Springer Proceedings in Mathematics

Irene Sabadini
Daniele C. Struppa *Editors*

The Mathematical Legacy of Leon Ehrenpreis

Springer

Springer Proceedings in Mathematics

The book series features volumes of selected contributions from workshops and conferences in all areas of current research activity in mathematics. Besides an overall evaluation, at the hands of the publisher, of the interest, scientific quality, and timeliness of each proposal, every individual contribution is refereed to standards comparable to those of leading mathematics journals. This series thus proposes to the research community well-edited and authoritative reports on newest developments in the most interesting and promising areas of mathematical research today.

Springer Proceedings in Mathematics

Volume 16

For further volumes:
www.springer.com/series/8806

Irene Sabadini • Daniele C. Struppa
Editors

The Mathematical Legacy of Leon Ehrenpreis

Springer

Editors
Irene Sabadini
Dipartimento di Matematica
Politecnico di Milano
Milano, Italy

Daniele C. Struppa
Schmid College of Science
and Technology
Chapman University
Orange, CA, USA

ISSN 2190-5614 ISSN 2190-5622 (electronic)
Springer Proceedings in Mathematics
ISBN 978-88-470-1946-1 ISBN 978-88-470-1947-8 (eBook)
DOI 10.1007/978-88-470-1947-8
Springer Milan Heidelberg New York Dordrecht London

Library of Congress Control Number: 2011945461

Mathematics Subject Classification: 11-06, 32-06, 42-06, 46-06

© Springer-Verlag Italia 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Leon Ehrenpreis at Stockholm

Preface

Like many other mathematicians around the world, we were saddened and shocked when news reached us that Leon Ehrenpreis had passed away on 16 August 2010. Our first instinct was to collect a volume of mathematical contributions by his many friends and collaborators as well as by many mathematicians whose mathematical career has been influenced by Leon's work. We are very appreciative for the immediate support that Springer and Dr. Francesca Bonadei have given to our idea and for the enthusiastic response of the many authors who have agreed to participate in what we consider as an act of respect, friendship, and affection for Leon. We are also indebted to Leon's daugther, Yael Ehrenpreis Meyer, who has shared with us the beautiful picture of Leon in Stockholm, a picture that so perfectly reflects Leon's zest for life. Finally, we are grateful to Professor Malgrange for sharing with us a personal letter that Leon wrote to him in June 1960 and which is appended to this volume.

As a way of introduction to the volume, we include, in the next few pages, three short essays that focus on three different periods of Leon Ehrenpreis' mathematical life.

Milan, Italy — Irene Sabadini
Orange, USA — Daniele C. Struppa

Contents

Part I
Introduction to the Volume

Leon Ehrenpreis: Some Old Souvenirs

Bernard Malgrange

In the years 1952–1953, I had finished my studies at École Normale Supérieure, and I had a position of research in CNRS, under the supervision of Laurent Schwartz. His book on the theory of distributions had been recently published; this book and his paper on mean periodic functions were full of open problems on linear differential equations, especially with constant coefficients and convolution equations. I was mainly interested in the problem of "elementary solutions": given a differential polynomial P with constant coefficients, does there exist a distribution f on $\mathbb{R}^n$ verifying $Pf = \delta$, δ the Dirac measure?

Schwartz suggested to solve this problem by finding a "tempered" f: by Fourier transform, this is equivalent to the problem of "division of a distribution" by a polynomial. I tried this method, but unsuccessfully (the problem was solved several years later, independently by Hörmander and Łojasiewicz). But I found that one can bypass the division of distributions: by duality, one is reduced to the following problem: if a family $\{P\varphi_\alpha\}$ (φ_α, functions $\mathcal{C}^\infty$ with compact support) tend to zero in a suitable sense, then the $\{\varphi_\alpha\}$ tend also to zero. Now, by Fourier transform P is transformed into a polynomial, and φ_α into an entire function with some growth conditions at infinity described by the Paley–Wiener theorem. And a simple argument of maximum modulus gave the required result.

There are a lot of convergence conditions which can be chosen. The simplest is perhaps the following one: if the φ_α's have a bounded support and if the $P\varphi_\alpha$ tend to zero in L^2, then the φ_α tend also to zero in L^2.

The same method, with a little more work, gives also the following results:

(i) Let f be a $\mathcal{C}^\infty$ function (resp. a distribution of finite order) in $\mathbb{R}^n$; then there exists another one g with $Pg = f$.
(ii) The exponential-polynomial solutions of $Pf = 0$ are dense in the $\mathcal{C}^\infty$, or in the distributions solutions.

B. Malgrange (✉)
Institut Fourier, BP 74, 38402 Saint Martin d'Hères, France
e-mail: bernard.malgrange@ujf-grenoble.fr

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_1, © Springer-Verlag Italia 2012

Furthermore, the same results are true for $\mathbb{R}^n$ replaced by an open convex set.

I published notes in *Comptes Rendus de l'Académie des Sciences* on these results. A short time later somebody, I think J. Dieudonné, told me that a young American mathematician, named Leon Ehrenpreis, had obtained also the same results. They were published in *American Journal* in 1954 under the title "Solution of some problems of division *I*".

This was the beginning of a kind of emulation, although this was essentially the only one time where we obtained independently similar results. To this emulation, I can perhaps add the name of Lars Hörmander, who namely reproved the existence theorem in his thesis by proving the required L^2 inequality directly with energy integrals, without Fourier transform; this permitted to him to get by the same method existence theorems for some equations with variable coefficients ("equations of principal type"), which could not be obtained by our complex methods.

Concerning the next period, I will mention mainly the series of papers by Ehrenpreis "Solution of some problem of division", especially the numbers III and IV. Let me indicate briefly the main results of these papers.

In III, he solves a problem left open by the preceding works: given a differential polynomial P with constant coefficients, and a distribution f in $\mathbb{R}^n$ (not necessarily of finite order), there exists another one g with $Pg = f$. The proof consists in a very precise analysis in terms of Fourier transform of the topology of the space $\mathcal{D}$ of Schwartz (i.e. the space of $\mathcal{C}^\infty$ functions with compact support). Later, I interpreted this analysis as giving a theorem of propagation of regularity for the solutions of equations with constant coefficients. For a more systematic study of this point of view, I refer to the book "Linear partial differential operators" by Hörmander.

I was much impressed by this paper. But I was even more impressed by the next one, number IV. This paper is devoted to convolutions equations $\mu * f = g$, μ a given distribution with compact support, f and g $\mathcal{C}^\infty$ functions or distributions. The main results are the following:

(i) A necessary and sufficient condition for $\mu *$ to be surjective in the space of $\mathcal{C}^\infty$ functions, or distributions, in $\mathbb{R}^n$. The condition, called by Ehrenpreis "slowly decreasing", is as follows:

    If $\hat{\mu}$ is the Fourier transform of μ (which is an entire function in $\mathbb{C}^n$), there exists $a > 0$ such that for each real z, there exists another z' with $|z' - z| \leq a \log(1 + |z|)$ and $|\hat{\mu}(z')| \geq (a + |z|)^{-a}$ (here $|z|$ is any norm in $\mathbb{C}^n$).

    If we replace "distribution" by "distribution of finite order", one needs a stronger condition: the first inequality should be replaced by $|z' - z| \leq a$.

(ii) A necessary and sufficient condition for "hypoellipticity" (called "ellipticity" after Schwartz): all distributions f verifying $\mu * f = 0$ are $\mathcal{C}^\infty$ functions.

    The condition generalizes the one obtained for differential polynomials by Hörmander in his thesis; but Ehrenpreis says that his own result was obtained independently.

    The condition is the following: first, $\hat{\mu}$ should be slowly decreasing; furthermore, on the variety of zeros of $\hat{\mu}$, one has an inequality $|\operatorname{Im} z| \geq a \log(1 + |z|)$.

But Ehrenpreis was soon after that work interested by a much more general situation: the overdetermined linear systems with constant coefficients. In 1960, he announces general results in this context.

For simplicity, I limit myself to systems with one unknown (the general case is similar). Also, I consider only the case of $\mathcal{C}^\infty$ functions in $\mathbb{R}^n$; in the case of distributions and convex open sets, the results are similar.

We give $P_1, \dots, P_m$, linear differential operators with constant coefficients. Then the results are as follows:

(i) Given functions $f_1, \dots, f_m$ $\mathcal{C}^\infty$, there exists a g $\mathcal{C}^\infty$ verifying $P_i g = f_i$ if and only if the f_i's verify the "trivial compatibility conditions": if $Q_1, \dots, Q_m$ are differential polynomials with constant coefficients satisfying $\Sigma Q_i P_i = 0$, then one has $\Sigma Q_i f_i = 0$

(ii) The exponential-polynomial solutions of $P_1 g = \cdots = P_m g = 0$ are dense in all solutions.

Actually, Ehrenpreis gives a much more precise statement, called "fundamental principle": roughly speaking, the $\mathcal{C}^\infty$ solutions of the system are Fourier transforms or "integrals" (in a suitable sense) of measures with support the complex variety of zeros of $\widehat{P_1}(z) = \cdots = \widehat{P_m}(z) = 0$, $\widehat{P_i}$ the polynomial associated to P_i.

I will just explain roughly how one can get (i) and (ii) (the fundamental principle requires some more work). By duality and Fourier transform, the problem is reduced to the following:

Let μ be a distribution with compact support, and $\hat{\mu}$ its Fourier transform. According to Paley–Wiener theorem, $\hat{\mu}$ is an entire function of exponential type with polynomial growth in any strip $|\operatorname{Im} z| \le a$, $a \in \mathbb{R}$, and conversely.

Now, suppose that, at every point $z_0 \in \mathbb{C}^n$, $\hat{\mu}$ is in the ideal of formal series in $(z - z_0)$ generated by $\widehat{P_1}, \dots, \widehat{P_m}$. Then, one has $\hat{\mu} = \Sigma \widehat{P_i} \hat{\nu}_i$, where ν_i are Fourier transforms of distributions with compact support (or, to abbreviate, entire functions with Paley–Wiener growth).

It is classical that, with these hypotheses, one has locally in $\mathbb{C}^n$, $\hat{\mu} = \Sigma \widehat{P_i} f_i$, f_i germs of holomorphic functions. Now the theory of Cartan–Oka proves that, in fact, one has a global result, i.e. $\hat{\mu} = \Sigma \widehat{P_i} f_i$, f_i entire. The problem is to prove that one can choose the f_i with Paley–Wiener growth.

The idea is to copy more or less the method of Cartan: first, get local bounds. Then, to globalize the result, use a theorem of vanishing of cohomology "with Paley–Wiener bounds". Note that, at this time, the idea of cohomology with bound was absolutely new.

As I said, these results were announced in 1960, in the paper "The fundamental principle for linear constant coefficients partial differential equations". A little more details were given in some monographed notes of lectures at Stanford. But it takes about 3 years to have a complete manuscript; and the final book "Fourier analysis in several complex variables" was not published before 1970. Needless to say that the book contains many more results on ellipticity, Cauchy problem, quasi-analytic classes, etc.

The 1960 announcement interested very much the (few) experts of the subject. At the first time, I was extremely surprised, may be a little bit sceptical. But, after two years, in the absence of complete proofs, I tried to give my own version. It differs from that of Ehrenpreis in two points:

First, I use a Dolbeault cohomology with bounds, instead of Čech cohomology as Ehrenpreis. In fact, a theorem of vanishing of Dolbeault cohomology "with Paley–Wiener growth" is surprisingly simple, much more that Čech cohomology with the same bound. Some time later, Hörmander got practically definitive results on Dolbeault cohomology with growth condition given by any plurisubharmonic function; he gave also an exposition of Ehrenpreis theory using this theorem.

The second difference, less important, is that I used (local) estimates on $\mathcal{C}^\infty$ functions, instead of holomorphic ones (the use of Dolbeault cohomology permits it). These estimates come from a development of the theory of division of distributions.

I note also that Palamodov gave also a version of the theory (his version is more close to that of Ehrenpreis).

These works finish essentially the subject. One could think of an extension to general systems of convolution equations, but this seems very difficult, or even almost impossible. The only one reasonable result to be expected was the density of exponential polynomials for general systems of homogeneous equations, a result obtained for one variable by Schwartz in his theory of mean periodic functions. But, in 1974 Gurevitch proved that the result is not true for several variables.

Let me finish by a few words about our personal relations. Actually, we met for the first time in Paris, in 1958 (if my memory is correct); this was a rather long time after our first works. Before that meeting, I thought of Ehrenpreis, with a little bit of tension, as a rather abstract person with whom I was more one less in competition. But, at our first meeting, he was so open and friendly that all tension disappeared totally. We became friends, although we did not meet so often. I remember especially a visit he made in Tunis, in 1970, where I stayed for one year. I think he was very pleased with this visit, except that the Jewish Tunisian food seemed not to fit him. Later, I met him several times in New-York, where I come often for familial reasons. He came to some lectures I gave to Courant Institute.

More recently, not a long time before his death, I had the surprise and pleasure to see him at a lecture I gave at Kolchin Seminar, in CUNY. I was especially happy, since I had not seen him since a rather long time, and we took the opportunity to remember old souvenirs. When leaving him, I could not imagine that it was our last meeting.

Leon Ehrenpreis, a Unique Mathematician

Daniele C. Struppa

1 Introduction

What made Ehrenpreis' mathematics so unique was his bold approach to classical problems, and his interest in finding an overarching and unifying framework for a variety of apparently unrelated problems. In this note I will try to highlight this characteristic, by looking at some of Ehrenpreis' papers which are not, strictly speaking, connected with either the Fundamental Principle or the Radon Transform.

Malgrange's section on the work that he, Hörmander, and Ehrenpreis accomplished in the context of systems of linear constant coefficient partial differential operators has illuminated a particularly intense period in the history of modern analysis: in this context, the contribution of Ehrenpreis is almost completely summarized in his first full length book [8].

The section authored by Kuchment, on the other hand, gives a beautiful picture of Ehrenpreis' involvement with integral geometry and its far reaching work on the Radon transform, as described in his pioneering work [15].[1]

My own involvement with Ehrenpreis stemmed from me being (from 1978 to 1981) a doctoral student of Carlos Berenstein, who himself was a former student of Leon. As such I came to meet Ehrenpreis many times during his frequent visits to College Park, Maryland. What I remember most from our conversations, and from his talks, was his overarching belief that one should consider the theory of holomorphic functions (in several complex variables) as a special case of a more

[1] As a somewhat amusing and personal note, I should mention that in the late 1980s I had founded a small publishing company in southern Italy, Mediterranean Press was its name; at that time Ehrenpreis was visiting my department, and he had accepted my invitation to write a book on the Radon transform for my company. During the next several years, I therefore saw several preliminary versions of the book, but by the mid-1990s I had left Italy, sold my equity in the company, and Ehrenpreis had found a much more appropriate outlet for his work.

D.C. Struppa (✉)
Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA
e-mail: struppa@chapman.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_2, © Springer-Verlag Italia 2012

general theory of overdetermined systems. As for the theory of holomorphic functions in one variable, one should be trying to think of it as a special case of a more general theory of mean-periodic functions. His belief in this general approach was illustrated by some of his most original and beautiful work. In this short note, I would like to focus on three specific instances in which his worldview allowed him to recreate classical theorems in a much more general setting, thus opening the way to fruitful and unexpected generalizations.

2 The Hartogs' Theorem

I will begin with the beautiful proof that Ehrenpreis gave in 1961 for the well-known Hartogs' theorem on the removability of compact singularities for holomorphic functions of more than one complex variable. The theorem states that if K is a compact set in $\mathbb{C}^n$, with $n \geq 2$, every holomorphic function outside of K can be extended (in a unique way) to a holomorphic function inside of K. This result, which was originally proved by Hartogs [17] in 1906 and was probably the first to demonstrate the unique flavor of complex analysis in several variables, has been given many different proofs and has been generalized to many settings [26, 28]. But it was only with Ehrenpreis' surprising [7] that it became clear that the result has little to do with holomorphic functions, but it is rather a consequence of an essentially algebraic property of the Cauchy–Riemann system. The actual statement of Ehrenpreis is as follows:

Theorem 1 *Let K be a compact set in $\mathbb{R}^n$, and let $P_1, \ldots, P_r$ be r polynomials in n complex variables with no common factors. Denote by $P_i(D)$ the differential operator that is obtained by replacing the complex variable $z = (z_1, \ldots, z_n)$ in $\mathbf{P}$ by the formal differential operator $D = (-i\partial/\partial x_1, \ldots, -i\partial/\partial x_n)$. Then every infinitely differentiable function on $\mathbb{R}^n \setminus K$ which is a solution, in $\mathbb{R}^n \setminus K$, of the system $\mathbf{P}(D)f = 0$, namely $P_1(D)f = \cdots = P_r(D)f = 0$, can be extended uniquely to an infinitely differentiable function on $\mathbb{R}^n$, solution everywhere of the same system. The new solution coincides with f on $\mathbb{R}^n \setminus K_\varepsilon$, where K_ε indicates a small ε neighborhood of K.*

The proof of the result is a brilliant (and early) example of the use of cohomology vanishing arguments. Essentially, one extends f in some arbitrary way to an infinitely differentiable function g and then notices that the collection $\{P_i(D)g\}$ is a compactly supported 1-cocycle with coefficients in the sheaf $\mathcal{R}$ of solutions of the system $\mathbf{P}(D)f = 0$. Using the Ehrenpreis–Malgrange division theorem [6, 23] (which essentially states that an entire quotient between a holomorphic function and a polynomial has the same growth order as the original holomorphic function), Ehrenpreis shows that the first cohomology group with compact support and with coefficients in the sheaf $\mathcal{R}$ vanishes, and therefore the 1-cocycle is a 1-coboundary, and the correction that this provides is sufficient to modify the original extension g into a global solution of the system.

This proof is beautiful on several counts: it is very simple (in fact, it can be given in full detail in just a few lines), it takes advantage of the equally beautiful Ehrenpreis–Malgrange lemma (in itself a powerful statement on the nature of polynomials), and finally it uncovers the fundamentally algebraic nature of the problem. The extension to the case of a rectangular system is technically more complicated and was given in detail, for example, in [25], as well as sketched in Ehrenpreis' own [8]. But the structure of the proof is so easy that it is in fact possible to generalize it further to infinite-order differential operators (see, for example, [20]) as well as to convolution equations as in [24]. We refer the reader to [26, 28] for a rather complete history of the various developments surrounding the various proofs of the Hartogs' theorem, and where more complete references (including the works of Kawai concerning the case of systems of variable coefficient differential equations) are given.

3 The Edge-of-the-Wedge Theorem

A second instance in which the theory of several complex variables is reinterpreted in a larger context is offered by Ehrenpreis' interest in a general approach to the question of extension of holomorphic functions. Clearly Hartogs' theorem is an example of such an interest, but Ehrenpreis was interested in a more general issue, in which the extension was not necessarily across a compact set. To this problem Ehrenpreis devoted a series of papers, [9–11, 13, 14, 16], whose focus, in a sense, is on the extension of the edge-of-the-wedge theorem, from the case of holomorphic functions to the case of more general solutions to overdetermined systems of linear constant coefficient differential equations.

This is not the place for a full discussion of the problem, but it is probably worth sketching at least the fundamental setting, which Ehrenpreis considered in his papers beginning with [9], but whose intellectual origins can once again be traced back to [8]. Consider s (not necessarily) different open sets $\Omega_1, \ldots, \Omega_s$ in $\mathbb{R}^n$ and r differential operators (once again not necessarily different) $\mathbf{D}_j = (D_{j1}, \ldots, D_{js})$ with constant coefficients. Suppose, furthermore, that there is a set X contained in every closed set $\overline{\Omega_j}$, which can be used to parameterize the solutions (in some suitable Analytically Uniform space) of $\mathbf{D}_j f_j = 0$ on each Ω_j. We use this term to indicate, in accordance with Chap. IX of [8], that a suitable Cauchy problem (determined by the operators $\mathbf{D}_j$ and initial values on X) is well posed. Suppose now that the solutions f_j satisfy on X some differential relations

$$\sum_{ij} a_{ij} \partial_i f_j = 0$$

generated by suitable constant coefficient differential operators ∂_i. Then one may ask what kind of consequences can be derived regarding the f_j. In particular, is it possible to extend them to being solutions of those same operators $\mathbf{D}_j$ on larger sets (this is a removability of singularities problem, of a very different nature from the one we examined in the section on the Hartogs' theorem, since the singularities are not confined, in this case, to compact sets)?

The study of such general Cauchy problems, and more precisely the conditions under which the problem is well posed (conditions on the spaces of functions involved, on the geometrical properties of the varieties associated to the operators, and on the specific geometry of X) are discussed in [8], but while the results of Ehrenpreis are extremely general, they are somewhat difficult to apply to specific conditions.

In the papers that begin with [9], however, Ehrenpreis fixes his attention on the way in which these results are far reaching generalizations of well-known function theory theorems. For example, if $s = 1$, $\Omega = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 > 0\}$, and if $r = 1$, with D being now the Laplacian, then one can consider a very special differential relation, say

$$\frac{\partial f}{\partial x_1} = 0,$$

and then any general theorem will end up being a generalization of what is known as the *reflection theorem*, namely the theorem that states that harmonic functions in the half-space $\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 > 0\}$, which satisfy $\partial f/\partial x_1 = 0$, can be extended to harmonic functions on all of $\mathbb{R}^n$.

By the same token, the edge-of-the-wedge theorem can be seen in this context. Take all the differential operators to be the Cauchy–Riemann systems in n variables, and take two open sets Ω_1, Ω_2 to be complex tubes over two convex cones in the real space. Then the differential relation is actually the request that the two functions f_1, f_2, holomorphic respectively on Ω_1 and Ω_2, coincide on the real boundary of the two tubes. The conclusion of the edge-of-the-wedge theorem then is that there is a holomorphic function f which extends the $f_j's$ to the convex hull of $\Omega_1 \bigcup \Omega_2$.

Once again, Ehrenpreis shows us here a very general approach to a variety of different problems in which holomorphicity (or harmonicity) are just special cases of functions which are solutions to more general systems of differential equations. I can only leave it to the reader to further explore these ideas in the articles cited in the references.

4 Infinite-Order Differential Operators and the Fabry Gap Theorem

Finally, I want to go back to the interest of Ehrenpreis in convolution equations, and in the role they could play in understanding some classical properties of holomorphic functions in one complex variable. As Malgrange has observed in his note, a full extension of the Ehrenpreis–Palamodov Fundamental Principle to (systems of) convolution equations is not possible, essentially because of the example of Gurevich to which Malgrange makes reference. This said, Ehrenpreis never abandoned the possibility that at least for some classes of convolutors, it may be possible to prove what is essentially a version of the Fundamental Principle. He first showed how to obtain a weak version of the Fundamental Principle in Chap. 11 of [8], but his result was somewhat hard to apply, and the restrictions on the convolutors are hard to decipher. But his intuition was in fact correct. That this was the case was shown first by Berenstein and Dostal [1] in a very special case, and later on by Berenstein and Taylor [5], at least for the case of systems of convolution equations with one unknown function. What Berenstein and Taylor show in [5] is that it is possible to construct a class of convolutors (which they call *slowly decreasing*, following a terminology already used by Ehrenpreis to indicate the condition that is necessary to establish surjectivity in suitable spaces) for which a reasonable analogue of the Fundamental Principle holds. Their theory was further extended to the case of rectangular systems of convolution equations in all LAU spaces in my dissertation [27]. It is worth pointing out (and in fact it is necessary in view of what will follow) that infinite-order differential operators on the space of holomorphic functions offer an example of slowly decreasing convolutors, and therefore the theory developed in [5, 27] can be applied to solutions of (systems of) such operators. One of the consequences of these extensions of the Fundamental Principle consists in the fact that convergent exponential sums, both in one and in several variables, can be considered as solutions to systems of slowly decreasing convolution equations and in particular (when holomorphic functions are considered) to systems of infinite-order differential equations.

This leads us to one of the most intriguing contributions of Ehrenpreis to classical complex analysis. In Chap. 12 of his monograph [8], as well as in [12], Ehrenpreis reconsiders the classical Fabry gap theorem. In brief, the theorem can be stated as follows: let z denote the complex variable, and, for complex numbers c_j and real numbers a_j, consider the series

$$\sum_{j=1}^{+\infty} c_j e^{ia_j z}.$$

Assume that, in the strip $|\mathrm{Im}\, z < 1|$, the series converges, uniformly on compact sets, to a function $f(z)$ which can be analytically continued to a neighborhood of some point z_0 on the boundary of the strip itself. Then, if the sequence $\{a_j\}$ is *lacunary* in the sense that

$$\frac{n}{a_n} \to 0 \quad \text{as } n \to \infty$$

and there exists a positive constant c such that

$$|a_n - a_m| \geq c|n-m|,$$

the function f can actually be continued analytically to an entire strip containing z_0, and, on the compact subsets of this new strip, the series $\sum_{j=1}^{+\infty} c_j e^{ia_j z}$ converges to the continuation of $f(z)$. There are several ways to look at this theorem, and maybe the most important classical reference is Levinson's important [22]. But Ehrenpreis offers in Chap. 12 of [8] two very unconventional approaches. One consists in noticing that every exponential $e^{ia_j z}$ is itself a solution of the particular differential equation

$$\frac{df}{dz} - ia_j f = 0,$$

and therefore it is not unreasonable to think of the series $\sum_{j=1}^{+\infty} c_j e^{ia_j z}$ as a series of solutions to different differential equations. Since exponentials, in several variables, appear as solutions to overdetermined systems of differential equations, this offers Ehrenpreis a way toward a very powerful generalization. More precisely, Ehrenpreis considers a sequence $\{\mathbf{D}_j\}$ of differential operators, with $\mathbf{D}_j = (D_{j1}, \ldots, D_{jn})$, and then seeks conditions on such operators that allow us to study overconvergence properties of the series $\sum f_j$, where the summands in the series are solutions to $\mathbf{D}_j f_j = 0$. The results that Ehrenpreis obtained in this direction are somewhat technical and probably ripe for further analysis. As far as I know, they have not yet been explored with the depth they seem to deserve.

But Ehrenpreis also offers another way of interpreting the series $\sum_{j=1}^{+\infty} c_j e^{ia_j z}$; specifically he points out that if $f(z) = \sum_{j=1}^{+\infty} c_j e^{ia_j z}$, then f can be thought of as a solution of the convolution equation $S * f = 0$, where S is the convolutor whose Fourier transform is, up to some converging factor, the entire function $\Pi(1 - z/a_j)$. It was this beautiful intuition that proved to be most fruitful and opened the way for a variety of interesting generalization. Most notable is probably Kawai's work [18, 19] on what he called the *Fabry–Ehrenpreis gap theorem*, and which stemmed from the interpretation of S as an infinite-order differential operator. Kawai's work is also extremely beautiful and brings into the picture the theory of hyperfunctions, as the natural environment for the study of infinite-order differential operators. As it often happens, new results open new doors, and Berenstein and the author pushed further some of these ideas and applied them to what they called now the *Fabry–Ehrenpreis–Kawai gap theorem* in a series of papers, which exploited the original intuition of Ehrenpreis and found its most general formulation in [2, 3, 21] and finally in [4]. In those papers, we believe that the original vision of Ehrenpreis on the role that convolution equations can play in understanding the overconvergence behavior of Dirichlet series (and generalized Dirichlet series) is carried out to a great extent.

