (4.1) converted to a single second-order differential equation]

$$y'' + cy' + ky = u(t), \quad y(0) = y_0, \quad y'(0) = y_0'. \tag{4.1*}$$

The *control* $u(t)$ is the effect of a servomechanism to be designed so that the system be brought to its rest position $y = 0$, $y' = 0$ in minimum time, at some later time t_1. Since the voltage available is limited, we have a constraint of the form of an inequality

$$|u(t)| \leq U_0 \qquad (U_0 \text{ constant}). \tag{4.2*}$$

If $y_0 > 0$, $y_0' > 0$, then u should initially be directed in the negative y-direction with greatest magnitude, $u(t) = -U_0$. At some instant we should switch to $u(t) = +U_0$, to avoid overshooting. A control that takes only the two values $\pm U_0$ is called a *bang-bang control.* Engineers believed on intuitive grounds in the time-optimality of this type of control with properly chosen switching times. Later theories (by Bushaw 1952 [Annals of Math. Studies 41 (1958), 29-52], Bellman et al. [Quart. Appl. Math. 14 (1956), 11-18] and others) confirmed this under relatively general assumptions for problems in the case that (4.1) involved is linear. Here, when $u = [u_1, \ldots, u_r]^T$, instead of (4.2*) we have

$$|u_j(t)| \leq U_{0j}, \qquad j = 1, \ldots, r, \tag{4.2**}$$

and *bang-bang* means that u_j takes only the two values $\pm U_{0j}$.

Furthermore, the *bang-bang principle* asserts that if a system can be transferred to the origin by some control, time-optimal or not, then there is a bang-bang control that transfers the system to the origin in the same time.

Engineers like bang-bang controls because these controls need only "on-off" servomechanisms; hence they are technically simpler than general controls.

A frame for a general theory, as motivated by our discussion and the nature of practical control problems, is obtained as follows. Instead of the special constraints (4.2*) or (4.2**), one often has constraints

$$C_\ell(t, y(t), u(t)) \geq 0, \qquad \ell = 1, \ldots, m. \tag{4.2}$$

Also, instead of a single initial condition $(t_0, y_0) \in \mathbf{R}^{n+1}$, one considers $(t_0, y_0) \in T_0 \subset \mathbf{R}^{n+1}$. The given set T_0 is called an *initial set*, $u = u(t)$ in (4.1) is called a *control*, and $y = y(t)$ (describing the time evolution of the system) a *trajectory*. The control is supposed to transfer the system from an initial state $(t_0, y_0) \in T_0$ to a terminal state (t_1, y_1) in a given *terminal set* $T_1 \subset \mathbf{R}^{n+1}$. A control u is called *admissible* if there exists a corresponding trajectory [solution y of (4.1)] such that the constraints (4.2) are satisfied and the system is transferred from a $(t_0, y_o) \in T_0$ to a $(t_1, y_1) \in T_1$. This y is called an *admissible trajectory* and (y, u) an *admissible pair.*

In the above time-optimal control problem we minimize

$$t_1 - t_0 = \int_{t_0}^{t_1} 1 \, dt. \tag{4.3*}$$

Instead of the integrand 1, with a control problem we more generally associate a *performance function* $F_0 = F_0(t, y, u)$ and a *performance index* (or *payoff*)

$$J[y, u] = \int_{t_0}^{t_1} F_0(t, y(t), u(t)) \, dt. \tag{4.3}$$

The *optimal control problem* then is the problem of finding an admissible pair that minimizes $J[y, u]$.

Finally, to be able to apply to our problem the theory of the Lagrange problem of the calculus of variations, as developed by Carathéodory and others, we must convert the inequality constraints (4.2) to the form of differential equation constraints. This can be done by introducing a new variable $z = [z_\ell]$ defined by

$$z_\ell'^2 = C_\ell(t, y, u) \qquad \ell = 1, \ldots, m,$$

with C_ℓ as in (4.2), as was first shown by Berkovitz [J. Math. Anal. Appl. **3** (1961), 145-169].

5. MINIMAL SURFACES AND BERGMAN OPERATORS

As a first motivating example in his doctoral dissertation, Carathéodory mentioned a special minimal surface (see below). This is not merely by chance, but many ideas in the calculus of variations were, and are still being, sparked by minimal surfaces, a large and active field of present research that has been attractive for a long time, apart from its intrinsic beauty mainly because of its relations to complex analysis (Weierstrass formulas!), to physics (Plateau's problem, crystals) and to partial differential equations, where the nonlinearity makes the minimal surface equation a particularly worthwhile object of study.

We begin by recalling some basic facts that we shall need in our investigation.

A *minimal surface* $S \subset \mathbf{R}^3$ is a surface whose *mean curvature* $H = \frac{1}{2}(\kappa_1 + \kappa_2)$ is identically zero; here κ_1, κ_2 are the principal curvatures (see, for instance[19],[91]). If S is represented in the usual parametric form

$$r = r(v^1, v^2), \tag{5.1}$$

$(v^1, v^2) \in \Omega$, of class $C^2(\Omega)$, with $g_{\alpha\beta}$ and $b_{\alpha\beta}$ denoting the coefficients of the first and second fundamental forms, respectively, then

$$H = \frac{1}{2} b_{\alpha\beta} g^{\alpha\beta}, \tag{5.2}$$

where we use Einstein's summation convention, and $g_{\alpha\beta} g^{\beta\gamma} = \delta_\alpha^\gamma$, the Kronecker tensor. We also recall that the *Gauss curvature* of S is

$$K = \frac{b}{g}, \tag{5.3}$$

with the discriminants $g = g_{11} g_{22} - g_{12}^2$ and $b = b_{11} b_{22} - b_{12}^2$.

Instead of "portion of a surface" we shall briefly say "surface"; this will cause no misunderstandings.

If $C \subset \mathbf{R}^3$ is a simple (rectifiable) closed curve and S minimizes the area functional J in the class of all C^2-surfaces bounded by C and homeomorphic to the unit disk, then S is necessarily a minimal surface. This explains the name as well as the connection with the calculus of variations. If S is represented by (5.1), then

$$J[r(v^1, v^2)] = \iint_\Omega \sqrt{g}\, dv^1 dv^2. \tag{5.4}$$

The corresponding Euler equation, called the *minimal surface equation*, is

$$b_{\alpha\beta} g^{\alpha\beta} = 0 \qquad \text{[cf. (5.2)]}. \tag{5.5}$$

A basic characterization of minimal surfaces by Bonnet and Christoffel is given in

Theorem 5.1. A surface S, not a sphere, is a minimal surface if and only if its *spherical mapping*

$$\begin{gathered} S \to S_0 \\ r(v^1, v^2) \mapsto n(v^1, v^2), \qquad n = g^{-\frac{1}{2}} r_{v^1} \times r_{v^2} \end{gathered} \tag{5.6}$$

into the unit sphere S_0 is conformal.

For a Cartesian representation

$$S : z = z(x, y) \tag{5.7}$$

the minimal surface equation takes the form

$$\left(\frac{z_x}{W}\right)_x + \left(\frac{z_y}{W}\right)_y = 0, \qquad W^2 = 1 + z_x^2 + z_y^2 \tag{5.8*}$$

already given by Lagrange in 1760, or

$$(1 + z_y^2)z_{xx} - 2z_x z_y z_{xy} + (1 + z_x^2)z_{yy} = 0. \tag{5.8}$$

Carathéodory began his doctoral dissertation by pointing to the earliest highlight, Euler's discovery (published in his 1744 book) of the minimal surface of revolution, the *catenoid*, whose meridian is a *catenary* (the curve of a hanging chain or cable). In 1776 followed Meusnier's discovery that a right *helicoid* (a "staircase surface") is a minimal surface. It was also Meusnier who showed that the left-hand side of (5.8) divided by $2W^2$ is geometrically the mean curvature. *Scherk's minimal surfaces*

$$z = \ln(\cos y \sec x) \quad \text{and} \quad z = \arcsin(\sinh x \sinh y) \tag{5.9}$$

(and three others) appeared in 1831 and 1835. This, together with the great general interest in minimal surfaces in view of the calculus of variations, physics (Plateau's problem) and complex analysis (Weierstrass's formulas, 1866), resulted in a great number of important discoveries in the field during the second half of the nineteenth century, which Nitsche[28],[4] calls "the Golden Age in the theory of minimal surfaces". For details, we can refer to Nitsche's book[28] or to Darboux[10].

Carathéodory mentioned minimal surfaces as an example of a problem in which extremals may have discontinuous tangent directions where Euler's equation or the integrand cease to satisfy conditions imposed. For instance, the meridian of a surface of revolution of minimum area may degenerate into two segments perpendicular to the axis and a joining segment of the axis itself.

We consider next an application of *Bergman operators* in the theory of minimal surfaces.

For a second-order partial differential equation

$$Lu = u_{zz^*} + b(z,\ z^*)u_{z^*} + c(z,\ z^*)u = 0, \tag{5.10}$$

from which we have eliminated one of the two first partial derivatives in the usual fashion, a Bergman integral operator T is defined by

$$u(z,z^*) = Tf(z,z^*) = \int_{-1}^{1} E(z,z^*,t) f\left(\frac{1}{2}z(1-t^2)\right)(1-t^2)^{-\frac{1}{2}}\, dt. \tag{5.11}$$

For u to be a solution of (5.10), the kernel E must satisfy the *kernel equation*

$$(1-t^2)E_{z^*t} - \frac{1}{t}E_{z^*} + 2ztLE = 0. \tag{5.12}$$

T is called a *class P operator* if its kernel E is a polynomial in t, with coefficients depending on z, z^*. The class of equations admitting such operators ("*equations of class P*") is rather large and can be characterized in a number of different ways; see[18],[45–107]. Thus a class P kernel is of the form

$$E(z,z^*,t) = \sum_{\mu=0}^{m} q_\mu(z,z^*)t^{2\mu}. \tag{5.13}$$

We have omitted odd-power terms in t, without restriction, since these would not contribute to the integral (5.11) representing u.

The equation

$$Lu = u_{zz^*} + \frac{2}{\omega^2}u = 0, \qquad \omega = 1 + zz^* = 1 + x^2 + y^2 \tag{5.14}$$

is of class P and plays a role in connection with minimal surfaces, as we shall prove.

From (5.12) it follows that the simplest kernel for (5.14) is

$$E(z,z^*,t) = 1 - \frac{4zz^*}{\omega}t^2. \tag{5.15}$$

Now solutions obtained from a class P operator can be cast into integral-free form [l.c., 51]. For (5.15) this yields

$$u(z, z^*) = \tilde{f}'(z) - \frac{2z^*}{\omega}\tilde{f}(z). \tag{5.16}$$

The relation to minimal surfaces is now accomplished as follows. Let x, y be the Gaussian coordinates on the unit sphere S_0 (with center at the origin) obtained by inverse stereographic projection of $(x, y) \in \mathbf{R}^2$ into S_0. Then the invariant second-order differential operator associated with the third fundamental form of a surface S (the first fundamental form of the spherical image of S in S_0) for any S is defined by (see[28],[51])

$$\Delta^{\mathrm{III}} u = \frac{1}{4}\omega^2 \Delta u. \tag{5.17}$$

Now if S is a minimal surface, one has the Jacobi condition

$$\Delta^{\mathrm{III}} u + 2u = 0. \tag{5.18}$$

Here u is the Minkowski *support function* of S (the distance of the tangent planes of S from the origin). Since

$$z = x + iy, \quad z^* = x - iy, \quad \Delta u = 4u_{zz^*},$$

we see that u satisfies (5.14), as we wanted to prove.

From (5.17) we have the result that the general form of the support function [i.e., the general solution of (5.14)] is obtained without the use of the Weierstrass formulas for minimal surfaces, by using class P operators converted to integral-free form.

Important minimal surfaces correspond to simple Bergman associated functions $f(z)$, for instance, *Enneper's surface* to z^3 and the helicoid to $iz/2(1-\ln z)$.

For the Dirichlet problem for (5.14) on a disk of radius R, Agostinelli[1] proved the interesting result that there is a unique solution when $R < 1$, no solution in general when $R = 1$, and infinitely many solutions when $R > 1$. To this there corresponds the following for minimal surfaces (see[28],[95],[104]).

Theorem 5.2. If the spherical image I_S of a minimal surface S is contained in the interior of a hemisphere, then its area is a relative weak minimum compared to the area of neighboring surfaces with the same boundary. If I_S contains a hemisphere in its interior, then the area of S is not minimum compared to that of other surfaces with the same boundary.

The second statement follows by associating with (5.14) the eigenvalue problem

$$u_{zz^*} + \left(\frac{2}{\omega^2} + \lambda\right) u = 0 \text{ in } \Omega, \qquad u = 0 \text{ on } \partial\Omega, \tag{5.19}$$

noting that

$$u(z, z^*) = (1 - zz^*)/\omega \tag{5.20}$$

is an eigenfunction of (5.19) for the unit disk and $\lambda = 0$, and that $\lambda_{\min}(\Omega) < 0$ when Ω contains the unit disk in its interior.

It should be noted that equation (5.14) is a special case of

$$L_n u = u_{zz^*} + \frac{n(n+1)}{\omega^2} u = 0, \qquad \omega = 1 + zz^*, \tag{5.21}$$

which has been investigated extensively and is of class P for every $n \in \mathbf{N}$ (see[18]). The Dirichlet problem for (5.21) on a disk can be solved by using class P operators as follows.

Theorem 5.3. For the operator defined by (5.11), (5.13) the associated function corresponding to boundary values

$$U(\theta) = \sum_{\sigma=0}^{\infty} A_{n\sigma} e^{i\sigma\theta} \tag{5.22}$$

on the unit disk is

$$f(z) = \sum_{\sigma=0}^{\infty} a_{n\sigma} z^{\sigma} \tag{5.23a}$$

where

$$a_{n\sigma} = \frac{A_{n\sigma}}{H_{n\sigma}}, \quad H_{n\sigma} = \frac{\pi(2\sigma)!}{2^{3\sigma}\sigma!(\sigma+n)!} \prod_{\lambda=0}^{n-1} (\sigma - n + 1 + 2\lambda). \tag{5.23b}$$

Proof. For the present equation the Bergman representation (5.11) is

$$u(z,z^*)=\int_{-1}^{1}\sum_{\mu=0}^{n}(-4)^{\mu}\binom{n+\mu}{2\mu}\left(\frac{\eta}{\omega}\right)^{\mu}t^{2\mu}f\left(\frac{z}{2}(1-t^2)\right)(1-t^2)^{-\frac{1}{2}}\,dt$$

where $\eta=zz^*$ and $\omega=1+zz^*$. On $|z|=1$, thus,

$$\begin{aligned}U(\theta)&=\sum_{\mu=0}^{n}(-2)^{\mu}\binom{n+\mu}{2\mu}\int_{-1}^{1}t^{2\mu}f\left(\frac{e^{i\theta}}{2}(1-t^2)\right)(1-t^2)^{-\frac{1}{2}}\,dt\\&=\sum_{\sigma=0}^{\infty}a_{n\sigma}\sum_{\mu=0}^{n}(-2)^{\mu}\binom{n+\mu}{2\mu}\frac{e^{i\sigma\theta}}{2^{\sigma}}\int_{-1}^{1}t^{2\mu}(1-t^2)^{\sigma-\frac{1}{2}}\,dt.\end{aligned}$$

The integral is $B(\mu+\frac{1}{2},\ \sigma+\frac{1}{2})$. This gives

$$U(\theta)=\sum_{\sigma=0}^{\infty}A_{n\sigma}e^{i\sigma\theta},\qquad A_{n\sigma}=H_{n\sigma}a_{n\sigma}\tag{5.24a}$$

where

$$H_{n\sigma}=\frac{\Gamma(\sigma+\frac{1}{2})}{2^{\sigma}}\sum_{\mu=0}^{n}(-2)^{\mu}\binom{n+\mu}{2\mu}\frac{\Gamma(\mu+\frac{1}{2})}{(\mu+\sigma)!}\tag{5.24b}$$

From this, (5.23) follows. This completes the proof.

The form of the coefficient of (5.21) suggests to search for solutions that are functions of ω alone. We show that all such solutions can be characterized explicitly, even when the constant in the equation is left unrestricted.

Theorem 5.4. All C^2-solutions $u=u(\omega)$ of

$$u_{zz^*}+\frac{M}{\omega^2}u=0,\qquad M=k(k-1),\quad k\in\mathbf{R},\tag{5.25}$$

are of the form

$$u(\omega)=\omega^k V(\omega)\tag{5.26}$$

where $V(s)$ is a solution of the hypergeometric equation

$$s(1-s)V''+[2k-(2k+1)s]\,V'-k^2V=0.\tag{5.27}$$

[Thus $k=2$ for (5.14).]

Proof. This follows by substitution, using $zz^* = \omega - 1$ and observing that the indicial equation has the roots k and $1 - k$.

Equation (5.25) has a uniqueness property among all equations of a certain family, in the sense of

Corollary 5.5. Equation (5.25) is the only equation among

$$u_{zz^*} + \frac{M}{\omega^p} u = 0, \qquad p \in \mathbb{N}, \tag{5.28}$$

that has solutions of the form (5.26) with V a solution of the hypergeometric equation.

Remark. The parameters in (5.27) are $\alpha = \beta = k$, $\gamma = 2k$; thus 2, 2, 4 in the case of (5.14). For integer k – the case of a positive integer k is precisely that in which the equation is of class P (see before) – these parameters are integer, so that the equation is degenerate in the sense of the theory of the hypergeometric equation; hence it has a polynomial solution, as follows from that theory. More precisely, from Erdélyi[12],I,72(20), we have in our case

$$\alpha = m + 1 = k, \quad \beta = m + \ell + 1 = k, \quad \gamma = m + n + \ell + 2 = 2k;$$

thus $m = k - 1$, $\ell = 0$, $n = k - 1$. Hence one of Kummer's 24 solutions, namely,

$$(-s)^{-\beta} F\left(\beta + 1 - \gamma, \beta, \beta + 1 - \alpha, \frac{1}{s}\right)$$

listed as formula (13) on p. 105 in[12],I, gives as polynomial solution of degree $k - 1$ in $1/\omega$

$$u = \omega^k V(\omega) = \omega^k (-\omega)^{-k} F\left(1 - k, k, 1, \frac{1}{\omega}\right). \tag{5.29}$$

For instance, for $k = 2$ [equation (5.14)] we get from (5.29), except for an irrelevant constant factor, the solution

$$u = 1 - 2\omega^{-1} \tag{5.30}$$

which agrees with (5.16) when $\tilde{f}(z) = -z$.

These results are in harmony with those on class P operators, for instance, on equation (5.21), in which $n = k - 1$ and one gets a polynomial solution for the kernel when cast back into integral form, which has degree n in t^2.

Agostinelli also obtained an interesting relation to the biharmonic equation. More generally we have

Theorem 5.6. If u satisfies

$$u_{zz^*} + \frac{M}{\omega^2}u = 0, \qquad \omega = 1 + zz^*, \quad M \in \mathbf{R}, \tag{5.31}$$

then $U = \omega u$ satisfies

$$U_{zzz^*z^*} + \frac{M_2}{\omega^4}U = 0, \qquad M_2 = M(2 - M). \tag{5.32}$$

Thus $M = 2$ [equation (5.14)] leads "just by chance" to the biharmonic equation. A generalization of Theorem 5.6 will be published elsewhere.

6. IMPACTS ON FUNCTIONAL ANALYSIS

The two most important formative factors in the early evolution of functional analysis (a name coined by P. Lévy in 1922) were the impacts of the calculus of variations and of integral equations, the former from the very beginning in Volterra's work of 1887, and the latter since the appearance of Fredholm's theory in 1900-1903.

Functional analysis originated in Italy, and it is generally agreed to regard 1887 as its birth year, the year in which Volterra [Opere I, 294-328] published five notes on classes of functionals. Those papers were intended to generalize Riemann's methods of complex analysis, but Volterra modeled his methods after those in the calculus of variations. Near the beginning, he said: "If ... y depends on all values of a function $\phi(x)\ldots$ in $(A\ldots B)$, we write

$$y\Big|\Big[\overset{B}{\underset{A}{\phi}}(x)\Big]\Big| \quad \text{or simply} \quad y\big|[\phi(x)]\big|."$$

He assumed ϕ to be continuously differentiable on (A, B). He then defined a "variation"

$$\delta y = y|[\phi + \theta]| - y|[\phi]|$$

as well as a "derivative"

$$y'|[\phi(x), t]| = \lim_{\substack{n-m\to 0 \\ \max|\theta|\to 0}} \frac{\delta y}{\sigma}, \qquad \sigma = \int_m^n \theta(x)\,dx,$$

where θ has constant sign and $[m, n]$ contracts to a point t. This theory, created at a time when topological tools were not yet available, was ad hoc, and was later criticized for that (Dieudonné[11],[86]), but also used in further work (Hamilton & Nashed[14]; see also A. Weil [Oeuvres II, 532]).

A very substantial effect on nascent functional analysis resulted from the breakdown of the Dirichlet principle considered in Sec. 2, along with the remark that new methods of existence proofs for the Dirichlet problem were invented by Schwarz, Poincaré, and C. Neumann. Most important of these in view of functional analysis was the latter "method of the arithmetic mean" (1870), which led directly to integral equations and helped to spark the great interest in Fredholm's fundamental work of 1900-1903 as well as Hilbert's activity on integral equations resulting from it.

Hilbert's proof of the Dirichlet principle (Sec. 2) was still by means of classical analysis, but it is interesting to note an earlier, not quite successful attempt of a functional analytic proof, namely, by Arzelà in 1896, based on the Arzelà-Ascoli theorem (Monna[24],[108–113]).

A characteristic feature of the further evolution over the next four decades was a deeper and deeper intuitive functional analytic understanding of the central concept of the calculus of variations, the functional. This began with an important step forward, made in 1903 by Hadamard [Oeuvres I, 405-408], who initiated the idea of a general representation of a well-defined class of functionals by a general formula. He gave such a representation in a very important case, namely, for all bounded linear functionals U on $C[a,\ b]$ by the formula

$$U[f] = \lim_{m\to\infty} \int_a^b f(x)H_m(x)\,dx, \qquad H_m \in C[a,\ b]. \tag{6.1}$$

Hadamard was very much impressed by Volterra's use of variational techniques in his new developments of 1887. The significance of (6.1) and its novelty is perhaps often not fully appreciated because of the two shortcomings that the H_m are not uniquely determined by U and that (6.1) involves the limit of an integral rather than an integral, shortcomings that were removed by F. Riesz in 1909 when he published his representation of the same class of functionals in terms of

a Riemann-Stieltjes integral,

$$U[f] = \int_a^b f(x)\, d\alpha(x), \tag{6.2}$$

where $\alpha = \alpha(x)$ is of bounded variation on $[a,\ b]$ and for given U is unique (if we require $\alpha(0) = 0$ and right continuity of α).

It seems practically unknown that Carathéodory influenced Riesz by his works[8],*III*,54−77 of 1907 and[8],*III*,78−110 of 1911, as Riesz [Oeuvres 819, 823-826] acknowledged in 1911, in a paper extending his work on (6.2) as well as Carathéodory's of 1907. This made the latter the earliest trace of the famous *Bochner-Weil-Raikov theorem* of 1932/1940 in abstract harmonic analysis, stating that a function $u(x)$ on a locally compact abelian group G is "positive definite" if and only if there is a nonnegative measure μ on the character group $\widehat{G}$ of G such that

$$u(x) = \int_{\widehat{G}} \chi(x)\, d\mu(\chi).$$

The calculus of variations continued to exercise its influence on developing functional analysis during the first decade of our century, which included as a landmark the appearance of Fréchet's thesis (1906) and, at the end, the publication of Hadamard's book on the calculus of variations, because of which Carathéodory[8],*V*,309 called 1910 a "historical date in the calculus of variations".

In his thesis "Sur quelques points du Calcul fonctionnel" [Rend. Circ. Mat. Palermo **22**, 1-74], Fréchet defined metric (called *écart*), metric space (Hausdorff's later term, 1914), completeness, (sequential) compactness, and separability, in connection with infinite dimensional function spaces, and gave various concrete special spaces to illustrate his abstract concepts. Most remarkable is that he defined metric axiomatically and by the same axioms that we use today.

The variational ideas of Volterra influenced Fréchet, perhaps more indirectly through Hadamard than directly, because Hadamard was Fréchet's high school teacher and later his professor at the university and his personal friend, working in the calculus of variations himself and emphasizing that problems such

as Bernoulli's brachistochrone problem was "au coeur même de ce Calcul [fonctionnel]" [Oeuvres I, 438]. Also, Fréchet remained fully aware of the variational roots of functional analysis, later stating that "l'Analyse fonctionnelle tire son origine du Calcul des Variations" ["Les espaces abstraits" (1928), 4]. In one of his notes [Comptes Rendus Paris **139** (1904), 848-850], in which he generalized Weierstrass's theorem of the existence of minimum of a continuous function on a compact set, he mentioned the Dirichlet principle. In his thesis [p. 31] he extended Hilbert's generalized Dirichlet principle (cf. Sec. 2) to functionals on metric space. In connection with $C[a, b]$ he mentioned Weierstrass [p. 36], but one should realize that it was a big step forward from Weierstrass's neighborhoods (Sec. 1) to Fréchet's system of axioms of metric.