References

1. Berenstein, C.A., Dostal, M.: Analytically Uniform Spaces and Their Applications to Convolution Equations. Springer Lecture Notes in Mathematics, vol. 256. Springer, Berlin (1972)
2. Berenstein, C.A., Struppa, D.C.: On the Fabry–Ehrenpreis–Kawai gap theorem. Publ. Res. Inst. Math. Sci. **23**, 565–574 (1987)
3. Berenstein, C.A., Struppa, D.C.: Dirichlet series and convolution equations. Publ. Res. Inst. Math. Sci. **24**, 783–810 (1988)
4. Berenstein, C.A., Kawai, T., Struppa, D.C.: Interpolation varieties and the Fabry–Ehrenpreis–Kawai gap theorem. Adv. Math. **122**, 280–310 (1996)
5. Berenstein, C.A., Taylor, B.A.: Interpolation problems in $\mathbb{C}^n$ with applications to harmonic analysis. J. Anal. Math. **38**, 188–254 (1980)
6. Ehrenpreis, L.: Solutions of some problems of division I. Am. J. Math. **76**, 883–903 (1954)
7. Ehrenpreis, L.: A new proof and an extension of Hartogs' theorem. Bull. Am. Math. Soc. **67**, 507–509 (1961)
8. Ehrenpreis, L.: Fourier Analysis in Several Complex Variables. Wiley-Interscience, New York (1970)

9. Ehrenpreis, L.: Edge of the wedge theorem for partial differential equations. Harmonic analysis in Euclidean spaces. Proc. Symp. Pure Math. **XXXV**, 203–212 (1979)
10. Ehrenpreis, L.: Reflection, removable singularities, and approximation for partial differential equations. I. Ann. Math. **112**, 1–20 (1980)
11. Ehrenpreis, L.: The edge-of-the-wedge theorem for partial differential equations. Ann. Math. Stud. **100**, 155–169 (1981)
12. Ehrenpreis, L.: Spectral gaps and lacunas. Bull. Sci. Math. **105**, 17–28 (1981)
13. Ehrenpreis, L.: Reflection, removable singularities, and approximation for partial differential equations. II. Trans. Am. Math. Soc. **302**, 1–45 (1987)
14. Ehrenpreis, L.: Extensions of solutions of partial differential equations. Geometrical and algebraical aspects in several complex variables. Semin. Conf. **8**, 361–375 (1991)
15. Ehrenpreis, L.: The Universality of the Radon Transform. With an Appendix by Peter Kuchment and Eric Todd Quinto. Oxford University Press, New York (2003)
16. Ehrenpreis, L.: Some novel aspects of the Cauchy problem. Harmonic analysis, signal processing, and complexity. Prog. Math. **238**, 1–14 (2005)
17. Hartogs, F.: Einige Folgerungen aux Cauchyschen Integralformel bei Funktionen Mehrer Veranderlichen. Munch. Sitzungber. **36**, 223–241 (1906)
18. Kawai, T.: The Fabry–Ehrenpreis gap theorem for hyperfunctions. Proc. Jpn. Acad., Ser. A, Math. Sci., **60**, 276–278 (1984)
19. Kawai, T.: The Fabry–Ehrenpreis gap theorem and systems of linear differential equations of infinite order. Am. J. Math. **109**, 57–64 (1987)
20. Kawai, T., Struppa, D.C.: On the existence of holomorphic solutions of systems of linear differential equations of infinite order and with constant coefficients. Int. J. Math. **1**, 63–82 (1990)
21. Kawai, T., Struppa, D.C.: Overconvergence phenomena and grouping in exponential representation of solutions of linear differential equations of infinite order. Adv. Math. **161**, 131–140 (2001)
22. Levinson, N.: Gap and Density Theorems. Am. Math. Society, New York (1940)
23. Malgrange, B.: Existence et approximation des solutions des équations aux dérivées partielles et des équations de convolution. Ann. Inst. Fourier (Grenoble) **6**, 271–355 (1956)
24. Meril, A., Struppa, D.C.: Phénomène de Hartogs et équations de convolution. Séminaire d'Analyse P. Lelong–P. Dolbeault–H. Skoda, Années 1985/1986. Springer Lecture Notes in Math., vol. 1295, pp. 146–156. Springer, Berlin (1987)
25. Palamodov, V.P.: Linear Differential Operators with Constant Coefficients. Springer, Berlin (1970)
26. Range, R.M.: Extension phenomena in multidimensional complex analysis: Correction of the historical record. Math. Intell. **24**, 4–12 (2002)
27. Struppa, D.C.: The fundamental principle for systems of convolution equations. Mem. Am. Math. Soc. **283**, 1–167 (1981)
28. Struppa, D.C.: The first eighty years of Hartogs' theorem. In: Geometry Seminars, pp. 127–209. Univ. Stud. Bologna, Bologna (1988)

Leon Ehrenpreis, Recollections from the Recent Decades

Peter Kuchment

Leon Ehrenpreis was an outstanding world-class mathematician and a wonderful, warm person. I had a privilege to consider myself his friend for the last two decades. It is hard to do justice to his manifold mathematics and personality, but I will try to at least add some recollections to this tribute volume.[1]

Leon Ehrenpreis has been one of my mathematical heroes for about 40 years. I first encountered his, Lars Hörmander's, Bernard Malgrange's, and Victor Palamodov's fundamental and beautiful works on systems of linear constant coefficient PDEs in early 1970s, when I was an undergraduate student and then a PhD candidate. They have had a profound impact on me, in particular when working on the Floquet theory of periodic PDEs, which we with Leonid Zelenko started developing in a few years. I am sure that Bernard Malgrange and Daniele Struppa have described this part of Leon's legacy much better than I ever could. I will only address some of the research Leon pursued in the last two decades of his life, which I was lucky to witness.

Some time around 1988, a medical industry contract forced me to learn the basics of a fascinating topic that I had never heard of before, the so-called computed tomography. This turned out to be fateful. Our research group in Voronezh found the mathematics of tomography so challenging and exciting that in the following decades several of us have being devoting a significant part of time working on tomographic problems. Appearance in the 1980s of the Russian translation of the cornerstone book on this topic by Frank Natterer [52] also helped. Interestingly enough, I discovered that several mathematicians whom I admired for their work in completely different areas (e.g., Carlos Berenstein, Simon Gindikin, and Victor Palamodov) had already been working on tomography-related issues. This is an instance of a strange effect that I have observed several times in my life, when several

[1] One can also read the AMS Notices article [30] for recollections of several Leon's friends and colleagues. A volume on tomography [10] is also dedicated to Leon's memory.

P. Kuchment (✉)
Texas A&M University, College Station, TX, 77843-3368, USA
e-mail: kuchment@math.tamu.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_3, © Springer-Verlag Italia 2012

people working in closely related areas suddenly and independently make a sharp turn to the same new direction.

The end of the 1980s was a fascinating time in the former Soviet Union, when contacts with the West have started to become somewhat possible. In particular, the existence of some of famous mathematicians could be checked experimentally (before that, the names like L. Ehrenpreis, L. Hörmander, P. Lax, L. Nierenberg, and many others seemed to me to belong to some deities rather than real people). In 1989 I had my first chance to travel abroad, and I spent about a month in the USA going to various universities and to an AMS tomography conference in Arcata, California. This is where I saw for the first time some of my scientific heroes (e.g., L. Ehrenpreis, S. Helgason, F. Natterer) in flesh.[2]

Meeting Leon in Arcata was a big surprise to me, since I had no clue that he had become interested in integral geometry or tomography. This was another instance of the simultaneous change of direction. He showed a polite interest in what I told him about my PDE work related to his, but it was clear that he was thinking in somewhat different (although not orthogonal) direction now. This was the first time when I heard Leon mentioning his book on Radon transform, which was "nearly finished." It did appear ... in 2003 [28]. In the 13–14 years in between, Leon had been sending generously the nth versions of his manuscript to anyone interested, and the ideas and problems contained in these texts have influenced many of us.

After emigrating later in 1989 to the USA, I found employment at the Wichita State University in Kansas. The year 1990 was a tough time for finding employment for a middle-age emigree mathematician with mediocre, at best, command of English. Having recommendation letters from colleagues such as L. Ehrenpreis was crucial, and I am indebted forever to them and many other mathematicians who supported me in various ways in these difficult times.

Settling down in Wichita was rather pleasant. My family loved the city. The mathematics department was quite good, including several prominent people in the areas of my interest, in particular in inverse problems (Victor Isakov and Ziqi Sun). When I started bringing in speakers and collaborators, Leon Ehrenpreis was one of the first invitees, and since then he had become a constant visitor of our department and then of the Mathematics department of Texas A&M, where I moved in 2001. His lectures and discussions that I and my graduate students had with him were extremely interesting, scientifically rewarding, and personally enjoyable.

I will skip some personal recollections, which one can find in [30] and concentrate rather on mathematics. One of the first topics that we discussed was a strange byproduct of the papers [48, 49] published a couple of years before. There we with S. Lvin described the range of the so-called exponential Radon transform, which

---

[2]The Arcata meeting was also the place where I met for the first time other colleagues, whom I now consider as long-time friends (J. Boman, D. Finch, A. Markoe, E. T. Quinto, G. Uhlmann, and many others). I could not even imagine that twelve years later I would have a privilege to work at the same department at Texas A&M with another group of researchers whose work I studied and admired as a young mathematician in Russia, such as Ron Douglas, Ciprian Foias, Carl Pearcy, and Gilles Pisier.

arises in the Single Photon Emission Computed Tomography (SPECT), an important medical imaging method [52]. I will not burden the reader with technicalities and just describe the result on a hand-waving level. It is known [34–36, 42, 43, 52] that the ranges of Radon-type transforms are usually of infinite codimension in natural function spaces. Knowing the description of the range plays an important role in integral geometry and tomography. After range conditions are found, it is usually straightforward to go back and check their necessity,[3] while a proof of their completeness is usually technical. Thus, when the conditions of [48] were found, we expected that reproving their necessity should be a piece of cake: just plug the transform of a function into these conditions and see immediately that they are satisfied. However, when we did this, we discovered an infinite and totally nonobvious to us set of nonlinear differential identities for the standard sine function: for any odd natural n,

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{d}{dx} - \sin x\right) \circ \left(\frac{d}{dx} - \sin x + i\right) \circ \cdots \circ \left(\frac{d}{dx} - \sin x + (k-1)i\right) \times \left((\sin x)^{n-k}\right) = 0, \tag{1}$$

where i is the imaginary unit, and $\circ$ denotes the composition of differential operators. The attempt to prove these identities directly (i.e., without any integral geometry and Fourier analysis) succeeded [49] but took a significant time. We are still puzzled by the meaning of these identities [50]. Several integral geometry and tomography experts devoted their time and effort to trying to understand better the meaning of these range conditions. This is also what we set out to do with my PhD student Valentina Aguilar and Leon Ehrenpreis. We succeeded in the following sense: we showed, in particular, that these identities are equivalent to an interesting theorem of separate analyticity type.

Theorem 1 ([8]) *Let D be a disk in $\mathbb{R}^2$, and f be a function in the exterior of D. Suppose that when restricted to any tangent line L to D, the function $f|_L$, as a function of one real variable, extends to an entire function on the complexification of L. Then f, as a function on $\mathbb{R}^2 \setminus D$, extends to an entire function on $\mathbb{C}^2$.*

Well, this fact also did not look obvious to us. Analyticity of f in a complex neighborhood of $\mathbb{R}^2 \setminus D$ follows from the old (and not that well-known) separate analyticity theorem by S. Bernstein (see [9]); however this theorem cannot produce statement about f being an entire function. Thus, since proving the above theorem, a couple of things about it kept bothering us for several years. First of all, this is a fact of several complex variables, while our proof did not look like an SCV argument at all. Is there a truly complex analysis proof? Another, related, question is whether such a theorem can be proven for a different convex body instead of a

[3] For instance, when the so-called *moment conditions* [35, 42] for the standard Radon transform are written, checking their necessity boils down to noticing that the kth power $(x \cdot \omega)^k$ of the inner product of two vectors is a homogeneous polynomial of degree k with respect to each of them.

disk D? An SCV proof was later provided in [54], although it was rather complicated and was not generalizable (at least, easily) to other convex curves. Leon has worked out some other examples of convex algebraic curves (unpublished), but general picture remained unclear to us. Finally, A. Tumanov presented recently [59] a beautiful short proof based on attachment of analytic disks (where Tumanov is a great expert), which works for any strictly convex body D with a mild condition on the smoothness of its boundary.

Another issue that we addressed with Leon and my Master student Alex Panchenko also originated from emission tomography. The exponential Radon transform in SPECT depends upon an "attenuation" parameter $\mu \geq 0$. In [29], we introduced and studied a "mother" exponential Radon transform, which had no free parameters, but by different restrictions of which one can obtain the exponential Radon transforms corresponding to all possible values of the attenuation. We also obtained the range description there, which was based upon the F. John's differential equations. In this particular case, the (ultrahyperbolic) John's equation could be recast as a boundary Cauchy–Riemann equation.

Although we have not done any joint research since 2000, we kept discussing (in person and by e-mail) various integral-geometric and PDE issues. One was the fascinating and surprisingly hard "strip problem" [1, 2, 4, 6, 37, 38, 57, 58], to which Leon has contributed [27, 28] and which he extended to a more general PDE setting (see, e.g., [6, 27, 28]). It was eventually resolved due to efforts of several mathematicians, including M. Agranovsky, J. Globevnik, and A. Tumanov (see the reference above).

Leon was also very much interested in the activity concerning the "restricted spherical means" operator, i.e., a version of Radon transform that integrates a given function over spheres of arbitrary radii, but with the centers restricted to a hypersurface S. The study of such operators was very active since the beginning of the 1990s, due first to needs of approximation theory , then self-sustained just due to the beauty and complexity of arising problems (see [3, 7] and references therein), and finally it received a huge boost in the last decade, due to the discovered relations to a newly developing method of medical imaging, the so-called thermo-/photo-acoustic tomography (see the surveys [5, 31–33, 47, 60] and references there).

The restricted spherical mean problem happens to be a very particular case of one of the questions raised by Leon in his book [28]. This brings us from the "small" problems discussed above to the much more general thinking Leon has been doing on transforms of Radon type and their very wide generalizations. This was reflected in his papers of the period and in the monograph [28]. The title of this book,"*The Universality of the Radon Transform*," and the wealth of topics and ideas covered and variety of open problems suggested shows how deeply Leon believed in wide range importance of this approach. He was not the first to realize such widespread applicability of transforms of Radon type, although probably the first to give such a bold name to a book. Fritz John in his book [46] showed how important this circle of ideas is for PDEs. Israel Gelfand, Simon Gindikin, Sigurdur Helgason, Victor Palamodov, and many other mathematicians studied in detail applications to PDEs, harmonic analysis, group representation theory, special functions, mathematical physics, etc. (e.g., [34–36, 42–45, 55]). Still, Leon's book is rather unique in terms of many nonstandard issues raised there. Leon also was unique in his writing style, introducing new notations and names for well-known objects, which does not help a reader. However, after getting through these hurdles, one opens a treasure chest of ideas.

The variety of things that Leon addressed in the book [28] and his other publications of the time [11–27], and which he considered inter-related, is enormous: "exotic" boundary-value problems for PDEs, Poisson summation formulas, Eisenstein and Poincare series on $SL(2,\mathbb{R})$ and $SL(3,\mathbb{R})$, various number-theoretic problems, Hartogs–Lewy extension, FBI transform (although it carries an unrecognizable name in [28], being an instance of what he called "nonlinear Fourier transform"), edge-of-the-wedge theorems, Phragmén–Lindelöf type theorems for PDEs, special functions, among others.

Notwithstanding the overarching title, a wide variety of topics covered, and large volume, [28] is neither a textbook on the "usual" Radon transform nor a comprehensive historical survey or reference manual; it is not designed for reading by an uninitiated; it does not cover many important developments, techniques, and results that one can find in [34–36, 39–45, 55, 56], such as curved manifolds case, κ-operator approach, Radon transforms of differential forms and tensors, projective geometry setting, most of the group representation relations, etc. At Leon's request, Todd Quinto and I contributed the appendix [51] to [28] devoted to a brief survey of some tomographic applications. Due to the natural size limitations, it also cannot be considered comprehensive. One can find a thorough discussion of tomographic issues in [52, 53].

In spite of all these omissions, this unique book [28] should occupy a space on the bookshelf of anyone working on PDEs, Fourier analysis, several complex variables, and integral geometry. I am sure it will be a source of inspiration for many mathematicians, who will take their time to get through the text.

The memory of Leon Ehrenpreis will stay with all who encountered his amazing mathematics and experienced his friendship. I am grateful to the fate for giving me the chance and privilege to meet Leon and to collaborate with him.

References

1. Agranovsky, M.: CR foliations, the strip-problem and Globevnik–Stout conjecture. C. R. Math. **343**(2), 91–94 (2006)
2. Agranovsky, M.: Propagation of boundary CR foliations and Morera type theorems for manifolds with attached analytic discs. Adv. Math. **211**(1), 284–326 (2007)
3. Agranovsky, M., Berenstein, C.A., Kuchment, P.: Approximation by spherical waves in L^p-spaces. J. Geom. Anal. **6**(3), 365–383 (1996)
4. Agranovsky, M., Globevnik, J.: Analyticity on circles for rational and real analytic functions of two real variables. J. Anal. Math. **91**, 31–65 (2003)
5. Agranovsky, M., Kuchment, P., Kunyansky, L.: On reconstruction formulas and algorithms for the thermoacoustic and photoacoustic tomography. In: Wang, L.H. (ed.) Photoacoustic Imaging and Spectroscopy, pp. 89–101. CRC Press, Boca Raton (2009)
6. Agranovsky, M., Narayanan, E.: Isotopic families of contact manifolds for elliptic PDE. Proc. Am. Math. Soc. **134**(7), 2117–2123 (2006)

7. Agranovsky, M.L., Quinto, E.T.: Injectivity sets for the Radon transform over circles and complete systems of radial functions. J. Funct. Anal. **139**, 383–413 (1996)
8. Aguilar, V., Ehrenpreis, L., Kuchment, P.: Range condition for the exponential Radon transform. J. Anal. Math. **68**, 1–13 (1996)
9. Akhiezer, N.I., Ronkin, L.I.: On separately analytic functions of several variables and theorems on “the edge of the wedge”. Russ. Math. Surv. **28**(3), 27–42 (1973)
10. Bal, G., Finch, D., Kuchment, P., Schotland, J., Stefanov, P., Uhlmann, G. (eds.) Tomography and Inverse Transport Theory. Contemp. Math., vol. 559, Am. Math. Soc., Providence (2011)
11. Ehrenpreis, L.: Special functions. Inverse Probl. Imaging **4**(4), 639–647 (2010)
12. Ehrenpreis, L.: Microglobal analysis. Adv. Nonlinear Stud. **10**(3), 729–739 (2010)
13. Ehrenpreis, L.: Eisenstein and Poincaré series on $SL(3,\mathbb{R})$. Int. J. Number Theory **5**(8), 1447–1475 (2009)
14. Ehrenpreis, L.: The Radon transform for functions defined on planes. In: Integral Geometry and Tomography. Contemp. Math., vol. 405, pp. 41–46. Am. Math. Soc., Providence (2006)
15. Ehrenpreis, L.: Some novel aspects of the Cauchy problem. In: Harmonic Analysis, Signal Processing, and Complexity. Progr. Math., vol. 238, pp. 1–14. Birkhäuser Boston, Boston (2005)
16. Ehrenpreis, L.: The role of Paley–Wiener theory in partial differential equations. In: The Legacy of Norbert Wiener: A Centennial Symposium, Cambridge, MA, 1994. Proc. Sympos. Pure Math., vol. 60, pp. 71–83. Am. Math. Soc., Providence (1997)
17. Ehrenpreis, L.: Parametric and nonparametric Radon transform. In: 75 Years of Radon Transform, Vienna, 1992. Conf. Proc. Lecture Notes Math. Phys., vol. IV, pp. 110–122. Int. Press, Cambridge (1994)
18. Ehrenpreis, L.: Some Nonlinear Aspects of the Radon Transform, Tomography, Impedance Imaging, and Integral Geometry (South Hadley, MA, 1993). Lectures in Appl. Math., vol. 30, pp. 69–81. Am. Math. Soc., Providence (1994)
19. Ehrenpreis, L.: Singularities, Functional Equations, and the Circle Method. The Rademacher Legacy to Mathematics (University Park, PA, 1992). Contemp. Math., vol. 166, pp. 35–80. Am. Math. Soc., Providence (1994)
20. Ehrenpreis, L.: Exotic parametrization problems. Ann. Inst. Fourier (Grenoble) **4**(5), 1253–1266 (1993)
21. Ehrenpreis, L.: Function theory for Rogers–Ramanujan-like partition identities. In: A Tribute to Emil Grosswald: Number Theory and Related Analysis. Contemp. Math., vol. 143, pp. 259–320. Am. Math. Soc., Providence (1993)
22. Ehrenpreis, L.: The Schottky relation in genus 4. In: Curves, Jacobians, and Abelian varieties, Amherst, MA, 1990. Contemp. Math., vol. 136, pp. 139–160. Am. Math. Soc., Providence (1992)
23. Ehrenpreis, L.: Extensions of solutions of partial differential equations. In: Geometrical and Algebraical Aspects in Several Complex Variables, Cetraro, 1989. Sem. Conf., vol. 8, pp. 361–375. EditEl, Rende (1991)
24. Ehrenpreis, L.: Hypergeometric functions. In: Special Functions (Okayama, 1990), ICM-90 Satell. Conf. Proc, pp. 78–89. Springer, Tokyo (1991)
25. Ehrenpreis, L.: Lewy unsolvability and several complex variables. Mich. Math. J. **38**(3), 417–439 (1991)
26. Ehrenpreis, L.: The Radon transform and tensor products. In: Integral Geometry and Tomography, Arcata, CA, 1989. Contemp. Math., vol. 113, pp. 57–63. Am. Math. Soc., Providence (1990)
27. Ehrenpreis, L.: Three Problems at Mount Holyoke. Contemp. Math., vol. 278. Am. Math. Soc., Providence (2001)
28. Ehrenpreis, L.: The Universality of the Radon Transform. Oxford Univ. Press, London (2003)
29. Ehrenpreis, L., Kuchment, P., Panchenko, A.: The exponential X-ray transform and Fritz John’s equation. I. Range description. In: Analysis, Geometry, Number Theory: the Mathematics of Leon Ehrenpreis (Philadelphia, PA, 1998). Contemporary Math., vol. 251, pp. 173–188. Am. Math. Soc., Providence (2000)

30. Farkas, H., Kawai, T., Kuchment, P., Quinto, T.E., Sternberg, S., Struppa, D., Taylor, B.A.: Remembering Leon Ehrenpreis: 1930–2010. Not. Am. Math. Soc. **58**(5), 674–681 (2011)
31. Finch, D., Patch, S., Rakesh: Determining a function from its mean values over a family of spheres. SIAM J. Math. Anal. **35**(5), 1213–1240 (2004)
32. Finch, D., Rakesh: The spherical mean value operator with centers on a sphere. Inverse Probl. **23**(6), S37–S50 (2007)
33. Finch, D., Rakesh: Recovering a function from its spherical mean values in two and three dimensions. In: Photoacoustic Imaging and Spectroscopy. CRC Press, Boca Raton (2009)
34. Gelfand, I., Gindikin, S., Graev, M.: Integral geometry in affine and projective spaces. J. Sov. Math. **18**, 39–167 (1980)
35. Gelfand, I., Gindikin, S., Graev, M.: Selected Topics in Integral Geometry. Transl. Math. Monogr., vol. 220. Am. Math. Soc., Providence (2003)
36. Gelfand, I., Graev, M., Vilenkin, N.: Generalized Functions. Integral Geometry and Representation Theory, vol. 5. Academic Press, San Diego (1965)
37. Globevnik, J.: Testing analyticity on rotation invariant families of curves. Trans. Am. Math. Soc. **306**, 401–410 (1988)
38. Globevnik, J.: Analyticity on translates of a Jordan curve. Trans. Am. Math. Soc. **359**, 5555–5565 (2007)
39. Guillemin, V.: Fourier Integral Operators from the Radon Transform Point of View. Proc. Symposia in Pure Math., vol. 27, pp. 297–300 (1975)
40. Guillemin, V.: On Some Results of Gelfand in Integral Geometry. Proc. Symposia in Pure Math., vol. 43, pp. 149–155 (1985)
41. Guillemin, V., Sternberg, S.: Geometric Asymptotics. Am. Math. Soc., Providence (1977)
42. Helgason, S.: The Radon Transform. Birkhäuser, Basel (1980)
43. Helgason, S.: Groups and Geometric Analysis. Am. Math. Soc., Providence (2000)
44. Helgason, S.: Geometric Analysis on Symmetric Spaces. AMS, Providence (2008)
45. Helgason, S.: Integral Geometry and Radon Transforms. Springer, Berlin (2010)
46. John, F.: Plane Waves and Spherical Means, Applied to Partial Differential Equations. Dover, New York (1971)
47. Kuchment, P., Kunyansky, L.: Mathematics of thermoacoustic and photoacoustic tomography. Eur. J. Appl. Math. **19**(2), 191–224 (2008)
48. Kuchment, P., Lvin, S.: Paley–Wiener theorem for exponential Radon transform. Acta Appl. Math. **18**, 251–260 (1990)
49. Kuchment, P., Lvin, S.: The range of the exponential Radon transform. Sov. Math. Dokl. **42**(1), 183–184 (1991)
50. Kuchment, P., Lvin, S.: Identities for $\sin x$ that came from medical imaging. Preprint, arXiv:1110.6109
51. Kuchment, P., Quinto, E.T.: Some problems of integral geometry arising in tomography. In: The Universality of the Radon Transform. Oxford Univ. Press, London (2003)
52. Natterer, F.: The Mathematics of Computerized Tomography. Wiley, New York (1986)
53. Natterer, F., Wübbeling, F.: Mathematical Methods in Image Reconstruction. Monographs on Mathematical Modeling and Computation, vol. 5. SIAM, Philadelphia (2001)
54. Öktem, O.: Extension of separately analytic functions and applications to range characterization of the exponential Radon transform. Ann. Pol. Math. **70**, 195–213 (1998)
55. Palamodov, V.P.: Reconstructive Integral Geometry. Birkhäuser, Basel (2004)
56. Sharafutdinov, V.A.: Integral Geometry of Tensor Fields. V.S.P. Intl Science (1994)
57. Tumanov, A.: A Morera type theorem in the strip. Math. Res. Lett. **11**, 23–29 (2004)
58. Tumanov, A.: Testing analyticity on circles. Am. J. Math. **129**(3), 785–790 (2007)
59. Tumanov, A.: Analytic continuation from a family of lines. J. Anal. Math. **105**, 391–396 (2008)
60. Wang, L. (ed.): Photoacoustic Imaging and Spectroscopy. CRC Press, Boca Raton (2009)

Part II
Invited Papers

Analyticity on Curves

Mark Agranovsky and Lawrence Zalcman

Abstract Under what conditions can one conclude that a continuous function on a plane domain Ω is holomorphic, given that its restrictions to a collection of Jordan curves in Ω which cover Ω admit holomorphic extensions? We survey progress on this problem over the past 40 years, with an emphasis on recent results.

1 Introduction

The circle of ideas discussed in this paper originates with the following:

Question *Let $f \in C(\mathbb{R}^2)$ and suppose that for each circle γ of (fixed) radius $r > 0$ in the plane, the restriction of f to γ has a continuous extension to the closed disc $\overline{D}_\gamma$ bounded by γ which is analytic in the open disc D_γ. Must f be an entire function?*

It is well known (and easy to see) that f extends from $\gamma_r(w) = \{z : |z - w| = r\}$ continuously to a function analytic on $D_{\gamma_r(w)}$ if and only if

$$\int_{\gamma_r(w)} f(z) z^n \, dz = 0, \quad n = 0, 1, 2, \ldots, \tag{1}$$

We dedicate this paper to the memory of our friend Leon Ehrenpreis. Leon was fascinated by the strip problem, contributed to its solution [13], and led the way in generalizing it from a result concerning analytic functions to solutions of elliptic equations [14]. Indeed, one of his last major addresses, the opening lecture of the conference Integral Geometry and Tomography, delivered at Stockholm University on August 12, 2008, was entitled "The Strip Theorem for PDE"; see [15, II–IV].

M. Agranovsky (✉) · L. Zalcman
Bar-Ilan University, Ramat-Gan 52900, Israel
e-mail: agranovs@macs.biu.ac.il

L. Zalcman
e-mail: zalcman@macs.biu.ac.il

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_4, © Springer-Verlag Italia 2012

or, equivalently,

$$\int_0^{2\pi} f\big(w + re^{i\theta}\big)e^{in\theta}\,d\theta, \quad n = 1, 2, \ldots. \tag{1'}$$

Thus, the question may be formulated as asking whether condition (1) or (1′) for all $w \in \mathbb{C}$ implies that f is entire.

Some forty years ago, this question was shown to have an affirmative answer [8, Proposition 1].

It is natural to ask how conditions (1) and (1′) might be weakened while still preserving the conclusion that f is globally analytic. Remarkably, it turns out that if $f \in C(\mathbb{R}^2)$ satisfies

$$\int_{\gamma_r(w)} f(z)\,dz = 0$$

for all $w \in \mathbb{C}$ and two fixed values r_1 and r_2 of r, then f is an entire function—so long as r_1/r_2 does not belong to the (countable) set of quotients of positive zeros of the Bessel function J_1 [27, Theorem 2]. It is not difficult to see that f need not be holomorphic anywhere if the integral condition holds only for circles of a single radius or if r_1/r_2 *is* a quotient of zeros of J_1 [27, p. 244].

Similarly, if $f \in C(\mathbb{R}^2)$ satisfies

$$\int_0^{2\pi} f\big(w + re^{i\theta}\big)e^{i\theta}\,d\theta = 0$$

and

$$\int_0^{2\pi} f\big(w + re^{i\theta}\big)e^{in\theta}\,d\theta = 0$$

for some fixed r and fixed $n \geq 2$ and all $w \in \mathbb{C}$, then f must be entire [28, Theorem 6]; cf. [12, Theorem 3.1].

Analogous results hold for analytic functions on the hyperbolic plane, i.e., the unit disc with the Poincaré metric [10, pp. 125–126] (cf. [11, Sect. 6] and [1, Theorem 2]).

These results, whose proofs involve the theory of mean-periodic functions, are cited principally to provide background. Our main concern in this paper is with conditions like (1), and it is to these matters that we now turn.

2 The Strip Problem and the Argument Principle

The results stated for circles in the previous section actually have generalizations to Jordan curves. In this paper, we are concerned with developments arising from the general version of the result of Agranovsky and Valsky cited above. This may be stated as follows.

Theorem 1 ([8, Proposition 1]) *Let* $\in C(\mathbb{R}^2)$, *and let* $\gamma \subset \mathbb{R}^2$ *be a smooth Jordan curve bounding the Jordan region* D. *Suppose that for any rigid motion* ω *of the plane, the restriction of* f *to* $\omega(\gamma)$ *extends continuously to a holomorphic function on* $\omega(D)$. *Then* f *is an entire function.*

Remark 1 Throughout, we take "smooth" to mean piecewise continuously differentiable, i.e., we assume that the contours considered are piecewise C^1, so that the existence of an analytic extension is equivalent to the analogue of (1) with $\gamma_r(w)$ replaced by $\omega(\gamma)$. The case of *arbitrary* Jordan curves (which might, for instance, have positive Lebesgue area) remains uninvestigated so far as we know and may yet yield some surprises.

The proof of Theorem 1 uses the following lemma, which can be regarded as a weak version of the argument principle.

Lemma 1 *Let* D *be a Jordan domain in the plane with piecewise* C^1 *boundary* $\gamma = \partial D$. *If* $f \in C^1(\gamma)$ *extends continuously into* D *as a holomorphic function and the mapping* $f : \gamma \mapsto f(\gamma)$ *has topological degree* 0, *then* f *is constant.*

Proof We claim that $f(\overline{D}) = f(\gamma)$. This implies that f is constant, since otherwise the left-hand side would have nonempty interior, while the right-hand side evidently has measure zero. Suppose then that $f(\overline{D}) \neq f(\gamma)$, so there exists $w_0 \in f(D) \setminus f(\gamma)$. Thus, the holomorphic function $f(z) - w_0$ does not vanish on γ but has zeros on D; so, by the argument principle,

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)-w_0}\,dz > 0.$$

On the other hand, since $\deg f = 0$, the change of variables $w = f(z)$ shows that the integral above equals

$$(\deg f)\cdot\frac{1}{2\pi i}\int_{f(\gamma)} \frac{1}{w-w_0}\,dw = 0.$$

This contradiction completes the proof. □

The proof of Theorem 1 follows easily from Lemma 1. Indeed, fix $z_0 \in \mathbb{C}$ and consider the averaged function

$$\hat{f}(z) = \frac{1}{2\pi}\int_0^{2\pi} f\big(z_0 + e^{i\varphi}(z-z_0)\big)\,d\varphi.$$

It satisfies the same condition of analytic extendibility inside the contours $\omega(\gamma)$ as f. However, $\hat{f}$ depends only on $|z - z_0|$ and hence has topological degree 0 on any closed curve $\omega(\gamma)$. By Lemma 1, f is constant on each such curve, which implies that $\hat{f}$ is identically constant. But this means that f satisfies the mean-value property at z_0. Since z_0 was arbitrary, f is harmonic. Since the function zf obviously satisfies the same extendibility condition as f, it too is harmonic. A simple computation, based on the identity $\Delta = 4\frac{\partial^2}{\partial z \partial \bar{z}}$, then shows that f is holomorphic.

Theorem 1 asserts that if

$$\int_{\omega(\gamma)} f(z) z^n dz = 0, \quad n = 0, 1, 2, \dots, \tag{2}$$

for every rigid motion ω, then f is analytic on $\bigcup_\omega \omega(\gamma) = \mathbb{C}$. If γ is a smooth Jordan curve other than a circle and condition (2) is weakened to

$$\int_{\omega(\gamma)} f(z)\, dz = 0$$

for every rigid motion ω of $\mathbb{C}$, i.e., if γ has the *Morera property* [29, p. 186], then it is very often the case that one can conclude that f is analytic in $\mathbb{C}$. In fact, not a single example to the contrary is known! This is closely related to the celebrated Pompeiu property, for which see [12, 25–27, 29, 30].

The averaging technique used in the proof of Theorem 1 suggests investigating what may be learned from considering the condition (2) when ω is restricted to belong to the compact group S^1 of rotations of the plane. This line of inquiry was initiated by Josip Globevnik in [17] and pursued by him in a number of subsequent publications. For instance, in [17] (see also [18, 19]), he takes γ to be a Jordan curve not surrounding the origin and allows ω to range over the group S^1 of all rotations about the origin, so that $\omega_t(\gamma) = e^{it}\gamma$. The region $\Omega = \bigcup_t \omega_t(\gamma)$ is then an annulus about 0, and (2) for all $\omega \in S^1$ implies that $f \in C(\Omega)$ is analytic in the interior of Ω.

It is natural to replace the group S^1 of rotations by the (noncompact) translation group R^1. Given a Jordan curve γ in the plane, consider the collection $\{\gamma_t : t \in \mathbb{R}\}$ of all translates $\gamma_t = t + \gamma$ of γ. Then the region covered by $\{\gamma_t\}$ is the strip

$$\Omega = \bigcup_{t \in \mathbb{R}} (t + \gamma) = \{z : \alpha \le \operatorname{Im} z \le \beta\},$$

where $\alpha = \min\{\operatorname{Im} z : z \in \gamma\}$ and $\beta = \max\{\operatorname{Im} z : z \in \gamma\}$. Denoting by D_t the Jordan domain bounded by γ_t and by Ω° the interior of Ω, we may consider the following question. □

Strip problem *Let γ be a Jordan curve in $\mathbb{C}$ and set $\Omega = \{z + t : z \in \gamma,\ t \in \mathbb{R}\}$. Let $f \in C(\Omega)$ and suppose that for each $t \in \mathbb{R}$, the restriction of f to γ_t extends continuously to D_t as a holomorphic function. Must f be holomorphic on Ω°?*

Of particular interest is the case in which $\gamma = \{z : |z| = 1\}$ is the unit circle and $\Omega = \{z : -1 \le \operatorname{Im} z \le 1\}$ is the strip of width 2 about the real axis. In this form, the strip problem seems first to have been proposed by Globevnik in a lecture entitled “Analyticity on families of curves,” delivered at Bar-Ilan University in November, 1987; see [20, p. 1921]. The designation “strip problem” is due to Agranovsky.

As noted previously, the condition of analytic extendibility in the strip problem is equivalent to the vanishing of all analytic moments

$$\int_{\gamma_t} f(z)z^n\,dz, \quad n=0,1,2,\ldots,$$

for all $t \in \mathbb{R}$. In striking contrast to the results discussed earlier, here we cannot drop even a single moment. For instance, if γ is the unit circle, so that Ω is the strip $\{z : -1 \le \operatorname{Im} z \le 1\}$, and $m \ge 1$ is odd, then $f(t+e^{i\theta}) = \sin m\theta$ is well defined on Ω and satisfies (2) for all $n \ge 0$, $n \ne m$, but is holomorphic nowhere in Ω°. (For m even, one takes instead $\cos m\theta$.)

In certain circumstances, the strip problem can be given an elementary solution. Such is the case, for instance, where $\gamma = \partial Q$, where Q is the square with vertices $(1,1)$, $(-1,1)$, $(-1,-1)$, and $(1,-1)$. In this instance, the difference between the extensions of f to Q_t and Q_s vanishes on an interval in $\gamma_t \cap \gamma_s$ when $|s-t|<2$ and so must vanish identically on $Q_t \cap Q_s$. Thus, these extensions define a single-valued holomorphic function in $\{z : -1 < \operatorname{Im} z < 1\}$. Unfortunately, this simple argument is no longer available when the curves of the family $\{\gamma_t\}$ meet one another in only a finite set.

For functions that do not grow too rapidly, it is possible to resolve the strip problem using the Fourier transform. Indeed, suppose that $f \in C(\Omega)$ satisfies the hypotheses of the strip problem and that

$$\int_{\mathbb{R}} \bigl|f(x,y)\bigr|\,dx < \infty$$

for each $\alpha < y < \beta$. Then f is holomorphic in Ω°.

For the proof, set

$$\hat{f}(x,y,\lambda) = \int_{\mathbb{R}} f(x+t,y)e^{-i\lambda t}\,dt.$$

Clearly,

$$\hat{f}(x,y,\lambda) = e^{i\lambda x}\int_{\mathbb{R}} f(x+t,y)e^{-i\lambda(x+t)}\,dt = e^{i\lambda x}\hat{f}(\lambda,y),$$

where

$$\hat{f}(\lambda,y) = \int_{\mathbb{R}} f(x,y)e^{-i\lambda x}\,dx$$

is the Fourier transform of f with respect to the x-variable. Since the property of holomorphic extendibility is invariant under translations $x \mapsto x+t$, the function $(x,y) \mapsto \hat{f}(x,y,\lambda)$ possesses the same property for each fixed $\lambda \in \mathbb{R}$. Multiplication by a holomorphic function obviously preserves the property; therefore,

$$g(x,y,\lambda) := f(x,y,\lambda)e^{-i\lambda(x+iy)} = \hat{f}(\lambda,y)e^{\lambda y}$$

also has the property. However, this last function evidently does not depend on x, so the mapping $g_\lambda(x, y) := g(x, y, \lambda)$ has topological degree 0. By Lemma 1, $g_\lambda = c(\lambda)$, and hence,

$$\hat{f}(\lambda, y) = g(x, y, \lambda)e^{-\lambda y} = c(\lambda)e^{-\lambda y}.$$

Thus, by Fourier inversion,

$$f(x, y) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\lambda, y)e^{i\lambda x}dx = \frac{1}{2\pi}\int_{\mathbb{R}} c(\lambda)e^{i\lambda(x+iy)}\,dx,$$

which is clearly holomorphic in $z = x + iy$.

When f is not integrable on lines parallel to the x-axis, this approach fails (but see [13]). In any case, Fourier analysis does not seem to help in finite versions of the strip problem, in which the growth of f plays no role at all. Specifically, one can restrict the parameter t in the strip problem to some finite interval $[a, b]$. In this case, as already observed by Globevnik, the strip must not be too short if the problem is to have an affirmative solution. As an example, take $\gamma = \{z : |z| = 1\}$ and let $t \in [-1, 1]$. Then the function $f(z) = z^2/\overline{z}$ (understood to be 0 at $z = 0$) is continuous on $\Omega = \bigcup_{t\in[-1,1]}(t + \gamma)$ and admits analytic extension to each domain D_t, but obviously fails to be analytic on Ω°; see [17, Example 1].

The problems discussed in this section are particular cases of the following general problem.

Generalized strip problem *Let $\{\gamma_t,\ t \in [a, b]\}$ be a continuous or smooth one-parameter family of Jordan curves in the complex plane. Set $\Omega = \bigcup_t \gamma_t$. Suppose $f \in C(\Omega)$ and that for each $t \in [a, b]$, the restriction of f to γ_t extends continuously as a holomorphic function on the region D_t bounded by γ_t. Under what conditions can we conclude that f is holomorphic on Ω°?*

The rest of this paper is devoted to a discussion of various aspects of this problem and progress toward its solution.

3 Generic Families of Circles and Rational Functions: Lifting the Problem in $\mathbb{C}^2$

It was observed in [6] that the strip problem can be viewed as a problem in the space $\mathbb{C}^2$ of two complex variables. Actually, this is evident from the presence of the additional parameter t. Indeed, one can define the function

$$F(\zeta, t) = F_t(t + \zeta),$$

where $F_t(\zeta)$ is the holomorphic extension of f in the domain $D_t := t + D$. The function $F(\zeta, t)$ is then defined on the solid cylinder $D \times \mathbb{R}$ and is holomorphic in ζ. If $\zeta \in \gamma = \partial D$, then $t + \zeta \in \Omega$, and we have $F(\zeta, t) = f(\zeta + t)$. Therefore, the function $F(\zeta, t)$ takes the same values at the points of intersection of the straight line $\zeta + t = c$ with the cylinder $\partial D \times \mathbb{R}$.

Now the question is whether this identification continues to hold on the segment between the points of intersection, i.e., whether $F(\zeta, t) = c$ on the segments $\{\zeta + t = c\} \cap (D \times \mathbb{R})$. Indeed, if this is the case, then $F(\zeta, t) = g(\zeta + t)$ for some function g, which is holomorphic since F is holomorphic in ζ. Moreover, when $\zeta \in \gamma$, $g(\zeta + t) = F(\zeta, t) = f(\zeta + t)$, and we conclude that f is holomorphic in $\Omega = \bigcup \gamma_t$. Unfortunately, it was not clear how to make use of the above model.

A different higher-dimensional model of the strip problem appeared in [6]. This model is based on an embedding into $\mathbb{C}^2$ and leads to a solution of the strip problem for rational functions of two real variables. Specifically, a test of holomorphy is obtained for generic one-parameter families of circles and for functions of the form $f(x, y) = P(x, y)/Q(x, y)$, where P and Q are polynomials in two variables.

To formulate the result from [6], we require some notation. Let $I = [t_0, t_1]$, and let $c : I \to \mathbb{C}$ and $r : I \to (0, \infty)$ be continuous mappings, where c is not constant. Denote by $C(t)$ the circle $\{z \in \mathbb{C} : |z - c(t)| = r(t)\}$ and by Ω the union

$$\Omega = \bigcup_{t \in I} C(t).$$

For $a, b \in \mathbb{C}$, $a \neq b$, define the special family of circles:

$$\mathscr{R}_{a,b} = \Big\{C(t) : c(t) = a + t(b - a), r(t) = |b - a|\sqrt{t(t+1)},\ t \in [0, \infty)\Big\}.$$

Theorem 2 ([6]) *Let*

$$f(x, y) = \frac{P(x, y)}{Q(x, y)}$$

be a rational function (of the real variables x and y) such that $Q(x, y) \neq 0$ for $(x, y) \in \Omega$. Suppose that $f|_{C(t)}$ extends analytically into the disc $D(t) := \{|z - c(t)| < r(t)\}$ for every $t \in I$. Then f is holomorphic in Ω, i.e., f is a rational function of $z = x + iy$ with poles outside Ω, unless the family $\{C(t) : t \in I\}$ is contained in an exceptional family $\mathscr{R}_{a,b}$ for some $a, b \in \mathbb{C}$.

Sketch of Proof First, we can assume that the polynomials P and Q are coprime (have no common polynomial factors) and also rewrite P and Q as polynomials in z and $\overline{z}$, writing again P and Q for the new polynomials $P(z, \overline{z})$, $Q(z, \overline{z})$.

Observe that the function $f(z, \overline{z}) := P(z, \overline{z})/Q(z, \overline{z})$ always admits continuous extension into the disc $D(t)$ as a rational function of z. This extension is obtained by substituting the expression $\overline{z} = \overline{c(t)} + r^2(t)/(z - c(t))$ into P and Q, as follows from the equation of the circle. Thus, the assumption on f means that this rational function has no poles in $D(t)$.

Now we lift the whole construction to $\mathbb{C}^2$ by embedding $\mathbb{C} \ni z \mapsto (z, \overline{z}) \in \mathbb{C}^2$. The complexification of the circle $C(t)$ is the quadric $(z - c(t))(w - \overline{c(t)}) = r^2(t)$ in $\mathbb{C}^2$. The circle $C(t)$ itself becomes the boundary of the semiquadric

$$\Lambda_t = \Big\{(z, w) \in \mathbb{C}^2 : \big(z - c(t)\big)\big(w - \overline{c(t)}\big) = r^2(t), \big|z - c(t)\big| < r(t)\Big\}.$$

The center $z = c(t)$ of the disc $D(t)$ corresponds to $w = \infty$.

The assumption on f now translates to the condition that for each fixed t, the rational function

$$h(z) = \frac{P\big(z, \overline{c} + \frac{r^2}{z-c}\big)}{Q\big(z, \overline{c} + \frac{r^2}{z-c}\big)}, \quad c = c(t),$$

has no poles on the disc $D(t)$.

These poles come from zeros of $Q(z, w)$, and there are only two possibilities:

Case 1. $Q(z, w)$ has at least one zero on each semiquadric Λ_t, $t \in I$;
Case 2. $Q(z, w)$ has no zeros on each semiquadric Λ_t, $t \in I$.

Indeed, to say that $Q(z, w)$ vanishes on some Λ_t means that the extension of the polynomial $Q(z, \overline{z})$ into the disc $\{|z - c(t)| < r(t)\}$ as a rational function of z has a zero different from $z_t = c(t)$. The above analytic extension can be represented as $\frac{Q_t(z)}{(z-c(t))^m}$, where $Q_t(z)$ is a polynomial and m is the degree of Q with respect to $\overline{z}$. Therefore, $Q_t(z)$ has a zero in the disc $|z - c(t)| < r(t)$. By assumption, Q does not vanish on the boundary $|z - c(t)| = r(t)$ of the disc; hence neither does the polynomial $Q_t(z)$. The dependence on t of the circles $C(t)$ and the polynomials Q_t is continuous, so by the argument principle, the number of zeros of Q_t inside the disc $D(t)$ is the same for all t. Therefore, Q vanishes on each Λ_t once it vanishes on one of them.

If Case 1 holds, then the algebraic curve $Q^{-1}(0)$ meets each Λ_t in at least one point (z_t, w_t). However, the rational function P/Q has no poles on Λ_t, and hence the zero of the denominator Q must cancel a zero of the numerator P, which means that the algebraic curve $P^{-1}(0)$ passes through the same point (z_t, w_t).

The polynomials P and Q are coprime; hence, by Bezout's theorem, the set $P^{-1}(0) \cap Q^{-1}(0)$ is finite. Using this fact, it is not hard to prove that all Λ_t must pass through the same point $(z_t, w_t) = (a, \overline{b})$. A simple calculation then shows that the family $\{C(t) : t \in I\}$ is included in the exceptional family $\mathscr{R}_{a,b}$.

Now suppose that Case 2 holds. Fix the parameter t. We know that Q does not vanish on Λ_t. Consider the family of semiquadrics

$$\Lambda'_s := \big\{(z, w) \in C^2 : \big(z - c(t)\big)\big(w - \overline{c}(t)\big) = s^2 r^2(t), \big|z - c(t)\big| < sr(t)\big\},$$
$$s \in (0, 1],$$

shrinking, as $s \to 0$, to the complex line $\{c(t)\} \times \mathbb{C}$. Using the argument principle, one shows that Q has no zeros on any Λ'_s and hence does not vanish on the limit complex line. The important point here is that the zeros do not go to the boundary of Λ'_s, because this boundary belongs to $\Omega \subset \{(z, \overline{z})\}$, where Q is free of zeros by the assumption.

Since $Q(c(t), w)$ is a polynomial in w, this means that $Q(c(t), w)$ is constant in w. Because this is true for all t, it follows that $Q(z, w) = Q(z)$ is a holomorphic polynomial (i.e., does not contain $w = \overline{z}$).

It is now easy to finish the proof. Since Q is holomorphic, the original condition is true also for the polynomial $P = fQ$. The holomorphic extension $P_t(z)$ of the polynomial $P(z,\overline{z}) = p_0(z) + p_1(z)\overline{z} + \cdots + p_m(z)\overline{z}^m$ from the circle $|z - c(t)| = r(t)$ can be written in the form

$$P_t(z) = \frac{a_m(z)(z - c(t))^m + \cdots + a_0(z)}{(z - c(t))^m},$$

where $a_j(z)$ are polynomials and $a_0(z) = p_m(z)$. If $m > 0$, then $a(c(t)) = p_m(c(t)) = 0$ because P_t has no pole at $z = c(t)$. This is true for all t, so by the uniqueness theorem, $p_m = 0$. Thus $P(z) = p_0(z)$ is a holomorphic polynomial, and $f = P/Q$ is a holomorphic rational function.

To show that the theorem fails for the exceptional families, it suffices to consider $\mathscr{R}_{0,1} = \{C(t) : t \in [0,\infty)\}$, where $C(t) = \{z : |z - t| = \sqrt{t(t+1)}\}$. The domain Ω covered by the circles is the half-plane $\operatorname{Re} z > -\frac{1}{2}$. The function $\frac{z}{\overline{z}+1}$ satisfies the conditions of the theorem and, in particular, extends analytically without poles from any circle $C(t)$; however, it is not holomorphic. □

Remark 2 The example $f(z) = z^2/\overline{z}$ and circles enclosing 0 cited in Sect. 2 does not contradict Theorem 2, since the denominator $Q(x, y) = x - iy$ vanishes at 0 and hence does not satisfy the condition $Q \neq 0$ in Ω.

An analogue of Theorem 2 for real-analytic functions was also proved in [6].

Theorem 3 ([6]) *Let $c : (0,1) \to \mathbb{C}$ be a C^1-function such that $c'(t) \neq 0$, $t \in (0,1)$, and $r : (0,1) \to [0,\infty)$ be a positive C^1 function. Suppose that $|r'(t)| < |c'(t)|$, $t \in (0,1)$. Let f be a real-analytic function in the neighborhood of the union Ω of the circles $C(t) = \{z : |z - c(t)| = r(t)\}$. If, for all $t \in (0,1)$, the restriction $f|_{C(t)}$ extends holomorphically into the disc bounded by $C(t)$, then f is holomorphic in Ω.*

Using different methods, based on the Fourier transform, Ehrenpreis [13] also solved the strip problem for real analytic functions when γ is a circle.

4 Generic Families of Circles, Continuous Functions

Shortly after the work described in the previous section was completed, Tumanov [23] obtained a solution to the strip problem for the case of circles in which no additional conditions (beyond continuity) are made on the functions under consideration. In [23] the case of circles centered on a line was considered, and in [24] generic families of circles were treated. Here we formulate (with slight editorial changes) the more general result from the article [24].

Theorem 4 ([24]) *Let $\{C(t) : t \in [0,1]\}$ be a family of circles whose centers $c(t)$ and radii $r(t)$ are continuous functions on $[0,1]$ and piecewise C^3 smooth functions on $(0,1)$. Assume that $c'(t) \neq 0$ and $|r'(t)| < |c'(t)|$, $t \in (0,1)$. Denote by $D(t)$ the disc bounded by $C(t)$. Suppose that*

(a) $\overline{D(0)} \cap \overline{D(1)} = \emptyset$, *and*
(b) *no circle* $C(t)$ *encloses another circle* $C(s)$.

Let $\Omega = \bigcup_{t\in[0,1]} D(t)$ *and* $f \in C(\overline{\Omega})$. *If, for every* $t \in [0,1]$, *the restricted function* $f|_{C(t)}$ *extends holomorphically into* $D(t)$, *then* f *is holomorphic in* Ω.

As in [6], the proof is based on viewing the problem in $\mathbb{C}^2$ via the embedding $\mathbb{C} \ni z \mapsto (z, \overline{z}) \in \mathbb{C}^2$. In [6], the function f is assumed rational, i.e., of the form $P(z,\overline{z})/Q(z,\overline{z})$. This function naturally defines a rational function $P(z,w)/Q(z,w)$ in $\mathbb{C}^2$, and the proof in [6] is based on studying the location of the zero sets of the polynomials P and Q with respect to the semiquadrics Λ_t obtained by the complexification of the circles $C(t)$. However, when the function f is merely continuous, there is no natural global extension to $\mathbb{C}^2$. Nevertheless, f extends to a certain three-dimensional CR-manifold in $\mathbb{C}^2$ as a CR-function. Namely, f possesses a holomorphic extension from each circle $C(t)$ to a complexified circle, the semi-quadric Λ_t. These extensions define a CR-function $F(z,w)$ on the union $\Lambda = \cup \Lambda_t$ of the semi-quadrics, which is a CR-manifold.

At this point, CR theory comes in. The main part of the proof in [24] is to show that the function $F(z,w)$ does not depend on w, which then implies that f is holomorphic. The main tool comes from Hans Lewy's proof of an extension theorem and uses Cauchy-type integration over loops belonging to the CR-manifold Λ.

5 Generic Families of Jordan Curves. The Argument Principle is Back

Soon after the case of circles was finally resolved in [23, 24], the problem was solved for generic families of Jordan curves [2, 3] under the condition that the functions tested are real-analytic.

Of particular interest is the fact that the argument principle, which appeared in the earlier works [8, 17] as a technical tool, returns now to the stage as a main player. It turns out that the strip problem itself has a topological meaning and can be regarded as a parametric version of the argument principle.