Mandelbrojt [Comptes Rendus Paris **277** (1973), Vie Académique 74] claimed that it was the lack of rigor in Volterra's method "de passage du fini à l'infini" that let Fréchet search for "une méthode directe, générale et très rigoureuse." It seems that this is at best only one factor motivating Fréchet's path-breaking work. A.E. Taylor [Archive Hist. Exact Sciences **27** (1982), 233-295] attempted to describe further factors, above all Hadamard's work, and furthermore the influence of Riemann, Weierstrass, Cantor, Hilbert, Borel, and Lebesgue, whose thesis had appeared in 1902. But Taylor actually hesitated to reach a conclusion of his sixty-page discussion by stating [p. 286] that

> "if Hadamard was fully frank in what he wrote about Fréchet for the Académie des Sciences, Fréchet's decision to go at things in a totally abstract way was his own decision ... There were, of course, certain trends lending to facilitate Fréchet's move into abstraction. The works of Riemann, Weierstrass and Hilbert contributed to that trend ... Abstraction was "in the air" at that time ..."

One could make these statements more precise by pointing to Dedekinds "Was sind und was sollen die Zahlen?" (1888) and Hilbert's "Grundlagen der Geometrie" (1899) as instrumental in paving the way to axiomatics. A factor completely missing in Taylor's paper is the influence of the ideas of Poincaré, in particular those in "La Science et l'hypothèse" (1902), Poincaré's most important book on the philosophy of science and mathematics, and "La valeur de Science" (1905).

The great effect of the latter work can be seen from F. Riesz's "Die Genesis des Raumbegriffs" [1906; published in German 1907; Oeuvres 110-161], a remarkable early axiomatic attempt into general topology, more promising than Fréchet's thesis, as far as non-metric properties are concerned[4],296–297. Unfortunately, that paper was published in an obscure journal and before the publication of Riesz's Oeuvres in 1960 became only partially known by a summary of a portion of it in 1908, as an abstract at the International Congress of Mathematicians.

Fréchet's work was very much encouraged by Hadamard, who in 1910 in his "Leçons sur le Calcul des Variations, recueillies par M. Fréchet" attempted to present the calculus of variations as a chapter of developing functional analysis. Along with Bolza's second edition entitled "Vorlesungen über Variationsrechnung" (1909), Carathéodory enthusiastically welcomed Hadamard's book in a detailed review[8],V,309–326. Defending the abstract approach, he pointed out (in another review, [p. 306]) that functional-analytic ideas were already present in lectures of Hilbert and Minkowski and in "the fundamental memoir of Hadamard of 1907 [Oeuvres II, 515-629] on boundary curves of plane domains which correspond to certain extremal properties of Green's functions" for the biharmonic equation. Although to Carathéodory, hopes for the calculus of variations looked perhaps greater than they actually were, the review shows that the new ideas were rapidly spreading, but details were not yet at the fingertips of the reader, not even in France, because we see that Carathéodory gave them in some detail in his review for the French Bulletin des Sciences mathématiques.

7. SOBOLEV SPACES IN THE CALCULUS OF VARIATIONS

On p. 316 of the review just discussed, Carathéodory mentioned that Hadamard used (and generalized) an ingenious idea of 1879 by du Bois-Reymond [Math. Ann. 15, 289-314] by which differentiability assumptions in connection with extremals can be reduced. Du Bois-Reymond proved that if $y = y(x) \in C^1([x_0, x_1])$ is an extremal of the functional $J[y]$ given by (1.1) whose integrand is a C^1-function, then $F_{y'}$ is differentiable and the Euler-Lagrange equation (1.2) makes sense. This is accomplished as follows. For y an extremal and any C^1-

function η, in (1.3) we have (1.7):

$$\int_{x_0}^{x_1} (F_y \eta + F_{y'} \eta') \, dx = 0. \tag{7.1}$$

Integration by parts to get rid of η' would give (1.2) because of $\eta(x_0) = \eta(x_1) = 0$ by assumption. But we can also integrate by parts so that η' is retained as a factor of the integrand of the resulting equation

$$\int_{x_0}^{x_1} h(x) \eta'(x) \, dx = 0,$$

where

$$h(x) = -\int_{x_0}^{x_1} F_y(t, y(t), y'(t)) \, dt + F_{y'}(x, y(x), y'(x)).$$

From the above assumption it follows that h is continuous, and du Bois-Reymond proved that h is constant, so that its derivative must exist. Hence the other integration by parts becomes permissible and leads to the Euler-Lagrange equation.

The functional-analytic significance of this work lies in the fact that this is an early appearance of the idea of weak solution, and in the proof, du Bois-Reymond used *test functions* (C^∞-functions with compact support), probably for the first time.

It is known that a more pressing need for generalized solutions had arisen much earlier in connection with the wave equation

$$u_{tt} = c^2 u_{xx}$$

where d'Alembert's solution of 1746,

$$u(x,t) = f(x + ct) + g(x - ct)$$

defines a "generalized solution" when f and g are no longer C^2-functions in the domain considered. Quite generally, whereas the initial phase of functional analysis was closely related to the calculus of variations, later other factors took the lead in the development, namely, integral equations beginning around 1900, ideas from algebra beginning around 1925, quantum mechanics in 1925, general topology about ten years later, and partial differential equations at about the

same time. Important to us is the progress on functionals during that period, by F. Riesz on Hilbert space in 1907 and 1934-1935, on $C[a,\ b]$ in 1909, and on $L^p[a,\ b]$ in 1910, and then by Hahn (1927) and Banach (1929) in connection with the Hahn-Banach theorem; and these results, basic in themselves, also reflect the growing intuitive understanding of functionals. We follow this process by dissecting ideas that led to Sobolev spaces and distributions, and afterwards show how it relates to the calculus of variations.

Preceded by some notes between 1933 and 1935, in 1936, in the hands of S.L. Sobolev [Mat. Sbornik (2) **1**, 39-72], functionals became the basic instrument in developing a systematic theory of generalized, or weak, solutions of initial and boundary value problems for linear partial differential equations, as follows. First, Sobolev defined the space Φ_s of functions $\phi \in C_0^s(\mathbf{R}^n)$ with compact support, with convergence $\phi_n \overset{s}{\to} \phi$ of $\{\phi_n\} \subset \Phi_s$ to mean convergence of the sequence $\{\phi_n\}$ to ϕ uniformly on $\mathbf{R}^n$ together with every sequence of derivatives to order s, the supports of all ϕ_n lying in a certain bounded domain. Then he defined the space Z_s of all linear functionals ρ on Φ_s continuous in the topology just defined, that is,

$$\langle \phi_n, \rho \rangle \to \langle \phi, \rho \rangle \quad \text{when} \quad \phi_n \overset{s}{\to} \phi; \tag{7.2}$$

here $\langle \phi, \rho \rangle$ is the value of ρ at ϕ. Clearly, every integrable function ρ generates such a functional according to

$$\langle \phi, \rho \rangle = \int \phi(x)\rho(x)\, dx,$$

but Z_s also contains elements not generated in this way.

Now if L is a linear partial differential operator with sufficiently smooth coefficients on a domain $\Omega \subset \mathbf{R}^n$ and L^* is its adjoint, then for a (classical) solution u of $Lu = 0$, multiplication by a $\phi \in \Phi_{s+k}$ (k the order of L) and integration gives

$$\langle Lu, \phi \rangle = 0$$

and Green's formula produces, because of the compactness of support,

$$\langle Lu, \phi \rangle = \langle u, L^*\phi \rangle.$$

Now for a $\rho \in Z_s$ the functional $\langle \rho, L^* \phi \rangle$ makes sense by what has just been said, and Sobolev called this generalized function ρ a *generalized solution* of $Lu = 0$ if

$$\langle \rho, L^* \phi \rangle = 0 \quad \text{for every} \quad \phi \in \Phi_{s+k}. \tag{7.3}$$

In his paper of 1936, Sobolev developed and applied these ideas to the existence and uniqueness of solutions of the Cauchy problem for a second-order linear differential equation.

This approach to Cauchy problems (as well as to Dirichlet problems for elliptic equations) motivated the important practical question for conditions under which a generalized solution is "more regular" than it follows from its definition, in particular for conditions under which it is (a.e. equal to) a classical solution. Answers to this and other questions were provided in 1938 by Sobolev [Mat. Sbornik (2) 4, 471-497] in his theory of Sobolev spaces and their relations to each other as expressed by the famous Sobolev embedding theorems. He first defined the *generalized derivative* (or distribution derivative)

$$D^\alpha u = \partial^{|\alpha|} u / \partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}, \qquad |\alpha| = \alpha_1 + \dots + \alpha_n,$$

of a real-valued function u, locally integrable on a domain $\Omega \subset \mathbf{R}^n$, to be the function v such that for every "test function" $\phi \in C_0^{|\alpha|}(\Omega)$, the zero meaning compact support in Ω,

$$\int_\Omega u D^\alpha \phi \, dx = (-1)^{|\alpha|} \int_\Omega v \phi \, dx, \tag{7.4}$$

a formula motivated by integration by parts (Green's theorem).

He then defined $W_p^\ell(\Omega)$ to be the Banach space of real-valued functions u whose generalized derivatives $D^\alpha u$, $|\alpha| \le \ell$, exist on Ω and are in $L_p(\Omega)$ with respect to Lebesgue measure on Ω, $p \ge 1$, with norm defined, for instance, by

$$\|u\|_{W_p^\ell} = \left[\sum_{0 \le |\alpha| \le \ell} \int_\Omega |D^\alpha u(x)|^p \, dx \right]^{1/p}. \tag{7.5}$$

Thus $W_p^0(\Omega) = L_p(\Omega)$, and W_2^ℓ is a Hilbert space. Other norms are possible, and Sobolev found general criteria for the equivalence of different norms on W_p^ℓ.

His most important discovery in this theory was his so-called *embedding theorems*, which give a special ordering of the spaces so that one space lies entirely in another (and the identity mapping of each function onto itself, regarded as an element of the "larger" space, is continuous, by the definition of "embedding"). Specifically, if $\ell p > n$, then $u \in W_p^\ell(\Omega)$ implies that u is continuous on Ω. More generally, if $\ell p > n + kp$, then

$$W_p^\ell(\Omega) \subset C^k(\Omega).$$

This embedding results from "Sobolev inequalities" between the norms of one and the same function when regarded as an element of different spaces. These inequalities contain special earlier integral inequalities by Poincaré, F. Riesz, Hardy-Littlewood, and others.

Whereas fully developed basic theories of Sobolev spaces and distributions by Sobolev in 1936-1938 and L. Schwartz in 1945 resulted from partial differential equations, the Heaviside calculus, and earlier work on Fourier transforms (by Plancherel, Wiener, and Bochner), it seems little known that the first traces of Sobolev spaces appeared in connection with the calculus of variations, at the time when Carathéodory had just completed his Habilitationsschrift. Indeed, in addition to Hadamard's functional-analytic approach to variational problems, culminating in his book of 1910, to which Carathéodory devoted a long review article (see Sec. 6), in 1906, B. Levi opened up another avenue to the calculus of variations that combined functional-analytic aspects with ideas of the "modern" (Borel-Lebesgue) theory of real functions. We recall that classically, admissible functions for (1.1) were assumed to be at least C^1, and that it was Carathéodory who in 1904 took the first step in extending $J[y]$ to a more satisfactory domain and developed a corresponding theory that included admissible functions $y(x)$ with bounded and piecewise continuous derivative $y'(x)$. As a next major step, in 1906 and 1907, in papers on the Dirichlet problem, B. Levi [Rend. Circ. Mat. Palermo **22**, 293-359] and Fubini [l.c., **23**, 58-84] used the Lebesgue integral and continuous functions of two variables which are absolutely continuous in each variable for almost all values of the other and have partial derivatives in L_2; these functions are elements of W_2^2. Essentially the same functions were employed by

Evans in 1920 [Rice Institute Pamphlet **7**, 252-329] in his generalized potential theory. In 1926, Tonelli [Atti Reale Accad. Lincei **6**, 633-638] used functions similar to those of Levi's in his work on surface area, with partial derivatives in L_1 (instead of L_2). A systematic application of Hilbert-space theory to functions those as Levi's was suggested in 1933 by Nikodým [Fund. Math. **21**, 129-150] and others.

The most remarkable link in the chain of steps of extending the generality of admissible functions was taken in 1940 by J.W. Calkin and C.B. Morrey, Jr. [Duke Math. J. **6**, 170-215] by introducing new function spaces, in order to get a more satisfactory existence theory of weak solutions. These new spaces can now be identified with *Sobolev spaces* $W_p^t(\Omega)$, which have thus found another important field of application, in addition to partial differential equations; in particular, this applies to $W_p^1(\Omega)$.

Furthermore, in 1952, Morrey, Jr. [Pacific J. Math. **2**, 24-53] related sequential lower semicontinuity of

$$J[y] = \int_\Omega F(Dy)\, dx, \qquad Dy = \left[\frac{\partial y_i}{\partial x_\alpha}\right]_{n\times N} \tag{7.6}$$

on an appropriate Sobolev space to quasiconvexity of F. Here, $\Omega \subset \mathbf{R}^n$ is open, bounded, and smooth, $y : \Omega \to \mathbf{R}^N$ is sufficiently regular, and *quasiconvexity* means that

$$F(A) \le \frac{1}{|G|}\int_G F(A + D\phi)\, dx \tag{7.7}$$

for all open $G \subset \mathbf{R}^n$, all matrices $A_{n\times N} \in M^{n\times N}$, and all $\phi \in C^1(G;\mathbf{R}^N)$ that are identically zero on ∂G. This result is basic for the existence theory when $n,\ N > 1$. To see this, we state a relatively recent result (Acerbi & Fusco [Arch. Rat. Mech. Anal. **86** (1984), 125-145]).

Theorem 7.1. Let $F : M^{n\times N} \to \mathbf{R}$ be continuous and let

$$0 \le F(B) \le C(1 + |B|^q) \quad \text{with } C > 0 \text{ and } 1 \le q < \infty \text{ given.}$$

Then $J[y]$ in (7.6) is weakly sequentially lower semicontinuous on Sobolev space $W_q^1(\Omega;\mathbf{R}^N)$ if and only if F is quasiconvex.

Note that in the scalar case $N = 1$, quasiconvexity is equivalent to convexity with respect to y', whereas for $N > 1$ it is weaker than convexity.

In conclusion it follows that the direct methods initiated by Hilbert in 1900 and popularized since 1911 in papers and books by Tonelli, have become the main tool to treat the problem of existence of minima, and the use of Sobolev spaces W_p^1 is related to the need of working in a function space with a sufficiently weak topology, so that minimizing sequences do converge. Now existence theorems for generalized solutions in themselves are of secondary interest, and conditions on C^1- or C^2-regularity are an important aspect of the whole theory, just as in the theory of partial differential equations. Briefly, a weak existence theory should be supplemented by a *regularity theory*. Such a theory was given by Morrey in 1940 and in a simplified form in 1960 [Ann. Scuola Norm. Pisa (III) 4 (1960), 1-16]. An intuitive reason for the suitability of Sobolev spaces in this respect results from the Sobolev embedding theorems stated before.

In regularity theory, the latest developments concern efforts toward direct approaches to regularity, by various authors. It would be impossible to present details here, but an impression of the present (still rather imperfect) state can be obtained, for instance, from a paper by Giaquinta [Proc. Int. Congr. Math. Berkeley 1986, II, 1072-1083] and its references.

From the stages of the developments investigated we may gain the overall impression that the observable increase in generality and abstraction in the calculus of variations resulted mainly for reasons of intrinsic necessities, rather than from external factors, and it seems most remarkable that also in this respect Carathéodory's work in the field, which initiated and directed much of the evolution over several decades, was typical and trend-setting.

REFERENCES

1. Agostinelli, C., Risoluzione per un campo circolare o sferico di un problema più generale di quello di Dirichlet. *Atti Accad. Sci. Torino, Cl. Sci. Fis. Mat. Natur.* **72** (1936-1937), 317-328.

2. Berkovitz, L.D., *Optimal Control Theory.* New York: Springer, 1974.

3. Birkhoff, G.D., *Dynamical Systems.* New York: American Mathematical Society, 1927.

4. Birkhoff, G., & E. Kreyszig, The establishment of functional analysis. *Hist. Math.* **11** (1984), 258-321.

5. Biermann, K.-R., *Die Mathematik und ihre Dozenten an der Universität Berlin 1810-1920.* Berlin: Akademie-Verlag, 1973.

6. Bolza, O., *Lectures on the Calculus of Variations.* New York: Chelsea, n.d. (Preface dated 1904.)

7. Bott, R., Marston Morse and his mathematical works. *Bull. Amer. Math. Soc.* (N.S). **3**, (1980), 907-950.

8. Carathéodory, C., *Gesammelte mathematische Schriften.* 5 vols. München: Beck, 1954.

9. Carathéodory, C., *Calculus of Variations and Partial Differential Equations of the First Order.* New York: Chelsea, 1982. (Original German edition Berlin, 1935.)

10. Darboux, G., *Leçons sur la théorie générale des surfaces.* Première partie. Bronx, NY: Chelsea, 1972. (Original 2nd French edition Paris, 1914.)

11. Dieudonné, J., *History of Functional Analysis.* Amsterdam: North-Holland, 1981.

12. Erdélyi, A., *Higher Transcendental Functions.* 3 vols. New York: McGraw-Hill, 1953-1955.

13. Funk, P., *Variationsrechnung und ihre Anwendung in Physik und Technik.* 2. Aufl. Berlin: Springer, 1970.

14. Hamilton, E.P., & M.Z. Nashed, Global and local variational derivatives and integral representations of Gâteaux differentials. *J. Funct. Anal.* **49** (1982), 128-144.

15. Hilbert, D., *Gesammelte Abhandlungen.* **3** vols. New York: Chelsea, 1981, 1965. (Original German edition 1932-1935.)

16. Jacobi, C.G.J., *Gesammelte Werke.* 8 vols. New York: Chelsea, 1969. (Original edition Berlin, 1866-1891.)

17. Kneser, A., *Lehrbuch der Variationsrechnung.* Braunschweig: Vieweg, 1900.

18. Kracht, M., & E. Kreyszig, *Methods of Complex Analysis in Partial Differential Equations with Applications.* New York: Wiley, 1988.

19. Kreyszig, E., *Introduction to Differential Geometry and Riemannian Geometry.* Toronto: Toronto Press, 1975.

20. Lebesgue, H., *Oeuvres scientifiques.* 5 vols. Geneva: Institut de mathématiques, Université de Genève, 1972-1973.

21. Lusternik, L., & L. Schnirelmann, *Topological Methods in Variational Problems.* Moscow: Gozidat, 1930. (In Russian; French translation Paris: Hermann, 1934.)

22. Lützen, J., *The Prehistory of the Theory of Distributions.* New York: Springer, 1982.

23. Migotti, A., *Variationsrechnung. Vorlesungen gehalten im Somm. Sem. 1882 v. Prof. Weierstrass.* Unpublished. (In the possession of the second author.)

24. Monna, A.F., *Dirichlet's Principle.* Utrecht: Oostoek, Scheltema & Holkema, 1975.

25. Morse, M., *The Calculus of Variations in the Large.* Providence, RI: American Mathematical Society, 1934.

26. Morse, M., *Functional Topology and Abstract Variational Theory.* Paris: Gauthier-Villars, 1939.

27. Morse, M., *Variational Analysis.* New York: Wiley, 1973.

28. Nitsche, J.C.C., *Vorlesungen über Minimalflächen.* Berlin: Springer 1975.

29. Panayotopoulos, A. (ed.), *C. Carathéodory: International Symposium, Athens, September 1973, Proceedings.* Athens: Greek Mathematical Society, 1974.

30. Poincaré, H., Sur les lignes géodésiques des surfaces convexes. *Trans. Amer. Math. Soc.* **6** (1905), 237-274.

31. Weierstrass, K., Mathematische Werke. 7 vols. Hildesheim: Olms, n.d. (Original edition Berlin & Leipzig, 1894-1927.)

32. Wussing, H. (ed.), *Die Hilbertschen Probleme.* Leipzig: Akademische Verlagsgesellschaft, 1971. (Translated from the Russian edition, Moscow: Nauka, 1969.)

33. Young, L.C., *Lectures on the Calculus of Variations and Optimal Control Theory.* Philadelphia: Saunders, 1969.

M. Kracht
Mathematisches Institut
Universitaet Duesseldorf
D-4 Duesseldorf 1
Universitaetsstr. 1
FED. REP. GERMANY

E. Kreyszig
Department of Mathematics
Carleton University
Ottawa, K1S 5B6
CANADA

Constantin Carathéodory: An International Tribute (pp. 629–635)
edited by Th. M. Rassias

AN EXISTENCE THEOREM FOR STRONGLY NONLINEAR EQUATIONS

D. Kravvaritis

1. Introduction

Let G be an open bounded subset of $\mathbb{R}^n$ such that the Sobolev Imbedding Theorem holds on G. We consider a semilinear partial differential equation of order $2m$ of the form

$$A(u) + B(u) = f \tag{1}$$

where

$$A(u) = \sum_{|\alpha|,|\beta| \leq m} (-1)^{|\alpha|} D^{\alpha} a_{\alpha\beta}(x) D^{\beta} u$$

is a linear differential operator and

$$B(u) = \sum_{|\gamma| \leq m-1} (-1)^{|\gamma|} D^{\gamma} B_{\gamma}(x, u, \dots, D^{m-1} u)$$

is a term of lower order which is strongly nonlinear in the sense that no growth restriction is imposed on the coefficient functions B_{γ}. In this paper we are concerned with the existence of solutions for equations of the form (1) on a closed subspace V of the Sobolev space $W^{m,p}(G)$, where $1 < p < +\infty$ and $p \neq 2$. Our result is based on an abstract existence theorem of Browder [1] for a class of mappings of monotone type which are not defined everywhere.

The case $p = 2$ was treated by Hess ([2], [3]). Boundary value problems for strongly nonlinear elliptic equations have been studied by many authors (cf. [1], [4], [5] and their references).

2. Notations

Let G be an open bounded subset of the Euclidean space $\mathbb{R}^n$, such that the Sobolev Imbedding Theorem holds on G. The points of G will be denoted by $x = (x_1, x_2, \dots, x_n)$.

If $\alpha = (\alpha_1, \dots, \alpha_n)$ is a multi-index of non-negative integers and $|\alpha| = \sum \alpha_i$, then D^α denotes the differential operator

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}.$$

Let $W^{m,p}(G)$ be the Sobolev space of real valued functions u defined on G whose distributional derivatives of order $\leq m$ belong to $L^p(G)$. The norm on $W^{m,p}(G)$ is

$$\|u\|_{m,p}^p = \sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^p}^p.$$

The expression $W_0^{m,p}(G)$ will denote the closure of testing functions in $W^{m,p}(G)$.

Let N be the number of multi-indices γ with $|\gamma| \leq m-1$. Then for each $\eta \in \mathbb{R}^N$ we write $\eta = \{\eta_\gamma : |\gamma| \leq m-1\}$ and $\eta(u)(x) = \{D^\gamma u(x) : |\gamma| \leq m-1\}$.

If V is a reflexive Banach space and V^* its dual, then (u^*, u) denotes the duality pairing between elements $u^* \in V^*$ and $u \in V$. The symbols "$\to$" and "$\rightharpoonup$" mean strong and weak convergence,respectively.

3. Statement of the Result

First, we give the following definition.

Definition 1. Let V be a Banach space, V' a dense linear subspace of V and T an operator of a subset $D(T)$ of V into V^*. T is called of type (M) with respect to V' if the following conditions hold:

(i) $V' \subset D(T)$ and for each finite-dimensional subspace F of $V', T : F \to V^*$ is demicontinuous.

(ii) If $\{u_n\}$ is a sequence in $V', u \in V$ and $w \in V^*$ such that $u_n \to u$ in $V, (T(u_n), v) \to (w, v)$ for all $v \in V'$ and $\limsup(T(u_n), u_n) \leq (w, u)$, we have $u \in D(T)$ and $T(u) = w$.

Our result is based on the following abstract existence theorem given by Browder in [1, Theorem 7].

Theorem 1. Let V be a reflexive Banach space, V_0 a separable dense linear subspace of V and T an operator from $D(T)$ into V^* such that for each dense subspace V' of V_0, T is of type (M) with respect to V'. Suppose that

$$(T(u), u) \cdot \|u\|_V^{-1} \to +\infty \quad \text{as} \quad \|u\|_V \to \infty \quad (u \in V_0),$$

then $R(T) = V^*$.

We impose the following conditions upon the linear part.

(A1) $a_{\alpha\beta} \in L^\infty(G)$ for all $|\alpha|, |\beta| \leq m$

(A2) $\sum_{|\alpha|,|\beta| \leq m} a_{\alpha\beta}(x)\xi_\alpha\xi_\beta \geq 0.$

The assumptions which we make upon the strongly nonlinear perturbing term $B(u)$ are:

(B1) For each γ with $|\gamma| \leq m-1, B_\gamma(x, \eta)$ is a function from $G \times \mathbb{R}^N$ to $\mathbb{R}$ satisfying the Caratheodory condition: $B_\gamma(x, \eta)$ is measurable in x for each fixed $\eta \in \mathbb{R}^N$ and continuous in η for almost $x \in G$.