Let us first formulate the result in question and then explain how it is related to the argument principle. We consider one-parameter families $\{\gamma_t\}$, $t \in [0,1]$, of real-analytic Jordan curves in the complex plane. The dependence on the parameter t is assumed to be real-analytic as well, i.e., the Riemann mappings

$$G(\cdot, t) : \Delta \to D_t$$

of the unit disc onto the domain D_t bounded by γ_t can be chosen to depend real-analytically on t.

Write $\Omega = \bigcup_{t\in[0,1]} \gamma_t$. We call the family γ_t *regular* if

$$\frac{\partial(G, \overline{G})}{\partial(\psi, t)}(\psi, t) \neq 0$$

for all (ψ, t) such that $G(e^{i\psi}, t) \notin \partial\Omega$.

Example In the strip problem, one has $G(\zeta, t) = \zeta + t$ and $\partial(G, \overline{G})/\partial(\psi, t) = 2i \cos\psi$. The boundary $\partial\Omega$ of the strip corresponds to $\psi = \pm\pi/2$; at all other points, the Jacobian is different from zero.

Theorem 5 ([2]) *Let $\gamma_t = \partial D_t$, $t \in [0, 1]$, be a real-analytic family of real-analytic Jordan curves in $\mathbb{C}$. Suppose that $\bigcap_{t\in[0,1]} \overline{D_t} = \emptyset$. Let $\Omega = \bigcup_{t\in[0,1]} \gamma_t$, and let f be a real-analytic function in a neighborhood of Ω. If for each $t \in [0, 1]$, f extends continuously from γ_t to a holomorphic function in D_t, then f is holomorphic on Ω°.*

Remark 3 The condition of empty intersection of the closed domains $\overline{D_t}$ is weaker than the condition in Theorem 4. On the other hand, Theorem 5 fails without the condition of empty intersection, as is shown, for example, by $f(z) = z^2/\overline{z}$. While this function is obviously not holomorphic, it extends holomorphically from any circle enclosing the origin. The closed discs bounded by these circles all have the point 0 in common.

Sketch of Proof For simplicity, we illustrate the idea of the proof for the case where the family of curves is closed, that is, $\gamma_0 = \gamma_1$. In that case, the parameterizing manifold can be taken to be the circle S^1. The curves γ_t sweep out an annulus-like domain in the complex plane.

Suppose that f is not holomorphic on Ω°. Choose a real-analytic parametrization $G(\zeta, t)$ of the family γ_t. For each t, the mapping $\zeta \mapsto G(\zeta, t)$ is a conformal mapping of the unit disc Δ onto the domain D_t bounded by γ_t, and $G(\zeta, t)$ is real-analytic in $(\zeta, t) \in \overline{\Delta} \times S^1$.

We have $G(\partial\Delta \times \{t\}) = \gamma_t$, and hence $\Omega = G(\partial\Delta \times S^1)$. The first important observation is that the topological degree of the mapping $G : \partial\Delta \times S^1 \to \Omega$ is zero, because G maps the manifold without boundary (the 2-torus $\partial\Delta \times S^1$) onto the domain $\Omega \subset \mathbb{C}$ having nonempty boundary.

Let $F_t \in H(D_t)$ be the holomorphic extension of f from the curve $\gamma_t = \partial D_t$. Set $F(\zeta, t) = F_t(G(\zeta, t))$, $\zeta \in \Delta$. Then we can construct the real-analytic mapping

$$\Phi = (F, G) : \overline{\Delta} \times S^1 \to \mathbb{C}^2$$

of the solid torus $\overline{\Delta} \times S^1$ into $\mathbb{C}^2$.

Since $F_t = f$ on γ_t, we have $F(\zeta, t) = f(G(\zeta, t))$ when $\zeta \in \partial\Delta$. Therefore, $\Phi(u) = (f(G(u)), G(u))$ for $u = (\zeta, t), |\zeta| = 1$, and Φ maps $\partial\Delta \times S^1$ onto the graph of f over Ω:

$$\Phi : \partial\Delta \times S^1 \mapsto \mathrm{Graph}_\Omega\, f.$$

Since for $u \in \partial\Delta \times S^1$, $\Phi = \pi \circ G$, where $\pi(z) = (f(z), z)$, we have for the topological degree $\deg\Phi = (\deg\pi)(\deg G) = 0$.

Step 1 (cocycle L). It is proved in [2] that the condition of empty intersection is equivalent to the curve $C := \Phi(\{0\} \times S^1)$ being homologically nontrivial in $X := \Phi(\overline{\Delta} \times S^1)$ This means that C is not the boundary of any 2-chain lying in X.

The proof is based on the covering homotopy theorem [22, pp. 61–66]. If $C = \partial Z$ for some 2-chain $Z \subset X$, then C can be homotopically deformed to a point z_0 within

X. Using the covering homotopy theorem, one shows that there is a homotopy of the curve (circle) of centers $A := \{0\} \times S^1$ to a closed curve $B \subset \overline{\Delta} \times \{0\}$. Like A, the curve B is homotopically nontrivial in $\overline{\Delta} \times S^1$. The mapping Φ maps B to a point z_0. But the curve B intersects each disc $\Delta_t := \overline{\Delta} \times \{t\}$, which implies that the image $\Phi(B) = \{z_0\}$ belongs to each closed domain $D_t = \Phi(\Delta_t)$. Contradiction.

By Poincaré duality, the homological nontriviality of the curve C is equivalent to the condition that there exists a (smooth) curve $L \subset \mathbb{C}^2$, whose endpoints lie outside $\Phi(\overline{\Delta} \times S^1)$, which intersects C transversally with positive intersection index, $\mathrm{Ind}(L \cap C) > 0$.

Step 2 (the function $J(\zeta, t)$**).** The next step is to construct a nonzero function $J(\zeta, t)$ in the cylinder $\Delta \times [0, 1]$ with the following properties:

(i) $J(\zeta, t)$ is holomorphic in $\zeta \in \Delta$;
(ii) $J(0, t) = 0$;
(iii) $\frac{J^2(u_1)}{|J(u_1)|^2} = \frac{J^2(u_2)}{|J(u_2)|^2}$ for any points $u_j = (\zeta_j, t_j)$, $j = 1, 2$, such that $J(u_1) \neq 0 \neq J(u_2)$, $|\zeta_j| = 1$, and $G(u_1) = G(u_2)$.

To construct such a function, observe that

$$F\big(e^{i\psi}, t\big) = f\big(G\big(e^{i\psi}, t\big)\big).$$

Differentiating both sides of this identity with respect to the angle ψ and the parameter t, one obtains by Cramer's rule, a representation of the $\overline{z}$-derivative of f as the ratio of two Jacobians:

$$\frac{\partial f}{\partial \overline{z}}\big(G(u)\big) = \frac{\frac{\partial(G,F)}{\partial(\psi,t)}(u)}{\frac{\partial(G,\overline{G})}{\partial(\psi,t)}(u)}, \quad u = G\big(e^{i\psi}, t\big). \tag{3}$$

Since F and G are holomorphic in ζ, we have

$$\frac{\partial F}{\partial \psi}(\zeta, t) = i\zeta \frac{\partial F}{\partial \zeta}(\zeta, t), \quad \zeta = re^{i\psi}.$$

Define

$$J(\zeta, t) := \zeta \frac{\partial(G, F)}{\partial(\zeta, t)}(u).$$

Then conditions (i) and (ii) are obvious. Condition (iii) follows from formula (3), which says that

$$J(\zeta, t) = \frac{\partial(G, \overline{G})}{\partial(\psi, t)}(u) \frac{\partial f}{\partial \overline{z}}\big(G(u)\big), \quad u = (\zeta, t) \in \partial\Delta \times S^1. \tag{4}$$

It follows from our assumptions that $J \not\equiv 0$. Then since $i \frac{\partial(G,\overline{G})}{\partial(\psi,t)}$ is real-valued, we have

$$\frac{J^2(u)}{|J(u)|^2} = -\frac{\frac{\partial f}{\partial \overline{z}}(G(u))}{\overline{\frac{\partial f}{\partial \overline{z}}(G(u))}}$$

whenever $J(u) \neq 0$, and (iii) follows.

Using real-analyticity, one shows that (iii) implies that $J^2/|J|^2$ can be represented on $\partial\Delta \times S^1$ as a composition

$$\frac{J^2}{|J|^2}(u) = (\sigma \circ G)(u), \quad u \in \partial\Delta \times S^1, \tag{5}$$

where σ is a smooth function.

Concluding step (topological). Let L be the dual 1-cocycle constructed in Step 1. Consider its preimage $S := \Phi^{-1}(L)$, which is a two-dimensional chain. The curves L and $C = \Phi(\{0\} \times S^1)$ have positive intersection index, which implies that $\mathrm{Ind}(S \cap (\{0\} \times S^1)) > 0$. We have $\{0\} \times S^1 \subset J^{-1}(0)$; and, using the fact that $J(\zeta, t)$ is holomorphic in ζ, one shows that $\mathrm{Ind}(S \cap J^{-1}(0)) \geq \mathrm{Ind}(S \cap (\{0\} \times S^1)) > 0$.

The latter inequality means that the sum of indices of the vector field J at the zeros on S is positive:

$$N = \sum_k \mathrm{ind}_{u=b_k} J(u) > 0,$$

where b_k are zeros of J on S. We recall that $\mathrm{ind}_{b_k} J$ is defined as the winding number of the normalized function $J/|J| : c_k \to S^1$, where c_k is a small topological circle on S surrounding b_k.

On the other hand, the doubled sum $2I$ of the indices of $J(u)$ inside S equals the winding number W_J of $J^2/|J|^2$ on the boundary $\partial S \subset \partial\Delta \times S^1$; hence $W_J > 0$. But the representation (5) and $\deg G = 0$ imply $W_J = 0$. We are led to this contradiction by the assumption that J is not identically 0. Therefore, $J \equiv 0$; hence $\frac{\partial f}{\partial \bar{z}} \equiv 0$ by (4). This completes the proof. $\square$

Comment on the proof. Essentially, what was used in the proof is that the holomorphic extensions $F_t(\zeta) = F(t, \zeta)$ identify the points from the same fiber $G^{-1}(z)$ on $|\zeta| = 1$. This fiber appears since the point z can belong to different curves γ_t, and the holomorphic extensions inside these curves take the same value $f(z)$ at the point z.

In the case in which the family of curves is invariant under a group (translations or rotations), the use of Fourier series or the Fourier transform, which is an averaging of the holomorphic extensions F_t with respect to the parameter t with certain weight, allows us to project the family of extensions to a single function in the disc $\{\zeta : |\zeta| < 1\}$. The constancy of the function $F(\zeta, t)$ on the G-fibers turns for the projected functions into a condition of the identification of points on the unit circle. Then one is led to the situation described by Lemma 1.

In the general situation, one does not have a group available to define, by averaging over t, a projection of the function f to functions depending only on the variable ζ. This leads to the necessity of developing an analogue of Lemma 1 in the parameter space (ζ, t).

The conclusion of the theorem is that the holomorphic extensions F_t coincide on intersections of the interiors of the curves γ_t. This is equivalent to the degeneracy of the mapping $\Psi = (G, F)$ constructed from the parameterization G of the family and the holomorphic extensions F. Thus, the aim is to check that the Jacobian J of Ψ is zero. It is observed that the constancy of F on the G-fibers is inherited not by the function J, but rather by its squared angular part $J^2/|J|^2$. Then the parametric analogue of Lemma 1, applied to $J^2/|J|^2$, leads to a contradiction with the assumption that J is not identically zero. Therefore, J vanishes identically, and then (4) implies $\frac{\partial f}{\partial \overline{z}} = 0$ in Ω. These are the principal ingredients and main idea of the proof in [3, 5].

6 The Strip Problem for Meromorphic Extensions. Back to Circles

Complex moment conditions can be used to characterize not only holomorphic functions, but also more general classes. Consider, for example, the collection of polyanalytic functions of order (at most) m on $\mathbb{C}$, i.e., solutions of the equation

$$\frac{\partial^m f}{\partial \overline{z}^m} = 0.$$

These functions can be written in the form

$$f(z) = h_1(z) + \overline{z}h_2(z) + \cdots + \overline{z}^{m-1}h_m(z), \tag{6}$$

where the h_j are entire functions.

Now, in close analogy with the case of analytic functions discussed in the Introduction, polyanalytic functions can be characterized in terms of appropriate moment conditions. Indeed, suppose $f \in C(\mathbb{R}^2)$ and that there exist $r_1, r_2 > 0$ such that

$$\int_{|z|=r_j} f(w+z)z^{m-1}\,dz = 0, \quad j = 1, 2,$$

for all $w \in \mathbb{C}$. Then f is polyanalytic on $\mathbb{C}$ of order m, unless r_1/r_2 is a quotient of positive zeros of the Bessel function J_m [27, Theorem 4]. Note that when $m = 1$, this reduces to the two-circle version of Morera's theorem cited in the Introduction. For an analogue on the disc, see [9, Corollary 1].

Similarly, if there exist positive integers $1 \le m < n$ such that for some fixed $r > 0$, the restriction of $f \in C(\mathbb{R}^2)$ to each circle of radius r has vanishing Fourier coefficients of order $-m$ and $-n$, i.e.,

$$\frac{1}{2\pi}\int_0^{2\pi} f\big(w + re^{i\theta}\big)e^{i\theta k}\,d\theta = 0, \quad k = m, n,$$

for all $w \in \mathbb{C}$, f is polyanalytic of order m [28, Theorem 6].

In analogy with (1), let us consider for fixed r and $m \geq 1$ the family of moment conditions

$$\int_{|z-c|=r} f(z) z^{k-1}\, dz = 0 \quad \text{for all } k \geq m. \tag{7}$$

Condition (7) means that f can be extended continuously to $\{z : |z-c| \leq r\}$ as a meromorphic function F whose only singularity is a pole of order at most m at the center $z = c$.

In this section, we present some recent results concerning meromorphic extensions into families of circles.

It is easy to see that a polyanalytic function (6) extends into any circle $C = \{z : |z-c| = r\}$ as a meromorphic function with a pole at c of order at most m, since $\overline{z} = \overline{c} + \frac{r^2}{z-c}$ on the circle C.

Now we want to characterize those one-parameter families of circles that are sufficient to test polyanalytic functions via moment conditions (7).

We consider a one-parameter C^1 family of circles $\mathscr{C} = \{C_t : t \in [0,1]\}$, where $C_t = \{z : |z - c(t)| = r(t)\}$. Here $c : [0,1] \to \mathbb{C}$ and $r : [0,1] \mapsto [0,\infty)$ belong to $C^1[0,1]$. We also assume that $|c'(t)|^2 + |r'(t)|^2 \neq 0$ for $t \in (0,1)$.

We call the quadratic polynomial

$$d(t,w) = \overline{c'(t)} w^2 + 2r'(t) w + c'(t)$$

the *discriminant polynomial* of the family $\{C_t\}$.

There are three possibilities for the roots w_1, w_2 of this polynomial: (1) if $c'(t) = 0$, then $w_1 = w_2 = 0$; while if $c'(t) \neq 0$, one must have either (2) $|w_1| = |w_2| = 1$ or (3) $|w_1| < 1, |w_2| > 1$ since $|w_1 w_2| = |\frac{c'(t)}{\overline{c'(t)}}| = 1$.

Definition 1 The discriminant curve $S(\mathscr{C})$ of the family of circles $\mathscr{C} = \{C_t,\ t \in [0,1]\}$ is the closure of the set $\{c(t) + r(t)w(t) : t \in [0,1]\}$, where $w(t)$ is the root of the discriminant equation $d(t,w) = 0$ with $|w(t)| < 1$ if such a root exists.

Definition 2 We call the family $\{C_t : t \in [0,1]\}$ a chain of circles with initial point a and endpoint b if $C_0 = \{a\}$ and $C_1 = \{b\}$.

Theorem 6 ([4]) *Let $\mathscr{C} = \{C_t : t \in [0,1]\}$ be a C^1-chain of circles in $\mathbb{C}$, starting at a and ending at b, where $a \neq b$, and set $\Omega = \bigcup_{t \in [0,1]} C_t$. Suppose that*

(∗) *the discriminant curve $S(\mathscr{C})$ contains no continuous curve joining the point a and b.*

Let f be real-analytic on $\overline{\Omega}$ and satisfying condition (7) for a fixed nonnegative integer m. Then f is a polyanalytic function of order at most m, i.e., f has the form (6).

Remark 4 The discriminant of the polynomial $d(t, w)$ equals $4(|r'(t)|^2 - |c'(t)|^2)$. If $|c'(t)| \geq |r'(t)|$, then both roots of $d(t, w)$ satisfy $|w_1(t)| = |w_2(t)| = 1$, and therefore $S(\mathscr{C}) = \emptyset$. Thus, the condition for the derivatives of $c(t)$ and $r(t)$ in Theorems 3 and 4 imply condition (*).

For real-analytic functions, the special case $m = 0$ of Theorem 6 implies a stronger test of analyticity than Theorem 4, since it avoids the condition of no enclosed circles required by Theorem 4.

The condition of shrinking the chain of circles to points a and b can be replaced by the weaker condition of the homological nontriviality of the family, which appears in the proof of Theorem 5; we omit the details.

Sketch of Proof The proof is based on an approach similar to that in Theorem 5 and again exploits the argument principle. The key idea is to study the dynamics of zeros $z_j(t)$ and poles $p_j(t)$ of the meromorphic extensions F_t into the discs bounded by the circles C_t. These zeros and poles can reach the boundary circle C_t and disappear, or they can remain inside the circle for all t. In the latter case, we call a curve of zeros $Z_j := \{z = z(t)\}$ a *travelling zero*. Analogously, one defines *travelling poles*.

Using the argument principle, it is shown in [3] that the number $N(f)$ of travelling zeros, counting multiplicities, coincides with the number $P(f)$ of travelling poles.

The next step is computing the numbers N and P for the derivative $\frac{\partial f}{\partial \bar{z}}$. The computation shows that the number of travelling poles at the center $w = 0$ of the circle $C_t : c(t) + r(t)w$, $|w| = 1$, decreases by one, but new poles can appear at zeros $w(t)$ of the discriminant polynomial $D(t, w)$, i.e., on the discriminant curve. However, condition $(*)$ prevents these new poles from forming a travelling pole, and hence $P(\frac{\partial f}{\partial \bar{z}}) = P(f) - 1$.

Thus, differentiating $m - 1$ times in $\bar{z}$, where m is the maximal order of the poles at $w = 0$, and assuming that at no step is the derivative identically zero (in which case f is polyanalytic of order at most m, and we are done), one eliminates all travelling poles. Therefore, the meromorphic extensions of the function $g := \frac{\partial^{m-1} f}{\partial \bar{z}^{m-1}}$ from the circles C_t are free of travelling poles.

However, the same computation shows that upon differentiating in $\bar{z}$ the holomorphic extension develops a zero at the center $w = 0$. Therefore, the number of travelling zeros $N(\frac{\partial g}{\partial \bar{z}}) > 0$. The travelling zeros must be compensated by travelling poles, as noted above. However, the only poles the extensions of the derivative $\frac{\partial g}{\partial \bar{z}}$ develop are those on the discriminant curve. By assumption, these poles are not travelling. Thus, $N(\frac{\partial g}{\partial \bar{z}}) > 0 = P(\frac{\partial g}{\partial \bar{z}})$. This contradiction implies that $\frac{\partial g}{\partial \bar{z}} = \frac{\partial^m f}{\partial \bar{z}^m} = 0$ for all $z \in \Omega$. □

Recently, Globevnik [21] proved Theorem 6 under milder conditions of continuity for the tested functions f, but for the specific family of concentric hyperbolic circles in the unit disc. His method uses the approach from [23, 24], mentioned in Sect. 4.

7 The Strip Problem for Elliptic PDE

Leon Ehrenpreis learned of the strip problem in 2000 from a talk given by Zalcman at the AMS–IMS–SIAM Joint Summer Research Conference on Radon Transforms and Tomography held at Mount Holyoke. Soon afterwards, he came up with the following nice "harmonic" version of the strip problem.

Theorem 7 ([13]) *Denote by B the open unit ball in $\mathbb{R}^n$ and consider the cylinder $\Omega = \partial B + \mathbb{R}e_1$, where $e_1 = (1, 0, \ldots, 0)$. Let f be a C^2-function in Ω. Suppose that for each $t \in \mathbb{R}$, there exists a harmonic function F_t in the ball $B_t := B + te_1$ such that $F_t = f$ on ∂B_t to first order, i.e., $F_t(z) = f(z)$, $\partial_{\nu_t} F_t(z) = \partial_{\nu_t} f(z)$ for all $z \in \partial B_t$. (Here ν_t is the normal derivative to ∂B_t.) Then f is harmonic in Ω.*

Ehrenpreis' original proof required certain additional technical assumptions. The short proof given below, based solely on Stokes' formula, is adapted from [7].

Proof Pick any function $h \in C(\overline{B})$ which is harmonic in B. (It suffices to take an arbitrary harmonic polynomial.) From the assumption and Stokes' formula, we have

$$0 = \int_B \big(\Delta F_t(x + te_1)h(x) - F_t(x + te_1)\Delta h(x)\big)\, dV(x) \tag{8}$$

$$= \int_{\partial B} \big(\partial_\nu f(x + te_1)h(x) - f(x + te_1)\partial_\nu h(x)\big)\, dA, \tag{9}$$

where ν is the exterior unit normal to the unit sphere ∂B.

Again by Stokes' formula, the surface integral on the right-hand side can be rewritten as a volume integral, and, since $\Delta h = 0$, we obtain

$$\int_B \Delta f(x + te_1)h(x)\, dV(x) = 0.$$

Differentiating in t and repeatedly using Stokes' formula yields

$$\int_B \partial_{x_1} \Delta f(x + te_1)h(x)\, dV = \int_{\partial B} x_1 \Delta f(x + te_1)h(x)\, dA(x) = 0.$$

Since $h|_{\partial B}$ are dense in $C(\partial B)$, one concludes that $x_1 \Delta f(x + te_1) = 0$ for $x \in \partial B$. Therefore, Δf vanishes on all the boundaries $\partial B_t = \partial B + te_1$ and hence on Ω. The proof is complete. □

Observe that this argument also works for truncated solid cylinders of the form $\Omega = \bigcup_{t\in[a,b]}(B + te_1)$, where $[a, b]$ is an arbitrary segment of the real line.

Shortly thereafter, Ehrenpreis generalized the result from the Laplace operator to any elliptic operator with sufficiently smooth coefficients, again for the translations of a ball. This result is presented in his fundamental monograph [14, 9.5]. It says that if the boundary values on the translations $\partial B + te_1$ of a function f coincide, to order m, with a solution F_t in $B + te_1$ of a given elliptic equation $LF_t = 0$, then f itself is a solution, $Lf = 0$ in the union of the spheres $\partial B + te_1$.

In [14], the problem of generalizing such results from Laplace's equation to more general PDE was posed. The following theorem presents a reasonably general result for elliptic equations. Before formulating this result, we need to introduce some notation. Consider a C^1-isotopy of domains in $\mathbb{R}^n$. This is a family $D_t = \omega_t(D),\ t \in I = [0, 1]$, where $D = D_0$ is the initial domain and $\omega_t : D \to D_t$ is a family of diffeomorphisms, continuously differentiable with respect to the parameter $t \in [0, 1]$ and such that $\omega_0 = \mathrm{id}$.

We say that the family D_t is *transversal* if for each $t \in I$,

$$\rho_t(u) := \big\langle \partial_t \omega_t(u), \nu_t\big(\omega_t(u)\big)\big\rangle \neq 0$$

for a dense set of $u \in \partial D$, where $\langle\ ,\ \rangle$ is the inner product in $\mathbb{R}^n$. Transversality of the isotopy means that the set of points where the direction of the transformation ω_t is tangent to the boundary ∂D_t is nowhere dense. A simple example of a nontransversal isotopy is a rotation of a ball; in this case, the vector of the motion is tangent to the boundary sphere at each point.

The following theorem was obtained (under some additional technical assumptions) in [14] for the case of translations of a ball $D = B$, $\omega_t(u) = u + te_1$, and in [7] for the general case.

Theorem 8 ([7]) *Let Ω_t, $t \in I$, be a C^1-transversal isotopy of domains in $\mathbb{R}^n$. Set $\Omega = \bigcup_{t\in I} \partial\Omega_t$, and let*

$$L = P(x, D) = \sum_{|\alpha|\le 2m} a_\alpha(x) D^\alpha$$

be an elliptic partial differential operator of order $2m$ with smooth coefficients defined in the domain Ω_t. Suppose that if the function $f \in C^{2m}(\Omega)$ coincides on each boundary $\partial\Omega_t$, to order at least m with a solution in the domain Ω_t, i.e., that for each $t \in I$, the Cauchy boundary problem

$$LF_t = 0, \quad x \in \Omega_t,$$

with boundary conditions

$$\partial^j_{\nu_t} F_t(x) = \partial^j_{\nu_t} f(x), \quad x \in \partial\Omega_t,\ j = 0, 1, \ldots, m,$$

has a solution $F_t \in C^m(\overline{\Omega_t})$. Then f is a global solution, i.e., $Lf(x) = 0,\ x \in \Omega^\circ$.

The proof is similar to that described above for the case of the Laplace operator.

8 Conclusions, Further Generalizations

- The problem of characterizing holomorphic functions in plane domains by holomorphic extendibility from closed curves is, in fact, part of a more general problem formulated for manifolds. Namely, the holomorphy of a function is equivalent to its graph, which is a real two-dimensional manifold in $\mathbb{C}^2$, being a complex (holomorphic) manifold. The condition of holomorphic extendibility inside a closed curve means that the graph of the extensions is attached by its boundary to the graph of the function. The graphs of the holomorphic extensions are analytic discs in $\mathbb{C}^2$. Therefore, the question may be reformulated as follows: which families of analytic discs attached to a two-dimensional real manifold in $\mathbb{C}^2$ imply that the manifold is complex? In turn, this question reduces to the problem of estimating the CR-dimension of a real manifold in $\mathbb{C}^n$ in terms of attached analytic discs, which is studied in [5].

- The question about analytic extensions from curves can also be asked for non-closed curves. Recently, Fridman and Ma [16] have proved that given a domain D in $\mathbb{C}^n$, there exists a continuous foliation of D into real curves such that any C^1-function which can be extended holomorphically into some neighborhood of each curve in the foliation is holomorphic on D [16].

Acknowledgements This work was partially supported by Israel Science Foundation Grants 688/08 and 395/07 and is a part of the European Science Foundation Networking Programme HCAA.