(B2) For each $\gamma, B_\gamma(x, \eta)$ is essentially bounded for $|\eta|$ bounded and $\psi(x, \eta) = \sum_{|\gamma| \leq m-1} B_\gamma(x, \eta)\eta_\gamma \geq 0$ for all $\eta \in \mathbb{R}^N$.

(B3) For each γ, there exists a function δ_γ and a constant K such that

$$|B_\gamma(x, \hat{\eta})| \leq \delta_\gamma(|\hat{\eta}|)\psi(x, \eta) + K,$$

where $\hat{\eta}$ denotes the variables which occur in $B_\gamma(x, \eta)$, and $\delta_\gamma(t) \to 0$ as $t \to \infty$.

Let V be a closed subspace of $W^{m,p}(G)$ with $W_0^{m,p}(G) \subset V$ such that the following condition holds:

(*) $V_0 = V \cap C^m(\overline{G})$ is dense in V.

The Dirichlet bilinear form

$$a(u,v) = \sum_{|\alpha|,|\beta| \le m} \int_G a_{\alpha\beta} D^\beta u D^\alpha v dx$$

associated with $A(u)$ is well defined for all $u \in V_0$ and $v \in V$. For a fixed $u \in V_0$ this form defines a functional $T_1(u) \in V^*$ by the relation

$$a(u,v) = (T_1(u), v) \quad \text{for all} \quad v \in V .$$

The Dirichlet form

$$b(u,v) = \sum_{|\gamma| \le m-1} \int_G B_\gamma(\eta(u)) D^\gamma v dx$$

is also well defined for all $u \in V_0$ and $v \in V$ and defines a functional $T_2(u) \in V^*$ by

$$b(u,v) = (T_2(u), v) \quad \text{for all} \quad v \in V .$$

Let V_1 be the subset of V defined by

$$V_1 = \{ u \in V : \text{for each } |\gamma| \le m-1 , B_\gamma(\eta(u)) \in L^1(G) , \psi(x, \eta(u)) \in L^1(G)$$
$$\text{and} \quad \exists u^* \in V^* : (u^*, v) = \sum_{|\alpha|,|\beta| \le m} \int_G a_{\alpha\beta} D^\beta u D^\alpha v dx$$
$$+ \sum_{|\gamma| \le m-1} \int_G B_\gamma(\eta(u)) D^\alpha v dx \quad \text{for all} \quad v \in V_0 \} .$$

Clearly $V_0 \subset V_1 \subset V$. For $u \in V_1$ we set $T(u) = u^*$ and $c(u,v) = (T(u), v) = (u^*, v)$ for all $v \in V_0$.

Definition 2. A function u is a variational solution of the boundary value problem for the equation $A(u) + B(u) = f$, ($f \in V^*$ given) if (i) $u \in V_1$ and (ii) $c(u,v) = (f,v)$ for all $v \in V_0$, i.e., if $T(u) = f$.

Theorem 2. The operator $T : V_1 \to V^*$ is of type (M) with respect to V' for each dense subspace V' of V_0.

Proof. Let V' be a dense linear subspace of V_0. The assertion (i) of Definition 1 follows immediately. In order to prove the assertion (ii), let

$\{u_n\}$ be a sequence in $V', u \in V$ and $w \in V^*$. Suppose that $u_n \rightharpoonup u$ in $V, (T(u_n), v) \to (w, v)$ for all $v \in V'$ and $\limsup(T(u_n), u_n) \leq (w, u)$. We shall prove that $u \in D(T) = V_1$ and $T(u) = w$.

For all $|\beta| \leq m$ the sequence $\{D^\beta u_n\}$ converges weakly to $D^\beta u$ in $L^p(G)$. It follows that

$$(T_1(u_n), v) = \sum_{|\alpha|,|\beta| \leq m} \int_G a_{\alpha\beta} D^\beta u_n D^\alpha v dx \to \sum_{|\alpha|,|\beta| \leq m} \int_G a_{\alpha\beta} D^\beta u D^\alpha v dx .$$

for all $v \in V_0$.

By the Sobolev Imbedding Theorem, u_n converges strongly to u in $W^{m-1,p}(G)$. Hence, we may find an infinite subsequence which we again denote by $\{u_n\}$ such that for each γ with $|\gamma| \leq m-1, D^\gamma u_n(x)$ converges almost everywhere to $D^\gamma u(x)$. Now, we have

$$\liminf (T_1(u_n), u_n) + \limsup(T_2(u_n), u_n) \leq \limsup(T(u_n), u_n) \leq (w, u) .$$

From this inequality and the condition (A_2) we get that

$$\limsup(T_2(u_n), u_n) \leq (w, u) = K_1 .$$

The sequence $\{\psi(x, \eta(u_n))\}$ converges almost everywhere to $\psi(x, \eta(u))$ and by Fatou's lemma we have

$$\int_G \psi(x, \eta(u)) dx \leq \liminf \int_G \psi(x, \eta(u_n)) dx \leq K_1 .$$

Hence $\psi(x, \eta(u)) \in L^1(G)$. For a fixed γ, there exists for each $\delta > 0$ a constant $C(\delta) > 0$ such that for any η and almost all x in G either

$$|B_\gamma(\hat{\eta}(u_n(x)))| \leq \delta \psi(x, \eta(u_n(x))) + K$$

or

$$|D^\gamma u_n(x)| \leq C(\delta) \quad (\text{by } (B_3)) .$$

Hence, for any measurable set A of G,

$$\int_A |B_\gamma(\hat{\eta}(u_n(x)))| dx \leq C_1(\delta) \operatorname{meas}(A) + \delta \int_G \psi(x, \eta(u_n)) dx .$$

Given $\varepsilon > 0$, let δ be such that $\delta K_1 < \frac{\varepsilon}{2}$ and let $\text{meas}(A) < \frac{\varepsilon}{2C_1(\delta)}$. Then Vitali's convergence theorem implies that $B_\gamma(\eta(u_n)) \to B_\gamma(\eta(u))$ strongly in $L^1(G)$ ($|\gamma| \leq m-1$). Hence, for any $v \in V_0$ we have

$$(T_2(u_n), v) = \sum_{|\gamma| \leq m-1} \int_G B_\gamma(\eta(u_n)) D^\gamma v dx \to \sum_{|\gamma| \leq m-1} \int_G B_\gamma(\eta(u)) D^\gamma v dx.$$

It then follows that

$$(T(u_n), v) \to \sum_{|\alpha|,|\beta| \leq m} \int_G a_{\alpha\beta} D^\beta u D^\alpha v dx + \sum_{|\gamma| \leq m-1} \int_G B_\gamma(\eta(u)) D^\gamma v dx$$

for all $v \in V_0$.

Since $(T(u_n), v) \to (w, v)$ for all $v \in V'$, it follows that $(T(u_n), v) \to (w, v)$ for all $v \in V_0$. Hence $u \in D(T)$ and $T(u) = w$, which completes the proof of the theorem.

We can now state our existence theorem.

Theorem 3. Let G be a bounded open subset of $\mathbb{R}^n$ such that the Sobolev Imbedding Theorem holds on G, V a closed subspace of $W^{m,p}(G)$ for which the assumption $(*)$ holds. Let $A(u)$ and $B(u)$ be differential operators which satisfy the hypotheses (A_1),(A_2) and (B_1)-(B_3), respectively.

Suppose that

$$\{a(u,u) + b(u,u)\} \cdot \|u\|_V^{-1} \to +\infty$$

as $\|u\|_V \to +\infty, u \in V_0$. Then the boundary value problem for $A(u) + B(u) = f$ has a variational solution for each $f \in V^*$.

References

1. F. Browder, *Existence theory for boundary value problems for quasilinear elliptic systems with strongly nonlinear lower order terms*, Proc. Sympos. Pure Math., Vol. 23, pp. 269–286, Amer. Math. Soc., Providence, R. I., 1973.
2. P. Hess, *On nonlinear mappings of monotone type with respect to two Banach spaces*, J. Math. Pure et Appl. **52** (1973), 13–26.

3. P. Hess, *Variational inequalities for strongly nonlinear elliptic operators*, J. Math. Pure et Appl. **52** (1973), 285–298.
4. R. Landes, *On Galerkin's Method in the Existence Theory of Quasilinear Elliptic Equations*, J. Funct. Anal. **39** (1980), 123–148.
5. J. R. L. Webb, *Boundary value problems for strongly nonlinear elliptic equations*, J. London Math. Soc. (2) **21** (1980), 123–132.
6. J. R. L. Webb, *Strongly nonlinear elliptic equations*, Lecture Notes in Math., Vol. 665, pp. 242–256, Springer Verlag, 1978.

D. Kravvaritis
Department of Mathematics
National Technical University of Athens
Zografou Campus
15773 Athens
Greece

Constantin Carathéodory: An International Tribute (pp. 636–675)
edited by Th. M. Rassias

THE PROBLEM OF OPTIMIZATION OF THE ENSURED RESULT: UNIMPROVABILITY OF FULL-MEMORY STRATEGIES

A. V. Kryazhimskii

Abstract

For a control system described by an ordinary differential equation the problem of optimization (minimization) of the ensured result [1] is considered. A subset W of the set V of all dynamical disturbances admissible for the system is fixed. The class of control strategies with full memory of trajectories is called unimprovable (with respect to W), if the optimal ensured result achieved in this class coincides with that achieved in the class of abstract control procedures with full memory of disturbances (provided disturbances lie in W). Unimprovability conditions for W compact in L^2 different from those for $W = V$ [2] are stated. Through the sup-operation over all L^2-compact subsets of V a new definition of the ensured result (with respect to V) is introduced, and corresponding unimprovability conditions are formulated.

1. Introduction

A controlled dynamical system described by an ordinary differential equation in a finite-dimensional space is the

classical object studied by the theory of optimal control [3]. The controllability is reflected by the fact that the right hand side of the system equation depends on the values of a finite-dimensional function of time (a control) which is an argument of the basic optimization problem. The concept of a Caratheodory solution (of the system equation) is fundamental for the theory. The tool of classical solutions is not adequate to the problem, for optimal controls found practically or provided by necessary conditions of optimality are ordinarily non-continuous, and corresponding optimal trajectories satisfy the system equation only almost everywhere (a.e.). The transition from piece-wise continuous controls to measurable ones makes Caratheodory solutions natural for describing trajectories. It can be noted that the class of measurable controls is in general not wide enough to ensure existence of optimal controls (even for simple optimality indexes); its expansion to the class of the so called relaxed controls [4, 5] is necessary. Though relaxed controls are no longer finite-dimensional functions of time, corresponding trajectories remain Caratheodory solutions of ordinary differential equations (we deal with relaxed controls in Sec. 6-8).

Another problem leading to the necessity of substituting non-continuous functions of time into the right hand side of the system equation (implying Caratheodory solutions) is that of guiding a system under dynamical disturbances. A disturbance is normally a finite-dimensional function of time acting upon a trajectory the way controls do. Disturbances can in particular be formed by a controller's opponent whoose goal is contrary to that of the controller; this is a situation of a differential game [6]. So the controller is forced to

react to each disturbance from a certain set estimated a priori.The latter contains non-continuous functions in a number of practical cases. At least, a natural way to counteract disturbances by feedback generates as usual non-continuous controls too.

The problem of constructing an optimal feedback (closed-loop) control law is treated by the theory of optimization of the ensured result. The ensured result is the worst (the maximal, if a minimization problem is considered) value of the optimality index on the trajectories generated by all disturbances with a fixed closed-loop control law. The basis of the theory is set forth in the monographs [7, 8] where fundamental game-theoretical aspects of the problem are investigated and special solution methods are worked out. A general solution methods based on stochastic program constructions is suggested in the monograph [1].

In the book [2] an important property of closed-loop control laws that can be called unimprovability is described. It says that the transition from the practically realizable closed-loop control laws to the "ideal" control procedures called quasi-strategies does not change the optimal value of the ensured result (controls formed by a quasi-strategy satisfy the single condition: their values in present do not depend on the future values of disturbances). If only the values of disturbances are constrained, the major condition of unimprovability is existence of a saddle point for the Hamiltonian of the system. Note that this case implies non-compactness of the set of disturbances in the space L^2. In this paper we obtain new unimprovability conditions for compact classes of disturbances. They differ from traditional ones; in particular the Hamiltonian saddle point condition is

removed, and continuity properties of the optimized functional are weakened; on the other hand, the role of the full memory of a trajectory is strengthened, and existence of special one-to-one "trajectory-disturbance" mappings is required. The importance of the conditions is illustrated by examples. Using maximization over all compact classes of disturbances we introduce a new definition of the ensured result (a c-uniform ensured result); its optimal value is in general better than that defined traditionally. The corresponding conditions of unimprovability are stated.

2. Notations and Basic Objects

We set $\mathbb{N} = \{ 1,2,\dots \}$, $[k : l] = \{ i \in \mathbb{N} : k \le i \le l \}$ ($k,l \in \mathbb{N}$), and fix : $n, p, q \in \mathbb{N}$, a segment $I = [t_o, \vartheta_o]$ where $t_o < \vartheta_o$, non-empty compacta $P \subset \mathbb{R}^p$, $Q \subset \mathbb{R}^q$, and a continuous function $f : I \times \mathbb{R}^n \times P \times Q \mapsto \mathbb{R}^n$. The Euklidean norm in $\mathbb{R}^k$ ($k \in \mathbb{N}$) is denoted by $|\cdot|$; $x^T y$ denotes the scalar product of vectors x and y in $\mathbb{R}^k$. Measurability, measure and integral are understood in the sense of Lebesgue. For the restrictions of a function y defined on I to intervals $[t_o, t[$ and $[t, \vartheta[$ ($[t, \vartheta]$, if $\vartheta = \vartheta_o$) notations $y|t$ and $y|t,\vartheta$ are used, respectively; for a set E of functions defined on I we put $E|t = \{ y|t : y \in E \}$, $E|t,\vartheta = \{ y|t,\vartheta : y \in E \}$. Spaces $C(I,\mathbb{R}^n)$, $C([t_o, t], \mathbb{R}^n)$ ($t \in I$) and $L^l(I,\mathbb{R}^k)$ ($k,l \in \mathbb{N}$) are denoted briefly by C, C_t and $L^{l,k}$, for their norms notations $|\cdot|_C$, $|\cdot|_{C_t}$ and $|\cdot|_{L^{l,k}}$ are used. Symbol $\dot{x}$ stands for the derivative of an (absolutely continuous) function $x : I \mapsto \mathbb{R}^k$.

3. The Problem of Optimization of Ensured Result

Let us consider a controlled system described by the ordinary differential equation

$$\dot{x}(t) = f(t, x(t), u(t), v(t)) \qquad (3.1)$$

in $\mathbb{R}^n$ at the time interval I ; here $x(t)$ is a state of the system at time t ; $u(t) \in P$ and $v(t) \in Q$ are values of a control parameter and an (uncontrolled) disturbance parameter at time t , respectively. The initial state is given by the condition

$$x(t_o) = x_o \ . \qquad (3.2)$$

Hereafter $x_o \in \mathbb{R}^n$ is fixed.

Each measurable function $u : I \longmapsto P$ (resp., $v : I \longmapsto Q$) will be called *a control* (resp., *a disturbance*). The set of all controls (resp., disturbances) will be denoted by U (resp., V). We assume that for each $\vartheta \in I$, $u \in U$ and $v \in V$ there exists the unique Caratheodory solution of the Cauchy problem (3.1), (3.2) at $[t_o , \vartheta]$; for $\vartheta = \vartheta_o$ we call this solution *the trajectory generated by* u and v, and denote it by $x(\cdot\,|u,v)$.

The problem in question is that of minimization of a functional J defined on trajectories, provided a disturbance v (belonging to a given set W) is not fixed a priori. The tool of minimization is a control law (strategy) forming values $u(t)$ in real time without using information of future values $x(\tau)$ and $v(\tau)$, $\tau > t$. Since a disturbance is not fixed, a strategy does not determine a single trajectory. According to the "minimax" approach, a maximal value of J over a class of all trajectories compatible with a chosen strategy (an ensured

result) is actually minimized. An optimal ensured result depends naturally on a class of admissible strategies playing the role of arguments of an extremal problem.

N.N. Krasovskii stated and investigated the problem in [1] for $W = V$. He considered the class of positional strategies. These strategies need minimal information : $u(t)$ is formed on the basis of a current position $(t, x(t))$, on the other hand the class of positional strategies is unimprovable (see Introduction) in many typical cases. This paper deals with W compact in $L^{2,q}$. Unimprovability conditions for strategies not depending explicitly on disturbances (strategies with full memory of trajectories) are studied.

Now we pass to formal definitions.

Each family

$$S = ((\tau_i , \mathcal{U}_i))_{i \in [0:m]} \tag{3.3}$$

where $m \in \mathbb{N}$, $t_o = \tau_o < \cdots < \tau_m < \vartheta_o = \tau_{m+1}$, and $\mathcal{U}_i : C_{\tau_i} \longmapsto U|\tau_i,\tau_{i+1}$ $(i \in [0:m])$ will be called *a (full-memory) strategy*; mappings $\mathcal{U}_i$ will be called *feedback controls* of the strategy S. The elements of S (3.3) have the following sense: τ_i is a time instant the controller takes a decision to correct his control, and $\mathcal{U}_i$ is an algorithm to form control values at $[\tau_i , \tau_{i+1}[$ having information of the history of a trajectory at $[t_o , \tau_i]$. This corresponds to the following definition of a trajectory. *A trajectory generated by a strategy* S *and a disturbance* v is a function $x \in C$ such that $x = x(\cdot|u,v)$ where $u \in U$ satisfies the equality $u|\tau_i,\tau_{i+1} = \mathcal{U}_i(x|\tau_i)$ for each $i \in [0:m]$; we denote this trajectory by $x(\cdot|S,v)$ (It obviously exists and is unique); the above mentioned function u (determined

uniquely) will be called *the control generated by* S *and* v. For any strategy S and any set $W \subset V$, we put

$$X(S,W) = \{ x(\cdot|S,v) : v \in W \} . \qquad (3.4)$$

The set of all strategies will be denoted by $\mathcal{S}$.

Let

$$X = \{ x(\cdot|u,v) : u \in U , v \in V \} \qquad (3.5)$$

and $J : X \mapsto \mathbb{R}^1$. In Sec. 3-7 a non-empty set $W \subset V$ is fixed.

The ensured result on W *for a strategy* S is defined to be the value

$$\rho(S,W) = \sup \{ J(x) : x \in X(S,W) \} . \qquad (3.6)$$

The optimal ensured result on W *for a* (non-empty) *class* $\mathcal{S}' \subset \mathcal{S}$ is defined to be the value

$$\rho_0(\mathcal{S}',W) = \inf \{ \rho(S,W) : S \in \mathcal{S}' \} . \qquad (3.7)$$

The discussed optimization problem is set formally as that of finding the optimal ensured result on W for a chosen class of strategies.

We fix the class of positional strategies. A strategy S (3.3) will be called *positional* (closed-loop), if for each $i \in [0:m]$ the feedback control u_i is a function of the "terminal point of a trajectory", i.e. $u_i(x) = u_i'(x(\tau_i))$ $(x \in C_{\tau_i})$ where $u_i' : \mathbb{R}^n \mapsto U|\tau_i,\tau_{i+1}$. The set of all positional strategies will be denoted by $\mathcal{S}_C$.

4. Quasi-Strategies. Unimprovable Classes of Strategies

A quasi-strategy (we follow the terminology of [2]) is a most general way of forming controls in real time without using information of future. It is connected with the concept of a Volterra operator [9] and was introduced in [10, 11].

The equality $w'|t = w''|t$, where w' and w'' are disturbances or controls and $t \in I$ will further be understood as $w'(\tau) = w''(\tau)$ a.e. $\tau \in [\tau_0, \tau]$. *A quasi-strategy on* W is a mapping $S : W \mapsto U$ such that for any $v', v'' \in W$ and $t \in I$ satisfying $v'|t = v''|t$ it holds $S(v')|t = S(v'')|t$. The trajectory $x(\cdot|S(v),v)$ where S is a quasi-strategy and v is a disturbance will be said to be *generated by* S *and* v and will also be denoted by $x(\cdot|S,v)$. The set of all quasi-strategies on W will be denoted by $\mathcal{Q}(W)$. For each quasi-strategy S on W introduce the notation (3.4) and define *the ensured result* by (3.6). *The optimal ensured result on* W *in the class of quasi-strategies* is defined to be the value

$$\rho_0(\mathcal{Q}(W)) = \inf \{ \rho(S,W) : S \in \mathcal{Q}(W) \} . \qquad (4.1)$$

Theorem 4.1.

$$\rho_0(\mathcal{S},W) \geq \rho_0(\mathcal{Q}(W)) . \qquad (4.2)$$

Proof. Let S be an arbitrary strategy. Define the mapping $S' : W \mapsto U$ setting $S'(v)$ $(v \in W)$ to be the control generated by S and v. It is easy to prove that S' is a quasi-strategy on W and $X(S',W) = X(S,W)$. The latter implies that (see (3.6)) $\rho(S',W) = \rho(S,W)$. Since

S is arbitrary, we have (4.2).

Corollary 4.1. *For each non-empty class* $\mathcal{S}' \subset \mathcal{S}$ *it holds* $\rho_0(\mathcal{S}', W) \geq \rho_0(Q(W))$.

A (non-empty) class $\mathcal{S}' \subset \mathcal{S}$ will be called *unimprovable on* W , if $\rho_0(\mathcal{S}', W) = \rho_0(Q(W))$.

Let us give several unimprovability results for the case $W = V$.

5. Classes of Strategies Unimprovable on V

Consider the following conditions.

Condition 5.1 (growth condition). There exists a $K > 0$ such that $|f(t,x,u,v)| \leq K(1 + |x|)$ for all $t \in I$, $x \in \mathbb{R}^n$, $u \in P$, $v \in Q$.

Condition 5.2 (local Lipschitz condition). For any bounded set $E \subset \mathbb{R}^n$ there exists a $K(E) > 0$ such that $|f(t,x',u,v) - f(t,x'',u,v)| \leq K(E)|x' - x''|$ for all $t \in I$, $x', x'' \in E$, $u \in P$, $v \in Q$.

Condition 5.3 (saddle point condition). For any $l \in \mathbb{R}^n$, $t \in I$ and $x \in \mathbb{R}^n$ it holds

$$\min_{u \in P} \max_{v \in Q} l^T f(t,x,u,v) = \max_{v \in Q} \min_{u \in P} l^T f(t,x,u,v) .$$

We shall say that the functional J is *uniformly C-continuous on a* (non-empty) *set* $X' \subset X$, if

$$\sup \left\{ |J(x) - J(y)| : x,y \in X' , |x - y|_C \le \beta \right\} \longrightarrow 0$$
$$\text{as} \quad \beta \longrightarrow 0 . \qquad (5.1)$$

Theorem 5.1. *Let Conditions* 5.1, 5.2, *and* 5.3 *be fulfilled and the functional* J *be uniformly* C*-continuous on* X. *Then*

1) *the class* $\mathcal{S}$ (of all strategies) *is unimprovable on* V,

2) *if* J *has the form*

$$J(x) = \varphi(x(\vartheta_0)) + \int_I \psi(t,x(t))dt , \qquad (5.2)$$

where $\varphi : \mathbb{R}^n \longmapsto \mathbb{R}^1$ *and* $\psi : I \times \mathbb{R}^n \longmapsto \mathbb{R}^1$ *are continuous, then the class* $\mathcal{S}_C$ (of all positional strategies) *is unimprovable on* V.

Statements 1) and 2) follow from [8, Lemma 96.1] and [2, Theorems 4.4.3 and 4.4.4], respectively.

Remark 5.1. The above results from [8] and [2] imply actually that statements 1) and 2) are true, if $\mathcal{S}$ and $\mathcal{S}_C$ are replaced by the classes of all strategies and all positional strategies, whoose feedback controls take constant values.

Remark 5.2. Condition 5.1 in Theorem 5.1 can be replaced by the more weak one: X is bounded in C.

Remark 5.3. Condition 5.2 in Theorem 5.1 can be replaced by the more weak one requiring uniqeness of a trajectory generated by an arbitrary relaxed "control-disturbance" input [5] (it follows from [12]).

Let us give two examples showing that Condition 5.3 and the uniform C-continuity of J on X are important for Theorem 5.1.

Example 5.1 (importance of Condition 5.3). Let $n = 1$, $I = [0,1]$, $P = Q = \{-1,1\}$, the system (3.1), (3.2) have the form

$$\dot{x}(t) = u(t)v(t), \quad x(0) = 0,$$

and $J(x) = x(1)$. Conditions 5.1 and 5.2 are fulfilled, Condition 5.3 is violated. It is easily seen that $\rho_0(Q(V)) = -1$ (consider the quasi-strategy $v \mapsto -v$ on V). On the other hand for any strategy S the set $X(S,V)$ (see (3.4)) contains the function $x : t \mapsto t$; hence $\rho_0(\mathcal{S},V) = 1$. The class $\mathcal{S}$ is not unimprovable and statements 1) and 2) of Theorem 5.1 are not true.

Example 5.2 (importance of the uniform C-continuity of J on X). Let $n = 1$, $I = [0,1]$, $P = Q = [-1,1]$, the system (3.1), (3.2) have the form

$$\dot{x}(t) = u(t) + v(t), \quad x(0) = 0,$$

and $J(x) = |\dot{x}|_L 2,1$. Conditions 5.1, 5.2 and 5.3 are fulfilled, but J is not uniform C-contnuous on X. It is easily seen that $\rho_0(Q(W)) = 0$ (consider the quasi-strategy $v \mapsto v$ on V). On the other hand for any strategy S there exists a $v \in V$ such that $|\dot{x}(t|S,v)| \geq 1$ a.e. $t \in I$; hence $\rho_0(\mathcal{S},V) \geq 1$. Statement 1) of Theorem 5.1 is not true.

6. Unimprovability Conditions for the Class $\mathcal{S}$ on W : W Compact in $L^{2,q}$

In this Section W compact (in general not closed) in $L^{2,q}$ will be considered (this obviously implies compactness of W in $L^{s,q}$ for each $s \in [1, \infty)$). Introduce the

Condition 6.1. The set

$$X(W) = \{ x(\cdot|u,v) : u \in U , v \in V \} \qquad (6.1)$$

is bounded in C .

Remark 6.1. Condition 6.1 implies by the Arzela's theorem that the set X(W) is compact in C . Note that Condition 5.1 is sufficient for Condition 6.1.

To formulate other conditions of unimprovability we use the concept of a relaxed control. Introduce it following [5, Ch.IV]. Let $\mathcal{B}$ be the set of all functions $b : I \times P \mapsto \mathbb{R}^1$ such that $b(\cdot,u)$ is measurable for each $u \in P$, $b(t,\cdot)$ is continuous for each $t \in I$, and $\sup \{ |b(t,u)| : u \in P \} \le \lambda(t)$ a.e. $t \in I$ for a certain integrable $\lambda : I \mapsto \mathbb{R}^1$. *A relaxed control* is a function μ on I taking values in the set of all Borel probability measures on P . Measurability of μ is understood in the following sense : for each $b \in \mathcal{B}$ the function $t \mapsto b(t,\mu(t))$ is measurable; hereafter

$$g(\mu(t)) = \int_P g(u)(\mu(t))(du)$$

for any continuous $g : P \mapsto \mathbb{R}^k$. The set of all relaxed

controls will be denoted by RU. *A trajectory generated by a relaxed control* μ *and a disturbance* v is a Caratheodory solution of the Cauchy problem

$$\dot{x}(t) = f(t,x(t),\mu(t),v(t)) \ , \ x(t_o) = x_o$$

on I. If δ_z $(z \in P)$ is the Borel probability measure on P such that $\delta_z(\{z\}) = 1$ (the measure consentrated at point z), then for each (ordinary) control u and each disturbance v the trajectory generated by u and v coincides with that generated by the relaxed control $t \mapsto \delta_{u(t)}$ and the disturbance v; this allows us to identify the above relaxed control with u [5, Ch. IV]. Thus, we consider U as a subset of RU. Assume also that the metric generated by the $*$ - weak norm of the space conjugate to $L^1(I,C(P))$ is fixed on RU [5, Ch. IV]. Note that a sequence (μ_i) converges to a μ in RU, if for each $b \in \mathcal{B}$

$$\int_I b(t,\mu_i(t)) \longrightarrow \int_I b(t,\mu(t))dt \ . \qquad (6.2)$$

Now formulate the conditions.

Condition 6.2. Each $\mu \in RU$ and each v from the closure of W in $L^{2,q}$ generate the single trajectory.

Condition 6.3. If $t \in I$, $u \in U$, $v',v'' \in W$ and $x(\cdot|u,v')|t = x(\cdot|u,v'')|t$, then $v'|t = v''|t$.

For each function $y : I \mapsto \mathbb{R}^k$, define its *positive* and *negative* δ*-shifts* $(\delta \geq 0)$ $y^{+\delta} : I \mapsto \mathbb{R}^k$ and $y^{-\delta} : I \mapsto \mathbb{R}^k$ by

$$y^{+\delta}(t) = \begin{cases} y(t+\delta) & , \ t \in [t_0 \quad , \vartheta_0 - \delta[\\ y_0 & , \ t \in [\vartheta_0 - \vartheta \ , \vartheta_0] \end{cases} , \tag{6.3}$$

$$y^{-\delta}(t) = \begin{cases} y_0 & , \ t \in]t_0 \quad , t + \delta] \\ y(t-\delta) & , \ t \in]t_0 + \delta \ , \vartheta_0] \end{cases} ;$$

here y_0 is a fixed element from $\mathbb{R}^k$; if y is a control ($k = p$) or a disturbance ($k = q$), then we denote y_0 by u_0 and v_0, respectively, and assume $u_0 \in P$ and $v_0 \in Q$; if $y = \dot{x}$ for $x \in X$ ($k = n$), we put $y_0 = 0$.

Fix a $r > 0$. We shall say that the functional J is *uniformly* (L^r,δ)*-continuous on a* (non-empty) *set* $X' \subset X$, if

$$\sup \left\{ |J(x) - J(y)| : x,y \in X' , \right.$$
$$\left. |\dot{x}^{+\delta} - \dot{y}|_{L^r,n} \le \beta , \ 0 \le \delta \le \beta \right\} \longrightarrow 0$$
$$\text{as} \quad \beta \longrightarrow 0 . \tag{6.4}$$

Example 6.1. Let $r = 2$, $J(x) = |\dot{x}|^2_{L^2,n}$ and a set $X' \subset X$ be bounded in C . Then J is uniformly (L^r,δ)-continuous on X' . Indeed, if $x,y \in X'$, $|\dot{x}^{+\delta} - \dot{y}|_{L^2,n} \le \beta$ and $0 \le \delta \le \beta$, then

$$\left| \ |\dot{x}|^2_{L^2,n} - |\dot{x}^{+\delta}|^2_{L^2,n} \ \right| \le K^2\delta \le K^2\beta$$

where $K = \sup \left\{ \operatorname{vraimax}_{t \in I} |\dot{z}(t)| : z \in X' \right\}$ ($K < +\infty$, since X' is bounded in C and f is continuous), and

$$\left| \ |\dot{x}^{+\delta}|^2_{L^2,n} - |\dot{y}|^2_{L^2,n} \ \right| \le \beta(|\dot{x}^{+\delta}|_{L^2,n} + |\dot{y}|_{L^2,n}) \le K_1\beta$$

where $K_1 = 2K(\vartheta_o - t_o)^{1/2}$. From these inequalities we get

$$| \, |\dot{x}|^2_{L^{2,n}} - |\dot{y}|^2_{L^{2,n}} \, | \le (K^2 + K_1)\beta \; ;$$

(6.4) is true. Note that J is in general not uniformly C-continuous on X' .

Remark 6.3. If the functional J is uniformly C-continuous on X', then it is uniformly (L^r,δ)-continuous on X' .

Below the basic theorem is formulated.

Theorem 6.1. *Let the set* W *be compact in* $L^{2,q}$, *Conditions* 6.1, 6.2 *and* 6.3 *be fulfilled, and the functional* J *be uniformly* (L^r,δ)*-continuous on* $X(W)$. *Then the class* $\mathcal{S}$ (of all strategies) *is unimprovable.*

Here we do several comments (the proof of Theorem 6.1 is given in Sec. 6). Statement 2) of Theorem 5.1 is in general not true under the conditions of Theorem 6.1 (the class $\mathcal{S}_c$ of positional strategies is not unimprovable for J having the form (5.1)).

Example 6.2. Let $n = 1$, $I = [-1,1]$, $P = Q = [-2, 2]$, the system (3.1), (3.2) have the form

$$\dot{x}(t) = g(t)u(t) - v(t) \quad , \quad x(-1) = 0 \, ,$$

$$g(t) = \begin{cases} 0 \, , \, t \le 0 \\ t \, , \, t > 0 \end{cases} , \qquad (6.5)$$

$W = \{ v_{-1}, v_1 \}$ where

$$v_1(t) = \begin{cases} 2, & -1 \le t < -1/2 \\ -2, & -1/2 \le t < 0 \\ 0, & 0 \le t < 1/2 \\ 2, & 1/2 \le t \le 1 \end{cases},$$

$v_{-1} = -v_1$ and $J(x) = |x(1)|$.

All conditions of Theorem 6.1 and of statement 2) of Theorem 5.1 are fulfilled. It is easy to calculate that $\rho_o(S_o, W) = 0$ for the quasi-strategy S_o on W determined by $(S_o(v_1))(t) = 2$ and $(S_o(v_{-1}))(t) = -2$. Since J is non-negative, $\rho_o(Q(W)) = 0$. Let S be an arbitrary positional strategy. Denote $y_1 = x(\cdot\,|S, v_1)$ and $y_{-1} = x(\cdot\,|S, v_{-1})$. It is easily seen that $y_1(0) = y_{-1}(0) = 0$. Taking into account that S is positional and $v_1|0,1/2 = v_{-1}|0,1/2$ we get $y_1|0,1/2 = y_{-1}|0,1/2$. Let $b = y_1(1/2) = y_{-1}(1/2)$. Assume that $b \ge 0$. Since $\dot{y}_{-1}(t) \ge -2t + 2$ a.e $t \in [1/2, 1]$, we have $y_{-1}(1) \ge 1/4$. If $b < 0$, the analogous inequality $y_1(1) \le -1/4$ is true. In both cases $\rho(S, W) \ge 1/4$. But S is arbitrary, thus, $\rho_o(\mathcal{S}_C, W) \ge 1/4$. The class $\mathcal{S}_S$ is not unimprovable.

Let us illustrate the importance of Conditions 6.2 and 6.3 for Theorem 6.1.

Example 6.3 (importance of Condition 6.3). Let $n = 4$, $I = [-1,1]$, $P = \{-1,1\}$, $Q = [0,1]$, and the system (3.1), (3.2) have the form

$$\dot{x}_1(t) = u(t) \quad ,$$

$$\dot{x}_2(t) = \begin{cases} x_1^2(t)|t| & , t > 0 \\ 0 & , t \le 0 \end{cases},$$

$$\dot{x}_3(t) = \begin{cases} x_1(t)|t|v(t) & , t < 0 \\ g(t, x_2(t), x_3(t)) & , t \ge 0 \end{cases},$$

$$\dot{x}_4(t) = v(t) \qquad , \qquad (6.6)$$

$$x_1(-1) = x_2(-1) = x_3(-1) = x_4(-1) = 0 ;$$

here

$$g(t, x_2, x_3) = \begin{cases} t x_3^{1/2} & , x_3 \ge |x_2| \\ 0 & , x_3 \le 0 \\ t x_3^{3/2}/|x_2| & , 0 < x_3 < |x_2| \end{cases}.$$

Let W be the set of all absolutely continuous disturbances v such that $v(-1) = 0$ and $|\dot{v}(t)| = 1$ a.e. $t \in 1$ (it is clear that W is compact in $L^{2,1}$). All assumptions from Section 3 are true as one can easily verify. Conditions 6.1 and 6.3 are fulfilled (see the last equation in (6.6) and Remark 6.2). Condition 6.2 is violated, since the relaxed control $t \mapsto 1/2(\delta_1 + \delta_{-1})$ (together with an arbitrary disturbance v) generates two trajectories x' and x'' having different third components

($x_3'(t) = 0$ and $x_3''(t) = \begin{cases} 0 & , t \le 0 \\ t^4/16 & , t > 0 \end{cases}$, respectively).

Put $J(x) = x_3(t)$. The functional J is uniformly C-continuous and consequently uniformly (L^r, δ)-continuous on $X(W)$ (see Remark 6.3). Let us show that Theorem 6.1 is

not true. Consider the quasi-stratagy $S_o : v \mapsto -\dot{v}$ on W For any $v \in W$ the trajectory $x = x(\cdot|S_o,v)$ satisfies the equalities $x_1(t) = -v(t)$ $(t \in I)$, $x_3(t) = -v^2(t)|t|$ $(t \le 0)$, $x_3(t) = x_3(0) \le 0$ $(t \le 0)$; therefore $J(x) = x_3(1) \le 0$. Hence $\rho_o(Q(W)) \le \rho(S_o,W) \le 0$. Take an arbitrary strategy S . It is clear that there exists a $v \in W$ such that the trajectory $x = x(\cdot|S,v)$ satisfies the equality $x_1 = v$. Then $x_2(t) = x_3(t) > 0$ $(t \in]-1,0])$. Hence, taking into account the differential equation for x_3 , we get $x_3(t) = (t^2/4 + x_3(0))$ $(t \ge 0)$. Consequently $\rho(S,W) \ge J(x) = x_3(1) \ge 1/16$. Since S is arbitrary, $\rho_o(\mathcal{S},W) \ge 1/16$. The class $\mathcal{S}$ is not unimprovable.

Note that if one change the places of u and v in the equation (6.6) and puts $J(x) = -x_3(1)$, all conditions of Theorem 6.1 are also fulfilled, except Condition 6.2 : the disturbance $v : t \mapsto 0$ belonging to the closure of W in $L^{2,1}$ generates (together with an arbitrary control) two different trajectories. Following the previous pattern, one can show that Theorem 6.1 is not true.

Example 6.4 (importance of Condition 6.3). Let $n = 1$, $I = [-1,1]$, $P = Q = (-1,1)$, and the system (3.1), (3.2) have the form

$$\dot{x}(t) = u(t) + g(t)v(t) \ , \ x(-1) = 0$$

where g is determined by (6.5). Let $W = \{v_{-1}, v_1\}$ where $v_{-1}(t) = -1$, $v_1(t) = 1$ and $J(x) = a(x(1))$ where a is a non-negative continuous function on $\mathbb{R}^1$ such that

$$a(y) = \begin{cases} 0, & |y| \ge 5/2 \\ 1, & |y| \le 3/2 \end{cases}$$

All conditions of Theorem 6.1 are fulfilled, except Condition 6.3 (it is violated due to (6.5)). For the quasi-strategy $S_o : v \mapsto v$ on W we have $X(S_o,W) = \{x_{-1},x_1\}$ and $|x_1(1)| = |x_{-1}(1)| = 5/2$. Thus, $\rho_o(\mathcal{Q}(W)) \le \rho(S_o,W) = 0$. Let S be an arbitrary strategy. Denote $y_{-1} = x(\cdot\,|S,v_{-1})$ and $y_1 = x(\cdot\,|S,v_1)$ and put $b = y_{-1}(0) = y_1(0)$. Assume that $b \ge 0$. Since $1 \ge y_{-1}(0) \ge 0$ and $1-t \ge \dot{y}_{-1}(t) \ge -1-t$ a.e. $t \in [0,1]$, it holds $3/2 \ge y_{-1}(1) \ge -3/2$. If $b < 0$, we have the analogous inequalities $3/2 \ge y_1(1) \ge -3/2$. In both cases (see the function a) $\rho(S,W) \ge 1$. But S is arbitrary, therefore $\rho_o(\mathcal{S},W) \ge 1$. Thus, the class $\mathcal{S}$ is not unimprovable.

7. Proof of Theorem 6.1.

Lemma 7.1.

1) [5, Theorem IV.2.1] *RU is a compactum.*

2) [5, Theorem VI.1.1] *If a sequence* $((\mu_i, v_i, x_i))$ *from* $RU \times V \times C$ *converges to* $(\mu, v, x) \in RU \times V \times C$ *in* $RU \times L^{2,q} \times C$, *and* x_i *is a trajectory generated by* μ_i *and* v_i $(i \in \mathbb{N})$, *then* x *is a trajectory generated by* μ *and* v.

Lemma 7.2. *Let* $\mu_i \longrightarrow \mu$ *in* RU, $\delta_i > 0$ $(i \in \mathbb{N})$, $\delta_i \longrightarrow 0$ *and* $\nu_i \in RU$ *be such that* $\nu_i(t) = \mu_i(t-\delta_i)$ *for* $t \in [t_o + \delta_i, \vartheta_o]$ $(i \in \mathbb{N})$. *Then* $\nu_i \longrightarrow \mu$ *in* RU.

Proof. Let $\mathcal{B}'$ be the set of all functions $b \in \mathcal{B}$ having the form $b(t,u) = \sum_{j=1}^{k} g_j(t)b_j(u)$ where $k \in \mathbb{N}$, $g : I \mapsto \mathbb{R}^1$ and $b_j : P \mapsto \mathbb{R}^1$ $(j \in [1:k])$ are continuous. Since $\mathcal{B}'$ is dense in $L^1(I,C(P)) = \mathcal{B}$ [5, Theorem I.5.18], it is sufficient to show that the convergence

(6.2) where μ_i is replaced with ν_i is true for all $b \in \mathcal{B}'$. Take an arbitrary $b \in \mathcal{B}'$. It is clear that

$$w(\delta) = \sup \left\{ |b(t',u) - b(t'',u)| : t',t'' \in I , |t' - t''| \leq \delta , u \in P \right\} \longrightarrow 0 \quad \text{as} \quad \delta \longrightarrow 0 .$$

Hence

$$\beta_i = \int_{t_o+\delta_i}^{\vartheta_o} (b(t,\nu_i(t)) - b(t - \delta_i,\nu_i(t)))dt \longrightarrow 0 .$$

Note that

$$\alpha_i = \int_{t_o}^{t_o+\delta_i} b(t,\nu_i(t))dt \longrightarrow 0 ,$$

$$\gamma_i = \int_{\vartheta_o-\delta_i}^{\vartheta_o} b(t,\mu_i(t))dt \longrightarrow 0 .$$

These convergences and the convergence (6.2) imply

$$\int_I b(t,\nu_i(t))dt = \alpha_i + \int_{t_o+\delta_i}^{\vartheta_o} b(t,\nu_i(t))dt =$$

$$= \alpha_i + \beta_i + \int_{t_o+\delta_i}^{\nu_o} b(t - \delta_i,\nu_i(t))dt =$$

$$= \alpha_i + \beta_i + \int_{t_o}^{\vartheta_o - \delta_i} b(t,\mu_i(t))dt =$$

$$= \alpha_i + \beta_i - \gamma_i + \int_I b(t,\mu_i(t))dt \longrightarrow \int_I b(t,\mu(t))dt \ .$$

Lemma 7.3. *Let a set* $W' \subset W$ *be non-empty and compact in* $L^{2,q}$ *, and Conditions* 6.1 *and* 6.2 *be fulfilled. Then*

$$\sup \left\{ |\dot{x}^{+\delta}(\cdot\,|u^{-\delta},v) - \dot{x}(\cdot\,|u,v)|_{L}r,n : u \in U,\ v \in W' \right\} \longrightarrow 0$$
$$\text{as } \delta \longrightarrow 0 \qquad (7.1)$$

Proof. Suppose that (7.1) is not true. Then there exist an $\varepsilon > 0$ and sequences (u_i) from U, (v_i) from W' and (δ_i) of positive numbers such that for $x_i = x(\cdot\,|u_i,v_i)$ and $x_i^+ = x^{+\delta_i}(\cdot\,|u_i^{-\delta_i}, v_i)$ it holds

$$|\dot{x}_i^+ - \dot{x}_i|_{L}r,n \geq \varepsilon \quad (i \in \mathbb{N}) \ . \qquad (7.2)$$

Since RU is a compactum (Lemma 7.1, 1)), W is compact in $L^{2,q}$ and $X(W')$ is compact in C (it follows from Condition 6.1 by the Arzela's theorem), assume without loss of generality that

$$u_i \longrightarrow \mu \in RU \ \text{ in } \ RU \ , \qquad (7.3)$$
$$v_i \longrightarrow v \in W \ \text{ in } \ L^{2,q} \ , \qquad (7.4)$$
$$x_i \longrightarrow x \in C \ \text{ in } \ C \ , \qquad (7.5)$$
$$x_i^+ \longrightarrow x^+ \in C \ \text{ in } \ C \ ; \qquad (7.6)$$

in (7.4) $\overline{W}$ stands for the closure of W in $L^{2,q}$. Let

y be the trajectory generated by μ and v (it is unique by Condition 6.2). The convergence (7.3) implies by Lemma 7.2 that

$$u_i^{-\delta_i} \longrightarrow \mu \quad \text{in} \quad RU \,. \tag{7.7}$$

By Lemma 7.1, 2) convergences (7.3), (7.4) and (7.5) imply $x = y$, and convergences (7.7), (7.4) and (7.6) imply $x^+ = y$. Consequently (see (7.5) and (7.6))

$$|x_i - x_i^+|_C \longrightarrow 0 \,. \tag{7.8}$$

Let E be a compactum in $\mathbb{R}^n$ such that $\{ x(t) : x \in X(W),\ t \in I \} \subset E$ (see Condition 6.1),

$$w(\beta) = \sup \left\{ |f(t',x',u',v') - f(t'',x'',u'',v'')| : t',t'' \in I,\ x',x'' \in E,\ u',u'' \in P,\ v',v'' \in Q,\ |t' - t''| \le \beta,\ |x' - x''| \le \beta,\ |v' - v''| \le \beta \right\} \quad (\beta \ge 0) \,,$$

$$c = \sup \left\{ \operatorname*{vraimax}_{t \in I} |\dot{x}(t)| : x \in X(W) \right\}$$

(Condition 6.1 ensures $c < +\infty$), and

$$\gamma_1(t) = \max \left\{ \delta_i,\ |x_i - x_i^+|_C,\ |v_i(t - \delta_i) - v_i(t)| \right\}. \tag{7.9}$$

Then (taking into account that $\dot{x}_i^+(t) = 0$ for $t \in [\vartheta_o - \delta\,,\ \vartheta_o]$) we have

$$|\dot{x}_i^+ - \dot{x}_i|_{L^{r,n}}^r \le \int_{t_o}^{\vartheta_o - \delta_i} |f(t+\delta_i,\ \dot{x}_i^+(t),\ u_i(t),\ v_i(t+\delta_i)) -$$

$$|f(t, x_i(t), u_i(t), v_i(t))|^r + c^r\delta_i \le$$

$$\int_{t_o}^{\vartheta_o-\delta_i} w^r(\gamma_i(t))dt + c^r\delta_i \ . \qquad (7.10)$$

Note that the last integral exists (the function $\beta \longrightarrow w(\beta)$ is continuous and the function $t \longrightarrow \gamma_i(t)$ is measurable and bounded; therefore the superposition $t \longmapsto w^r(\gamma(t))$ is integrable [5, Theorem I.4.22]). Fix a $\delta > 0$ such that

$$(c_1^r + c^r)\delta < \varepsilon^r/4 \ , \qquad (7.11)$$

where $c_1 = \sup \{ w(\beta) : \beta \ge 0 \}$ (obviously, $c_1 < +\infty$). The $L^{2,q}$-compactness of W' yields $|v_i^{-\delta_i} - v_i|_{L^{2,q}} \longrightarrow 0$. Therefore, assume with no loss of generality (taking a subsequence, if it is necessary) that $|v_i(t-\delta_i) - v_i(t)| \longrightarrow 0$ a.e. $t \in [t_o, \vartheta_o-\delta]$. Then (7.8) and (7.9) give the convergence $\gamma_i(t) \longrightarrow 0$ a.e. $t \in [t_o, \vartheta_o-\delta]$. By the Lebesgue's theorem

$$\int_{t_o}^{\vartheta_o-\delta} w^r(\gamma_i(t))dt \longrightarrow \int_{t_o}^{\vartheta_o-\delta} w^r(0)dt = 0 \ . \qquad (7.12)$$

From (7.10), (7.11) and (7.12) we get that for all sufficiently large i

$$|\dot{x}_i^+ - \dot{x}_i|_{L^{r,n}}^r \le \int_{t_o}^{\vartheta_o-\delta} w^r(\gamma_i(t))dt + (c_1^r + c^r)\delta < \varepsilon^r/2 \ .$$

This contradicts the assumption (7.2).

The following notion is basic for the proof of Theorem 6.1. Define *the negative δ-shift* ($\delta > 0$) of a quasi-strategy S on a set $W' \subset V$ to be the mapping $S^{-\delta} : v \longmapsto S^{-\delta}(v) = (S(v))^{-\delta} : W' \longmapsto U$; it is easily seen that $S^{-\delta}$ is a quasi-strategy on W' .

Lemma 7.4. *Let a set $W' \subset W$ be non-empty and compact in $L^{2,q}$, and Conditions* 6.1 *and* 6.2 *be fulfilled. Then for each quasi-strategy S on W'*

$$\sup_{x \in X(S^{-\delta},W')} \inf_{y \in X(S,W')} |\dot{x}^{+\delta} - \dot{y}|_L r,n \longrightarrow 0$$

$$\text{as} \quad \delta \longrightarrow 0 . \qquad (7.13)$$

Proof. Let S be a quasi-strategy on W' , $\delta > 0$ and x be an arbitrary element from $X(S^{-\delta},W')$, i.e. $x = x(\cdot\,|(S(v))^{-\delta}, v)$ for a certain $v \in W'$. Then

$$\inf_{y \in X(S,W)} |\dot{x}^{+\delta} - \dot{y}|_L r,n \le |\dot{x}^{+\delta}(\cdot\,|(S(v))^{-\delta}, v) -$$

$$x(\cdot\,|S(v),v)|_L r,n \le \sigma(\delta)$$

where $\sigma(\delta)$ is the value given in (7.1). So far as x is arbitrary, the value given in (7.13) is no larger than $\sigma(\delta)$ too. By Lemma 7.3 $\sigma(\delta) \longrightarrow 0$ as $\delta \longrightarrow 0$; this completes the proof.