References

1. Agranovsky, M.L.: The Fourier transform on $SL_2(\mathbb{R})$ and theorems of Morera type. Sov. Math. Dokl. **19**, 1522–1525 (1978)
2. Agranovsky, M.L.: CR foliations, the strip-problem and Globevnik–Stout conjecture. C. R. Acad. Sci. Paris **343**, 91–94 (2006)
3. Agranovsky, M.L.: Propagation of boundary CR-foliations and Morera type theorems for manifolds with attached analytic discs. Adv. Math. **211**, 284–326 (2007)
4. Agranovsky, M.L.: Characterization of polyanalytic functions by meromorphic extensions in chains of circles. J. Anal. Math. **113**, 305–329 (2011)
5. Agranovsky, M.L.: Complex dimensions of real manifolds, attached analytic discs and parametric argument principle. arXiv:0704.2871v2 [math]
6. Agranovsky, M.L., Globevnik, J.: Analyticity on circles for rational and real-analytic functions of two real variables. J. Anal. Math. **91**, 31–65 (2003)
7. Agranovsky, M.L., Narayanan, E.K.: Isotopic families of contact manifolds for elliptic PDE. Proc. Am. Math. Soc. **134**, 2117–2123 (2006)
8. Agranovsky, M.L., Valsky, R.E.: Maximality of invariant subalgebras of functions. Sib. Math. J. **12**, 1–7 (1971)
9. Agranovsky, M.L., Volchkov, V.V., Zalcman, L.: Conical uniqueness sets for the spherical Radon transform. Bull. Lond. Math. Soc. **31**, 231–236 (1999)
10. Berenstein, C.A., Zalcman, L.: Pompeiu's problem on spaces of constant curvature. J. Anal. Math. **30**, 113–130 (1976)
11. Berenstein, C.A., Zalcman, L.: Pompeiu's problem on symmetric spaces. Comment. Math. Helv. **55**, 593–621 (1980)
12. Brown, L., Schreiber, B.M., Taylor, B.A.: Spectral synthesis and the Pompeiu problem. Ann. Inst. Fourier (Grenoble) **23**(3), 125–154 (1973)
13. Ehrenpreis, L.: Three problems at Mount Holyoke. In: Radon Transforms and Tomography. Contemp. Math., vol. 278, pp. 123–130 (2001)

14. Ehrenpreis, L.: The Universality of the Radon Transform. Clarendon, Oxford (2003)
15. Ehrenpreis, L.: The strip problem for PDE. Inverse Probl. Theor. Imaging **4**(4), III–IV (2010)
16. Fridman, B.L., Ma, D.: Testing holomorphy on curves. arXiv:1012.4329v2
17. Globevnik, J.: Analyticity on rotation invariant families of circles. Trans. Am. Math. Soc. **280**, 247–254 (1983)
18. Globevnik, J.: Testing analyticity on rotation invariant families of curves. Trans. Am. Math. Soc. **306**, 401–410 (1988)
19. Globevnik, J.: Holomorphic extensions and rotation invariance. Complex Var. Theory Appl. **24**, 49–51 (1994)
20. Globevnik, J.: Holomorphic extensions from open families of circles. Trans. Am. Math. Soc. **355**, 1921–1931 (2003)
21. Globevnik, J.: Meromorphic extensions from small circles and holomorphic extensions from spheres. arXiv:1101.0136v2
22. Hu, S.T.: Homotopy Theory. Academic Press, San Diego (1959)
23. Tumanov, A.: A Morera type theorem in the strip. Math. Res. Lett. **11**, 23–29 (2004)
24. Tumanov, A.: Testing analyticity on circles. Am. J. Math. **129**, 785–790 (2007)
25. Volchkov, V.V.: Integral Geometry and Convolution Equations. Kluwer Academic, Dordrecht (2003)
26. Volchkov, V.V., Volchkov, V.V.: Harmonic Analysis of Mean Periodic Functions on Symmetric Spaces and the Heisenberg Group. Springer, London (2009)
27. Zalcman, L.: Analyticity and the Pompeiu problem. Arch. Ration. Mech. Anal. **47**, 237–254 (1972)
28. Zalcman, L.: Mean values and differential equations. Isr. J. Math. **14**, 339–352 (1973)
29. Zalcman, L.: A bibliographic survey of the Pompeiu problem. In: Approximation by Solutions of Partial Differential Equations, pp. 185–194. Kluwer Academic, Dordrecht (1992)
30. Zalcman, L.: Supplementary bibliography to "A bibliographic survey of the Pompeiu problem". In: Radon Transforms and Tomography. Contemp. Math., vol. 278, pp. 69–74 (2001)

On Local Injectivity for Generalized Radon Transforms

Jan Boman

Abstract We consider a class of weighted plane generalized Radon transforms $Rf(\gamma) = \int f(x, u(\xi, \eta, x))m(\xi, \eta, x)\,dx$, where the curve $\gamma = \gamma_{(\xi,\eta)}$ is defined by $y = u(\xi, \eta, x)$, and $m(\xi, \eta, x)$ is a given positive weight function. We prove local injectivity for this transform across a given curve γ^0 near a given point (x^0, y^0) on γ^0 for classes of curves and weight functions that are invariant under arbitrary smooth coordinate transformations in the plane.

1 Introduction

I met Leon Ehrenpreis already in 1961, when he presented his celebrated "Fundamental Principle" at a summer school at Stanford University. Much later I was fortunate to meet Leon again at many Radon transform meetings and during several visits to Temple University. Leon's enthusiasm and generosity in sharing mathematical ideas, his broad outlook, and his original way of looking at problems was a great source of inspiration for many of us.

It was shown by Strichartz [15] that the classical Radon transform is *locally injective* in the following sense, here for simplicity formulated for the case of $\mathbf{R}^2$. If the continuous function $f(x, y)$ is supported in $y \geq x^2$ and

$$\int f(x, \xi x + \eta)\,dx = 0 \tag{1}$$

for all (ξ, η) in a neighborhood of $(0, 0)$, then f must vanish in some neighborhood of the origin. It is known [2] that the corresponding statement is not always true if a smooth positive weight function $m(\xi, \eta, x)$ is introduced, so that (1) is replaced by

$$\int f(x, \xi x + \eta)m(\xi, \eta, x)\,dx = 0. \tag{2}$$

Dedicated to the memory of Leon Ehrenpreis.

J. Boman (✉)
Stockholm University, 10691 Stockholm, Sweden
e-mail: jabo@math.su.se

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_5, © Springer-Verlag Italia 2012

On the other hand, if $m(\xi,\eta,x)$ is positive and real analytic, then local injectivity again holds [5]. Thus the set of weight functions $m(\xi,\eta,x)$ for which local injectivity holds is dense in the space of smooth weight functions. In [4] we recently extended the construction in [2] by showing that the set of smooth positive weight functions for which local injectivity does not hold is also dense. In [3] we proved local injectivity for a class of smooth weight functions that was introduced by Gindikin [12]. In this note we shall extend the result from [3] by replacing the family of straight lines by a two-dimensional family of curves satisfying a certain condition, thus obtaining a statement that is invariant under local diffeomorphisms in the xy space and in the $\xi\eta$ space of curves.

On the surface our Radon transform (2) looks like a parametric Radon transform in the sense of Ehrenpreis [7]. However, replacing dx by $ds/\sqrt{1+\xi^2}$, where ds denotes arc length measure on the line, we can of course just as well regard our transform as a restriction of a nonparametric Radon transform, expressed in affine coordinates.

With a suitable choice of coordinates in xy space and $\xi\eta$ space, an arbitrary curve family in a neighborhood of the origin in $x\,y$ space can be written $y = u(\xi,\eta,x)$ for (ξ,η) near $(0,0)$, where

$$u(\xi,\eta,x) = \eta + \xi x + \mathscr{O}\big(x^2\big) \quad \text{as } x \to 0.$$

The *dual* curve family is the family of curves in $\xi\eta$ space that is defined by the equation $y - u(x,\xi,\eta) = 0$ with x and y playing the role of parameters. The condition on the curve family in our main theorem (Theorem 2) is somewhat implicit and reads as follows: the curves in the dual family are solution curves to a second-order differential equation of the form

$$\eta'' = \Psi\big(\xi,\eta,\eta'\big),$$

where $\Psi(\xi,\eta,p)$ is a polynomial of degree at most 3 in p. This condition is known to be independent of the choice of coordinates in the $\xi\eta$ plane (Proposition 2). These curve families have been known for a long time. In fact, Eli Cartan showed in 1924 that those curves are the geodesics of torsion-free projective connections. But the condition has also played a role in integral geometry. I.M. Gelfand and his collaborators studied what they called *admissible families* of lines or curves, in some cases curve families with densities, in real or complex spaces in a long series of papers (see [8, 10, 11], and references given there). In the case of curve families in $\mathbf{R}^2$ it was found that a curve family with densities was admissible precisely when both curve family and densities satisfied the conditions considered here [10]. It would be interesting to understand why our local injectivity problem leads to the same conditions on curve families and densities as the admissibility property of Gelfand, Gindikin, and Shapiro.

In Sect. 2 we introduce the double fibration defined by a hypersurface Z in the product of two two-dimensional manifolds, we define a weighted Radon transform on the two-dimensional manifold of curves defined by this fibration, and we formulate the problem of local injectivity for this transform. In Sect. 3 we state and prove a local injectivity theorem (Theorem 1) under assumptions that are expressed in local coordinate systems in X and Γ. In Sect. 4 we prove that our condition on the weight function (actually a *density* on Z) is independent of the chosen coordinate systems by expressing it in coordinate-free terms. In Sect. 5 we give a couple of auxiliary results on curve families defined by an ordinary differential equation of the second order, and in Sect. 6 we extend Theorem 1 by replacing the condition on Z there by a weaker condition that is invariant under coordinate changes in X and Γ (Theorem 2).

2 A Weighted Radon Transform Associated to $Z \subset X \times \Gamma$

To define our Radon transform in an invariant way we shall use the double fibration setup introduced by Helgason and Gelfand [9], [14]. Let X and Γ be two-dimensional manifolds, and let $Z \subset X \times \Gamma$ be a smooth hypersurface in $X \times \Gamma$ such that both projections

$$\pi_X : Z \to X, \qquad \pi_\Gamma : Z \to \Gamma$$

have surjective differential at a point $(\mathbf{x}^0, \gamma^0) \in Z$. Using local coordinate systems in X and Γ, we can define Z near $(\mathbf{x}^0, \gamma^0)$ by

$$F(x, y, \xi, \eta) = 0, \tag{3}$$

where $\mathbf{x} = (x, y)$ denotes points of X, $\gamma = (\xi, \eta)$ denotes points of Γ, and F is a smooth function satisfying

$$d_{(x,y)}F \neq 0 \quad \text{and} \quad d_{(\xi,\eta)}F \neq 0 \quad \text{at } \big(\mathbf{x}^0, \gamma^0\big). \tag{4}$$

Furthermore we assume that the natural projections from the conormal

$$N^*(Z) \to T^*(X) \quad \text{and} \quad N^*(Z) \to T^*(\Gamma) \quad \text{are local diffeomorphisms}$$
$$\text{near the point } \big(\mathbf{x}^0, \gamma^0; dF\big(\mathbf{x}^0, \gamma^0\big)\big) \in N^*(Z). \tag{5}$$

We may assume that $\mathbf{x}^0 = (x^0, y^0) = (0, 0)$ and $\gamma^0 = (\xi^0, \eta^0) = (0, 0)$. We can also rotate the coordinate systems in X and Γ so that

$$F'_x = 0, \quad F'_y \neq 0, \quad F'_\xi = 0, \quad F'_\eta \neq 0 \quad \text{at the origin.} \tag{6}$$

Then we can solve y or η from (3) and obtain respectively

$$y = u(\xi, \eta, x), \qquad \eta = \rho(x, y, \xi)\,. \tag{7}$$

Differentiating the identity $F(x, u(\xi, \eta, x), \xi, \eta) = 0$, we obtain

$$F'_x + F'_y u'_x = 0, \quad F'_\xi + F'_y u'_\xi = 0, \quad F'_\eta + F'_y u'_\eta = 0 \tag{8}$$

in a neighborhood of the origin. From here and from (6) we obtain

$$u'_x = u'_\xi = 0 \quad \text{and} \quad u'_\eta \neq 0 \quad \text{at the origin,} \tag{9}$$

and similarly,

$$\rho'_x = \rho'_\xi = 0 \quad \text{and} \quad \rho'_y \neq 0 \quad \text{at the origin.}$$

If we choose (x, y, ξ) as coordinates on Z and (x, y, ξ, λ) as coordinates on $N^*(Z)$ and denote the conormal by $\lambda\, dF$, then the projection $N^*(Z) \to T^*(X)$ can be represented as

$$(x, y, \xi, \lambda) \quad \mapsto \quad (x, y, \theta_x, \theta_y) = \big(x, y, \lambda F'_x, \lambda F'_y\big).$$

This map has a nonsingular differential if and only if the map

$$(\xi, \lambda) \quad \mapsto \quad (\theta_x, \theta_y) = \big(\lambda F'_x, \lambda F'_y\big)$$

has a nonsingular differential, which is the case if and only if

$$\det \begin{vmatrix} \lambda F''_{x\xi} & F'_x \\ \lambda F''_{y\xi} & F'_y \end{vmatrix} = \lambda \det \begin{vmatrix} F''_{x\xi} & F'_x \\ F''_{y\xi} & F'_y \end{vmatrix} \neq 0.$$

Hence the assumption that the projection $N^*(Z) \to T^*(X)$ has a nonsingular differential is equivalent to

$$F''_{x\xi} \neq 0 \quad \text{at the origin.} \tag{10}$$

The fact that condition (10) is symmetric with respect to interchange of the spaces X and Γ shows that both projections (5) have nonsingular differentials if one of them does and (4) holds.

Differentiating the second equation of (8) with respect to x and using (9) and (10), we find that

$$u''_{x\xi} \neq 0 \quad \text{at the origin.} \tag{11}$$

This shows that the maps

$$\begin{aligned} (\xi, \eta) &\mapsto \big(u(\xi, \eta, 0), u'_x(\xi, \eta, 0)\big) \quad \text{and} \\ (x, \xi, \eta) &\mapsto \big(x, u(\xi, \eta, x), u'_x(\xi, \eta, x)\big) \end{aligned} \tag{12}$$

are local diffeomorphisms near the origin.

Let μ be a given positive, smooth density on Z. For instance, in the (ξ, η, x) coordinates on Z, let the density μ be given as

$$\mu = m(\xi, \eta, x)\, d\xi\, d\eta\, dx.$$

If f is a continuous function on X, then $f \circ \pi_X$ is a function on Z, so $(f \circ \pi_X)\mu$ is a density on Z. If φ is a compactly supported continuous function on Γ, then $\varphi \circ \pi_\Gamma$

is a compactly supported continuous function on Z, so we can form

$$\big\langle (f\circ\pi_X)\mu, \varphi\circ\pi_\Gamma\big\rangle = \iiint f\big(x, u(\xi,\eta,x)\big) m(\xi,\eta,x)\,\varphi(\xi,\eta)\,dx\,d\xi\,d\eta,$$

and we can define Rf on Γ as the measure

$$\varphi \mapsto \big\langle (f\circ\pi_X)\mu, \varphi\circ\pi_\Gamma\big\rangle = \big\langle \mu, (f\circ\pi_X)(\varphi\circ\pi_\Gamma)\big\rangle. \tag{13}$$

The situation is symmetric in f and φ, so if we define $R^*\varphi$ as the measure on X

$$f \mapsto \big\langle \mu, (f\circ\pi_X)(\varphi\circ\pi_\Gamma)\big\rangle, \tag{14}$$

we have $\langle Rf, \varphi\rangle = \langle f, R^*\varphi\rangle$. To obtain an explicit expression for $R^*\varphi$, we express the density μ and the function $\varphi\circ\pi_\Gamma$ in terms of the (x,y,ξ) coordinates on Z,

$$\mu = n(x,y,\xi)\,dx\,dy\,d\xi, \qquad (\varphi\circ\pi_\Gamma)(x,y,\xi) = \varphi\big(\xi, \rho(x,y,\xi)\big).$$

This gives

$$\begin{aligned}\langle Rf, \varphi\rangle &= \big\langle f, R^*\varphi\big\rangle = \big\langle \mu, (f\circ\pi_X)(\varphi\circ\pi_\Gamma)\big\rangle \\ &= \iiint f(x,y)\varphi\big(\xi, \rho(x,y,\xi)\big) n(x,y,\xi)\,dx\,dy\,d\xi.\end{aligned}$$

Since the Jacobian of the transformation $x = x', \xi = \xi', \eta = \rho(x', y, \xi')$ is ρ'_y, we have

$$n(x,y,\xi) = m\big(\xi, \rho(x,y,\xi), x\big)\,\rho'_y(x,y,\xi), \tag{15}$$

so

$$R^*\varphi(x,y) = \int \varphi\big(\xi, \rho(x,y,\xi)\big) n(x,y,\xi)\,d\xi.$$

It will be convenient to extend the definition of R and R^* as follows. If u is a continuous function on Z, then we define $R(uf)$ and $R^*(u\varphi)$ by replacing the right-hand sides of (13) and (14) by

$$\big\langle \mu, (f\circ\pi_X)(\varphi\circ\pi_\Gamma)u\big\rangle.$$

Then we have

$$\big\langle R(uf), \varphi\big\rangle = \big\langle f, R^*(u\varphi)\big\rangle. \tag{16}$$

In geometric terms the transform R integrates over the fibers $\pi_X(\pi_\Gamma^{-1}(\gamma)) \subset X$, which we will sometimes (by abuse of language) denote by γ or $\gamma_{(\xi,\eta)}$. The adjoint R^* integrates over the fibers $\pi_\Gamma(\pi_X^{-1}(\mathbf{x})) \subset \Gamma$.

Our problem can now be formulated in invariant terms as follows. Let $(\mathbf{x}^0, \gamma^0) \in Z \subset X \times \Gamma$, and let f be a continuous function defined in an open neighborhood U of $\mathbf{x}^0$. After shrinking U, if needed, we may assume that $U \setminus \gamma^0$ has precisely two components, which we denote by U_+ and U_-. Assume that f is supported in $U_+ \cup \{\mathbf{x}^0\}$ and that $Rf = 0$ in some neighborhood of γ^0. We want to conclude that $f = 0$ in some neighborhood of $\mathbf{x}^0$ (under suitable conditions on μ and Z).

3 Local Injectivity

We now formulate our assumptions on the density μ and the manifold Z (which determines the curve family) in terms of the coordinates (x, y, ξ) on Z. We assume as always that the manifold Z is defined by an equation $F(x, y, \xi, \eta) = 0$ for which (4) and (5) hold and that coordinates are chosen so that (6) holds.

The assumption on the density $\mu = n(x, y, \xi)\, dx\, dy\, d\xi$ will be as follows:

$$\text{there exist two functions } a_1 \text{ and } b_1 \text{ that are constant on the fibers } \pi_\Gamma^{-1}(\gamma),$$
$$\text{such that} \quad n'_\xi = \big(a_1 \rho'_\xi + b_1\big) n. \tag{17}$$

Recall that a function $a(\xi, \eta, x)$ on Z expressed in terms of the coordinates (ξ, η, x) is constant on the fibers $\pi_\Gamma^{-1}(\gamma)$ if it is independent of x. If the function is expressed in terms of the coordinates (x, y, ξ), then it is constant on the fibers $\pi_\Gamma^{-1}(\gamma)$ if it is of the form $a(x, y, \xi) = a_0(\xi, y - \xi x)$ for some function $a_0(\xi, \eta)$. The assumption on Z will be that Z is given by an equation $\eta = \rho(x, y, \xi)$, where ρ satisfies the differential equation

$$\rho''_{\xi\xi} = a_2\big(\rho'_\xi\big)^2 + b_2 \rho'_\xi + c_2 \tag{18}$$

for some functions a_2, b_2, and c_2 that are constant on the fibers $\pi_\Gamma^{-1}(\gamma)$.

As usual, we shall assume that coordinates in X and Γ are chosen such that $\mathbf{x}^0 = (0, 0)$ and $\gamma^0 = (0, 0)$. In this section we shall also assume that coordinates in X are chosen so that γ^0 is equal to the x-axis, in other words,

$$u(0, 0, x) = 0, \quad \text{or equivalently} \quad \rho(x, 0, 0) = 0, \tag{19}$$

in some neighborhood of $x = 0$.

Theorem 1 *Assume that Z is defined by $\eta = \rho(x, y, \xi)$, where ρ satisfies* (18) *and* (19), *and that the positive measure μ on Z satisfies* (17). *Let f be a continuous function defined in some neighborhood of* $(0, 0) \in X$ *and supported in a compact set contained in* $\{(x, y);\ y > 0\} \cup \{(0, 0)\}$, *and assume that $Rf(\xi, \eta) = 0$ in some neighborhood of* $(0, 0)$. *Then $f = 0$ in some neighborhood of* $(0, 0)$.

In the special case considered in [3] the manifold Z is defined by $y = \xi x + \eta$, so $\rho(x, y, \xi) = y - \xi x$, which gives $\rho'_\xi = -x$ and $\rho''_{\xi\xi} = 0$, so (18) holds with $a_2 = b_2 = c_2 = 0$. The weight function $m(\xi, \eta, x)$ in [3] was assumed to satisfy

$$m'_\xi(\xi, \eta, x) - x\, m'_\eta(\xi, \eta, x) = \big(x\, a(\xi, \eta) - b(\xi, \eta)\big) m(\xi, \eta, x) \tag{20}$$

for some $a(\xi,\eta)$ and $b(\xi,\eta)$ that are independent of x. By (15) we have in this case $n(x,y,\xi)=m(\xi,y-\xi x,x)$; hence $n'_\xi=m'_\xi-x\,m'_\eta$, which shows that for this Z, (17) is equivalent to (20).

Using the coordinates (ξ,η,x) on Z, we introduce the function $q(\xi,\eta,x)$ by

$$q(\xi,\eta,x)=\frac{u'_\xi(\xi,\eta,x)}{u'_\eta(\xi,\eta,x)}. \tag{21}$$

Since $u'_\eta(0,0,0)\neq 0$, the function q is well defined and smooth in some neighborhood of $(0,0,0)\in Z$. Choose a neighborhood U of $(0,0)\in X$ and a neighborhood V of $(0,0)\in \Gamma$ such that $u'_\eta\neq 0$ in $(U\times V)\cap Z$. Let f be a continuous function supported in U and set

$$G_k(\xi,\eta)=(-1)^k\int q(\xi,\eta,x)^k f\big(x,u(\xi,\eta,x)\big)m(\xi,\eta,x)\,dx,\quad k=0,1,\dots. \tag{22}$$

Note that all $G_k(\xi,\eta)$ are well defined in V and that $G_0(\xi,\eta)=Rf(\xi,\eta)$. The idea of the proof of Theorem 1 is to show that all G_k vanish in a fixed neighborhood of the origin. We shall see that this easily implies that $f=0$ in some neighborhood of the origin. In the special case where $u(\xi,\eta,x)=\xi x+\eta$, we have $q=-\rho'_\xi=x$.

The main ingredient in the proof of Theorem 1 is the following fact.

Proposition 1 *Let f be a continuous function supported in U, and let the functions $G_k(\xi,\eta)$ be defined in V by (22) as described above. Assume that the density μ and the hypersurface Z satisfy (17) and (18). Then there exist functions $a(\xi,\eta)$, $b(\xi,\eta)$, $c(\xi,\eta)$ such that the following differential equations are satisfied in distribution sense in V:*

$$\begin{aligned}&(\partial_\eta-a)G_1+(\partial_\xi-b)G_0=0,\quad \textit{and}\\ &(\partial_\eta-a)G_{k+1}+(\partial_\xi-b)G_k-cG_{k-1}=0,\quad k\ge 1.\end{aligned} \tag{23}$$

Proof If φ is an arbitrary smooth function on Γ that is supported in V and $n(x,y,\xi)$ is the density on Z as above, we have for every $k\ge 0$, the trivial identity

$$\int \partial_\xi\Big(\big(\rho'_\xi(x,y,\xi)\big)^k\varphi\big(\xi,\rho(x,y,\xi)\big)n(x,y,\xi)\Big)\,d\xi=0. \tag{24}$$

If $k\ge 1$, the integrand can be written

$$k\big(\rho'_\xi\big)^{k-1}\rho''_{\xi\xi}\,\varphi\,n+\big(\rho'_\xi\big)^k\big(\varphi'_\xi+\varphi'_\eta\rho'_\xi\big)n+\big(\rho'_\xi\big)^k\varphi\,n'_\xi.$$

Inserting the expressions for $\rho''_{\xi\xi}$ and n'_ξ from (17) and (18) and rearranging terms, we obtain

$$\big(\varphi'_\xi+\varphi'_\eta\rho'_\xi\big)\big(\rho'_\xi\big)^k n+\big(\rho'_\xi\big)^k\varphi n\big(a\rho'_\xi+b\big)+c\big(\rho'_\xi\big)^{k-1}\varphi n,$$

where $a = ka_2 + a_1$, $b = kb_2 + b_1$, and $c = kc_2$. Note that this expression for the integrand is also valid for $k = 0$, since $c = 0$ then. Thus (24) can be written

$$R^*\big((\rho'_\xi)^k\varphi'_\xi + (\rho'_\xi)^{k+1}\varphi'_\eta + (\rho'_\xi)^{k+1}a\varphi + (\rho'_\xi)^k b\varphi + (\rho'_\xi)^{k-1}c\varphi\big) = 0.$$

Multiplying by $f(x, y)$, integrating, and using (16) with $u = (\rho'_\xi)^j = (-q)^j$, we obtain

$$\big\langle R\big((-q)^k f\big), \varphi'_\xi\big\rangle + \big\langle R\big((-q)^{k+1} f\big), \varphi'_\eta\big\rangle + \big\langle R\big((-q)^{k+1} f\big), a\varphi\big\rangle$$
$$+ \big\langle R\big((-q)^k f\big), b\varphi\big\rangle + \big\langle R\big((-q)^{k-1} f\big), c\varphi\big\rangle = 0.$$

By virtue of the definition of G_k, this means the same as (23). □

The following simple observation will also be needed in the proof of Theorem 1.

Lemma 1 *If coordinate systems in X and Γ are chosen so that* (6) *holds*, *then*

$$\partial_x\big(u'_\xi/u'_\eta\big) \neq 0 \quad \textit{at the origin.}$$

Proof Since $u'_\eta \neq 0$, it is sufficient to observe that

$$u'_\eta u''_{x\xi} - u'_\xi u''_{x\eta} \neq 0,$$

which follows from (11) and the fact that $u'_\xi = 0$ at the origin. □

Proof of Theorem 1 We are going to prove that all the functions $G_k(\xi, \eta)$ vanish in a fixed neighborhood of the origin. To do this, we first need to fix a region in (ξ, η)-space for which the curve $\gamma_{(\xi,\eta)}$ does not meet the support of f.

Choose $\varepsilon > 0$ such that (23) holds in

$$\Omega_\varepsilon = \big\{(\xi, \eta);\ |\xi| < \varepsilon, |\eta| < \varepsilon\big\}.$$

By (6) we know that $u'_\eta(0, 0, 0) \neq 0$, and we may assume that $u'_\eta(0, 0, 0) > 0$. Hence we can choose $\delta > 0$ and κ such that $u'_\eta(0, 0, x) \geq \kappa > 0$ for $|x| \leq \delta$. By possibly replacing ε by a smaller number and recalling assumption (19), we can then achieve that

$$u(\xi, \eta, x) \leq -d < 0, \quad \text{for } -\varepsilon/2 < \eta < -\varepsilon,\ |\xi| < \varepsilon,\ |x| < \delta.$$

Set $K = \operatorname{supp} f$. Since $\gamma_{(0,0)} \cap K = \{(0, 0)\}$, it is clear that

$$\gamma_{(\xi,\eta)} \cap K \cap \big\{(x, y);\ |x| > \delta\big\} = \emptyset$$

if $|\xi|$ and $|\eta|$ are sufficiently small. Hence, by possibly choosing ε still smaller we can achieve that

$$\gamma_{(\xi,\eta)} \cap K = \emptyset \quad \text{if} \quad -\varepsilon/2 < \eta < -\varepsilon \text{ and } |\xi| < \varepsilon.$$

We now prove that all G_k vanish in Ω_ε. By the assumption $G_0 = 0$ in Ω_ε. By Proposition 1 this implies that the function $\eta \mapsto G_1(\xi, \eta)$ satisfies the ordinary differential equation

$$\partial_\eta G_1(\xi, \eta) - a(\xi, \eta) G_1(\xi, \eta) = 0$$

for $|\xi| < \varepsilon$. But $G_1(\xi, \eta)$ is obviously equal to zero whenever the curve $y = u(\xi, \eta, x)$ does not meet the support of f, which as we have just seen is certainly the case if $(\xi, \eta) \in \Omega_\varepsilon$ and $\eta < -\varepsilon/2$. Hence $G_1 = 0$ in Ω_ε.