Lemma 7.5. *Let Condition* 6.3 *be fulfilled, S be a quasi-strategy on W , and $\delta \in]0,\vartheta_o - t_o[$. Then for any $t \in I$, $x \in X(S^{-\delta},W)|t$, $v',v'' \in W$ such that*

$$x = x(\cdot\,|S^{-\delta}(v'),v')|t = x(\cdot\,|S^{-\delta}(v''),v'')|t \qquad (7.14)$$

It holds

$$v'(\tau) = v''(\tau) \quad \text{a.e. } \tau \in [t_o,t] \,, \qquad (7.15)$$

$$(S^{-\delta}(v'))(\tau) = (S^{-\delta}(v''))(\tau) \quad \text{a.e. } \tau \in [t_o,t+\delta] \cap I \,. \qquad (7.16)$$

Proof. We put

$$\tau_i = t_o + i\delta \quad (i \in \mathbb{N}), \quad m = \max\{\, i \in \mathbb{N} : \tau_i < \vartheta_o \,\} \,. \qquad (7.17)$$

Let us show that the statement of the Lemma is true for $t \le \tau_1$. By the definition of $S^{-\delta}$

$$(S^{-\delta}(v'))(\tau) = (Sv')^{-\delta}(\tau) = u_o = (S(v''))^{-\delta}(\tau) = (S^{-\delta}(v''))(\tau) \quad (\tau \in [t_o,\tau_1]) \,. \qquad (7.18)$$

Let $x \in X(S^{-\delta},W)|t$ and $v',v'' \in W$ satisfy (7.14). Introduce the control $u : \tau \longrightarrow u_o$. Rewrite (7.14) in the form

$$x(\cdot\,|u,v')|t = x(\cdot\,|u,v'')|t \,. \qquad (7.19)$$

Applying Condition 6.3 we get (7.15).
Hence

$$(S(v'))(\tau) = S(v''))(\tau) \quad \text{a.e. } \tau \in [t_o,t] \,. \qquad (7.20)$$

Consequently

$$(S^{-\delta}(v'))(\tau) = (S(v'))(\tau-\delta) = (S(v''))(\tau-\delta) = (S^{-\delta}(v''))(\tau) \quad \text{a.e. } \tau \in [\tau_1,t+\delta] \cap I \,. \qquad (7.21)$$

This condition and (7.18) imply (7.16).

Now we use induction. Suppose that the statement is true for all $t \le \tau_i$, where $i \in [1 : m]$. Let us show that it is true for $t \in [\tau_i, \tau_{i+1}]$. Let $x \in X(S^{-\delta}, W)|t$ and $v', v'' \in W$ satisfy (7.14). By the assumption, $v'(\tau) = v''(\tau)$ a.e. $\tau \in [t_o, \tau_i]$. It implies that $(S(v'))(\tau) = (S(v''))(\tau)$ a.e. $\tau \in [t_o, \tau_i]$. Hence by the definition of $S^{-\delta}$

$$(S^{-\delta}(v'))(\tau) = (S(v'))(\tau)(\tau-\delta) = (S(v''))(\tau-\delta) = (S^{-\delta}(v''))(\tau)$$
$$\text{a.e. } \tau \in [\tau_i, \tau_{i+1}] .$$

Taking into account (7.16) we get

$$(S^{-\delta}(v'))(\tau) = (S^{-\delta}(v''))(\tau) \quad \text{a.e. } \tau \in [t_o, \tau_{i+1}] .$$

Introduce a control u such that $u(\tau)$ is equal to both sides of this equality a.e. $\tau \in [t_o, \tau_{i+1}]$. Rewrite (7.14) in the form (7.19) and applying Condition 6.3 obtain (7.15). The latter leads (as in the case $t \le \tau_1$) to (7.20) and (7.21) which imply (7.16).

Lemma 7.6. *Let Condition* 6.3 *be fulfilled,* S *be a quasi-strategy on* W *and* $\delta \in]0, \vartheta_o - t_o[$. *Then there exists a strategy* S' *such that for any* $v \in W$

$$x(\cdot|S', v) = x(\cdot|S^{-\delta}, v) . \qquad (7.22)$$

Proof. Here we use notations (7.17). Introduce a strategy $S' = ((\tau_i, u_i))_{i \in [0:m]}$, whoose feedback controls u_i ($i \in [0 : m]$) satisfy the following conditions : If $x \in X(S^{-\delta}, W)|\tau_i$, then $u_i(x)$ is an arbitrary element from

$U|\tau_i,\tau_{i+1}$: if

$$x \in X(S^{-\delta},W)|\tau_i \ , \tag{7.23}$$

then

$$u_i(x) = S^{-\delta}(v_*)|\tau_i,\tau_{i+1} \tag{7.24}$$

where $v_* \in W$ is such that

$$x = x(\cdot\,|S^{-\delta}(v_*),v_*)|\tau_i \ . \tag{7.25}$$

Fix an arbitrary $v \in W$. Let $y = x(\cdot\,|S',v)$ and u be the control generated by S' and v. So, $y = x(\cdot\,|u,v)$. To prove (7.22), it is sufficient to show that $u = S^{-\delta}(v)$. We shall show by induction that

$$u(\tau) = (S^{-\delta}(v))(\tau) \quad \text{a.e.} \quad \tau \in [t_o\ ,\tau_i] \tag{7.26}$$

for each $i \in [1 : m+1]$. Let $i = 1$. Since $y(\tau_o) = x_o$, we have (7.23) for $x = y|\tau_o$. Consequently,

$$u|\tau_1 = u_o(x) = S^{-\delta}(v_*)|\tau_1 \tag{7.27}$$

where $v_* \in W$ satisfies (7.25) (for $i = 1$). By the definition of $S^{-\delta}$,

$$(S^{-\delta}(v_*))(\tau) = (S(v_*))^{-\delta}(\tau) = u_o = (S(v))^{-\delta}(\tau) = (S^{-\delta}(v))(\tau) \quad \text{a.e.} \quad \tau \in [t_o,t_o+\delta] = [\tau_o,\tau_1] \ .$$

This and (7.27) imply (7.26) (for $i = 1$).

Suppose now that (7.26) is true for a certain $i \in [1 : m]$. Let us show that

$$u(\tau) = (S^{-\delta}(v))(\tau) \quad \text{a.e.} \quad \tau \in [t_0, \tau_{i+1}] \qquad (7.28)$$

(it will complete the proof). It follows from the assumption (7.26) that

$$y|\tau_i = x(\cdot\,|u,v)|\tau_i = x(\cdot\,|S^{-\delta}(v),v)|\tau_i \,. \qquad (7.29)$$

Denote the function given in (7.29) by x. We see that x satisfies the inclusion (7.23). Therefore according to (7.24)

$$u(\tau) = (\mathcal{U}_i(x))(\tau) = (S^{-\delta}(v_*))(\tau)$$
$$\text{a.e.} \quad \tau \in [\tau_i, \tau_{i+1}] \qquad (7.30)$$

where $v_* \in W$ satisfies (7.25). From (7.25), (7.29) and (7.30) follows

$$x(\cdot\,|S^{-\delta}(v),v)|\tau_i = x(\cdot\,|S^{-\delta}(v_*),v_*)|\tau_i \,.$$

Thus, by Lemma 7.5

$$(S^{-\delta}(v))(\tau) = (S^{-\delta}(v_*))(\tau) \quad \text{a.e.} \quad \tau \in [t_0, \tau_{i+1}] \,.$$

The last equality, (7.30) and (7.26) imply (7.28).

The next lemma will be given in a form more general than it is necessary for the proof of Theorem 6.1 (we shall use it in Section 9 considering another variant of the problem). The restriction of a quasi-strategy S on W to a (non-empty) set $W' \subset W$ will be denoted by $S|W'$ (it is clear that $S|W'$ is a quasi-strategy on W'); the notation $\rho(S,W') = \rho(S|W',W')$ will also be used. If S'

is a strategy, S is a quasi-strategy on W, $\varnothing \neq W' \subset W$, $\varepsilon > 0$, and $\rho(S,W') = -\infty$ (the latter can not be excluded, in general), then the inequality

$$\rho(S',W') \le \rho(S,W') + \varepsilon \qquad (7.31)$$

will mean that its left hand side is $-\infty$.

Note that the following lemma does not require W to be $L^{2,q}$-compact.

Lemma 7.7. *Let Conditions* 6.1, 6.2 *and* 6.3 *be fulfilled, and the functional* J *be uniformly* (L^r,δ)*-continuous on* $X(W)$. *Then there exists a mapping* $\tau_W :]0,\vartheta_0-t_0[\times Q(W) \longmapsto \mathcal{T}$ *with the following property: for any* $\varepsilon > 0$ *and non-empty* $L^{2,q}$*-compact set* $W' \subset W$ *there exists a* $\delta_0 \in]0,\vartheta_0-t_0[$ *such that for any* $\delta \in]0,\delta_0]$ *and* $S \in Q(W)$ *the strategy* $S' = \tau_W(\delta,S)$ *satisfies the inequality* (7.31).

Proof. Let for any $\delta \in]0,\vartheta_0-t_0[$ and $S \in Q(W)$, $\tau_W(\delta,S) = S'$ where S' is a strategy satisfying the equality (7.22) for each $v \in W$; Lemma 7.6 ensures the existence of S'. Fix an arbitrary $\varepsilon > 0$ and a non-empty $L^{2,q}$ compact set $W' \subset W$. Putting $X' = X(W')$ denote the values written out in (6.4) and (7.1) by $\alpha(\beta)$ and $\sigma(\delta)$, respectively. Take an $\beta > 0$ such that $\alpha(\beta) < \varepsilon$ (β exists, since J is uniformly (L^r,δ)-continuous on X'), and choose a $\delta_0 \in]0,\vartheta_0-t_0[$ such that $\sup\{\sigma(\delta) : \delta \in]0,\delta_0]\} < \beta$ (δ_0 exists by Lemma 7.3). Consider arbitrary $\delta \in]0,\delta_0]$ and $S \in Q(W)$. Let y be an arbitrary element from $X(S|W',W')$, i.e. $y = x(\cdot\,|S(v),v)$ for a certain $v \in W'$. Then $x = x(\cdot\,|(S(v))^{-\delta},v) = x(\cdot\,|(S|W')^{-\delta}(v),v) \in X((S|W')^{-\delta},W')$; by the definition of $\sigma(\delta)$, $|\dot{x}^{+\delta} - \dot{y}|_{L^r,n}$

$\leq \sigma(\delta) < \beta$. Hence $|J(x) - J(y)| \leq \alpha(\beta) < \varepsilon$. But by the definition of $S' = \mathcal{T}_W(\delta,S)$ (see (7.22)) $x \in X(S',W')$. Therefore, due to the arbitrary choice of y we have the inequality (7.31).

Proof of Theorem 6.1. Let all assumptions of Theorem 6.1 be fulfilled. Taking into consideration that W is compact in $L^{2,q}$ and using Lemma 7.7 we conclude that for any $\varepsilon > 0$ and $S \in Q(W)$ there exists a strategy S' ($S' = \mathcal{T}_W(\delta, S)$ for a sufficiently small δ) such that $\rho(S',W) \leq \rho(S,W) + \varepsilon$. Since S and ε are arbitrary, we have the inequality $\rho(\mathcal{T}, W) \leq \rho(Q(W))$ which completes the proof (recall Corollary 4.1).

8. c-Uniform Ensured Result

Denote by $comp(V)$ the set of all non-empty $L^{2,q}$-compact subsets of V , put $Q = Q(V)$, and denote by $\mathcal{T}^\Delta$ (Q^Δ, resp.) the set of all families $\mathbb{S} = (S_\delta)_{\delta>0}$ of elements of $\mathcal{T}$ (Q , resp.).

The value

$$\rho^\Delta(\mathbb{S}) = \sup_{W \in comp(V)} \overline{\lim_{\delta \to 0}} \rho(S_\delta, W') \tag{8.1}$$

where $\mathbb{S} = (S_\delta)_{\delta>0} \in \mathcal{T}^\Delta \cup Q^\Delta$ will be called *the c-uniform ensured result* for $\mathbb{S}$. The values

$$\rho_o^\Delta(\mathcal{T}) = \inf \left\{ \rho^\Delta(\mathbb{S}) : \mathbb{S} \in \mathcal{T}^\Delta \right\} \tag{8.2}$$

and

$$\rho_o^{\Delta}(\mathcal{Q}) = \inf \left\{ \rho^{\Delta}(\mathbb{S}) : \mathbb{S} \in \mathcal{Q}^{\Delta} \right\} \qquad (8.3)$$

will be called *optimal c-uniform results* for the classes $\mathcal{S}$ (of strategies) and $\mathcal{Q}$ (quasi-strategies), respectively.

Theorem 8.1.

$$\rho_o^{\Delta}(\mathcal{Q}) \le \rho_o(\mathcal{Q}) \quad , \qquad (8.4)$$

$$\rho_o^{\Delta}(\mathcal{S}) \le \rho_o(\mathcal{S},V) \quad , \qquad (8.5)$$

$$\rho_o^{\Delta}(\mathcal{S}) \ge \rho_o^{\Delta}(\mathcal{Q}) \quad . \qquad (8.6)$$

Proof. For an arbitrary quasi-strategy $\mathbb{S}$ on V and an $\mathbb{S} = (S_\delta)_{\delta>0}$ where $S_\delta = S$, it holds $\rho^{\Delta}(\mathbb{S}) \le \rho(S,V)$ (see (8.1) and (3.6)). Hence (see (8.3)) $\rho_o^{\Delta}(\mathcal{Q}) \le \rho(S,V)$, which due to the arbitrary choice of S implies (8.4). The inequality (8.5) can be proved in a similar way. Take an arbitrary strategy S. Like in the proof of Theorem 4.1, we build up a quasi-strategy S' on V such that $X(S',W') = X(S,W')$ for any $W' \subset V$. This obviously implies (8.6).

The inequalities (8.4) and (8.5) may in general be strict, if the functional J is uniformly (L^r,δ)-continuous on $X(W')$ for each $W' \in \mathrm{comp}(V)$.

Example 8.1. Let $n = 3$, $I = [-1,1]$, $P = Q = \{-1, 1\}$, the system (3.1), (3.2) be of the form

$$\dot{x}_1(t) = u(t) \ ,$$

$$\dot{x}_2(t) = \begin{cases} |x_1(t)| \ , & t \le 0 \\ 0 \ , & t > 0 \end{cases} ,$$

$$\dot{x}_3(t) = v(t) \ ,$$

$$x_1(-1) = x_2(-1) = x_3(-1) = 0 \ .$$

For each function $x = (x_1, x_2, x_3) \in X$ we put

$$J(x) = \int_0^1 |\dot{x}_3(t+|x_2(0)|) - \dot{x}_1(t)|\,dt + x_2(0) \ .$$

It can easily be proved that J is uniformly (L^1,δ)-continuous on $X(W')$ for each $W' \in \mathrm{comp}(V)$. Let us show that

$$\rho_o^{\Delta}(\mathcal{Q}) = \rho_o^{\Delta}(\mathcal{S}) = 0 \ . \tag{8.7}$$

For each $\delta > 0$ let $m(\delta) = \min \{ m \in \mathbb{N} : 1/m > \delta \}$ and the strategy $S_\delta = ((\tau_{\delta,i}, u_{\delta,i}))_{i \in [0:2m(\delta)-1]}$ be determined by the following conditions: $\tau_{\delta,i} = -1 + i/m(\delta)$, $(u_{\delta,i}(x))(t) = (-1)^i$ $(i < m(\delta))$, $u_{\delta,i}(x) = \dot{x}_3|\tau_{\delta,i-1}$, $\tau_{\delta,i}$ $(i \ge m(\delta))$. For any $v \in V$ the trajectory $x = x(\cdot\,|S_\delta, v)$ satisfies the relations $|x_1(t)| \le \delta$ $(t \in [0,1])$, $|x_2(0)| \le \delta$, $\dot{x}_1(t) = \dot{x}_3(t-\delta) = v(t-\delta)$ a.e. $t \in [0, 1]$. Thus, for each $W' \in \mathrm{comp}(V)$ we have

$$\rho(S_\delta, W') = \sup_{v \in W'} J(x(\cdot\,|S_\delta, v)) \le \sup_{v \in W'} \sup_{\varepsilon \in [0,\delta]} \int_0^1 |v(t+\varepsilon) -$$

$$v(t-\delta)|\,dt + \varepsilon \longrightarrow 0 \quad \text{as} \quad \delta \longrightarrow 0 \ .$$

Therefore, $\rho_o^{\Delta}(\mathcal{S}) = 0$ which due to (8.6) implies (8.7).

Let us show that $\rho_0(Q) \geq 1/2$ (the inequality (8.4) and consequently the inequality (8.5) (see (8.6)) are strict). Consider an arbitrary quasi-strategy S on V and show that

$$\rho(S,V) \geq 1/2 . \qquad (8.8)$$

Introduce finite sequences $(v_j)_{j=0}^k$ of disturbances and $(u_j)_{j=0}^k$ of controls. Put $v_0(t) = 1$ $(t \in I)$, $u_0 = S(v_0)$, and $\delta = |x_2(0)|$ where x_2 is the second component of $x = x(\cdot\,|u_0,v_0)$ (it is clear that $\delta \in]0, 1/2]$). Determine $k \in \mathbb{N}$ by $k\delta \geq 1$, $(k-1)\delta < 1$ and denote $\tau_j = j\delta$ $(j \in [0 : k-1])$, $\tau_k = 1$. If $v_j \in V$ and $u_j \in U$ for a $j \in [0 : k-1]$ are defined, then determine $v_{j+1} \in V$ and $u_{j+1} \in U$ by the conditions $v_{j+1}(t+\delta) = -u_j(t)$ $(t \in]\tau_{j-1}, \tau_j])$, $v_{j+1}|\tau_j = v_j|\tau_j$, $u_{j+1} = S(v_{j+1})$. It follows from the first condition that $|v_{j+1}(t+\delta) - v_j(t)| = 2$ $(t \in]\tau_{j-1}, \tau_j]$; two other conditions imply $u_{j+1}|\tau_j = u_j|\tau_j$. From these relations we deduce the following inequalities for $x = x(\cdot\,|S,v_k)$:

$$J(x) \geq \int_0^{\tau_{k-1}} |v_k(t+\delta) - u_k(t)|\,dt + \delta \geq$$

$$\sum_{j=1}^{k-1} \int_{\tau_{j-1}}^{\tau_j} |v_{j+1}(t+\delta) - u_j(t)|\,dt + \delta \geq$$

$$2(k-2)\delta + \delta = 2k\delta - 3\delta \geq 1/2 .$$

This implies (8.8).

Let us provide conditions ensuring the inequality

(8.4) to turn into the equality. In this Section we assume that the functional J is defined on the closure X of the set X in C; its uniform C-continuity on X is defined by (5.1) where $X' = X$. Introduce the following strengthened variant of Conditions 6.1 and 6.2

Condition 8.1. The set X is bounded in C, and each $\mu \in RU$ and $v \in V$ generate the single trajectory.

Now we give several definitions assuming that Condition 8.1 is fulfilled; the trajectory generated by a $\mu \in RU$ and a $v \in V$ will be denoted by $x(\cdot|\mu,v)$ (note that, since U is dense in RU [5, Theorem IV.2.6], $x(\cdot|\mu,v) \in X$). Denote by $\mathcal{P}(RU)$ the set of all non-empty closed subsets of RU. The closure of a set E in RU will be denoted by clE. A mapping $H : V \mapsto \mathcal{P}(RU)$ such that for each $v',v'' \in V$ and $t \in I$ $v'|t = v''|t$ implies $H(v')|t = H(v'')|t$ will be called *a generalized quasi-strategy* (on V). The set of all generalized quasi-strategies will be denoted by $R\mathcal{Q}$. *The ensured result for a generalized quasi-strategy* H define as

$$\rho(H) = \sup \{ J(x(\cdot|\mu,v)) : \mu \in H(v), v \in V \};$$

the optimal ensured result in $R\mathcal{Q}$ is

$$\rho_o(R\mathcal{Q}) = \inf \{ \rho(H) : H \in R\mathcal{Q} \}.$$

It is clear that $\rho_o(R\mathcal{Q}) \le \rho_o(\mathcal{Q})$.

Theorem 8.2. *Let Condition* 8.1 *be fulfilled and the functional* J *be uniformly* C*-continuous on* X. *Then* $\rho_o(R\mathcal{Q}) = \rho_o(\mathcal{Q})$.

Proof. If Conditions 5.1, 5.2 and 5.3 are fulfilled (this implies Condition 8.1), then the theorem follows from [8, Lemma 96.1]. If Condition 8.1 is fulfilled, then a statement analogous to the above Lemma (implying the statement of the theorem) can be proved by the method of [12]; we omit the details.

Lemma 8.1. *Let Condition* 8.1 *be fulfilled, the functional* J *be uniformly* C*-continuous on* X , $\mathcal{S} = (S_\delta)_{\delta>0} \in \mathcal{Q}^\Delta$ *and* $H : V \longmapsto \mathcal{P}(RU)$ *be of the form*

$$H(v) = \bigcap_{S\in\mathbb{N}} \mathrm{cl}\{ S_\delta(v) : \delta \in [0, 1/s] \} \quad (v \in V) \tag{8.9}$$

Then $H \in R\mathcal{Q}$ *and* $\rho^\Delta(\mathcal{S}) \geq \rho(H)$.

Proof. Prove the first relation. Let $v', v'' \in V$, $t \in I$ and

$$v'|t = v''|t \ . \tag{8.10}$$

Let us show that $H(v')|t = H(v'')|t$. Consider an arbitrary $\mu' \in H(v')$. So far as v' and v'' are arbitrary, it is sufficient to prove that

$$\mu'|t \in H(v'')|t \ . \tag{8.11}$$

According to (8.9), there exist $\delta_s \in]0, 1/s]$ $(s \in \mathbb{N})$ such that $u'_s = S_{\delta_s}(v') \longrightarrow \mu'$ in RU . Consider the sequence (u''_s) , $u''_s = S_{\delta_s}(v'')$. Since it is compact in RU (Lemma 7.1, 1)), we have $u''_{s_j} \longrightarrow \mu''$ in RU for a

certain subsequence. As it is seen from (8.9), $\mu'' \in H(v'')$. The equality (8.10) yields $u'_S|t = u''_S|t$, thus, $\mu'|t = \mu''|t$, and (8.11) is proved.

Put

$$\varkappa(v) = \overline{\lim_{\delta \to 0}} J(x(\cdot\,|S_\delta(v),v)) \quad (v \in V). \tag{8.12}$$

Let us show that

$$\rho^\Delta(S) \geq \sup_{v \in V} \varkappa(v) \geq \rho(H) \tag{8.13}$$

(it will complete the proof). The first inequality (8.13) follows obviously from (8.1) and (8.12). Now it is sufficient to show that for any $v \in V$

$$\varkappa(v) \geq \varkappa^*(v) \tag{8.14}$$

where

$$\varkappa^*(v) = \sup \{ J(x(\cdot\,|\mu,v)) : \mu \in H(v) \}. \tag{8.15}$$

Taking (8.12) into account we have

$$\varkappa(v) = \lim_{s \to \infty} \sup_{\delta \in]0,1/s]} J(x(\cdot\,|S_\delta(v),v)) =$$
$$\lim_{s \to \infty} \sup \left\{ J(x(\cdot\,|u,v)) : u \in \{ S_\delta(v) : \delta \in]0,1/s] \} \right\} =$$
$$\lim_{s \to 0} \sup \left\{ J(x(\cdot\,|\mu,v)) : \mu \in \mathrm{cl}\{S_\delta(v) : \delta \in]0,1/s]\} \right\}.$$

The last equality follows from the uniform C-continuity of J on X and Lemma 7.1, 2). Take arbitrary $\varepsilon > 0$ and $s \in \mathbb{N}$ such that

$$\varkappa(v) \geq \sup \left\{ J(x(\cdot\,|\mu,v) : \mu \in \mathrm{cl}\{S_\delta(v) : \delta \in]0,1/s[\} \right\} - \varepsilon$$

The right side of this inequality is no smaller than $\varkappa^*(v) - \varepsilon$ as it is seen from (8.15) and (8.9). Hence due to the arbitrary choice of ε we have (8.14).

Theorem 8.3. *Let Condition* 8.1 *be fulfilled and the functional* J *be uniformly* C*-continuous on* X . *Then*

$$\rho_o^{\Delta}(Q) = \rho_o(Q) . \qquad (8.16)$$

Proof. Lemma 8.1 implies $\rho_o^{\Delta}(Q) \geq \rho_o(RQ)$. This inequality and Theorem 8.2 lead to the inequality opposite to (8.4). Since (8.4) is true (Theorem 8.1), we have (8.16).

9. Conditions for c-Uniform Unimprovability of the Class

The class $\mathcal{S}$ will be called c*-uniformly unimprovable*, if the inequality (8.6) turns into the equality. Introduce the following strengthened variant of Condition 6.3.

Condition 9.1. For each $t \in I$, $u \in U$, $v', v'' \in V$ such that $x(\cdot\,|u,v')|t = x(\cdot\,|u,v'')|t$ it holds $v'|t = v''|t$.