To proceed, we shall use induction over k. Assume that $G_k = 0$ in Ω_ε for all $k \leq p$, where $p \geq 1$. By Proposition 1 the function $\eta \mapsto G_{p+1}(\xi, \eta)$ must then satisfy the differential equation

$$\partial_\eta G_{p+1}(\xi, \eta) - a(\xi, \eta) G_{p+1}(\xi, \eta) = 0$$

in Ω_ε. Reasoning as before, we can conclude that $G_{p+1} = 0$ in Ω_ε, which proves the assertion.

To complete the proof of Theorem 1, we note that in particular

$$G_k(0, \eta) = (-1)^k \int q(0, \eta, x)^k f\big(x, u(0, \eta, x)\big) m(0, \eta, x)\, dx = 0 \qquad (25)$$

for all k and all $\eta < \varepsilon$. Since $q'_x \neq 0$ by Lemma 1, we can make the change of variable $q(0, \eta, x) = t$ in the integral (25) for an arbitrary fixed η with $\eta < \varepsilon$. It follows that (with obvious notation)

$$\int t^k f\big(x(t, \eta), u\big(0, \eta, x(t, \eta)\big)\big) m\big(0, \eta, x(t, \eta)\big) \frac{dx}{dt}\, dt = 0$$

for all k and for all η with $\eta < \varepsilon$. Since $m > 0$ and $dx/dt \neq 0$, it follows that f vanishes on the curve $y = u(0, \eta, x)$, and since those curves certainly cover a neighborhood of the origin in the (x, y) plane, we can conclude that $f = 0$ in a neighborhood of the origin. This completes the proof of Theorem 1. □

4 The Condition on the Density μ

We now describe condition (17) on the density μ in intrinsic terms. Let V be a vector field on Z that is everywhere tangent to the fibers $\pi_X^{-1}(\mathbf{x})$. In the coordinates (ξ, η, x) this vector field can be written $V = V_1 \partial_\xi + V_2 \partial_\eta$, where V_1 and V_2 are functions on Z. In fact, since the equation for the fiber $\pi_X^{-1}(x^0, y^0)$ is

$$y^0 = u\big(\xi, \eta, x^0\big),$$

we see that the condition for $v_1 \partial_\xi + v_2 \partial_\eta + v_3 \partial_x$ to be tangent to this curve is $v_3 = 0$ and $v_1 u'_\xi + v_2 u'_\eta = 0$. A vector field V on Z can be invariantly defined as a linear map from $C^\infty(Z)$ into itself such that $V(\varphi\psi) = \varphi V(\psi) + \psi V(\varphi)$. The operation of V on the density μ will be defined by $\langle V(\mu), \varphi\rangle = -\langle \mu, V(\varphi)\rangle$. If $\sigma = \sigma_1\, d\xi + \sigma_2\, d\eta + \sigma_3\, dx$ is a 1-form on Z and $V = V_1 \partial_\xi + V_2 \partial_\eta + V_3 \partial_x$ is a vector field, then the "contraction" $\sigma \lrcorner V = \sigma_1 V_1 + \sigma_2 V_2 + \sigma_3 V_3$ is a function on Z. If σ is a 1-form on Γ, then we denote by $\pi_\Gamma^*(\sigma)$ the pullback of σ to Z under the projection π_Γ.

Consider the following condition on μ:

$$\text{there exists a 1-form } \sigma \text{ on } \Gamma \text{ such that } V(\mu) = \big(\pi_\Gamma^*(\sigma) \lrcorner V\big)\mu \tag{26}$$

for all vector fields V that are tangent to the fibers $\pi_X^{-1}(\mathbf{x})$.

If condition (26) holds for one nonvanishing vector field V that is everywhere tangent to the fibers $\pi_X^{-1}(\mathbf{x})$, then it holds for all such vector fields. In fact, if $\tilde{V}$ is another vector field that is everywhere tangent to those fibers, then $\tilde{V} = \phi V$ for some function ϕ on Z, so multiplying (26) by ϕ, we obtain $\tilde{V}(\mu) = (\pi_\Gamma^*(\sigma) \lrcorner \tilde{V})\mu$. In this way we also see that the 1-form σ is independent of the vector field V.

We now show that (17) is the same as (26). In the (x, y, ξ) coordinates $V = \partial_\xi$ is tangent to the fibers $\pi_X^{-1}(\mathbf{x})$. Thus,

$$\begin{aligned} \big\langle V(\mu), \varphi\big\rangle = -\big\langle \mu, V(\varphi)\big\rangle &= -\iiint n(x, y, \xi)\partial_\xi \varphi(x, y, \xi)\, dx\, dy\, d\xi \\ &= \iiint n'_\xi(x, y, \xi)\varphi(x, y, \xi)\, dx\, dy\, d\xi. \end{aligned} \tag{27}$$

Let $\sigma = a(\xi, \eta)\, d\xi + b(\xi, \eta)\, d\eta$ be a 1-form on Γ. To compute $\pi_\Gamma^*(\sigma)$ on Z, we note that

$$\pi_\Gamma : (x, y, \xi) \mapsto (\xi, \eta),$$

where $\eta = \rho(x, y, \xi)$. Thus,

$$\pi_\Gamma^*(\sigma) = a\, d\xi + b\, d\rho = a\, d\xi + b\big(\rho'_x\, dx + \rho'_y\, dy + \rho'_\xi\, d\xi\big).$$

With $V = \partial_\xi$ this gives

$$\pi_\Gamma^*(\sigma) \lrcorner V = a + b\rho'_\xi.$$

Hence,

$$\big\langle \big(\pi_\Gamma^*(\sigma) \lrcorner V\big)\mu, \varphi\big\rangle = \iiint \big(a + b\, \rho'_\xi\big) n(x, y, \xi)\varphi(x, y, \xi)\, dx\, dy\, d\xi. \tag{28}$$

Combining (27) and (28), we now see that (26) is equivalent to (17).

It is possible to prove Theorem 1 using the coordinates (ξ, η, x) instead of the (x, y, ξ) coordinates. Recall that the function $q(\xi, \eta, x)$ is defined by (21). The assumptions on $\mu = m(\xi, \eta, x)\, d\xi\, d\eta\, dx$ and Z then read as follows. The function $m(\xi, \eta, x)$ must satisfy

$$\partial_\eta(qm) - \partial_\xi m = (a_1 q + b_1)m, \tag{29}$$

where a_1 and b_1 are functions that depend only on (ξ, η). The condition on Z is expressed by the following condition on the function q: there exist functions a_2, b_2, and c_2 that depend only on (ξ, η) such that

$$q\,\partial_\eta q - \partial_\xi q = a_2 q^2 + b_2 q + c_2. \tag{30}$$

The proof of Theorem 1 using those assumptions is quite parallel to the proof of Theorem 1 in [3], the factor q playing the role of the factor x in [3]. Recall that $q(\xi, \eta, x) = x$ if $u(\xi, \eta, x) = \xi x + \eta$.

It is instructive to verify that (29) is also the same as (26). Choose $V = \partial_\xi - q\,\partial_\eta$, where $q = u'_\xi / u'_\eta$ as above. Then

$$\begin{aligned}\big\langle V(\mu), \varphi\big\rangle &= -\big\langle \mu, V(\varphi)\big\rangle = -\big\langle \mu, (\partial_\xi - q\,\partial_\eta)\varphi\big\rangle \\ &= -\iiint \big((\partial_\xi - q\,\partial_\eta)\varphi(\xi, \eta, x)\big) m(\xi, \eta, x)\, d\xi\, d\eta\, dx \\ &= \iiint \big(\partial_\xi m - \partial_\eta (qm)\big)\, \varphi\, d\xi\, d\eta\, dx.\end{aligned}$$

Assume that $\sigma = a(\xi, \eta)\, d\xi + b(\xi, \eta)\, d\eta$. Then

$$\pi^*_\Gamma(\sigma) \lrcorner V = aV_1 + bV_2 = a - bq.$$

Thus condition (26) means that

$$\iiint \big(\partial_\xi m - \partial_\eta (qm)\big)\varphi\, d\xi\, d\eta\, dx = \iiint (a - bq) m\varphi\, d\xi\, d\eta\, dx$$

for all φ, or

$$\partial_\xi m - \partial_\eta (qm) = (a - bq)m$$

for some a and b, which is condition (29).

5 Curve Families and Ordinary Differential Equations

In the next section we shall see that condition (18) on Z is not invariant with respect to smooth coordinate transformations in Γ, and we shall replace it by an invariant condition. To do this, we shall use two well-known results from the geometric theory of ordinary differential equations.

Proposition 2 *Consider an ordinary differential equation of the second order*

$$y'' = \Phi(x, y, p), \quad p = dy/dx, \tag{31}$$

where $\Phi(x, y, p)$ *is a polynomial in* p *of degree* ≤ 3 *with coefficients that are smooth functions of* x *and* y. *Under an arbitrary smooth variable transformation in the plane*

$$x = F(X, Y), \quad y = G(X, Y), \tag{32}$$

(31) *is transformed into an equation of the same form*, *that is*,

$$Y'' = \Psi(X, Y, P), \quad P = dY/dX, \tag{33}$$

where $\Psi(X, Y, P)$ *is a polynomial in* P *of degree at most* 3.

This is Theorem 2 of Chap. 1 §6 E in [1].

Corollary 1 *Let a two-parameter family of curves in a neighborhood of* $(0, 0)$ *in* xy *space be given by* $F(x, y, \xi, \eta) = y - u(\xi, \eta, x) = 0$ *for parameters in a neighborhood of* $(0, 0)$ *in* $\xi\,\eta$ *space*, *and assume that* F *satisfies* (4) *and* (5). *Assume that there exist coordinates in* $x\,y$ *space in which the curves are represented as straight lines*. *Then the curves of the given family and its dual family both have the property that the curves of the family are solution curves of some differential equation* (31) *where* $\Phi(x, y, p)$ *is a polynomial in* p *of degree at most* 3.

Proof The assumption means that we can choose coordinates in $x\,y$ space such that $u(\xi, \eta, x) = A(\xi, \eta)x + B(\xi, \eta)$ for all (ξ, η) in some neighborhood of $(0, 0)$. Since the map (12) is a local diffeomorphism, we can choose coordinates in a neighborhood of the origin in $\xi\,\eta$ space such that $A(\xi, \eta) = \xi$ and $B(\xi, \eta) = \eta$, and hence in the new coordinates $u(\xi, \eta, x) = \xi x + \eta$. Thus we can find coordinates so that both the given family and its dual are solution curves of the differential equation $y'' = 0$. The zero function is a polynomial of degree at most 3, so the statement now follows from Proposition 2. □

Remark The converse of the statement in the corollary is also true. In other words, the condition that both the curves of the given family and its dual family are solution curves of a differential equation (31), where $\Phi(x, y, p)$ is a polynomial in p of degree at most 3, is a necessary and sufficient condition for the given curve family to be diffeomorphic to a family of straight lines. See [1], Chap. 1 §6 G.

Proposition 3 *Assume that* $\Phi(x, y, p)$ *is a polynomial in* p *of degree* ≤ 3 *as in Proposition* 2. *Then there exists a smooth variable transformation*

$$x = F(X, Y), \quad y = Y, \tag{34}$$

in some neighborhood of the origin, $F(0, 0) = 0$, $F'_X(0, 0) \neq 0$, *such that the differential equation* (31) *is transformed to* (33) *where* $\Psi(X, Y, P)$ *is a polynomial in* P *of degree at most* 2.

For the proof, we shall use the following lemma from [1], Chap. 1 §6 B.

Lemma 2 *The substitution* (34) *transforms* (31) *into the equation*

$$Y'' = \widehat{\Phi}(X, Y, P), \quad P = dY/dX, \tag{35}$$

where

$$\widehat{\Phi}(X, Y, P) = \frac{\Delta^3}{F'_X} \Phi(X, Y, P/\Delta) + \frac{P}{F'_X}\left(F''_{XX} + 2PF''_{XY} + P^2 F''_{YY}\right), \tag{36}$$

and $\Delta = F'_X + PF'_Y$.

Proof of Proposition 3 By Proposition 2 we know that Ψ will be a polynomial in P of degree at most 3, so it will be enough to show that we can choose $F(X, Y)$ so that the coefficient of P^3 in (36) vanishes. Assume that

$$\Phi = \Phi_0 + p\Phi_1 + p^2\Phi_2 + p^3\Phi_3,$$

where Φ_j are functions of (x, y). Third-degree terms in P will only occur in the terms

$$\frac{1}{F'_X}\left(\Delta^3\Phi_0 + P\Delta^2\Phi_1 + P^2\Delta\Phi_2 + P^3\Phi_3\right) + P^3\frac{F''_{YY}}{F'_X}.$$

The coefficient of P^3 will therefore be

$$\frac{1}{F'_X}\left(\Phi_0\left(F'_Y\right)^3 + \Phi_1\left(F'_Y\right)^2 + \Phi_2 F'_Y + \Phi_3\right) + \frac{F''_{YY}}{F'_X},$$

which can be written

$$\frac{1}{F'_X}\left(\Lambda\left(X, Y, F'_Y\right) + F''_{YY}\right),$$

where we have denoted the polynomial $P^3\Phi(X, Y, 1/P)$ by $\Lambda(X, Y, P)$. Note that $F'_X \neq 0$, since the Jacobian of the transformation is F'_X. Now choose $H(X, Y)$ in a neighborhood of the origin as a solution of the ordinary differential equation

$$H'_Y(X, Y) = -\Lambda\left(X, Y, H(X, Y)\right), \quad H(X, 0) = 0,$$

and then choose

$$F(X, Y) = X + \int_0^Y H(X, t)\,dt.$$

Then the coefficient of P^3 will vanish identically, and the proposition is proved. □

6 Local Injectivity: Invariant Statement

We can now formulate a statement that is invariant under separate coordinate transformations in X and Γ.

As always, the manifold Z is defined by $F(x, y, \xi, \eta) = 0$ and is assumed to satisfy (4) and (5), and the coordinates are assumed to be chosen so that (6) holds.

Theorem 2 *Let $Z \subset X \times \Gamma$ be defined by $F(x, y, \xi, \eta) = \eta - \rho(x, y, \xi) = 0$ in a neighborhood V of a point $(\mathbf{x}^0, \gamma^0) = (0, 0, 0, 0) \in X \times \Gamma$. Assume that the function $\xi \mapsto \rho(x, y, \xi)$ for all (x, y) in a neighborhood of $(0, 0) \in X$ satisfies a differential equation*

$$\rho''_{\xi\xi} = \Phi\big(\xi, \eta, \rho'_\xi\big)$$

in some neighborhood of $(\xi^0, \eta^0) = (0, 0) \in \Gamma$, where $\Phi(\xi, \eta, p)$ is a polynomial of degree ≤ 3 in p with coefficients that are smooth functions of (ξ, η), and that the density μ on Z satisfies (17) in V. Let U be a neighborhood of $\mathbf{x}^0$ in X such that $U \setminus \gamma^0$ has just two components that we denote by U_+ and U_-. Let f be a continuous function on U such that $\operatorname{supp} f$ is contained in $U_+ \cup \{\mathbf{x}^0\}$. Assume that $Rf = 0$ in some neighborhood of $\gamma^0 \in \Gamma$. Then $f = 0$ in some neighborhood $\mathbf{x}^0$.

Proof Choose a coordinate system in X such that (19) holds. Since this coordinate change obviously preserves the condition on $\rho(x, y, \xi)$ in the theorem, we can then use Proposition 3 to choose coordinates in Γ such that $\rho(x, y, \xi)$ satisfies (18). Those coordinate changes preserve the validity of (17), since we have proved that this condition is invariant under coordinate changes in X and Γ. The assertion now follows from Theorem 1. □

We remark that Proposition 2 shows that the condition on $\rho(x, y, \xi)$ in Theorem 2 is invariant under coordinate changes in Γ, and we have already observed that it is trivially invariant under coordinate changes in X.

Finally, we give an example of a curve family satisfying the condition in Theorem 2, which cannot be transformed to a family of straight lines. Following Gindikin [13], we consider for this purpose the set of horocycles in the hyperbolic plane with the right half-plane as a model for the latter. Then the horocycles are the circles that are tangent to the y-axis

$$F(x, y, \xi, \eta) = (x - \xi)^2 + (y - \eta)^2 - \xi^2 = 0. \tag{37}$$

The full set of horocycles is unsuitable for us, since there are in general two horocycles with the same tangent direction through a given point. But if we restrict (x, y, ξ, η) to a neighborhood of, for instance, the point $(x^0, y^0) = (1, 0)$, $(\xi^0, \eta^0) = (1, 1)$, we get rid of this ambiguity, and F will satisfy (4) and (5). Differentiating (37) twice with respect to ξ and eliminating x and y leads to the differential equation

$$\eta'' = \big(1 + \eta'^2\big)/2\eta,$$

which shows that the condition of Theorem 2 is satisfied. On the other hand, a similar computation with (x, y) and (ξ, η) interchanged shows that the curve family in the $x\,y$ plane satisfies the differential equation

$$y'' = \frac{1}{x}\left(1 + y'^2\right)\left(y' + \sqrt{1 + y'^2}\right).$$

The last expression is not a polynomial in y', so by the Corollary our curve family cannot be transformed to straight lines.

For further information on the geometric theory of differential equations, we refer to [6] and references given there.

References

1. Arnold, V.I.: Geometric Methods in the Theory of Ordinary Differential Equations, 2nd edn. Springer, Berlin (1988)
2. Boman, J.: An example of non-uniqueness for a generalized Radon transform. J. Anal. Math. **61**, 395–401 (1993)
3. Boman, J.: A local uniqueness theorem for weighted Radon transforms. Inverse Probl. Imaging **4**, 631–637 (2010)
4. Boman, J.: Local non-injectivity for weighted Radon transforms. Contemp. Math. **559**, 39–47 (2011)
5. Boman, J., Quinto, E.T.: Support theorems for real-analytic Radon transforms. Duke Math. J. **55**, 943–948 (1987)
6. Bryant, R., Griffiths, P., Hsu, L.: Toward a geometry of differential equations. In: Geometry, Topology, & Physics, Conf. Proc. Lecture Notes Geom. Topology, vol. IV. Int. Press, Cambridge (1995)
7. Ehrenpreis, L.: The Universality of the Radon Transform. Clarendon Press, Oxford (2003)
8. Gelfand, I.M., Graev, M.I.: Admissible n-dimensional complexes of curves in $\mathbf{R}^n$. Funct. Anal. Appl. **14**, 274–281 (1980)
9. Gelfand, I.M., Graev, M.I., Shapiro, Z.Ya.: Differential forms and integral geometry. Funct. Anal. Appl. **3**, 101–114 (1969)
10. Gelfand, I.M., Gindikin, S.G., Shapiro, Z.Ya.: A local problem of integral geometry in a space of curves. Funct. Anal. Appl. **13**, 87–102 (1979)
11. Gindikin, S.G.: Reduction of manifolds of rational curves and related problems of the theory of differential equations. Funct. Anal. Appl. **18**, 278–298 (1984)
12. Gindikin, S.G.: A remark on the weighted Radon transform on the plane. Inverse Probl. Imaging **4**, 649–653 (2010)
13. Gindikin, S.G.: Remarks on infinitesimally Desarguian families of curves. Funct. Anal. Appl. **45**, 265–270 (2011)
14. Helgason, S.: A duality in integral geometry on symmetric spaces. In: Proc. U.S.–Japan Seminar on Differential Geometry, Kyoto 1965, pp. 37–56. Nippon Hyoronsha, Tokyo (1966)
15. Strichartz, R.S.: Radon inversion—variations on a theme. Am. Math. Mon. **89**, 377–384 (1982), see also 420–423

Deconvolution for the Pompeiu Problem on the Heisenberg Group, I

Der-Chen Chang, Wayne Eby, and Eric Grinberg

Abstract We consider variations on the Pompeiu transform for the Heisenberg group $\mathbf{H}^n$ and focus on cases where the transform is known to be injective; in particular the cases of averages over a sphere and a ball, or two balls of appropriate radii. In these cases we develop a method which provides for the reconstruction of the function f from its integrals.

In addition, we consider these issues in connection with the Weyl calculus and group Fourier transform. We furthermore explore issues of convergence for this method of deconvolution and related issues of size of the Gelfand transform near the zero sets. Finally, given a set of deconvolvers which work for Euclidean space $\mathbf{C}^n$, we consider the problem of how to extend the deconvolution to the Heisenberg group, and we provide the extension in special cases.

1 Introduction

Dedication and a Mathematical Moment

This paper is dedicated to the memory of Leon Ehrenpreis. His insight, enthusiasm, and energy will continue to inspire us for many years to come. We take this opportunity to recall a brief mathematical moment with Leon. In one of countless sessions at the blackboard, someone commented that, while a real hypersurface allows for the concept of a *side*, and often *inside* and *outside*, the same does not hold

D.-C. Chang (✉)
Department of Mathematics and Department of Computer Science, Georgetown University, Washington D.C. 20057, USA
e-mail: chang@georgetown.edu

W. Eby
Department of Mathematical Sciences, Cameron University, Lawton, OK 73505, USA
e-mail: weby@cameron.edu

E. Grinberg
Department of Mathematics, University of Massachusetts Boston, Boston, MA 02125, USA
e-mail: eric.grinberg@umb.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_6, © Springer-Verlag Italia 2012

for a complex hypersurface, but it would be nice if the concept could be extended. Leon thought about it a bit and said that this may be possible, "probably by extending the notion of the Cauchy Problem."

The interplay between real and complex analysis was of great interest to Leon. This paper focuses on analysis on the Heisenberg group, which is a fertile group for this interplay.

The Pompeiu problem asks for conditions which guarantee the uniqueness of a function in terms of its integral averages. This problem may also be expressed as a question of injectivity of a certain integral transformation. In particular, given a function $f \in C(\mathbf{R}^n)$ and for a specified set $S \Subset \mathbf{R}^n$, consider the integral averages

$$\int_{\sigma S} f(x)\, d\mu_\sigma(x)$$

for all rigid motions σ of $\mathbf{R}^n$, where μ_σ is the area measure on the set σS. The Pompeiu problem asks whether the vanishing of all of these integrals is sufficient to conclude $f \equiv 0$. The same question may also be asked in the context of an integral transform. Define the Pompeiu transform as

$$\mathscr{P} f(\sigma) = \int_{\sigma S} f(x)\, d\mu_\sigma(x)$$

for each $\sigma \in \mathscr{M}(n)$, the Euclidean motion group. The Pompeiu problem may then be considered as the issue of injectivity for the Pompeiu transform. When the transform is injective, it is then possible to consider the problem of inversion.

We consider the problem of inversion of the Pompeiu transform at the level of establishing a method to recover the original function f from the transformed function $\mathscr{P} f$. Note that we do not give an explicit inverse to the operator $\mathscr{P}$, but we do establish a limiting procedure that allows the recovery of f from a given $\mathscr{P} f$. The problem of deconvolution is then the reconstruction of the function f from these integral averages. In the paper [7], Berenstein and Yger addressed this problem in the setting of the Euclidean space $\mathbf{R}^n$. They form explicit formulas to determine the deconvolvers $\nu_1, \ldots, \nu_n$ from the bounded measures $\mu_1, \ldots, \mu_n$ using the operations of derivation, integration, convolution, and summation. In this paper we address the problem of deconvolution for the Pompeiu problem in the setting of the Heisenberg group. We consider the Pompeiu problem for functions in $C \cap L^\infty(\mathbf{H}_n)$ and consider integration over a set $S \Subset \mathbf{C}^n \times \{0\} \subset \mathbf{H}_n$, such as described in [1]. Although the issue of deconvolution can be considered for any set S, or collection of sets, for which the Pompeiu transform is injective, we focus on two cases for which the zeros of the associated transforms are well known. The first case considers the set S as a ball and a sphere of the same radius, while the second considers S as two balls of appropriate radii. The general approach established in this paper provides a method to construct sequences of deconvolvers which allows reconstruction of f through a limiting process. In addition we address the issue of when these deconvolving sequences will converge to individual deconvolvers, defined as tempered distributions.

This issue of convergence relates directly to the location of the zeros of $\hat{\mu}_1, \ldots, \hat{\mu}_n$ and the issues discussed below.

In Euclidean space, a large part of the problem of deconvolution is directly related to a condition known as *Hörmander's strongly coprime condition* [17]. Consider the integral conditions as a set of convolution equations. In the case of two balls of radii r_1, r_2 the integral conditions may be written as

$$f * T_{r_1} = 0, \qquad f * T_{r_2} = 0,$$

where T_{r_j} are defined by

$$\langle \phi, T_{r_j} \rangle = \int_{|x|<r_j} \phi(x)\, d\mu_j(x)$$

for $j = 1, 2$, where μ_j is the area measure on the ball of radius r_j. Now observe that the problem of deconvolution can be solved by finding $\nu_1, \nu_2 \in \mathscr{E}$ which satisfy the analytic Bezout equality

$$\widehat{T_{r_1}}\widehat{\nu_1} + \widehat{T_{r_2}}\widehat{\nu_2} \equiv 1. \tag{1}$$

Hörmander's strongly coprime condition tells us that such ν_1, ν_2 exist as compactly supported distributions if and only if $\widehat{T_{r_1}}$ and $\widehat{T_{r_2}}$ satisfy the following estimate, related to Paley–Weiner estimates,

$$\left|\widehat{T_{r_1}}(\zeta)\right| + \left|\widehat{T_{r_2}}(\zeta)\right| \geq C \frac{1}{(1+|\zeta|)^N} e^{-B|\mathrm{Im}\,\zeta|} \tag{2}$$

for some $C, N, B > 0$. In addition to $\widehat{T_{r_1}}$ and $\widehat{T_{r_2}}$ not having common zeros, this condition describes a maximum rate of decay for a combination of $\widehat{T_{r_1}}$ and $\widehat{T_{r_2}}$, ensuring that where one of these becomes zero, the other must not become too small too quickly.

In the specific case of the Pompeiu problem for two balls of distinct radii, the issue of injectivity is determined by an arithmetic condition on the radii. In this case, the transform for balls B_{r_1} and B_{r_2} will be injective when the following condition on quotients of zeros of *Bessel* functions J holds:

$$\frac{r_1}{r_2} \notin \mathscr{Q}(J_{\frac{n}{2}}) = \left\{ \frac{\gamma x}{\gamma y} : \gamma \in \mathbf{R}^*,\, J_{\frac{n}{2}}(x) = J_{\frac{n}{2}}(y) = 0 \right\}.$$

When, in addition, the quotient of radii $\frac{r_1}{r_2}$ is poorly approximated by quotients of zeros of $J_{\frac{n}{2}}$, then Hörmander's strongly coprime condition can be shown to be satisfied. It then follows that the deconvolution problem for T_{r_1} and T_{r_2} can be solved with compactly supported distributions ν_1 and ν_2.