Theorem 9.1. *Let Conditions* 8.1 *and* 9.1 *be fulfilled and the functional* J *be uniformly* (L^r,δ)*-continuous on* X . *Then the class* $\mathcal{S}$ *is* c*-uniformly unimprovable.*

Proof. Note that Conditions 8.1 and 9.1 imply that

the set $W = V$ satisfies Conditions 6.1, 6.2 and 6.3. Therefore, for this set the statement of Lemma 7.6 is true. Fix a mapping $\tau_V :]0, \vartheta_0 - t_0[\times Q \mapsto \mathcal{P}$ given by statement of Lemma 7.7. Consider an arbitrary $S = (S_\delta)_{\delta>0} \in Q^\Delta$ and an $\varepsilon > 0$. Determine the family $S' = (S'_\delta)_{\delta>0} \in \mathcal{P}^\Delta$ by $S'_\delta = \tau_V(\delta, S_\delta)$ $(\delta \in]0, \vartheta_0 - t_0[)$. By the definition of τ_V for any $W' \in \mathrm{comp}(V)$ there exists a $\delta_0 \in]0, \vartheta_0 - t_0[$ such that for each $\delta \in]0, \delta_0[$ it holds $\rho(S'_\delta, W') \le \rho(S_\delta, W') + \varepsilon$. Thus,

$$\overline{\lim_{\delta \to 0}} \rho(S'_\delta, W') \le \overline{\lim_{\delta \to 0}} \rho(S_\delta, W') .$$

Hence (see (8.1)) due to the arbitrary choice of W', $\rho^\Delta(S') \le \rho^\Delta(S)$. This implies $\rho_0^\Delta(\mathcal{P}) \le \rho_0^\Delta(Q)$, since $S \in Q^\Delta$ is arbitrary. The last inequality and (8.6) complete the proof.

Theorem 9.1 (together with Remark 6.3), Theorem 8.3 and Theorem 8.1 (see (8.5)) yield

Corollary 9.1. *Let Conditions* 8.1 *and* 9.1 *be fulfilled and the functional* J *is uniformly* C*-continuous on* X. *Then*

$$\rho_0^\Delta(\mathcal{P}) = \rho_0^\Delta(Q) = \rho_0(Q) \le \rho(\mathcal{P}, V) .$$

Remark 9.1. Under the conditions of Corollary 9.1 the last inequality is strict, if and only if the class $\mathcal{P}$ is not unimprovable on V (note that $\mathcal{P}$ is c-uniformly unimprovable) ; such a situation is described in Example 5.1.

REFERENCES:

1. Krasovskii, N.N., *Controlling of a dynamical system. The problem of the minimum of the ensured result.* Moscow, Nauka, 1985 (Russian).
2. Subbotin, A.I. and Chentsov, A.G., *Optimization of a guarantee in the problems of control.* Moscow, Nauka, 1981 (Russian).
3. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mistchenko, E.F., *Mathematical theory of control processes.* Moscow, Nauka, 1969 (Russian).
4. Gamkrelidze, R.V., *On chattering optimal controls,* Dokl. Akad. Nauk SSSR, Vol. 143 (1962), No 6, 1243-1245 (Russian).
5. Warga, J., *Optimal control of differential and functional equations,* Academic Press, New York and London (1972).
6. Isaacs, R., *Differential games,* Academic Press, New York, (1965).
7. Krasovskii, N.N., *Game problems on meeting of motions,* Moscow, Nauka, 1970 (Russian).
8. Krasovskii, N.N. and Subbotin, A.I., *Game-theoretical control problems,* Spinger, New York etc., 1988.
9. Tikhonov, A.N., *On functional Volterra equations and applications to certain problems of mathematical pysics.* Bull. of Moscow Univ. (A), 1 (1938) (Russian).
10. Rull-Nardzevski, C., *A theory of pursuit and evasion,* Advances in Game Theory, Princeton Univ. Press, 1964, 113-126.
11. Roxin, E., *Axiomatic approach in differential games,* J.Opt. Theory Appl., Vol.3 (1969), No 3, 153-163.
12. Kryazhimskii, A.V., *On the theory of positional*

differential games of convergence-evasion, Soviet Math. Dokl. Vol. 19 (1978), No 2, 408–412.

A.V. Kryazhimskii
Institute of Mathematics
and Mechanics
Kovalevskoi 16
Sverdlovsk, 620219
USSR

Constantin Carathéodory: An International Tribute (pp. 676–712)
edited by Th. M. Rassias

LIMITS OF RANDOM MEASURES INDUCED BY AN ARRAY OF INDEPENDENT RANDOM VARIABLES

Hiroshi Kunita

Let $\{\xi_{n,j}\}$, $n,j=1,2,\ldots$ be an array of random variables such that for each n, $\xi_{n,j}$, $j=1,2,\ldots$ are independent with identical distributions. The object of this paper is to study the weak covergences of three sequences of random measures induced by $\{\xi_{n,j}\}$. The first two are

$$B_n(t,E) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \{\chi_E(\xi_{n,j}) - E[\chi_E(\xi_{n,j})]\}, \quad N_n(t,F) = \sum_{j=1}^{[nt]} \chi_F\left(\frac{\xi_{n,j}}{\sqrt{n}}\right),$$

where χ_E is the indicator function of the Borel set E. Under some conditions we will show that the sequence $\{B_n(t,E)\}$ converges weakly to a Gaussian random measure and the sequence $\{N_n(t,E)\}$ converges weakly to a Poisson random measure. These weak convergences will be applied to the the study of the limit theorem for sums of nonlinear functions of $\xi_{n,j}$:

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} f_n(\xi_{n,j}) - a_n(t),$$

where $\{a_n(t)\}$ are certain centralizing functions.

0. Introduction

There are extensive works on the limit theorems for sums of independent random variables. Let $\{\xi_{n,j}\}$, $n,j=1,2,\ldots$ be an

array of $\mathbb{R}^m$- valued random variables such that for each n, $\xi_{n,j}$, j=1,2,... are independent with the identical distribution π_n (i.i.d. random variables). Then one of the typical limit theorems states that under a suitable conditions on the sequence $\{\pi_n\}$, the distributions of the linear sums $(1/\sqrt{n})\Sigma_{j=1}^{n}\xi_{n,j}-a_n$ converge to an infinitely divisible distribution, where $\{a_n\}$ is a sequence of suitable centralizing constants. See Gnedenko-Kolmogorov's book [2]. More strongly, the sequence of stochastic processes $Y_n(t)=(1/\sqrt{n})\Sigma_{j=1}^{[nt]}\xi_{n,j}-a_nt$, $t\geq 0$ converges to a time homogeneous process with independent increments. Here [nt] is the integer part of the number nt. See Jacod-Klopotowski-Memin [5] and Jacod-Shiryaev [6].

In this article, we will study the limit theorem for sums of non-linear functions of $\xi_{n,j}$:

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} f_n(\xi_{n,j}) - a_n(t), \tag{0.1}$$

where $\{f_n\}$ are continuous $\mathbb{R}^d$-valued functions and $\{a_n(t)\}$ are certain centralizing $\mathbb{R}^d$-valued deterministic functions. Since $f_n(\xi_{n,j})$, j=1,2,... are i.i.d. random variables for each n, it is clear that the limit should be a process with independent increments if it exists. We wish to know how the limit is related to the functions $\{f_n\}$ and the array of random variables $\{\xi_{n,j}\}$. To study this problem, we will introduce three sequences of random measures induced by the array $\{\xi_{n,j}\}$ and discuss their weak convergences.

We first consider a sequence of random measures with time parameter t defined by

$$(0.2) \qquad B_n(t,E) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \left(\chi_E(\xi_{n,j}) - \pi_n(E) \right),$$

where $\chi_E(\lambda)$ is the indicator function of the Borel set E. It may be regarded as a stochastic process with values in additive set functions. It's mean is 0 and covariance is

$$E(B_n(t,E_1)B_n(t,E_2)) = t\left(\pi_n(E_1 \cap E_2) - \pi_n(E_1)\pi_n(E_2)\right).$$

Then, setting

$$a_n(t) \equiv \frac{[nt]}{\sqrt{n}} \int_{\mathbb{R}^m} f(\lambda)\pi_n(d\lambda),$$

$X_n(t)$ is represented as the integral of the function $f_n(\lambda)$ by the random measure $B_n(t,d\lambda)$, i.e.,

$$(0.3) \qquad X_n(t) = \int_{\mathbb{R}^m} f_n(\lambda) B_n(t,d\lambda).$$

We can show that if the sequence $\{\pi_n\}$ converges weakly to a distribution π, then the sequence of stochastic processes $\{B_n(t,E)\}$ converges weakly and the limit $B(t,E)$ is a Gaussian random measure for any t, whose mean is 0 and covariance is $t(\pi(E_1 \cap E_2) - \pi(E_1)\pi(E_2))$. See Section 2, Theorem 2.1.

Now suppose that the sequence $\{f_n\}$ converges to a function

f uniformly on compact sets. A simple question is whether the weak limit of stochastic processes $\{X_n(t)\}$ exists and the limit $X(t)$ is represented by

$$(0.4) \qquad X(t) = \int_{\mathbb{R}^m} f(\lambda)B(t,d\lambda).$$

The answer is yes if $f(\lambda)$ is a bounded function or more generally if $\lim_{|\lambda|\to\infty}|f(\lambda)|/|\lambda|=0$. In this case $X(t)$ is a Gaussian random variable for any t and moreover, the stochastic process $X(t)$, $t\geq 0$ is a Wiener process. However if the above limit is not zero, the representation (0.4) does not hold in general. At least a Poisson random measure will be needed for the representation of $X(t)$, if the distribution of $X(t)$ is not Gaussian.

We will obtain an integral representation similar to (0.4) by introducing another two random measures. The second random measure is defined on $\mathbb{R}^m-\{0\}$ by

$$(0.5) \qquad N_n(t,F) = \sum_{j=1}^{[nt]} \chi_F\left(\frac{\xi_{n,j}}{\sqrt{n}}\right).$$

The expectation of $N_n(1,F)$ is given by

$$(0.6) \qquad \mu_n(F) \equiv nP\left(\frac{\xi_{n,j}}{\sqrt{n}}\in F\right) = n\pi_n(\sqrt{n}F).$$

It is a Radon measure on $\mathbb{R}^m-\{0\}$. We call that the sequence $\{\mu_n\}$

converges vaguely to a Radon measure μ if $\int f d\mu_n$ converges to $\int f d\mu$ for any bounded continuous function f on $\mathbb{R}^m$ such that its support is included in $\mathbb{R}^m-\{0\}$. Now if the seqeunce $\{\mu_n\}$ converges vaguely to μ, then we can show that the sequence of stochastic processes $\{(B_n(t,E),N_n(t,E))\}$ converges weakly to $(B(t,E),N(t,F))$ where $N(t,F)$ is a Poisson random measure with mean $t\mu(F)$ for any t. Moreover $B(t,E)$ and $N(t,F)$ are independent processes. See Section 2, Theorem 2.1.

In order to introduce the third random measure, it is useful to represent points of $\mathbb{R}^m-\{0\}$ by polar coordinate (r,θ), where $r\in(0,\infty)$ and $\theta\in S^{m-1}$ (m-1 dimensional unit sphere with center 0). Then $\mathbb{R}^m-\{0\}$ is homeomorphic to $(0,\infty)\times S^{m-1}$ and $(0,\infty]\times S^{m-1}\cup\{0\}$ is a compactification of $\mathbb{R}^m$. We can regard $\{\infty\}\times S^{m-1}$ as the boundary of $\mathbb{R}^m$ and denote it by $\partial\mathbb{R}^m$. Now for $\varepsilon>0$ let $\chi_\varepsilon(t)$, $t\geq 0$ be the function such that $\chi_\varepsilon(t)=1$ if $t<\varepsilon$, $=0$ if $t\geq\varepsilon$. Define the random measures by

$$K_n^\varepsilon(t,G) = \int_G (1+|\lambda|)\chi_\varepsilon\left(\frac{|\lambda|}{\sqrt{n}}\right)B_n(t,d\lambda). \tag{0.7}$$

Its mean is 0 and covariance of $K_n^\varepsilon(t,G_1)$ and $K_n^\varepsilon(t,G_2)$ is $t(\nu_n^\varepsilon(G_1\cap G_2)-\xi_n^\varepsilon(G_1)\xi_n^\varepsilon(G_2))$, where

$$\nu_n^\varepsilon(G) \equiv \int_G (1+|\lambda|)^2\chi_\varepsilon\left(\frac{|\lambda|}{\sqrt{n}}\right)\pi_n(d\lambda), \tag{0.8}$$

$$\xi_n^\varepsilon(G) \equiv \int_G (1+|\lambda|)\chi_\varepsilon\left(\frac{|\lambda|}{\sqrt{n}}\right)\pi_n(d\lambda). \tag{0.9}$$

Assuming that the sequence $\{\nu_n^{\varepsilon}\}$ converges weakly in $\bar{\mathbb{R}}^m$ as $n\to\infty$, we can show that the sequence $\{(K_n^{\varepsilon}(t,G),N_n(t,F)\}$ converges weakly to $(K^{(\varepsilon)}(t,G),N(t,F))$ and the measure $K^{\varepsilon}(t,G)$ is decomposed to the sum of the three:

$$(0.10)\qquad K^{\varepsilon}(t,G) = \int_{G\cap\mathbb{R}^m}(1+|\lambda|)^{-1}\tilde{B}(t,d\lambda) + \int_{G\cap\partial\mathbb{R}^m}K^{(\infty)}(t,d\lambda)$$

$$+ \int_{G\cap(\mathbb{R}^m-\{0\})}|\lambda|\chi_{\varepsilon}(|\lambda|)\tilde{N}(t,d\lambda),$$

where $\tilde{B}(t,d\lambda)$ is equivalent to $B(t,d\lambda)$ in the sense of law, $K^{(\infty)}(t,d\lambda)$ is a Gaussian random measure supported by the boundary $\partial\mathbb{R}^m$ and $\tilde{N}(t,E)= N(t,E)-t\mu(E)$. See Section 2, Theorem 2.2.

Now suppose that the sequence $\{f_n\}$ defining (0.1) converges to f uniformly on compact sets and that the asymptotic function

$$(0.11)\qquad h(\lambda) = \lim_{n\to\infty}\frac{f_n(\sqrt{n}\lambda)}{\sqrt{n}|\lambda|+1}$$

exists (uniformly on comapct sets of $\mathbb{R}^m-\{0\}$) and $h(0,\theta)=\lim_{r\to\infty}h(r,\theta)$ exists. Our goal is to show that $\{(K_n^{(\varepsilon)}(t),N_n(t),X_n(t))\}$ converges weakly to $(K^{(\varepsilon)}(t),N(t),X(t))$ and $X(t)$ is represented by

$$(0.12)\qquad X(t) = \int_{\mathbb{R}^m}f(\lambda)\tilde{B}(t,d\lambda) + \int_{S^{m-1}}h(0,\theta)K^{(\infty)}(t,d\theta)$$

$$+ \int_{\mathbb{R}^m-\{0\}} |\lambda| h(\lambda) \chi_\varepsilon(|\lambda|) \tilde{N}(t,d\lambda)$$

$$+ \int_{\mathbb{R}^m-\{0\}} |\lambda| h(\lambda)(1-\chi_\varepsilon(\lambda)) N(t,d\lambda).$$

The proof will be given at Section 3.

In particular, if the random variables $\xi_{n,j}$ satisfy $\sup_n E[|\xi_{n,j}|^{2+\delta}]<\infty$ for some $\delta>0$, then both $K^{(\infty)}$ and $N(t,\cdot)$ are identically zero. Hence we again obtain the representation (0.4). This will be shown at the end of Section 3.

Finally we will consider the special case where $f_n(\lambda)\equiv f(\lambda) \equiv \lambda$. We have $f(\lambda)=|\lambda|h(\lambda)=\lambda$. Therefore $X(t)$ is represented by

$$(0.13) \qquad X(t) = \int_{\mathbb{R}^m} \lambda \tilde{B}(t,d\lambda) + \int_{S^{m-1}} \theta K^{(\infty)}(t,d\theta)$$

$$+ \int_{\mathbb{R}^m-\{0\}} \lambda \chi_\varepsilon(|\lambda|) \tilde{N}(t,d\lambda)$$

$$+ \int_{\mathbb{R}^m-\{0\}} \lambda(1-\chi_\varepsilon(\lambda)) N(t,d\lambda).$$

Set

$$(0.14) \qquad X_c(t) = \int_{\mathbb{R}^m} \lambda \tilde{B}(t,d\lambda) + \int_{S^{m-1}} \theta K^{(\infty)}(t,d\theta).$$

It is a Brownian motion and the above (0.12) is written as

$$(0.15) \qquad X(t) = X_c(t) + \int_{\mathbb{R}^m-\{0\}} \lambda \chi_\varepsilon(|\lambda|) \tilde{N}(t,d\lambda)$$

$$+ \int_{\mathbb{R}^m-\{0\}} \lambda(1-\chi_\varepsilon(\lambda))N(t,d\lambda).$$

This coincides with the Lévy-Itô's representation of the process with independent increments. See [4].

The representation (0.12) provides us an additional information on the limit distribution of sums of independent random variables. In fact (0.12) shows that the Gaussian part is related to the function $f(\lambda)$ directly and the Poisson part is related to the asymptotic function $h(\lambda)$ defined by (0.11). However in representation (0.13) this fact is not clear since $f(\lambda)$ and $|\lambda|h(\lambda)$ are the same function.

The next section is a preliminary part. We define the orthogonal and relatively orthogonal random measures. Then we define the law of a process with values in additive set functions.

1. Orthogonal random measures

Let Λ be a locally compact separable metric space. Let $\mathcal{S}(\Lambda)$ be a ring of subsets of Λ consisting of countable sets such that the σ-field generated by the ring $\mathcal{S}(\Lambda)$ coincides with the topological Borel field $\mathcal{B}(\Lambda)$ of Λ. We assume that any element of $\mathcal{S}(\Lambda)$ ie relatively compact.

Let $\{Z(E);E\in\mathcal{S}(\Lambda)\}$ be a family of real random variables defined on a probability space $(\Omega,\mathcal{F},P)$ having the finite additive property $Z(E_1\cup E_2)= Z(E_1)+Z(E_2)$ a.s. if $E_1\cap E_2=\phi$. Suppose

that $Z(E)$ is integrable for any E of $\mathcal{S}(\Lambda)$. Set $\mu(E)=E[Z(E)]$. Then it is an additive set function on $\mathcal{S}(\Lambda)$. We assume that μ is extended uniquely to a Radon (signed) measure on Λ. Then the system $\{Z(E);E\in\mathcal{S}(\Lambda)\}$ is called a *random measure with the mean measure* μ.

Now assume that $Z(E)$ is square integrable for any E of $\mathcal{S}(\Lambda)$ and satisfies the orthogonal property:

$$(1.1)\qquad E[(Z(E)-\mu(E))(Z(F)-\mu(F))] = 0 \quad \text{if} \quad E\cap F=\phi.$$

Set

$$(1.2)\qquad \pi(E) = E[(Z(E)-\mu(E))^2].$$

Then π is a positive additive set function because of the orthogonal property. We have further

$$(1.3)\qquad E[(Z(E)-\mu(E))(Z(F)-\mu(F))] = \pi(E\cap F), \qquad \forall E,F\in\mathcal{S}(\Lambda).$$

If the above π is extended to a Radon measure on Λ, Z is called an *orthogonal random measure with the characteristic* (μ,π).

We denote by $|\mu|$ the measure of the total variation of μ. The following proposition is more or less known.

Proposition 1.1. (1) Let $\{Z(E);E\in\mathcal{S}(\Lambda)\}$ be an orthogonal random measure with characteristic (μ,π). Then $Z(E)$ can be extended continuously to any E of $\mathcal{B}(\Lambda)$ such that $|\mu|(E)<\infty$ and

$\pi(E)<\infty$. Further,

$$(1.4) \qquad Z(\bigcup_{i=1}^{\infty} E_i) = \lim_{n\to\infty} \sum_{i=1}^{n} Z(E_i) \qquad \text{(in probability)}$$

holds for any disjoint sets E_i, $i=1,2,\ldots$ such that $|\mu|(\cup_{i=1}^{\infty}E_i)$ and $\pi(\cup_{i=1}^{\infty}E_i)$ are finite.

(2) If $f(\lambda)$ belongs to $L^1(|\mu|)\cap L^2(\pi)$, $\int f(\lambda)dZ(\lambda)$ is well defined. It satisfies

$$(1.5) \qquad E[\int f(\lambda)dZ(\lambda)] = \int f(\lambda)d\mu(\lambda),$$

$$(1.6) \qquad E[|\int f(\lambda)dZ(\lambda) - \int f(\lambda)d\mu(\lambda)|^2] = \int f(\lambda)^2 d\pi(\lambda).$$

We shall extend the orthogonal random measure. Let π and ν be Radon measures on Λ satisfying $\pi(E)\geq\nu(E)^2$ for any E of $\mathfrak{B}(\Lambda)$. A random measure $Z(E)$ is called *relatively* (π,ν)*-orthogonal* if $Z'(E)\equiv Z(E)-\mu(E)$ satisfies

$$(1.7) \qquad E[Z'(E)Z'(F)] = \pi(E\cap F)-\nu(E)\nu(F), \quad \forall E,F\in\mathfrak{B}(\Lambda).$$

A relatively (π,ν)-orthogonal random measure $Z(E)$ has properties similar to those of Proposition 1.1. Indeed, property (1) of Proposition 1.1 is valid. Further if f is a function belonging to $L^1(|\mu|)\cap L^2(\pi)\cap L^1(\nu)$, then $\int f(\lambda)dZ(\lambda)$ is well defined and it satisfies

$$(1.8)\qquad E[\,|\int f(\lambda)dZ(\lambda) - \int f(\lambda)d\mu(\lambda)|^2]$$

$$= \int f(\lambda)^2 d\pi(\lambda) - (\int f(\lambda)d\nu(\lambda))^2.$$

The triple (μ,π,ν) is called the *characteristic* of Z. Note that in the case where $\nu=0$, Z is an orthogonal random measure.

We give two important examples of relatively (π,ν)-orthogonal random measures. An relatively (π,ν)-orthogonal random measure $\{G(E);E\in\mathcal{B}(\Lambda)\}$ is called *Gaussian* if it is a Gaussian system of random variables. For a Gaussian orthogonal random measure, $G(E_1),\ldots,G(E_n)$ are independent whenever $E_1,\ldots,E_n$ are disjoint.

Next, let $\{N(E);E\in\mathcal{B}(\Lambda)\}$ be a random measure with values in nonnegative integers with mean measure μ. It is called a *Poisson random measure* if it satisfies two properties. (a) For any $E\in\mathcal{B}(\Lambda)$, $N(E)$ is subject to a Poisson distribution with intensity $\mu(E)$. (b) If $E_1,\ldots,E_n$ are disjoint, then $N(E_1),\ldots,N(E_n)$ are independent. It is then an orthogonal random measure. Since the mean and variance are the same for Poisson distributions, we have $\mu=\pi$.

Let $\{Z(t,E);t\in[0,\infty),E\in\mathcal{B}(\Lambda)\}$ be a family of real random variables such that for each fixed t, it is a relatively orthogonal random measure. Of course for fixed $E_1,\ldots,E_n$, $\{(Z(t,E_1),\ldots,Z(t,E_n));\ t\in[0,\infty)\}$ can be regarded as an $\mathbb{R}^n$ valued stochastic process. If it is continuous in probability and has independent increments with respect to time t for any $E_1,\ldots,E_n$, $\{Z(t,E);E\in\mathcal{B}(\Lambda)\}$ is called a *Lévy process with values in*

relatively orthogonal measures.

Let $Z(t,E)$ be a Lévy process with values in relatively orthogonal measures. We assume that $Z(t)$ is stationary, i.e., the law of $\{Z(t+h,E)-Z(h,E);E\in\mathcal{B}(\Lambda)\}$ does not depend on h and $Z(0,E)\equiv 0$ a.s. Its characteristic is $(t\mu,t\pi,t\nu)$, where (μ,π,ν) is the characteristic of $\{Z(1,E);E\in\mathcal{B}(\Lambda)\}$. Set $\tilde{Z}((s,t]\times E)=Z(t,E)-Z(s,E)$. Then $\tilde{Z}$ is extended uniquely to a random measure on the product space $[0,\infty)\times\Lambda$, whose characteristic is $(m\otimes\mu,m\otimes\pi,m\otimes\nu)$, where m is the Lebesgue measure on $[0,\infty)$. It is called the *time-space random measure associated with* $Z(t,E)$.

Finally we define the law of a random measure. Let $\mathcal{M}(\Lambda)$ be the set of all finitely additive set functions on $\mathcal{B}(\Lambda)$. It is a complete mertic space by the metric

$$(1.9)\qquad d(\mu,\nu)=\sum_{n=1}^{\infty}\frac{1}{2^n}\frac{|\mu(E_n)-\nu(E_n)|}{1+|\mu(E_n)-\nu(E_n)|},\qquad \{E_n\}=\mathcal{B}(\Lambda).$$

The law of the random measure Z is defined on $\mathcal{M}(\Lambda)$ by

$$(1.10)\qquad P(A)=P(\{\omega;Z\in A\}),\qquad A\in\mathcal{B}(\mathcal{M}(\Lambda)),$$

where $\mathcal{B}(\mathcal{M}(\Lambda))$ is the topological Borel field of $\mathcal{M}(\Lambda)$.

The law of a Gaussian random measure is determined by its characteristic (μ,π,ν). The law of a Poisson random measure is determined by its characteristic $(0,\mu,\mu)$.

Let $Z(t,E)$, $t\in[0,\infty)$, $E\in\mathcal{B}(\Lambda)$ be a stochastic process with

values in $\mathcal{M}(\Lambda)$, cádlág (right continuous with the left hand limits) with respect to t. Let $\mathbb{D}=\mathbb{D}([0,\infty);\mathcal{M}(\Lambda))$ be the set of all cádlág maps from $[0,\infty)$ into $\mathcal{M}(\Lambda)$. We associate $\mathbb{D}$ the Skorohod's J_1-topology (See Billingsley [1]). The law of $Z(t,E)$ is then defined on $\mathbb{D}$ in the same way as (1.10).

2. Weak convergence of random measures induced by i.i.d. random variables

Let $\{\xi_{n,j}\}$, $n,j=1,2,\ldots$ be an array of $\mathbb{R}^m$ valued random variables such that for each n $\xi_{n,j}$, $j=1,2,\ldots$ are independent, identically distributed (i.i.d). We will use the same notations as those in Introduction. We first introduce conditions for the sequences $\{\pi_n\}$ and $\{\mu_n\}$.

Condition (C.1) The sequence $\{\pi_n\}$ converges weakly to a probability distribution π.

Condition (C.2) The sequence $\{\mu_n\}$ converges vaguely to a Radon measure μ on $\mathbb{R}^m-\{0\}$.

Let $\mathcal{S}(\mathbb{R}^m)$ be a ring of $\mathbb{R}^m$ generating the Borel field of $\mathbb{R}^m$ For the later discussion it is convenient to assume that any element E of $\mathcal{S}(\mathbb{R}^m)$ satisfy $\lim_{n\to\infty}\pi_n(E)=\pi(E)$. (The condition is always satisfied if $\pi_n=\pi$ for any n or all E of $\mathcal{S}(\mathbb{R}^m)$ is a π-continuity set i.e. $\pi(\partial E)=0$ holds for any $E\in\mathcal{S}(\mathbb{R}^m)$ if $\pi_n\neq\pi$.) Simlarly let $\mathcal{S}(\mathbb{R}^m-\{0\})$ be a ring of $\mathbb{R}^m-\{0\}$ generating the Borel field of $\mathbb{R}^m-\{0\}$ such that any F of $\mathcal{S}(\mathbb{R}^m-\{0\})$ satisfies

$\lim_{n\to\infty}\mu_n(E)=\mu(E)$. The set of all fintely additive set functions on $\mathcal{B}(\mathbb{R}^m)$ (or $\mathcal{B}(\mathbb{R}^m-\{0\})$) is denoted by $\mathcal{M}(\mathbb{R}^m)$ (or $\mathcal{M}(\mathbb{R}^m-\{0\})$). Then the law of the pair (B_n,N_n) can be defined on the Skorohod space $\mathbb{D}=\mathbb{D}([0,\infty);\mathcal{M}(\mathbb{R}^m)\times\mathcal{M}(\mathbb{R}^m-\{0\}))$. The typical element of $\mathbb{D}$ is denoted by $B=B(t,E)$ and $N=N(t,F)$. We denote by P_n the law of (B_n,N_n). If the sequence $\{P_n\}$ converges weakly in $\mathbb{D}$, the sequence of pairs $\{(B_n,N_n)\}$ is said to converge weakly.

We wish to prove the following.

Theorem 2.1. Assume Conditions (C.1) and (C.2). Then the sequence of pairs $\{(B_n,N_n)\}$ converges weakly. Let $(B(t,E),N(t,F),P_\infty)$ be its weak limit. Then

(1) $\{B(t,E)\}$ is a Lévy process with values in relatively $(t\pi,t\pi)$-orthogonal measures. The associated time-space measure of $\{B(t,E)\}$ is Gaussian with characteristic $(0,m\otimes\pi,m\otimes\pi)$.

(2) $\{N(t,F)\}$ is a Lévy process with values in orthogonal measures. The associated time-space measure of $\{N(t,F)\}$ is a Poisson random measure with characteristic $(m\otimes\mu,m\otimes\mu)$.

(3) $\{B(t,E)\}$ and $\{N(t,F)\}$ are independent.

In order to establish the theorem, we need to prove two facts. The first is to prove that the sequence $\{P_n\}$ is tight, i.e., for any $\eta>0$ there exists a compact subset K of $\mathbb{D}$ such that $P_n(K)>1-\eta$ holds for all n. If it is shown, then the sequence $\{P_n\}$ contains a subsequence converging weakly. Let P_∞ be any limit measure. The second problem to show is that P_∞ is a solution of a certain martingale problem (See Proposition 2.3).

The theorem will then be proved by showing that the limit measure is unique and it has the property required in the theorem.

Proposition 2.2. The sequence $\{P_n\}$ is tight.

Proof. Obviously $(B_n(t,E),N_n(t,F))$ is a semimartingale adapted to the filtration $\mathcal{F}_t^n = \sigma(\xi_{n,j}; j \le [nt])$. We may apply a tightness criterion for a sequence of semimartingales. See Jacod-Shiryaev [6], Chapter VI, Section 4. We omit the detail since it is not difficult.

Proposition 2.3. Let P_∞ be an arbitrary weak limit of $\{P_n\}$. Set $B(t,\mathbb{E})=(B(t,E_1),\ldots,B(t,E_M))$ and $N(t,\mathbb{F})=(N(t,F_1),\ldots,N(t,F_N))$ where $E_p \in \mathcal{B}(\mathbb{R}^m), p=1,\ldots,M$ and $F_j \in \mathcal{B}(\mathbb{R}^m-\{0\}), j=1,\ldots,N$. Then for any C_b^2-function $F(x,y)$ on $\mathbb{R}^M \times \mathbb{R}^N$, the following is a martingale with respect to P_∞.

$$
\begin{aligned}
(2.1)\qquad & F(B(t,\mathbb{E}),N(t,\mathbb{F})) \\
& - \frac{1}{2} \sum_{p,q} \left(\pi(E_p \cap E_q) - \pi(E_p)\pi(E_q)\right) \int_0^t \frac{\partial^2 F}{\partial x_p \partial x_q}(B(s-,\mathbb{E}),N(s-,\mathbb{F}))ds \\
& - \int_0^t \int_{\mathbb{R}^m-\{0\}} \left(F(B(s-,\mathbb{E}),N(s-,\mathbb{F})+I_{\mathbb{F}}(\lambda)) - F(B(s-,\mathbb{E}),N(s-,\mathbb{F}))\right) \\
& \qquad \times \mu(d\lambda)ds.
\end{aligned}
$$

Proof. We prove the case M=N=1 only for simplicity. Taking a subsequence of $\{P_n\}$ if necessary we may assume that $\{P_n\}$ converges weakly to P_∞. We denote $B_n(t,E)$, $N_n(t,F)$ by $B_n(t)$, $N_n(t)$ and $B(t,E)$,$N(t,F)$ by $B(t)$, $N(t)$, respectively. Set

$$J = \{t: P_\infty((\Delta B(t),\Delta N(t)) \neq 0) > 0\},$$

where $\Delta B(t)=B(t)-B(t-)$ and $\Delta N(t)=N(T)-N(t-)$. It is at most a countable set. Then the finite dimensional distribution of $((B_n(t_1),N_n(t_1)),\ldots,(B_n(t_k),N_n(t_k)))$ converges if $t_1,\ldots,t_k$ do not belong to J. See Billingsley [1].

Set $t_j=j/n$. By the mean value theorem, we have

$$\begin{aligned}(2.2)\qquad & F(B_n(t),N_n(t)) - F(0,0) \\ &= \sum_{j=1}^{[nt]} F(B_n(t_j),N_n(t_j)) - F(B_n(t_{j-1}),N_n(t_{j-1})) \\ &= \sum_{j=1}^{[nt]} \frac{\partial F}{\partial x}(B_n(t_{j-1}),N_n(t_{j-1}))\Delta B_n(t_j) \\ &\quad + \frac{1}{2}\sum_{j=1}^{[nt]} \frac{\partial^2 F}{\partial x^2}(\eta_j,N_n(t_{j-1}))\Delta B_n(t_j)^2 \\ &\quad + \sum_{j=1}^{[nt]} \int_{\mathbb{R}^m-\{0\}} \Big(F(B_n(t_j),N_n(t_{j-1})+I_F(\lambda)) \\ &\qquad - F(B_n(t_j),N_n(t_{j-1}))\Big)\Delta N_n(t_j,d\lambda),\end{aligned}$$

where

$$\Delta B_n(t_j) = B_n(t_j)-B_n(t_{j-1}) = \frac{1}{\sqrt{n}}\Big(I_E(\xi_{n,j})-\pi_n(E)\Big),$$

$$\Delta N_n(t_j,F) = N_n(t_j,F)-N_n(t_{j-1},F) = I_F\Big(\frac{\xi_{n,j}}{\sqrt{n}}\Big),$$

and η_j are random variables satisfying $|\eta_j-B_n(t_{j-1})|\le |B_n(t_j)-B_n(t_{j-1})|$. Denote the first, second and third terms of the right hand side of (2.2) by $I_n^{(1)}(t)$, $I_n^{(2)}(t)$ and $I_n^{(3)}(t)$, respectively.

Let $\varphi(x_1,\ldots,x_1, y_1,\ldots,y_1)$ be a bounded continuous function on $\mathbb{R}^1\times\mathbb{R}^1$. Set

$$\Phi_n = \varphi(B_n(s_1),\ldots,B_n(s_1), N_n(s_1),\ldots,N_n(s_1)),$$

$$\Phi = \varphi(B(s_1),\ldots,B(s_1), N(s_1),\ldots,N(s_1)),$$

where $s_1<s$. Note that $(\partial F/\partial x)(B_n(t_{j-1}),N_n(t_{j-1}))$, $\Delta B_n(t_j)$ and Φ_n are independent if $j\ge[ns]+1$ and that the expectation of $\Delta B_n(t_j)$ is 0. Then we have

$$E[\Big(I_n^{(1)}(t) - I_n^{(1)}(s)\Big)\Phi_n] = 0, \qquad \forall n.$$

Next set

$$\hat{I}_n^{(2)}(t) = \frac{1}{2}\sum_{j=1}^{[nt]} \frac{\partial^2 F}{\partial x^2}(B_n(t_{j-1}),N_n(t_{j-1}))\Delta B_n(t_j)^2.$$

Since $E[\Delta B_n(t_k)^2]=n^{-1}\left(\pi_n(E)-\pi_n(E)^2\right)$, we have

$$E[\left(\hat{I}_n^{(2)}(t)-\hat{I}_n^{(2)}(s)\right)\Phi_n]$$

$$= \frac{1}{2}E[\left(\sum_{j=[ns]+1}^{[nt]} \frac{\partial^2 F}{\partial x^2}(B_n(t_{j-1}),N_n(t_{j-1}))\right)\frac{1}{n}\left(\pi_n(E)-\pi_n(E)^2\right)\Phi_n].$$

Since

$$\sum_{j=[ns]+1}^{[nt]} \frac{\partial^2 F}{\partial x^2}(B_n(t_{j-1}),N_n(t_{j-1}))\frac{1}{n} = \int_s^t \frac{\partial^2 F}{\partial x^2}(B_n(u-),N_n(u-))du$$

and $\{(B_n(u),N_n(u))\}$ converges weakly for any $u\in J$, the above converges to

$$\frac{1}{2}E_\infty[\left(\int_s^t \frac{\partial^2 F}{\partial x^2}(B(u-),N(u-))du\right)\Phi]\left(\pi(E)-\pi(E)^2\right).$$

On the other hand, we have $E[|I_n^{(2)}(t)-\hat{I}_n^{(2)}(t)|]\longrightarrow 0$ as $n\longrightarrow\infty$. Therefore we get

$$\lim_{n\to\infty} E[\left(I_n^{(2)}(t)-I_n^{(2)}(s)\right)\Phi_n]$$

$$= \frac{1}{2}E_\infty[\left(\int_s^t \frac{\partial^2 F}{\partial x^2}(B(u-),N(u-))du\right)\Phi]\{\pi(E)-\pi(E)^2\}.$$

Now set

$$\hat{I}_n^{(3)}(t) = \sum_{j=1}^{[nt]} \int_{\mathbb{R}^m-\{0\}} \Big(F(B_n(t_{j-1}), N_n(t_{j-1})+I_F(\lambda)) - F(B_n(t_{j-1}), N_n(t_{j-1}))\Big) \Delta N_n(t_j, d\lambda).$$

Then we have $E[|I_n^{(3)}(t)-\hat{I}_n^{(3)}(t)|] \longrightarrow 0$ as $n \longrightarrow 0$. Further,

$$E\Big[\Big(\hat{I}_n^{(3)}(t) - \hat{I}_n^{(3)}(s)\Big)\Phi_n\Big]$$

$$= E\Big[\Big(\sum_{j=[ns]+1}^{[nt]} \int_{\mathbb{R}^m-\{0\}} \{F(B_n(t_{j-1}), N_n(t_{j-1})+I_F(\lambda)) - F(B_n(t_{j-1}), N_n(t_{j-1}))\} \frac{1}{n} \mu_n(d\lambda)\Big)\Phi_n\Big].$$

It converges to

$$E_\infty\Big[\Big(\int_s^t \int_{\mathbb{R}^m-\{0\}} \Big(F(B(u-), N(u-)+I_F(\lambda)) - F(B(u-), N(u-))\Big)\mu(d\lambda)du\Big)\Phi\Big].$$

These computations imply that if $s,t \in J^c$,

$$E_\infty[(F(B(t), N(t)) - F(B(s), N(s)))\Phi]$$

$$= E_\infty\Big[\Big(\frac{1}{2}\int_s^t \frac{\partial^2 F}{\partial x^2}(B(u-), N(u-))du\Big(\pi(E) - \pi(E)^2\Big)$$

$$+ \int_s^t \int_{\mathbb{R}^m-\{0\}} \Big(F(B(u-),N(u-)+I_F(\lambda))-F(B(u-),N(u-))\Big)\mu(d\lambda)du\Big)\Phi].$$

The equality is valid for any s,t since J^c is dense in $[0,\infty)$ and both sides of the equality are right continuous with respect to s and t. Therefore the proposition is established in case M=N=1.

Proof of Theorem 2.1. Let $(B(t),N(t),P_\infty)$ be any weak limit. We will prove properties (1)-(3) of the theorem. Then this implies the uniqueness of the limit measure P_∞ since properties (1)-(3) determines the law of $(B(t),N(t))$ uniquely. The following discussion is close to Kunita-Watanabe [7]. We shall apply Proposition 2.3 to the function

$$F(x,y) = \exp i\{(\alpha,x)+(\beta,y)\},$$

where $(x,y)\in\mathbb{R}^M\times\mathbb{R}^N$ and $(\alpha,\beta)\in\mathbb{R}^M\times\mathbb{R}^N$. Then we find that

$$\exp i\{(\alpha,B(t,E))+(\beta,N(t,F))\}$$

$$= \psi(\alpha,\beta)\int_0^t \exp i\{(\alpha,B(u,E))+(\beta,N(u,F))\}du + M(t) \tag{2.3}$$

where

$$\psi(\alpha,\beta) = -\frac{1}{2}\sum_{p,q}\alpha_p\alpha_q\Big(\pi(E_p\cap E_q)-\pi(E_p)\pi(E_q)\Big) + \sum_p (e^{i\beta_p}-1)\mu(F_p)$$

and M(t) is a martingale. Denote the left hand side of (2.3) by $\Phi(t)$. Then we have

$$\Phi(t) - \Phi(s) = \psi(\alpha,\beta)\int_s^t \Phi(u)du + M_t - M_s.$$

Therefore

$$\Phi(t)\Phi(s)^{-1} = 1 + \psi(\alpha,\beta)\int_s^t \Phi(u)\Phi(s)^{-1}du + (M_t - M_s)\Phi(s)^{-1}.$$

Taking the conditional expectation with respect to P_∞, we obtain

$$E_\infty[\Phi(t)\Phi(s)^{-1}|\mathcal{F}_s] = 1 + \psi(\alpha,\beta)\int_s^t E_\infty[\Phi(u)\Phi(s)^{-1}|\mathcal{F}_s]du.$$

It may be regarded as a linear integral equation. The solution is an exponential function of $\psi(\alpha,\beta)$. Therefore,

$$(2.4)\qquad E_\infty[\exp i\{(\alpha,B(t,\mathbf{E})-B(s,\mathbf{E}))+(\beta,N(t,\mathbf{F})-N(s,\mathbf{F}))\}|\mathcal{F}_s]$$

$$= \exp\Big\{-\frac{1}{2}(t-s)\sum_{p,q}\alpha_p\alpha_q\Big(\pi(E_p\cap E_q)-\pi(E_p)\pi(E_q)\Big)\Big\}$$

$$\times \exp\,(t-s)\Big\{\sum_p (e^{i\beta_p}-1)\mu(F_p)\Big\}.$$

The above formula shows that $(B(t,\mathbf{E})-B(s,\mathbf{E}),N(t,\mathbf{F})-N(s,\mathbf{F}))$ is independent of $\mathcal{F}_s$. Consequently both B(t) and N(t) are Lévy

processes. Furthermore, setting $\beta=0$ in the above formula, we find that $B(t,E)-B(s,E)$ has a Gaussian distribution with mean 0 and covariance matrix $((t-s)\{\pi(E_p\cap E_q)-\pi(E_p)\pi(E_q)\})_{p,q=1,\ldots,M}$. Therefore it induces a time-space Gaussian random measure with characteristic $(0,m\otimes\pi,m\otimes\pi)$. Next, set $\alpha=0$ in formula (2.4). Then we find that $N(t,F_p)-N(s,F_p)$, $p=1,\ldots,N$ are independent and Poisson distributed with intensities $(t-s)\mu(F_p)$, respectively. Therefore $N(t,F)$ induces a time-space Poisson random measure.

Finally formula (2.4) implies

$$E_\infty[\exp i\{(\alpha,B(t,E)-B(s,E))+(\beta,N(t,F)-N(s,F))\}|\mathcal{F}_s]$$

$$= E_\infty[\exp i(\alpha,B(t,E)-B(s,E))|\mathcal{F}_s]E_\infty[\exp i(\beta,N(t,F)-N(s,F))|\mathcal{F}_s].$$

This proves that two processes $\{B(t,E)\}$ and $\{N(t,F)\}$ are independent. The proof is complete.

Let $\{K_n^\varepsilon(t,G)\}$ be a sequence of random measures defined by (0.7) and let $\{\nu_n^\varepsilon\}$ be a sequence of measures defined by (0.8). We iontroduce a condition.

Condition (C.3) For any ε, the sequence $\{\nu_n^\varepsilon\}$, $n=1,2,\ldots$ converges weakly to a measure ν^ε on $\mathbb{R}^m$.

Obviously the restriction of the measure ν^ε to $\mathbb{R}^m$ coincides with $(1+|\lambda|)^2\pi(d\lambda)$. Furthermore, under Condition (C.3), the sequence $\{\xi_n^\varepsilon\}$ defined by (0.9) converges weakly in $\mathbb{R}^m$ and the

limit ζ satisfies $\zeta(d\lambda)=(1+|\lambda|)\pi(d\lambda)$. Indeed, we have

$$\int_{\mathbb{R}^m} f(\lambda)\zeta_n^{\varepsilon}(d\lambda) = \int_{\mathbb{R}^m} \frac{f(\lambda)}{1+|\lambda|}\nu_n^{\varepsilon}(d\lambda)$$

$$\longrightarrow \int_{\mathbb{R}^m} \frac{f(\lambda)}{1+|\lambda|}\nu^{\varepsilon}(d\lambda) = \int_{\mathbb{R}^m} f(\lambda)(1+|\lambda|)\pi(d\lambda)$$

for any bounded continuous function f. Further if G is a ν^{ε}-continuity set, it is a ζ-continuity set.

Now the family of measures $\{\nu^{\varepsilon};\varepsilon>0\}$ decreases as $\varepsilon\downarrow 0$, since the same property is valid for $\{\nu_n^{\varepsilon};\varepsilon>0\}$ for all n: We can define the measure ν by

$$\nu = \lim_{\varepsilon\downarrow 0} \nu^{\varepsilon}. \tag{2.5}$$

Let $\mathcal{S}(\bar{\mathbb{R}}^m)$ be a field on $\bar{\mathbb{R}}^m$ generating Borel sets of $\bar{\mathbb{R}}^m$ such that any element of $\mathcal{S}(\bar{\mathbb{R}}^m)$ is a ν^{ε}-continity sets for any $\varepsilon>0$. Let $\mathcal{M}(\bar{\mathbb{R}}^m)$ be the set of all additive set functions on $\mathcal{S}(\bar{\mathbb{R}}^m)$. We may define the law of (K_n^{ε},N_n) on the space $\mathbb{D}=\mathbb{D}([0,\infty);\mathcal{M}(\bar{\mathbb{R}}^m)\times\mathcal{M}(\mathbb{R}^m-\{0\}))$. We denote it by $P_n^{(\varepsilon)}$. If the sequence $\{P_n^{(\varepsilon)}\}$ converges weakly as $n\to\infty$, the sequence $\{(K_n^{\varepsilon},N_n)\}$ is said to converge weakly.

Theorem 2.4. Assume Conditions (C.1)-(C.3). Let ε be a positive number such that $\{|\lambda|\le\varepsilon\}$ is a μ-continuity set. Then the sequence (K_n^{ε},N_n), $n=1,2,\ldots$ converges weakly. Let $(K^{\varepsilon},N,P_{\infty}^{(\varepsilon)})$ be its limit law. Then it is a Lévy process with

values in random measures. Further:

(1) The associated time-space measure of $\{N(t,F)\}$ is a Poisson random measure with characteristic $(m\otimes\mu, m\otimes\mu)$.

(2) Set

$$(2.6)\quad K(t,G) \equiv K^{\varepsilon}(t,G) - \int_{G\cap(\mathbb{R}^m-\{0\})} |\lambda|\chi_{\varepsilon}(|\lambda|)\{N(t,d\lambda)-t\mu(d\lambda)\}.$$

Then $\{K(t,G)\}$ is a Lévy process with values in relatively (ν,ξ)-orthogonal measures. Further the associated time-space measure of K is Gaussian with characteristic $(0, m\otimes\nu, m\otimes\xi)$.

(3) $\{N(t,F)\}$ and $\{K(t,G)\}$ are independent.

Remark Define

$$(2.7)\quad \tilde{B}(t,E) = \int_E (1+|\lambda|)^{-1}K(t,d\lambda), \quad E\subset\mathbb{R}^m.$$

$$(2.8)\quad K^{(\infty)}(t,G) = K(t,G) - B(t,G\cap\mathbb{R}^m).$$

Then $\tilde{B}(t,E)$ is Gaussian with characteristic $(0, m\otimes\pi)$ and $K^{(\infty)}$ is Gaussian with characterristic $(0,\nu^{(\infty)})$ where $\nu^{(\infty)}$ is the restriction of ν to the boundary $\partial\mathbb{R}^m$.

Proof. We again omit the proof of the tightness of $\{P_n^{(\varepsilon)}\}$. Let $P_{\infty}^{(\varepsilon)}$ be any weak limit of $\{P_n^{(\varepsilon)}\}$. We shall prove that $(K^{\varepsilon}, N, P_{\infty}^{(\varepsilon)})$ satisfies (1)-(3). Then this implies the uniqueness of the limit measure $P_{\infty}^{(\varepsilon)}$. Set $K^{\varepsilon}(t,G)=$

$(K^{\varepsilon}(t,G_1),\ldots,K^{\varepsilon}(t,G_M))$ where $G_p\in\mathcal{G}(\mathbb{R}^m)$, $p=1,\ldots,M$. Then for any C_b^2-function $F(x,y)$,

$$(2.9)\qquad F(K^{\varepsilon}(t,\mathbb{G}),N(t,\mathbb{F}))$$

$$-\frac{1}{2}\sum_{p,q}\Big(\nu(G_p\cap G_q)-\xi(G_p)\xi(G_q)\Big)\int_0^t\frac{\partial^2 F}{\partial x_p\partial x_q}(K^{\varepsilon}(s-,\mathbb{G}),N(s-,\mathbb{F}))ds$$

$$-\int_0^t\int_{\mathbb{R}^m-\{0\}}\Big(F(K^{\varepsilon}(s-,\mathbb{G})+|\lambda|\chi_{\varepsilon}(|\lambda|)I_{\mathbb{G}}(\infty,\frac{\lambda}{|\lambda|}),N(s-,\mathbb{F})+I_{\mathbb{F}}(\lambda))$$

$$-F(K^{\varepsilon}(s-,\mathbb{G}),N(s-,\mathbb{F}))-\sum_p|\lambda|\chi_{\varepsilon}(|\lambda|)I_{G_p}(\infty,\frac{\lambda}{|\lambda|})\frac{\partial F}{\partial x_p}\Big)\mu(d\lambda)ds$$

is a martingale with respect ot $P_{\infty}^{(\varepsilon)}$. We omit the proof. (c.f. Proposition 3.2).