When such ν_1 and ν_2 can be found, this provides a solution to the problem of deconvolution as follows. First note that $T_{r_1} * \nu_1 + T_{r_2} * \nu_2 = \delta$. Then convolve the convolution expressions $f * T_{r_1}$ and $f * T_{r_2}$ with ν_1 and ν_2 to obtain

$$f * T_{r_1} * \nu_1 + f * T_{r_2} * \nu_2 = f * (T_{r_1} * \nu_1 + T_{r_2} * \nu_2)$$
$$= f * \delta = f.$$

One of the goals of this paper is to establish a similar approach to deconvolution that works in the Heisenberg setting. We develop a method that works for the two specific cases considered in Sect. 3, the first case of a ball and sphere of the same radius and the second case of two balls of distinct radii. In both cases a sequence of deconvolvers is presented which in the limit accomplishes the deconvolution to return the original function f. Section 4 considers these issues in the context of the Weyl calculus. Issues of convergence for the deconvolving sequences are considered in Sect. 5. In the case of two balls of distinct radii, these considerations lead to an interesting interaction between Hörmander's strongly coprime condition, the issue of N-well approximation of the radii, and the space in which the sequences of deconvolvers will converge. Section 6 deals with the extension of a given deconvolution for the space $\mathbf{C}^n$ to the space $\mathbf{H}^n$. We plan to pursue these issues further in a later paper.

The inspiration for this paper comes from the desire to find deconvolvers ν_1 and ν_2 that satisfy the analytic Bezout equation (1). When considering this issue in the setting of the Gelfand transform for the Heisenberg group, we realized that the method of proving the Tauberian theorem can be adapted to construct ν_1 and ν_2 as needed. Note that the construction given in this paper involves a "local inversion" of $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ away from their zero sets and thus relates to the problem of division considered in the pioneering works of Ehrenpreis [12–14].

2 Heisenberg Group, Problem of Deconvolution

The Heisenberg group $\mathbf{H}^n$ can be given by coordinates $[\mathbf{z}, t] \in \mathbf{C}^n \times \mathbf{R}$ with group law defined by

$$[\mathbf{z}, t] \cdot [\mathbf{w}, s] = [\mathbf{z} + \mathbf{w}, t + s + 2\mathrm{Im}\, \mathbf{z} \cdot \bar{\mathbf{w}}].$$

This group can be realized as the boundary of the Siegel upper half-space $\mathscr{U}_{n+1}$ in $\mathbf{C}^{n+1}$

$$\partial \mathscr{U}_{n+1} = \left\{ (\mathbf{z}, z_{n+1}) \in \mathbf{C}^{n+1} : \mathrm{Im}\, z_{n+1} = |\mathbf{z}|^2 \right\},$$

where the group law gives a group action on the hypersurface. As usual, $|\mathbf{z}|^2 = \sum_{k=1}^n |z_k|^2$. The left-invariant vector fields associated to $\mathbf{H}^n$ are spanned by the basis

$$Z_j = \frac{\partial}{\partial z_j} + i \bar{z}_j \frac{\partial}{\partial t},$$

$$\bar{Z}_j = \frac{\partial}{\partial \bar{z}_j} - i z_j \frac{\partial}{\partial t},$$

for $j = 1, \ldots, n$, and

$$T = \frac{\partial}{\partial t}.$$

The only nonzero brackets are given by $[Z_j, \bar{Z}_j] = -2iT$ for $j = 1, \ldots, n$. The sub-Laplacian is given by

$$\mathscr{L} = \sum_{j=1}^{n} \bar{Z}_j Z_j + Z_j \bar{Z}_j.$$

Harmonic analysis on $\mathbf{H}^n$ for radial functions $f \in L_0^\infty(\mathbf{H}^n)$ utilizes the Gelfand transform defined using the bounded $U(n)$-spherical functions. These can be defined as joint eigenfunctions of $\mathscr{L}$ and iT and are given by

$$\psi_k^\lambda(\mathbf{z}, t) = ce^{2\pi i\lambda t} e^{-2\pi|\lambda||\mathbf{z}|^2} L_k^{(n-1)}\left(4\pi|\lambda||\mathbf{z}|^2\right) \quad \text{for } (k, \lambda) \in \mathbf{Z}_+ \times \mathbf{R}^*$$

and

$$\mathscr{J}_\rho(\mathbf{z}) = c\frac{J_{n-1}(\rho|\mathbf{z}|)}{(\rho|\mathbf{z}|)^{n-1}} \quad \text{for } \rho \in \mathbf{R}_+.$$

These bounded spherical functions are used to form the Gelfand transform for $L_0^1(\mathbf{H}^n)$, defined for $f \in L_0^1(\mathbf{H}^n)$ as $\widetilde{f}(p)$ for $p \in \mathscr{H}$ by

$$\widetilde{f}(\lambda, k) = \int_{\mathbf{H}^n} f(\mathbf{z}, t)\overline{\psi_k^\lambda(\mathbf{z}, t)}\, dm(\mathbf{z}, t)$$

and

$$\widetilde{f}(0; \rho) = \int_{\mathbf{H}^n} f(\mathbf{z}, t)\overline{\mathscr{J}_\rho(\mathbf{z})}\, dm(\mathbf{z}, t).$$

The spectrum is given by the Heisenberg fan $\mathscr{H}$ composed of a central Bessel ray $\mathscr{H}_\rho$ and infinitely many Laguerre rays $\mathscr{H}_{k,\pm}$ converging to the central Bessel ray. Denote $\mathscr{H}$, $\mathscr{H}_\rho$, and $\mathscr{H}_{k,\pm}$ as follows:

$$\begin{aligned} \mathscr{H} &= \mathscr{H}_\rho \cup \left(\bigcup_{k=1}^{\infty} \mathscr{H}_{k,+} \cup \mathscr{H}_{k,-} \right) \\ &= \left\{(0, \rho) : \rho \geq 0\right\} \cup \left(\bigcup_{k=1}^{\infty} \left\{ \left(\lambda, 4|\lambda| \left(k + \frac{n}{2}\right)\right) : \lambda \in \mathbf{R}^* \right\} \right). \end{aligned} \tag{3}$$

In application of the Gelfand transform, we will need to define $\Psi_k^{(n-1)}(x)$ as

$$\Psi_k^{(n-1)}(x) = \int_0^x e^{-t/2} L_k^{(n-1)}(t) t^{n-1}\, dt.$$

In addition, we use the notation j_n to represent the function $j_n(x) = \frac{J_n(x)}{x^n}$.

Part of our analysis of this problem will include the perspective of the Weyl calculus and the group Fourier transform. We first define the position and momentum operators $P = (P_1, \dots, P_n)$ and $Q = (Q_1, \dots, Q_n)$, given by $P_j u(\mathbf{x}) = x_j u(\mathbf{x})$ and $Q_j u(\mathbf{x}) = \frac{1}{i}\frac{\partial u}{\partial x_j}(\mathbf{x})$. As stated in [16], the group of unitary operators on L^2 are defined by

$$\rho(p,q,t) = e^{2\pi i(pD+qX+tI)} = e^{2\pi it}e^{2\pi i(pD+qX)},$$

which act on $f \in L^2$ as

$$\rho(p,q,t)f(x) = e^{2\pi it+2\pi iqx+\pi ipq}f(x+p),$$

forming the *Schrödinger representation of* $\mathbf{H}^n$. Through the nonisotropic Heisenberg dilation by $\lambda^{1/2}$, we obtain the representations used in the Weyl transform, which is then based on the infinite-dimensional representations

$$\pi_{\pm\lambda} = e^{2\pi i(\pm\lambda t \pm \lambda^{1/2}\mathbf{x}\cdot P + \lambda^{1/2}\mathbf{y}\cdot Q)} \quad \text{for } \lambda \in \mathbf{R}_+ \setminus \{0\}$$

and the one-dimensional representations

$$\pi_{(\xi,\eta)} = e^{2\pi i(x\cdot\xi + y\cdot\eta)} \quad \text{for } (\xi,\eta) \in \mathbf{R}^n \times \mathbf{R}^n$$

that can be attained in the limit as $\lambda \to 0$. The group Fourier transform on $\mathbf{H}^n$ is then given by

$$\pi_{\pm\lambda}(f) = \int_{\mathbf{R}^{2n+1}} f(\mathbf{x},\mathbf{y},t)\pi_{\pm\lambda}(\mathbf{x},\mathbf{y},t)\,d\mathbf{x}\,d\mathbf{y}\,dt$$

and

$$\pi_{(\xi,\eta)}(f) = \int_{\mathbf{R}^{2n+1}} f(\mathbf{x},\mathbf{y},t)\pi_{(\xi,\eta)}\,d\mathbf{x}\,d\mathbf{y}\,dt.$$

Both of these can be attained from $\mathscr{F}_{2n+1}$ by the following

$$\pi_{(\xi,\eta)}(f) = \mathscr{F}_{2n+1}(f)(-\eta,-\xi,0)$$

and

$$\pi_{\pm\lambda}(f) = \mathscr{F}_{2n+1}(f)\big(\mp\lambda^{1/2}P, -\lambda^{1/2}Q, \mp\lambda\big).$$

For additional details on this material, please see [1, 5], where this theory is applied for the Pompeiu transform on spheres in $\mathbf{H}^n$.

In Euclidean space $\mathbf{R}^n$, deconvolution takes place on the side of the Fourier transform, either in the explicit construction of deconvolvers $\nu_1, \dots, \nu_n$ satisfying the analytic Bezout equation

$$\widehat{\mu}_1\widehat{\nu}_1 + \cdots + \widehat{\mu}_n\widehat{\nu}_n \equiv 1$$

or in verification of existence of such $\nu_1, \ldots, \nu_n$ through verification of the Hörmander strongly coprime condition

$$\left|\widehat{\mu_1}(\zeta)\right| + \cdots + \left|\widehat{\mu}_n(\zeta)\right| \geq C\frac{1}{(1+|\zeta|)^N}e^{-B|\mathrm{Im}\,\zeta|}$$

for some $C, N, B > 0$. Observe that the process of deconvolution takes place on the Fourier transform side; the same will be true for deconvolution on the Heisenberg group, $\mathbf{H}^n$. When the measures $\mu_1, \ldots, \mu_n$ are $U(n)$-radial, it is possible to consider Gelfand transforms $\widetilde{\mu}_1, \ldots, \widetilde{\mu}_n$, directly related to Fourier–Bessel transforms $\widehat{\mu}_1(\rho), \ldots, \widehat{\mu}_n(\rho)$ with the radial variable $\rho = \sqrt{|\zeta_1|^2 + \cdots + |\zeta_n|^2}$. In both of the cases appearing in Sect. 3, which involve $U(n)$-radial spaces, the construction of deconvolvers is completed on the Gelfand transform side. As part of this study, we will also consider the connection to the group Fourier transform through the Weyl calculus in Sect. 4. We focus primarily on two specific cases where the measures are $U(n)$-radial in this paper and will consider more general cases in a later work.

We do not yet have a version of Hörmander's strongly coprime condition for the Gelfand transform on $\mathbf{H}_n$. However, given radial measures $\mu_1, \mu_2 \in L^1_{\mathbf{0}}(\mathbf{H}_n)$, deconvolution is still possible through explicit construction of ν_1, ν_2 satisfying

$$\widetilde{\mu}_1(p)\widetilde{\nu}_1(p) + \widetilde{\mu}_2(p)\widetilde{\nu}_2(p) \equiv 1 \tag{4}$$

for every $p \in \mathscr{H}$, i.e., for every $p = (\lambda, k) \in \mathbf{R}^* \times \mathbf{Z}_+$ or $p = (0, \rho) \in \mathbf{R}_+$. Although we cannot accomplish this in one step for the entire Heisenberg brush $\mathscr{H}$, we do accomplish this in the limit through use of a compact exhaustion $\{K_j\}$ of $\mathscr{H}$. That is to say, we construct sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ satisfying (4) for all $p \in K_j$ for each j. The process of deconvolution is then accomplished in passing to the limit.

Once this process of deconvolution is accomplished, we also consider the relation to the group Fourier transform using the Weyl calculus.

3 Procedure for Deconvolution: Two Specific Cases

Here we consider deconvolution of the Pompeiu transform for the function space $C \cap L^\infty(\mathbf{H}_n)$ as described in previous papers [1, 5]. It is at this level where two sets are required, and we first consider the case of a ball and sphere of the same radius in Sect. 3.1, followed by the case of two balls of appropriate radii in Sect. 3.2. Although both cases follow the same general procedure, the first is somewhat more direct because we can find a uniform separation of the zeros of the Gelfand transforms for the two sets. In each case we consider a limiting procedure, which allows us to construct a sequence of functions converging to the original function f. To do this, we will construct a sequence of deconvolvers $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ such that

$$\widetilde{T}_1\widetilde{\nu}_{1,j} + \widetilde{T}_2\widetilde{\nu}_{2,j} \to 1,$$

in the sense that

$$\widetilde{T}_1\widetilde{\nu}_{1,j} + \widetilde{T}_2\widetilde{\nu}_{2,j} \equiv 1$$

on some compact set K_j, where the sequence $\{K_j\}$ forms a compact exhaustion of the Heisenberg fan.

3.1 Ball and Sphere

Here we consider the Pompeiu transform defined in terms of the integral averages

$$\int_{|\mathbf{z}|=r} L_{\mathbf{g}} f(\mathbf{z}, 0)\, d\sigma_r(\mathbf{z}) \quad \text{for all } \mathbf{g} \in \mathbf{H}^n$$

and

$$\int_{|\mathbf{z}|<r} L_{\mathbf{g}} f(\mathbf{z}, 0)\, d\mu_r(\mathbf{z}) \quad \text{for all } \mathbf{g} \in \mathbf{H}^n.$$

These may also be written as convolution equations $f * S_r$ and $f * T_r$, where

$$\langle \phi, S_r \rangle = \int_{|\mathbf{z}|=r} \phi(\mathbf{z}, 0)\, d\sigma_r(\mathbf{z}) \quad \text{and} \quad \langle \phi, T_r \rangle = \int_{|\mathbf{z}|<r} \phi(\mathbf{z}, 0)\, d\mu_r(\mathbf{z}).$$

To define sequences of sets $\{K_j\}$ and $\{V_j\}$, we first list the zeros of $J_n(rx)$ sequentially as $\{\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots\}$ as well as the zeros of $J_{n-1}(rx)$, listed sequentially as $\{\lambda_1^0, \lambda_2^0, \ldots, \lambda_n^0, \ldots\}$. Then, letting $N_j = \frac{3\lambda_j + \lambda_{j+1}^0}{4}$ and $N_j^+ = \frac{\lambda_j + \lambda_{j+1}^0}{2}$, we form

$$K_j = \left\{p = (x, y) \in \mathscr{H} : x^2 + y^2 \le N_j^2\right\},$$

where $(x, y) = (\lambda, |\lambda|(4k + 2))$ or $(x, y) = (0, \rho^2)$, and

$$V_j = \left\{p = (x, y) \in \mathscr{H} : x^2 + y^2 < \left(N_j^+\right)^2\right\},$$

where $(x, y) = (\lambda, |\lambda|(4k + 2))$ or $(x, y) = (0, \rho^2)$.

We claim the following.

Theorem 1 *Let S_r and T_r be the distributions defined above. Consider the sequence of compact sets $\{K_j\} \subset \mathscr{H}$, which forms a compact exhaustion of the Heisenberg fan $\mathscr{H}$, as given above. There exist sequences of functions $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ with the property that*

$$\widetilde{S}_r\widetilde{\nu}_{1,j} + \widetilde{T}_r\widetilde{\nu}_{2,j} \equiv 1 \quad \text{on } K_j.$$

It is also true that

$$\widetilde{S}_r\widetilde{\nu}_{1,j} + \widetilde{T}_r\widetilde{\nu}_{2,j} \equiv 0 \quad \text{on } V_j^c,$$

where each V_j is an open set defined above such that $K_j \subset V_j \subset K_{j+1}$.

In the following proof, we furthermore give a method for constructing these sequences of functions $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$. In Sect. 5 we further illustrate how these sequences are used in the problem of deconvolution and address convergence issues.

Proof Recall that for $f \in L^1_{\mathbf{0}}(\mathbf{H}_n)$, the Gelfand transform $\widetilde{f}$ is defined on the Heisenberg fan $\mathscr{H}$ which was defined in (3). This is realized as a subset of the upper half-plane of $\mathbf{R}^2$. We use the compact exhaustion $\{K_j\}$ given before the statement of this theorem. Note that each K_j is compact and $\cup_j K_j = \mathscr{H}$. We will construct the sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ so that $\nu_{1,j}$ and $\nu_{2,j}$ satisfy the needed relation

$$\widetilde{S}_r \widetilde{\nu}_{1,j} + \widetilde{T}_r \widetilde{\nu}_{2,j} = 1$$

on K_j for $j = 1, \dots, n$. Let us recall the values of $\widetilde{S}_r$ and $\widetilde{T}_r$. First,

$$\widetilde{S}_r(\lambda, k) = c e^{-2\pi|\lambda|r^2} L_k^{(n-1)}\big(4\pi|\lambda|r^2\big),$$

and

$$\widetilde{S}_r(0, \rho) = c\frac{J_{n-1}(r\rho)}{(r\rho)^{n-1}}.$$

Likewise,

$$\widetilde{T}_r(\lambda, k) = c\int_0^r e^{-2\pi|\lambda|s^2} L_k^{(n-1)}\big(4\pi|\lambda|s^2\big)s^{2n-1}\,ds = c'\Psi_k^{(n-1)}\big(4\pi|\lambda|r^2\big),$$

and

$$\widetilde{T}_r(0, \rho) = c\frac{J_n(r\rho)}{(r\rho)^n}.$$

We consider the two sets $V_1 = \{\text{zeros of } \widetilde{S}_r\}$ and $V_2 = \{\text{zeros of } \widetilde{T}_r\}$. At times it will be necessary to focus on the zeros in the Bessel part of the spectrum. For this purpose, we define $U_1 = V_1 \cap \mathscr{H}_\rho$ and $U_2 = V_2 \cap \mathscr{H}_\rho$. In order to focus on those zeros of V_1 and V_2 which are contained in the set K_j of the compact exhaustion $\{K_j\}$, we define the sequences $\{V_{1,j}\}$ and $\{V_{2,j}\}$ by $V_{1,j} = V_1 \cap K_j$ and $V_{2,j} = V_2 \cap K_j$. Similarly, we may want to focus on the zeros inside of K_j which are in Bessel part of the spectrum and thus form the sequences $\{U_{1,j}\}$ and $\{U_{2,j}\}$, defined as $U_{1,j} = U_1 \cap K_j$ and $U_{2,j} = U_2 \cap K_j$.

The next goal is to construct appropriate neighborhoods of the elements of V_1 and V_2. These neighborhoods will be used together with "local identities" on each of the neighborhoods in the deconvolution procedure. We work outward from the central Bessel ray $\mathscr{H}_\rho$, containing the zeros U_1 and U_2, since each neighborhood of one of these zeros will contain an infinite number of zeros in the Laguerre rays. We index the sets of Bessel zeros

$$U_1 = \{M_1, M_2, \dots, M_n, \dots\}$$

and

$$U_2 = \{N_1, N_2, \ldots, N_n, \ldots\}.$$

We know, because of the relation of zeros of Bessel functions of consecutive indices, that these zeros are interlacing, i.e.,

$$M_1 < N_1 < M_2 < N_2 < \cdots .$$

We form sequences of neighborhoods $\{C_i\}$ and $\{C'_i\}$, where

$$C_i = \big[(N_{i-1} + M_i)/2, (M_i + N_i)/2\big]$$

is a neighborhood of M_i, and

$$C'_i = \big[(N_i + M_i)/2, (N_i + M_{i+1})/2\big]$$

is a neighborhood of N_i. For C_1, use $C_1 = [0, (M_1 + N_1)/2]$. We next list some important properties of these sequences. First, they cover the central Bessel ray, $\mathscr{H}_\rho$:

$$\left(\bigcup_{i=1}^{\infty} C_i\right) \cup \left(\bigcup_{i=1}^{\infty} C'_i\right) \supset \mathscr{H}_\rho.$$

Next, they are nearly disjoint, with consecutive neighborhoods intersecting only at the endpoints:

$$C_i \cap C'_i = \big\{(M_i + N_i)/2\big\} \quad \text{and} \quad C'_i \cap C_{i+1} = \big\{(N_i + M_{i+1})/2\big\}.$$

Finally, they separate the zero sets of $\widetilde{S}_r$ and $\widetilde{T}_r$. In particular, for all j,

$$N_j \cap \left(\bigcup_{i=1}^{\infty} C_i\right) = \emptyset \quad \text{and} \quad M_j \cap \left(\bigcup_{i=1}^{\infty} C'_i\right) = \emptyset.$$

Letting $\delta_i = (N_i - M_i)/2$ and $\delta'_i = (M_{i+1} - N_i)/2$, where for δ'_0 we use $N_0 = 0$, we also form the smaller neighborhoods $B_i = [M_i - \delta'_{i-1}/2, M_i + \delta_i/2]$ of M_i, and similarly, $B'_i = [N_i - \delta_i/2, N_i + \delta'_i/2]$ of N_i. This gives sequences of neighborhoods $\{B_i\}$ and $\{B'_i\}$ satisfying $B_i \subset C_i$ and $B'_i \subset C'_i$. We furthermore claim that there exist sequences of slightly larger neighborhoods $\{V_i\}$ and $\{V'_i\}$ where V_i satisfy $C_i \subset V_i$ and $V_i \cap (B'_{i-1} \cup B'_i) = \emptyset$. Likewise, V'_i satisfy $C'_i \subset V'_i$ and $V'_i \cap (B_{i-1} \cup B_i) = \emptyset$.

The next step is to extend the neighborhoods from the Bessel ray $\mathscr{H}_\rho$ to all of the Heisenberg fan $\mathscr{H}$. Recall that, as $k \to \infty$, the Laguerre rays $\mathscr{H}_{k,\pm}$ converge to $\mathscr{H}_\rho$. For any given i, we expand the neighborhood B_i in $\mathscr{H}_\rho$ to a wider neighborhood $B_{i,j}$ in the Heisenberg fan $\mathscr{H}$ such that $B_{i,j} \cap \mathscr{H}_\rho = B_i$. Note that in $\mathscr{H}$, $\{B_i\}$ can be written as $\{\{0\} \times [M_i - \delta_i^-/2, M_i + \delta_i^+/2]\}$ and that the Laguerre rays $\mathscr{H}_{k,\pm}$ can be expressed as

$$\mathscr{H}_{k,\pm} = \big\{\big(\lambda, 4|\lambda|(k + n/2)\big) : \lambda \in \mathbf{R}^*\big\}.$$

We then define $B_{i,j}$ by

$$B_{i,j} = \left\{(x, y) : \left(M_i - \delta_i^-/2\right)^2 \le x^2 + y^2 \le \left(M_i + \delta_i^+/2\right)^2 \text{ and } \left|\frac{y}{x}\right| \ge 4(j + n/2)\right\}.$$

We then want to choose $j_i \in \mathbf{Z}_+$ with the property that, for each $j \ge j_i$, exactly one of the Laguerre zeros on the ray $\mathscr{H}_{j,\pm}$ is inside of $B_{i,j} \cap \mathscr{H}_{j,\pm}$. We will also choose j_i to minimize all possible j_i satisfying this property and consider the neighborhood B_{i,j_i}. Using the same choice for j_i, we also consider the larger neighborhoods C_{i,j_i} and V_{i,j_i}. These yield sequences of neighborhoods $\{B_{i,j_i}\}$, $\{C_{i,j_i}\}$, and $\{V_{i,j_i}\}$. We use the identical construction for sequences of neighborhoods $\{B'_{i,j'_i}\}$, $\{C'_{i,j'_i}\}$, and $\{V'_{i,j'_i}\}$.

Considering the sequences $\{C_{i,j_i}\}$ and $\{C'_{i,j'_i}\}$, it remains true that

$$\mathscr{H}_\rho \subset \left(\bigcup_{i=1}^{\infty} C_{i,j_i}\right) \cup \left(\bigcup_{i=1}^{\infty} C'_{i,j'_i}\right).$$

However, in general, the above union will not cover all of the Heisenberg fan $\mathscr{H}$, and furthermore, there will be some of the zeros in the sets V_1 and V_2 which are not covered. However there can only be a finite number of such zeros on each Laguerre ray. Furthermore these remaining zeros will be locally finite. In order to cover these zeros, we first denote them as

$$\{P_1, P_2, \ldots, P_n, \ldots\} \quad \text{and} \quad \left\{P'_1, P'_2, \ldots, P'_n, \ldots\right\}.$$

There exist neighborhoods $\{D_1, D_2, \ldots, D_n, \ldots\}$ and $\{D'_1, D'_2 \ldots, D'_n, \ldots\}$, as well as neighborhoods $\{W_1, W_2 \ldots, W_n, \ldots\}$ and $\{W'_1, W'_2, \ldots, W'_n, \ldots\}$ satisfying

$$\left(\bigcup_{i=1}^{\infty} D_i\right) \cup \left(\bigcup_{i=1}^{\infty} D'_i\right) \supset \mathscr{H} \setminus \left[\left(\bigcup_{i=1}^{\infty} C_{i,j_i}\right) \cup \left(\bigcup_{i=1}^{\infty} C'_{i,j'_i}\right)\right],$$

where

1. $D_i \subset W_i$ and $D'_i \subset W'_i$,
2. $P_i \cap (\cup_{i=1}^{\infty} W'_i) = \emptyset$ and $P'_i \cap (\cup_{i=1}^{\infty} W_i) = \emptyset$,
3. $P_i \notin W_j$ for $i \neq j$ and $P'_i \notin W'_j$ for $i \neq j$.

We have found neighborhoods $\{C_{i,j_i}\}$ and $\{D_i\}_{i=1}^{\infty}$ of V_1, the zeros of $\widetilde{S}_r$, as well as neighborhoods $\{C'_{i,j'_i}\}$ and $\{D'_i\}_{i=1}^{\infty}$ of V_2, the zeros of $\widetilde{T}_r$, such that

$$\mathscr{H} \subset (\cup C_{i,j_i}) \cup \left(\bigcup_{i=1}^{\infty} D_i\right) \cup \left(\cup C'_{i,j'_i}\right) \cup \left(\bigcup_{i=1}^{\infty} D'_i\right)$$

and such that there is separation between V_1 and $(\cup C'_{i,j'_i}) \cup (\cup_{i=1}^{\infty} D'_i)$, as well as between V_2 and $(\cup C_{i,j_i}) \cup (\cup_{i=1}^{\infty} D_i)$. We next find sequences of "local identities" which equal 1 on these neighborhoods. These are used to "invert" $\widetilde{S}_r$ on neighborhoods of V_2 and $\widetilde{T}_r$ on neighborhoods of V_1, away from their zero sets. These are combined to "invert" the transform on all of $\mathscr{H}$.