Apply the above property to the exponential function. Then similarly as in the proof of Theorem 2.1, we obtain

$$(2.10)\qquad E_{\infty}^{(\varepsilon)}[\exp i\{(\alpha,K^{\varepsilon}(t,\mathbb{G})-K^{\varepsilon}(s,\mathbb{G}))+(\beta,\tilde{N}(t,\mathbb{F})-\tilde{N}(s,\mathbb{F}))\}|\mathcal{F}_s]$$

$$=\exp(t-s)\psi(\alpha,\beta),$$

where $\tilde{N}(t,\mathbb{F})=N(t,\mathbb{F})-t\mu(\mathbb{F})$ and

$$(2.11)\qquad \psi(\alpha,\beta)=-\frac{1}{2}\sum_{p,q}\Big(\nu(G_p\cap G_q)-\xi(G_p)\xi(G_q)\Big)\alpha_p\alpha_q$$

$$+\int_{\mathbb{R}^m-\{0\}}\Big(\exp\{i(\sum_j I_{F_j}(\lambda)\beta_j+\sum_p|\lambda|\chi_{\varepsilon}(|\lambda|)I_{G_p}(\infty,\frac{\lambda}{|\lambda|})\alpha_p)\}-1$$

$$- i\{\sum_j I_{F_j}(\lambda)\beta_j + \sum_p |\lambda|\chi_\varepsilon(|\lambda|) I_{G_p}(\infty, \frac{\lambda}{|\lambda|})\alpha_p\}\Big)\mu(d\lambda).$$

Set $\alpha=0$. Then we get the first assertion (1) of the theorem.

Next, define

$$\tilde{M}(t,G) = \int_{R^m-\{0\}} |\lambda|\chi_\varepsilon(|\lambda|) I_G(\infty, \frac{\lambda}{|\lambda|})\tilde{N}(t,d\lambda). \tag{2.12}$$

It is approximated by linear sums of $\tilde{N}(t,F_1),\ldots,\tilde{N}(t,F_N)$. Then from (2.10), we arrive at

$$E_\infty^{(\varepsilon)}[\exp i\{(\alpha,K^\varepsilon(t,G)-K^\varepsilon(s,G))+(\tilde{\beta},\tilde{M}(t,G)-\tilde{M}(s,G))\}|\mathcal{F}_s]$$

$$= \exp -\frac{1}{2}(t-s)\sum_{p,q}\Big(\nu(G_p\cap G_q)-\xi(G_p)\xi(G_q)\Big)\alpha_p\alpha_q$$

$$\times\exp\Big\{\int\Big(\exp i\{\sum_p |\lambda|\chi_\varepsilon(|\lambda|) I_{G_p}(\infty, \frac{\lambda}{|\lambda|})(\alpha_p+\tilde{\beta}_p)\} - 1$$

$$- \{\sum_p |\lambda|\chi_\varepsilon(|\lambda|) I_{G_p}(\infty, \frac{\lambda}{|\lambda|})(\alpha_p+\tilde{\beta}_p)\}\Big)\mu(d\lambda)\Big\}.$$

Setting $\tilde{\beta}=-\alpha$, we obtain

$$E_\infty^{(\varepsilon)}[\exp i(\alpha,K^\varepsilon(t,G)-\tilde{M}(t,G)-K^\varepsilon(s,G)+\tilde{M}(s,G))|\mathcal{F}_s]$$

$$= \exp -\frac{1}{2}(t-s)\sum_{p,q}\Big(\nu(G_p\cap G_q)-\xi(G_p)\xi(G_q)\Big)\alpha_p\alpha_q.$$

This proves the second assertion (2).

The assertion (3) will be obvious.

3. Weak convergence for sum of non-linear functions of i.i.d. random variables

Let $\{X_n(t)\}$ be the sequence of stochastic processes defined by (0.1). The law of $(K_n^{\varepsilon}, N_n, X_n)$ is defined on $\mathbb{D}=\mathbb{D}([0,\infty); \mathcal{M}(\bar{\mathbb{R}}^m)\times\mathcal{M}(\mathbb{R}^m-\{0\})\times\mathbb{R}^d)$. We denote it by $\tilde{P}_n^{\varepsilon}$. We will discuss the weak convergence of $\tilde{P}_n^{\varepsilon}, n=1,2,\ldots$.

We first introduce an assumption to the sequence of functions $\{f_n(\lambda)\}$.

Condition (C.4) (1) The sequence $\{f_n\}$ converges to f uniformly on compact sets.

(2) The sequence $\{f_n(\sqrt{n}\lambda)/(1+\sqrt{n}|\lambda|)\}$ converges to $h(\lambda)$ uniformly on compact subsets of $\mathbb{R}^m-\{0\}$. Further $h(0,\theta)\equiv\lim_{r\to 0}h(r,\theta)$ exists and is a continuous function of θ.

Theorem 3.1. Assume Conditions (C.1)-(C.4). Set

$$a_n(t) = \frac{[nt]}{\sqrt{n}}E\left[f_n(\xi_{n,1})\chi_{\varepsilon}\left(\frac{\xi_{n,1}}{\sqrt{n}}\right)\right], \tag{3.1}$$

where ε is a positive number such that $\{|\lambda|\leq\varepsilon\}$ is a μ-continuity set. Then the sequence $\{(K_n^{\varepsilon}, N_n, X_n)\}$ converges weakly. Let $(K^{\varepsilon}, N, X, \tilde{P}_{\infty}^{(\varepsilon)})$ be its limit. Then $X(t)$ is represented by (0.12), where $\tilde{B}(t,.)$ and $K^{(\infty)}(t,.)$ are defined through (2.6)-(2.8) and

$\tilde{N}(t,d\lambda)=N(t,d\lambda)-t\mu(d\lambda)$.

For the proof of Theorem 3.1, we need to prove that the sequence of laws of $\{(K_n^{(\varepsilon)},N_n,X_n)\}$ is tight. Since the proof is similar to that of Proposition 2.2, it is omitted. In the next proposition, we characterize any limit measure as a solution of a martingale problem.

Proposition 3.2. For a C_b^2-function $F(x,y,z)$ on $\mathbb{R}^m\times\mathbb{R}^N\times\mathbb{R}^d$, define

$$
\begin{aligned}
(3.2)\qquad \mathcal{L}F(x,y,z) &= \frac{1}{2}\sum_{p,q}\Big(\nu(G_p\cap G_q)-\xi(G_p)\xi(G_q)\Big)\frac{\partial^2 F}{\partial x_p\partial x_q}\\
&+\frac{1}{2}\sum_{k,l}\Big(\int g_k g_l\,d\nu-\int g_k\,d\xi\cdot\int g_l\,d\xi\Big)\frac{\partial^2 F}{\partial z_k\partial z_l}\\
&+\sum_{k,p}\Big(\int_{G_p} g_k\,d\nu-\xi(G_p)\int g_k\,d\xi\Big)\frac{\partial^2 F}{\partial z_k\partial x_p}\\
&+\int(1-\chi_\varepsilon(|\lambda|))\Big(F(x+|\lambda|\chi_\varepsilon(|\lambda|)I_G(\infty,\tfrac{\lambda}{|\lambda|}),y+I_F(\lambda),z+|\lambda|h(\lambda))\\
&\qquad -F(x,y,z)\Big)\mu(d\lambda)\\
&+\int\chi_\varepsilon(|\lambda|)\Big(F(x+|\lambda|\chi_\varepsilon(|\lambda|)I_G(\infty,\tfrac{\lambda}{|\lambda|}),y+I_F(\lambda),z+|\lambda|h(\lambda))\\
&-F(x,y,z)-|\lambda|\chi_\varepsilon(|\lambda|)\sum_p I_{G_p}(\infty,\tfrac{\lambda}{|\lambda|})\frac{\partial F}{\partial x_p}-|\lambda|\sum_k h_k(\lambda)\frac{\partial F}{\partial z_k}\Big)\mu(d\lambda),
\end{aligned}
$$

where $g=(g_1,\dots,g_d)$ is defined by

$$g(\lambda) = \frac{f(\lambda)}{1+|\lambda|} \qquad \text{if } \lambda \in \mathbb{R}^m,$$

(3.3)

$$= h(0,\theta) \qquad \text{if } \lambda = (\infty,\theta) \in \partial\mathbb{R}^m,$$

and $I_{\mathbb{G}}=(I_{G_1},\dots,I_{G_M})$ etc. Then for any weak limit $\tilde{P}_\infty^{(\varepsilon)}$ of $\{\tilde{P}_n^{(\varepsilon)}\}$,

(3.4) $$F(K^\varepsilon(t,\mathbb{G}),N(t,\mathbb{F}),X(t)) - \int_0^t \mathcal{L}F(K^\varepsilon(s-,\mathbb{G}),N(s-,\mathbb{F}),X(s-))ds$$

is a martingale with respect to $\tilde{P}_\infty^{(\varepsilon)}$.

Proof. For simplicity we prove the case M=N=0, d=1 only. The general case will be proved similarly, combining the technique used in the proof of Proposition 2.3.

Let $F(z)$ be a C_b^2-function. Set $t_j=j/n$. Then

(3.5) $$F(X_n(t))-F(X_n(s))$$

$$= \sum_{j=[ns]+1}^{[nt]} \Big(F(X_n(t_j))-F(X_n(t_{j-1}))\Big)\Big(1-\chi_\delta(\tfrac{1}{\sqrt{n}}|\xi_{n,j}|)\Big)$$

$$+ \sum_{j=[ns]+1}^{[nt]} F'(X_n(t_{j-1}))\frac{1}{\sqrt{n}}(f_n(\xi_{n,j})-b_n^\varepsilon)\chi_\delta(\tfrac{1}{\sqrt{n}}|\xi_{n,j}|)$$

$$+ \sum_{j=[ns]+1}^{[nt]} F''(\eta_{n,j})\frac{1}{n}(f_n(\xi_{n,j})-b_n^\varepsilon)^2\chi_\delta(\tfrac{1}{\sqrt{n}}|\xi_{n,j}|),$$

where

$$(3.6)\qquad b_n^{\varepsilon} \equiv \int f_n(\lambda)\chi_{\varepsilon}\left(\frac{|\lambda|}{\sqrt{n}}\right)d\pi_n(\lambda),$$

and $\eta_{n,j}$ are random variable such that $|\eta_{n,j}-X_n(t_j)|\le |\xi_{n,j}-b_n^{\varepsilon}|/\sqrt{n}$. Set

$$(3.7)\qquad g_n(\lambda) = \frac{f_n(\lambda)}{1+|\lambda|}.$$

The sum of the first and second terms of the right hand side of (3.5) is written as

$$\int_s^t\int\left(F\left(X_n(u-)+\left(|\lambda|+\frac{1}{\sqrt{n}}\right)g_n(\sqrt{n}\lambda)-\frac{1}{\sqrt{n}}b_n^{\varepsilon}\right) - F(X_n(u-))\right)$$

$$\times(1-\chi_{\varepsilon}(|\lambda|))N_n(du,d\lambda)$$

$$+\left(\int_s^t\int\left(F\left(X_n(u-)+\left(|\lambda|+\frac{1}{\sqrt{n}}\right)g_n(\sqrt{n}\lambda)-\frac{1}{\sqrt{n}}b_n^{\varepsilon}\right) - F(X_n(u-))\right)\right.$$

$$\left.\times(\chi_{\varepsilon}(|\lambda|)-\chi_{\delta}(|\lambda|))N_n(du,d\lambda)\right)$$

$$-\sum_{j=[ns]+1}^{[nt]}F'(X_n(t_{j-1}))\frac{1}{n}$$

$$\times\int\left(|\lambda|+\frac{1}{\sqrt{n}}\right)g_n(\sqrt{n}\lambda)\left(\chi_{\varepsilon}(|\lambda|)\chi_{\delta}\left(\frac{1}{\sqrt{n}}|\xi_{n,j}|\right)-\chi_{\delta}(|\lambda|)\right)\mu_n(d\lambda)\Big)$$

$$+ \sum_{j=[ns]+1}^{[nt]} F'(X_n(t_{j-1}))$$

$$\times\left(\frac{1}{\sqrt{n}} f_n(\xi_{n,j})\chi_\delta\left(\frac{1}{\sqrt{n}}|\xi_{n,j}|\right) - \frac{1}{\sqrt{n}}\int f_n(\lambda)\chi_\delta\left(\frac{|\lambda|}{\sqrt{n}}\right)d\pi_n(\lambda)\right)$$

$$= J_n^{(1)} + J_n^{(2)} + J_n^{(3)}.$$

Set

$$\tilde{\Phi}_n = \varphi(X_n(s_1),\ldots,X_n(s_1)), \quad \tilde{\Phi} = \varphi(X(s_1),\ldots,X(s_1))$$

where $s_1 \le s$. Then $E[J_n^{(1)}\tilde{\Phi}_n]$ is equal to

$$E\left[\int_s^t\int\left(F\left(X_n(u-)+(|\lambda|+\frac{1}{\sqrt{n}})g_n(\sqrt{n}\lambda)-\frac{1}{\sqrt{n}}b_n^\varepsilon\right) - F(X_n(u-))\right)\right.$$

$$\left.\times(1-\chi_\varepsilon(|\lambda|))\mu_n(d\lambda)du\right)\tilde{\Phi}_n].$$

Let $n\to\infty$. Then since $g_n(\sqrt{n}\lambda)\to h(\lambda)$ uniformly on compact sets of $\mathbb{R}^m-\{0\}$ and $b_n^\varepsilon/\sqrt{n}\to 0$ as $n\to\infty$, the above converges to

$$\tilde{E}_\infty^{(\varepsilon)}\left[\left(\int_s^t\int\left(F(X(u-)+|\lambda|h(\lambda))-F(X(u-))\right)(1-\chi_\varepsilon(|\lambda|))\mu(d\lambda)du\right)\tilde{\Phi}\right].$$

Similarly $\lim_{\delta\to 0}\lim_{n\to\infty}E[J_n^{(2)}\tilde{\Phi}_n]$ exists and is equal to

$$\tilde{E}_{\infty}^{(\varepsilon)}[\left(\int_s^t\int\left(F(X(u-)+|\lambda|h(\lambda))-F(X(u-))-|\lambda|h(\lambda)F'(X(u-))\right)\right.$$

$$\left.\times\chi_{\varepsilon}(|\lambda|)\mu(d\lambda)du\right)\tilde{\Phi}].$$

Obviously we have $E[J_n^{(3)}\tilde{\Phi}_n]=0$.

Next we have

$$E[\left(\sum_{j=[ns]+1}^{[nt]}F''(X_n(t_{j-1}))\frac{1}{n}(f_n(\xi_{n,j})-b_n^{\varepsilon})^2\chi_{\delta}(\frac{1}{\sqrt{n}}|\xi_{n,j}|)\right)\tilde{\Phi}_n]$$

$$= E[(\int_s^t F''(X_n(u-))du)\tilde{\Phi}_n]$$

$$\times\left(\int|g_n(\lambda)|^2d\nu_n^{\delta}(\lambda) - 2b_n^{\delta}b_n^{\varepsilon} + (b_n^{\varepsilon})^2\int\chi_{\delta}(\frac{1}{\sqrt{n}}|\lambda|)d\pi_n(\lambda)\right).$$

Since $b_n^{\varepsilon}=\int g_n(\lambda)d\xi_n^{\varepsilon}(\lambda)$ holds, the sequence $b_n^{\varepsilon}, n=1,2,\ldots$ converges to $\int g(\lambda)d\xi(\lambda)$ as $n\to 0$. Therefore the above converges to

$$\frac{1}{2}\tilde{E}_{\infty}^{(\varepsilon)}[(\int_s^t F''(X(u-))du)\tilde{\Phi}]\left(\int g(\lambda)^2d\nu(\lambda) - (\int g(\lambda)d\xi(\lambda))^2\right)$$

as $n\to\infty$ and $\delta\to 0$. Further,

$$E[\sum_{j=[ns]+1}^{[nt]}|F''(\eta_{n,j})-F''(X_n(t_{j-1}))|\frac{1}{n}(f_n(\xi_{n,j})-b_n^{\delta})^2\chi_{\delta}\left(\frac{|\xi_{n,j}|}{\sqrt{n}}\right)]$$

converges to 0. Putting together these computations, we arrive at

$$\tilde{E}_{\infty}^{(\varepsilon)}[(F(X(t))-F(X(s)))\tilde{\Phi}] = \tilde{E}_{\infty}^{(\varepsilon)}[(\int_s^t \mathcal{L}F(X(u-))du)\tilde{\Phi}]$$

if $s,t\in\tilde{J}^c$, where $\tilde{J}=\{t: \tilde{P}_{\infty}^{(\varepsilon)}((\Delta K^{\varepsilon}(t),\Delta N(t),\Delta X(t))\neq 0)>0\}$. Since $\tilde{J}^c$ is dense and both sides of the above are right continuous with respect to t and s, the equality is valid for any s,t. Therefore $F(X(t))-\int_0^t \mathcal{L}F(X(u-))du$ is a martingale. The proof is complete.

Proof of Theorem 3.1. Let $\tilde{P}_{\infty}^{(\varepsilon)}$ be any limit of $\{\tilde{P}_n^{(\varepsilon)}\}$. We shall prove that X(t) satisfies (0.12). Then this implies the uniqueness of the limit $\tilde{P}_{\infty}^{(\varepsilon)}$. Now apply Proposition 3.3. We can prove similarly as in the proof of Theorem 2.1 that

$$(3.8)\qquad \tilde{E}_{\infty}^{(\varepsilon)}[\exp i\{(\alpha,K^{\varepsilon}(t,G)-K^{\varepsilon}(s,G))+(\beta,N(t,F)-N(s,F))$$

$$+(\gamma,X(t)-X(s))\}|\mathcal{F}_s]$$

$$= \exp (t-s)\{\psi_1(\alpha,\gamma)+\psi_2(\alpha,\beta,\gamma)\},$$

where

$$(3.9)\qquad \psi_1(\alpha,\gamma) = -\frac{1}{2}\sum_{p,q}\left(\nu(G_p\cap G_q)-\zeta(G_p)\zeta(G_q)\right)\alpha_p\alpha_q$$

$$-\sum_{p,k}\left(\int_{G_p} g_k d\nu - \zeta(G_p)\int g_k d\zeta\right)\alpha_p\gamma_k$$

$$- \frac{1}{2}\sum_{k,l}\left(\int g_k g_l d\nu - \int g_k d\xi \cdot \int g_l d\xi\right)\gamma_k\gamma_l,$$

(3.10) $$\psi_2(\alpha,\beta,\gamma) = \int (1-\chi_\varepsilon(|\lambda|))\Big(\exp i\{|\lambda|\chi_\varepsilon(|\lambda|)\sum_p I_{G_p}(\infty,\tfrac{\lambda}{|\lambda|})\alpha_p + \sum_j I_{F_j}(\lambda)\beta_j + |\lambda|\sum_k h_k(\lambda)\gamma_k\} - 1\Big)\mu(d\lambda)$$

$$+ \int \chi_\varepsilon(|\lambda|)\Big(\exp i\{|\lambda|\chi_\varepsilon(|\lambda|)\sum_p I_{G_p}(\infty,\tfrac{\lambda}{|\lambda|})\alpha_p + \sum_j I_{F_j}(\lambda)\beta_j + |\lambda|\sum_k h_k(\lambda)\gamma_k\} - 1 - i\{|\lambda|\chi_\varepsilon(|\lambda|)\sum_p I_{G_p}(\infty,\tfrac{\lambda}{|\lambda|})\alpha_p + \sum_j I_{F_j}(\lambda)\beta_j + |\lambda|\sum_k h_k(\lambda)\gamma_k\}\Big)\mu(d\lambda).$$

Therefore $(K^\varepsilon(t),N(t),X(t))$ is a Lévy process. Let $X_c(t)$ and $X_d(t)$ be the continuous and discontinuous part of $X(t)$, respectively. Then $(K(t),0,X_c(t))$ is the continuous part of (K^ε,N,X). Its characteristic function is given by $\exp t\psi_1(\alpha,\gamma)$. Therefore $(K(t),X_c(t))$ is a Brownian motion with mean 0 and covariances

$$\tilde{E}_\infty^{(\varepsilon)}[K(t,G_1)K(t,G_2)] = t\Big(\nu(G_1\cap G_2) - \xi(G_1)\xi(G_2)\Big),$$

$$\tilde{E}_\infty^{(\varepsilon)}[K(t,G)X_c(t)] = t\int_G g d\nu,$$

$$\tilde{E}_\infty^{(\varepsilon)}[X_c^k(t)X_c^l(t)] = t\int g_k g_l d\nu.$$

Setting $\tilde{X}_c(t)=\int g(\lambda)K(t,d\lambda)$, the covariance of $X_c(t)$ and $\tilde{X}_c(t)$ are given by

$$E[X_c(t)X_c(t)'] = E[X_c(t)\tilde{X}_c(t)'] = E[\tilde{X}_c(t)\tilde{X}_c(t)']$$

$$= \int gg'd\nu - \int gd\xi\cdot\int g'd\xi,$$

where g' denotes the transpose of the column vector g. Therefore we have $E[|X_c(t)-\tilde{X}_c(t)|^2]=0$, proving $X_c(t)=\tilde{X}_c(t)$.

Next, $(K^{\varepsilon}-K,N,X_d)$ is the discontinuous part of (K^{ε},N,X). Therefore

$$\tilde{E}_{\infty}^{(\varepsilon)}[\exp i\{(\beta,N(t))+(\gamma,X_d)\}] = \exp t\psi_2(0,\beta,\gamma). \tag{3.11}$$

Set

$$\tilde{X}_d(t) = \int|\lambda|h(\lambda)(1-\chi_{\varepsilon}(|\lambda|))N(t,d\lambda)$$

$$+ \int|\lambda|h(\lambda)\chi_{\varepsilon}(|\lambda|)\tilde{N}(t,d\lambda).$$

It is approximated by linear sums of $N(t,F)$. Then from (3.11) we arrive at

$$\tilde{E}_{\infty}^{(\varepsilon)}[\exp i\{(\tilde{\beta},\tilde{X}_d(t))+(\gamma,X_d(t))\}] = \exp t\tilde{\psi}_2(\tilde{\beta},\gamma), \tag{3.12}$$

where

$$\tilde{\Psi}_2(\tilde{\beta},\gamma) = \int(1-\chi_\varepsilon(|\lambda|))\{\exp i(\tilde{\beta}+\gamma,|\lambda|h(\lambda))-1\}\mu(d\lambda)$$

$$+ \int\chi_\varepsilon(|\lambda|)\{\exp i(\tilde{\beta}+\gamma,|\lambda|h(\lambda))-1-i(\tilde{\beta}+\gamma,|\lambda|h(\lambda))\}\mu(d\lambda).$$

This proves that

$$\tilde{E}_\infty^{(\varepsilon)}[\exp i(\tilde{\beta},\tilde{X}_d(t)-X_d(t))] = 1, \quad \forall\,\tilde{\beta},$$

proving $\tilde{X}_d(t)=X_d(t)$. We have thus obtained the representation (0.12). The proof is complete.

Finally we consider the case where $\{\xi_{n,j}\}$ satisfies $\sup_n E[|\xi_{n,j}|^{2+\delta}]<\infty$ for some $\delta>0$. Then the measures $\{\nu_n^\varepsilon : n=1,2,\ldots\}$ converge weakly to the measure $(1+|\lambda|)^2\pi(d\lambda)$. Therefore the limit measure ν^ε of (2.4) are not supported by the boundary $\partial\mathbb{R}^m$. Then the random measure $K^{(\infty)}$ defined by (2.6) is zero. Furthermore the measure μ_n of (0.6) satisfies

$$\mu_n(|\lambda|>\varepsilon) = nP\left(\frac{|\xi_{n,j}|}{\sqrt{n}}>\varepsilon\right) \le n\,\frac{1}{\varepsilon^{2+\delta}}\left(\frac{1}{\sqrt{n}}\right)^{2+\delta}E[|\xi_{n,j}|^{2+\delta}].$$

It converges to 0 as $n\to\infty$. Therefore the Poisson random measure $N(t,\cdot)$ is also 0. Consequently Theorem 3.1 tells us that the limit $X(t)$ is represented by (0.4).

REFERENCES

[1] P. Billingsley, *Convergence of probability measures*, John Wiley and Sons, New York, 1968.

[2] B.V. Gnedenko and A.N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wisley, Reading, Mass., 1954.

[3] P.R. Halmos, *Measure theory*, Springer-Verlag, New York 1970

[4] K.Itô, *Lectures on stochastic processes*, Tata Institute of Fundamental Research, Bombay, 1960.

[5] J. Jacod, A. Klopotowski and J. Mémin , Théorème de la limite centrale et convergence fonctionelle vers un processus à accroissements indépendents, Ann. Inst. Henri Poincare (B), **18**(1982),1-45.

[6] J. Jacod and A.N. Shiryaev, *Limit theorems for stochastic processes*, Springer-Verlag Berlin Heidelberg 1987.

[7] H.Kunita and S. Watanabe, On square integrable matingales, Nagoya Math. J., **30**(1967), 209-245

Hiroshi Hunita
Department of Applied Science
Kyushu University 36
Fukuoka 812, Japan

World Scientific

ISBN 981-02-0229-6