We use the result on "local identities" that for K a compact subset of $\mathscr{H}$ and F an open subset of $\mathscr{H}$ such that $K \cap F = \emptyset$, there exists $\rho \in L^1_{\mathbf{0}}(\mathbf{H}^n)$ such that $\widetilde{\rho}|_K \equiv 1$ and $\widetilde{\rho}|_F \equiv 0$. We also require that $0 \leq \widetilde{\rho} \leq 1$. We thus find the sequences of functions $\{\rho_{1,i}\}$ and $\{\pi_{1,i}\}_{i=1}^{\infty}$ which satisfy

$$\widetilde{\rho}_{1,i} = \begin{cases} 1 & \text{on } C'_{i,j'_i}, \\ 0 & \text{on } F'_{i,j'_i}, \end{cases}$$

where $F'_{i,j'_i} = (V'_{i,j'_i})^c$, and, in addition,

$$\widetilde{\pi}_{1,i} = \begin{cases} 1 & \text{on } D'_i, \\ 0 & \text{on } (W'_i)^c. \end{cases}$$

Similarly, we will need the sequences $\{\rho_{0,i}\}$ and $\{\pi_{0,i}\}_{i=1}^{t}$ satisfying

$$\widetilde{\rho}_{0,i} = \begin{cases} 1 & \text{on } C_{i,j_i}, \\ 0 & \text{on } F_{i,j_i}, \end{cases}$$

where $F_{i,j_i} = (V_{i,j_i})^c$, and, in addition, let

$$\widetilde{\pi}_{0,i} = \begin{cases} 1 & \text{on } D_i, \\ 0 & \text{on } (W_i)^c. \end{cases}$$

We begin with $\widetilde{\rho}_{0,1}$, which is identically 1 on C_{1,j_1}, a set which includes the neighborhood along the Bessel ray $\mathscr{H}_\rho$ containing the origin, i.e.,

$$\big(0, \big[0, (M_1 + N_1)/2\big]\big) \subset \mathscr{H}_\rho.$$

Next, we extend to a function whose transform is identically 1 on $C_{1,j_1} \cup C'_{1,j'_1}$ using the construction $\rho_1 = \rho_{1,j} + \rho_{0,j} - \rho_{1,j} * \rho_{0,j}$. Then

$$\widetilde{\rho}^1 = \widetilde{\rho}_{1,1} + \widetilde{\rho}_{0,1} - \widetilde{\rho}_{1,1}\widetilde{\rho}_{0,1} = \begin{cases} 1 & \text{on } C_{1,j_1} \cup C'_{1,j'_1}, \\ 0 & \text{on } F_{1,j_1} \cap F'_{1,j'_1}. \end{cases}$$

The above procedure demonstrates how "local identities" for the sets C_{1,j_1} and C'_{1,j'_1} are joined to form a "local identity" on their union. Note also that these cover the Bessel zeros, as well as an infinite number of the Laguerre zeros. Thus only a finite number of Laguerre zeros are left, and these can be covered by neighborhoods $\{D_1, \ldots, D_{s_1}\}$ and $\{D'_1, \ldots, D'_{t_1}\}$.

Next expand to form a "local identity" for K_1 by adjoining the "local identities" $\pi_{0,1},\ldots,\pi_{0,s_1}$ of $D_1,\ldots,D_{s_1}$ and $\pi_{1,1},\ldots,\pi_{1,t_1}$ of $D'_1,\ldots,D'_{t_1}$, respectively. Then $\pi_0^1=\sum_{i=1}^{s_1}\pi_{0,i}$ is an identity for $D^1=\cup_{i=1}^{s_1}D_i$, while $\pi_1^1=\sum_{i=1}^{t_1}\pi_{1,i}$ is an identity for $D^{1'}=\cup_{i=1}^{t_1}D'_i$. First, let $\pi^1=\pi_0^1+\pi_1^1-\pi_0^1*\pi_1^1$, and note that this will satisfy

$$\widetilde{\pi}^1=\widetilde{\pi}_0^1+\widetilde{\pi}_1^1-\widetilde{\pi}_0^1\widetilde{\pi}_1^1=\begin{cases}1 & \text{on } D^1\cup D^{1'},\\ 0 & \text{on } (W^1)^c\cap(W^{1'})^c,\end{cases}$$

where $(W^1)^c=\cap_{i=1}^{s_1}(W_i)^c$ and $(W^{1'})^c=\cap_{i=1}^{t_1}(W'_i)^c$. Noting that

$$K_1\subset\left(C_{1,j_1}\cup C'_{1,j'_1}\right)\cup\left(D^1\cup D^{1'}\right),$$

we then expand the "identity" to K_1 by forming $\rho^{1*}=\rho^1+\pi^1-\rho^1*\pi^1$. The Gelfand transform of ρ^{1*} will satisfy

$$\widetilde{\rho}^{1*}=\widetilde{\rho}^1+\widetilde{\pi}^1-\widetilde{\rho}^1\widetilde{\pi}^1=\begin{cases}1 & \text{on } (C_{1,j_1}\cup C'_{1,j'_1})\cup(D^1\cup D^{1'}),\\ 0 & \text{on } (F_{1,j_1}\cap F'_{1,j'_1})\cap((W^1)^c\cap(W^{1'})^c),\end{cases}$$

thus forming an "identity" for K_1. Finally note that $\rho_1^{1*}=\rho_{1,1}+\pi_1^1$ is an "identity" on $C'_{1,j'_1}\cup D^{1'}$, a neighborhood of all the zeros of $\widetilde{S}_r$ in K_1, that is also separated from the zeros of $\widetilde{T}_r$. We can form $\rho_2^{1*}=\rho^{1*}-\rho_1^{1*}$; note this function will similarly give an "identity" on $A^1=(C_{1,j_1}\cup D^1)\setminus(V'_{1,j'_1}\cup D^{1'})$, a neighborhood of all zeros of $\widetilde{T}_r$ in K_1, also separated from the zeros of $\widetilde{S}_r$.

This construction is next extended to form "identities" for the K_j in the compact exhaustion of $\mathscr{H}$. Form neighborhoods $C_m=\cup_{i=1}^m C_{i,j_i}$ and $C'_m=\cup_{i=1}^m C'_{i,j'_i}$ for the zeros of $\widetilde{T}_r$ and $\widetilde{S}_r$, respectively, inside of K_m and near the Bessel ray $\mathscr{H}_\rho$. Note that "identities" for C_m and C'_m are given by $\rho_0^m=\sum_{i=1}^m\rho_{0,i}$ and $\rho_1^m=\sum_{i=1}^m\rho_{1,i}$. Letting $\rho^m=\rho_0^m+\rho_1^m-\rho_0^m*\rho_1^m$, we have that

$$\widetilde{\rho}^m=\widetilde{\rho}_0^m+\widetilde{\rho}_1^m-\widetilde{\rho}_0^m\widetilde{\rho}_1^m=\begin{cases}1 & \text{on } C_m\cup C_{m'},\\ 0 & \text{on } F^m\cap F^{m'},\end{cases}$$

where $F^m=\cap_{i=1}^m F_{i,j_i}$ and $F^{m'}=\cap_{i=1}^m F'_{i,j'_i}$. Then let $D^m=\cup_{i=1}^{s_m}D_i$ and $D^{m'}=\cup_{i=1}^{t_m}D'_i$. We also define $(W^m)^c=\cap_{i=1}^{s_m}(W_i)^c$ and $(W^{m'})^c=\cap_{i=1}^{t_m}(W'_i)^c$. Expanding to pick up the remaining Laguerre zeros in K_m, we let $\pi_0^m=\sum_{i=1}^{s_m}\pi_{0,i}$, $\pi_1^m=\sum_{i=1}^{t_m}\pi_{1,i}$, and

$$\pi^m=\pi_0^m+\pi_1^m-\pi_0^m*\pi_1^m.$$

Thus,

$$\widetilde{\pi}^m=\widetilde{\pi}_0^m+\widetilde{\pi}_1^m-\widetilde{\pi}_0^m\widetilde{\pi}_1^m=\begin{cases}1 & \text{on } D^m\cup D^{m'},\\ 0 & \text{on } (W^m)^c\cap(W^{m'})^c.\end{cases}$$

Finally, we can expand to all of K_m by forming $\rho^{m*} = \rho^m + \pi^m - \rho^m * \pi^m$, where

$$\widetilde{\rho}^{m*} = \widetilde{\rho}^m + \widetilde{\pi}^m - \widetilde{\rho}^m \widetilde{\pi}^m = \begin{cases} 1 & \text{on } C_m \cup C'_m \cup (D^m \cup D^{m'}), \\ 0 & \text{on } (F^m \cap F^{m'}) \cap ((W^m)^c \cap (W^{m'})^c). \end{cases}$$

Noting that $K_m \subset (C_m \cup C'_m) \cup (D^m \cup D^{m'})$, we see that ρ^{m*} forms an "identity" for K_m. As above, we split ρ^{m*} into two "identities" ρ_1^{m*} and ρ_2^{m*} which are "identities" on neighborhoods of the zeros of $\widetilde{S}_r$ and $\widetilde{T}_r$, respectively. First, $\rho_1^{m*} = \rho_1^m + \pi_1^m$ is an "identity" on $C^{m'} \cup D^{m'}$, a neighborhood of all the zeros of $\widetilde{S}_r$ in K_m, that is also separated from the zeros of $\widetilde{T}_r$. We can form $\rho_2^{m*} = \rho^{m*} - \rho_1^{m*}$ and note that this function will similarly give an "identity" on $A^m = (C^m \cup D^m) \setminus (V^{m'} \cup D^{m'})$, a neighborhood of all zeros of $\widetilde{T}_r$ in K_m, also separated from the zeros of $\widetilde{S}_r$.

We next form the deconvolving sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$. Here we will rely on the separation of the zeros of $\widetilde{S}_r$ from those of $\widetilde{T}_r$, and we utilize the neighborhoods and "local identities" formed above in the process of inversion. In particular, $\widetilde{S}_r$ does not vanish on $\bar{V}' = (\cup_{i=1}^{\infty} V'_{i,j'_i}) \cup (\cup_{i=1}^{\infty} D'_i)$, and similarly $\widetilde{T}_r$ does not vanish on $\bar{V} = (\cup_{i=1}^{\infty} V_{i,j_i}) \cup (\cup_{i=1}^{\infty} D_i)$. We first invert these on the set K_j, and for this purpose, we form $\bar{V}'_j = \bar{V}' \cap K_j$ and $\bar{V}_j = \bar{V} \cap K_j$. Let $M_{j,s} = \min_{x \in \bar{V}'_j} |\widetilde{S}_r(x)|$ and $M_{j,t} = \min_{x \in \bar{V}_j} |\widetilde{T}_r(x)|$. Then let $M_j = \min\{M_{j,s}, M_{j,t}\}$. There exists $\phi_j \in C^\infty$ such that $\phi_j(t) = \frac{1}{t}$ for $|t| \geq M_j$, while $\phi(t) = 0$ for $|t| \leq M_j/2$. Then form $\phi_j \circ S_r$ and $\phi_j \circ T_r$, and we rely on Lemma 6.4.4 of [5] to describe the Gelfand transforms $(\phi_j \circ S_r)\widetilde{}$ and $(\phi_j \circ T_r)\widetilde{}$. Notice that these invert $\widetilde{S}_r$ on the set $\bar{V}'_j$ and $\widetilde{T}_r$ on the set $\bar{V}_j$, away from their zero sets, as follows:

$$(\phi_j \circ S_r)\widetilde{}\,|_{\bar{V}'_j} = \phi_j(\widetilde{S}_r)|_{\bar{V}'_j} = 1/\widetilde{S}_r,$$

and

$$(\phi_j \circ T_r)\widetilde{}\,|_{\bar{V}_j} = \phi_j(\widetilde{T}_r)|_{\bar{V}_j} = 1/\widetilde{T}_r.$$

The deconvolving sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ are now formed by $\nu_{1,j} = \rho_{1,j} * (\phi_j \circ S_r)$ and $\nu_{2,j} = \rho_{2,j} * (\phi_j \circ T_r)$. We claim that these sequences satisfy the properties claimed in the theorem. Considering $\{S_r * \nu_{1,j} + T_r * \nu_{2,j}\}$, we form the Gelfand transforms

$$\begin{aligned}(S_r * \nu_{1,j} + T_r * \nu_{2,j})\widetilde{} &= \widetilde{S}_r \widetilde{\nu}_{1,j} + \widetilde{T}_r \widetilde{\nu}_{2,j} \\ &= \widetilde{S}_r \phi_j(\widetilde{S}_r) \widetilde{\rho}_{1,j} + \widetilde{T}_r \phi_j(\widetilde{T}_r) \widetilde{\rho}_{2,j}.\end{aligned}$$

Since $\phi_j(\widetilde{S}_r)|_{\mathrm{supp}(\widetilde{\rho}_{1,j}) \cap K_j} \equiv 1/\widetilde{S}_r$ and $\phi(\widetilde{T}_r)|_{\mathrm{supp}(\widetilde{\rho}_{2,j}) \cap K_j} \equiv 1/\widetilde{T}_r$, we have

$$\begin{aligned}(S_r * \nu_{1,j} + T_r * \nu_{2,j})\widetilde{}\,|_{K_j} &= \widetilde{\rho}_{1,j}|_{K_j} \cdot \widetilde{S}_r \cdot (1/\widetilde{S}_r) + \widetilde{\rho}_{2,j}|_{K_j} \cdot \widetilde{T}_r \cdot (1/\widetilde{T}_r) \\ &= \widetilde{\rho}_{1,j}|_{K_j} + \widetilde{\rho}_{2,j}|_{K_j}.\end{aligned}$$

Thus by design of $\{\rho_{1,j}\}$ and $\{\rho_{2,j}\}$, we have that

$$(S_r * \nu_{1,j} + T_r * \nu_{2,j})^{\sim}|_{K_j} = (\widetilde{\rho}_{1,j} + \widetilde{\rho}_{2,j})|_{K_j} \equiv 1$$

and

$$(S_r * \nu_{1,j} + T_r * \nu_{2,j})^{\sim}|_{(U_j)^c} = (\widetilde{\rho}_{1,j} + \widetilde{\rho}_{2,j})|_{(U_j)^c} \equiv 0.$$

Thus the claim of the theorem has been verified. □

This case yields a nice representative example on which we have demonstrated the more general method. Although there was a uniform separation of zeros of $\widetilde{S}_r$ and $\widetilde{T}_r$ in the above case, more generally, these zero sets may coalesce. In the case of two balls of appropriate radii considered in the next example, the zeros of $\widetilde{T}_{r_1}$ and $\widetilde{T}_{r_2}$ will coalesce. However, we will be able to extend the above method by taking enough care with the neighborhoods around these coalescing zeros.

3.2 Two Balls of Appropriate Radii

We now consider the Pompeiu transform defined in terms of the integral averages

$$\int_{|\mathbf{z}|<r_1} L_{\mathbf{g}} f(\mathbf{z}, 0)\, d\mu_{r_1}(\mathbf{z}) \quad \text{for all } \mathbf{g} \in \mathbf{H}_n$$

and

$$\int_{|\mathbf{z}|<r_2} L_{\mathbf{g}} f(\mathbf{z}, 0)\, d\mu_{r_2}(\mathbf{z}) \quad \text{for all } \mathbf{g} \in \mathbf{H}_n.$$

These may also be written as convolutions $f * T_1$ and $f * T_2$, where $\langle \phi, T_1\rangle = \int_{|\mathbf{z}|<r_1} \phi(\mathbf{z}, 0)\, d\mu_r(\mathbf{z})$ and $\langle \phi, T_2\rangle = \int_{|\mathbf{z}|<r_2} \phi(\mathbf{z}, 0)\, d\mu_r(\mathbf{z})$. We consider the case where r_1 and r_2 are such that the transform is injective. We claim the following.

Theorem 2 *We assume that r_1 and r_2 satisfy the conditions*

1. $(\frac{r_1}{r_2}) \notin \mathscr{Q}(J_n) = \{\frac{\gamma x}{\gamma y} : J_n(x) = J_n(y) = 0, \gamma \in \mathbf{R}^*\}$,
2. $(\frac{r_1}{r_2})^2 \notin \mathscr{Q}(\Psi_k^{(n-1)}) = \{\frac{\gamma x}{\gamma y} : \Psi_k^{(n-1)}(x) = \Psi_k^{(n-1)}(y) = 0, \gamma \in \mathbf{R}^*\}$ *for all* $k \in \mathbf{Z}_+$.

Then $\widetilde{T}_1$ and $\widetilde{T}_2$ do not have any common zeros. Consider the sequence of compact sets $\{K_j\} \subset \mathscr{H}$ given below which forms a compact exhaustion of the Heisenberg fan $\mathscr{H}$. There exist sequences of functions $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ with the property that

$$\widetilde{T}_1\widetilde{\nu}_{1,j} + \widetilde{T}_2\widetilde{\nu}_{2,j} \equiv 1 \quad \text{on } K_j$$

and

$$\widetilde{T}_1\widetilde{\nu}_{1,j} + \widetilde{T}_2\widetilde{\nu}_{2,j} \equiv 0 \quad \text{on } V_j^c,$$

where each V_j is an open set such that $K_j \subset V_j \subset K_{j+1}$.

Recall that $\Psi_k^{(n-1)}(x)$ was defined above by

$$\Psi_k^{(n-1)}(x) = \int_0^x e^{-t/2} L_k^{(n-1)}(t) t^{n-1}\, dt.$$

In this case the $\{K_j\}$ and $\{V_j\}$ will be given in the proof of the theorem. It is more convenient to make their definition after the zeros have been grouped appropriately.

Proof We first recall the values of the Gelfand transforms of T_1 and T_2:

$$\widetilde{T}_1(\lambda, k) = c \int_0^{r_1} e^{-2\pi|\lambda|s^2} s^{2n-1} L_k^{(n-1)}\big(4\pi|\lambda|s^2\big)\, ds = c' \Psi_k^{(n-1)}\big(4\pi|\lambda|r_1^2\big),$$
$$\widetilde{T}_1(0, \rho) = c\frac{J_n(\rho r_1)}{(\rho r_1)^n},$$

and

$$\widetilde{T}_2(\lambda, k) = c \int_0^{r_2} e^{-2\pi|\lambda|s^2} s^{2n-1} L_k^{(n-1)}\big(4\pi|\lambda|s^2\big)\, ds = c' \Psi_k^{(n-1)}\big(4\pi|\lambda|r_2^2\big),$$
$$\widetilde{T}_2(0, \rho) = c\frac{J_n(\rho r_2)}{(\rho r_2)^n}.$$

The procedure is close to that of the previous theorem; however there are some additional complications we will address. The interaction between zero sets of $\widetilde{T}_1$ and $\widetilde{T}_2$ plays a larger role. In general, the zeros will not alternate, and furthermore there is not a uniform separation between the zeros of $\widetilde{T}_1$ and the zeros of $\widetilde{T}_2$. Due to the coalescing zeros of these sets, the values of $\widetilde{T}_1$ can be very small near the zeros of $\widetilde{T}_2$. However, by being careful with the size of the sets enclosing the zeros, we can still construct the appropriate sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$. Nevertheless, the size of $\widetilde{T}_1$ near the zeros of $\widetilde{T}_2$ can still be an issue in the larger problem of deconvolution, as addressed in Sect. 5.

The procedure is just like the above case, and we begin by forming appropriate neighborhoods of the zeros of $\widetilde{T}_1$ and $\widetilde{T}_2$ along the Bessel ray. Letting $V_1 = \{\text{zeros of } \widetilde{T}_1\}$ and $V_2 = \{\text{zeros of } \widetilde{T}_2\}$, we have the Bessel zeros $U_1 = V_1 \cap \mathscr{H}_\rho$ and $U_2 = V_2 \cap \mathscr{H}_\rho$. Setting $U_1 = \{M_1, M_2, \ldots\}$ and $U_2 = \{N_1, N_2, \ldots\}$, we then let $U = U_1 \cup U_2 = \{Z_1, Z_2, \ldots\}$, where, in each case, these are listed in increasing order. We next form sequences of neighborhoods $\{C_k\}$ and $\{C_k'\}$ such that $\cup_{k=1}^{\infty}(C_k' \cup C_k) = \mathscr{H}_\rho$ and such that $\cup_{k=1}^{\infty} C_k'$ covers U_1, while $\cup_{k=1}^{\infty} C_k$ covers U_2. Beginning with the first zero $Z_1 = M_1 \in U_1$, we find the next Z_{j_1} equal to $N_1 \in U_2$. Then $Z_1, \ldots, Z_{j_1-1}$ are grouped as zeros of $\widetilde{T}_1$, and $C_1 = [0, \frac{Z_{j_1-1}+Z_{j_1}}{2}]$ is a neighborhood of these zeros of $\widetilde{T}_1$ that does not contain any zeros of $\widetilde{T}_2$. Since $Z_{j_1} \in U_2$ and we know that $Z_{j_1+1} = M_{j_1} \in U_1$, we also form $C_1' = [\frac{Z_{j_1-1}+Z_{j_1}}{2}, \frac{Z_{j_1}+Z_{j_1+1}}{2}]$ as a neighborhood of $N_1 \in U_2$ not containing any zeros of $\widetilde{T}_1$. Using $Z_{j_1+1} = M_{j_1}$, we then find the next Z_{j_2} that equals $N_2 \in U_2$. Then the zeros $Z_{j_1+1}, \ldots, Z_{j_2-1}$ are grouped as zeros of $\widetilde{T}_1$ not separated by any zeros of $\widetilde{T}_2$, and we may form a neighborhood $C_2 = [\frac{Z_{j_1}+Z_{j_1+1}}{2}, \frac{Z_{j_2-1}+Z_{j_2}}{2}]$. Then let $C_2' = [\frac{Z_{j_2-1}+Z_{j_2}}{2}, \frac{Z_{j_2}+Z_{j_2+1}}{2}]$ to give a neighborhood of $N_2 = Z_{j_2}$. Extending this procedure to all integers yields the desired collection of neighborhoods $\{C_k\}$ covering U_1 and $\{C_k'\}$ covering U_2 such that $\mathscr{H}_\rho = \cup_{k=1}^{\infty}(C_k \cup C_k')$. Similarly to the above proof, these neighborhoods also separate the zero sets of $\widetilde{T}_1$ and $\widetilde{T}_2$. In particular, for all j,

$$N_j \cap \left(\bigcup_{i=1}^{\infty} C_i\right) = \emptyset \quad \text{and} \quad M_j \cap \left(\bigcup_{i=1}^{\infty} C_i'\right) = \emptyset.$$

These collections of neighborhoods are also nearly disjoint, intersecting only at the endpoints, as above. This point in the discussion will also be convenient to define the sequences of sets $\{K_j\}$ and $\{V_j\}$. Let $N_i = \frac{3Z_{j_i}+Z_{j_i+1}}{4}$ and $N_i^+ = \frac{Z_{j_i}+Z_{j_i+1}}{2}$. We then let

$$K_j = \{p = (x, y) \in \mathscr{H} : x^2 + y^2 \le N_j^2\},$$

where $(x, y) = (\lambda, |\lambda|(4k + 2))$ or $(x, y) = (0, \rho^2)$, and

$$V_j = \{p = (x, y) \in \mathscr{H} : x^2 + y^2 < (N_j^+)^2\},$$

where $(x, y) = (\lambda, |\lambda|(4k + 2))$ or $(x, y) = (0, \rho^2)\}$, as above.

For an additional distance in the separation, we also form smaller neighborhoods $\{B_i\}$ and $\{B_i'\}$, satisfying $B_i \subset C_i$ and $B_i' \subset C_i'$, and also such that B_i is a neighborhood of zeros $\{Z_{j_{i-1}+1}, \dots, Z_{j_i-1}\} \in U_1$, where $j_0 = 0$, while B_i' is a neighborhood of the zeros $Z_{j_i} \in U_2$. This can be accomplished by letting $\delta_i = \frac{Z_{j_i}-Z_{j_i-1}}{2}$ and $\delta_i' = \frac{Z_{j_i+1}-Z_{j_i}}{2}$. Then form neighborhoods $B_i = [Z_{j_{i-1}+1} - \delta_{i-1}'/2, Z_{j_i-1} + \delta_i/2]$ and $B_i' = [Z_{j_i} - \delta_i/2, Z_{j_i} + \delta_i'/2]$. These neighborhoods have the desired properties, and they further guarantee that $\operatorname{dist}(U_1, B_i') = \varepsilon_i$ and $\operatorname{dist}(U_2, B_i) = \varepsilon_i'$, where $\varepsilon_i = \min(\delta_i/2, \delta_i'/2)$ and $\varepsilon_i' = \min(\delta_i/2, \delta_{i-1}'/2)$, giving a local separation of zeros. Although this local separation of zeros is not as strong as the uniform separation of zeros, we find above in Sect. 3.1, it is good enough for the purpose of forming the desired deconvolving sequences. In particular, the distance in the local separation of the zeros allows us to extend these neighborhoods beyond the Bessel ray into the Heisenberg fan, as above. But first we observe the existence of the collections of larger neighborhoods $\{V_i\}$ and $\{V_i'\}$. The V_i satisfy the properties $C_i \subset V_i$ and $V_i \cap (B_{i-1}' \cup B_i') = \emptyset$. Likewise, the V_i' satisfy $C_i' \subset V_i'$ and $V_i' \cap (B_{i-1} \cup B_i) = \emptyset$. Using the collection of neighborhoods $\{B_i\}$, $\{C_i\}$, and $\{V_i\}$ of the zeros $U_1 \in \mathscr{H}_\rho$ and the collection of neighborhoods $\{B_i'\}$, $\{C_i'\}$, and $\{V_i'\}$ of the zeros $U_2 \in \mathscr{H}_\rho$, it is possible to complete the process of forming the deconvolving sequence by a sequence of steps beginning with extension from the Bessel ray, $\mathscr{H}_\rho$. The key point in making the extension from the Bessel ray is the local separation by a set ε_i between the zeros U_1 and neighborhoods B_i', as well as between the zeros U_2 and neighborhoods B_i. We note that it is straightforward to see how the neighborhoods for this specific case could be extended to a more general case in which the zero sets are more irregularly distributed. Once the appropriate neighborhoods have been established, the remainder of the formation of the deconvolting sequences proceeds directly as described.

From here we complete the process of forming the deconvolving sequences using the same method as given in Sect. 3.1 for Theorem 1. It consists of extension of the above neighborhoods from the Bessel ray to cover an infinite number of Laguerre zeros, followed by formation of additional neighborhoods to cover the remaining Laguerre zeros and any remaining regions in the Heisenberg fan $\mathscr{H}$. From here, inversion of $\widetilde{T}_1$ and $\widetilde{T}_2$ are completed by use of "local identities" and a process of "local inversion." The local separation between the zeros U_1, U_2 and the neighborhoods B_i and B_i' described in the preceding paragraph extends to the entire Heisenberg fan $\mathscr{H}$. This is used to form the ϕ_i which locally invert $\widetilde{T}_1$ and $\widetilde{T}_2$. In addition the system of neighborhoods are used to form the sequences $\{\rho_{1,j}\}$ and $\{\rho_{2,j}\}$ with the property that $\widetilde{\rho}_{1,j} + \widetilde{\rho}_{2,j}|_{K_j} \equiv 1$. Putting these together will yield the desired deconvolving sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$.

The method for extending the neighborhoods $\{B_{i,k_i}\}$ and $\{B'_{i,k_i'}\}$, as well as $\{C_{i,k_i}\}$ and $\{C'_{i,k_i'}\}$ plus $\{V_{i,k_i}\}$ and $\{V'_{i,k_i'}\}$, is identical to that given above for Theorem 1. We outline this procedure again here. Recall the definition of $B_{i,j}$, now expanded to enclose the group of zeros $Z_{j_{i-1}+1}, \ldots, Z_{j_i-1}$:

$$B_{i,k} = \Bigg\{(x, y) \in \mathscr{H} : \left(Z_{j_{i-1}+1} - \frac{\delta'_{i-1}}{2}\right)^2 \leq x^2 + y^2 \leq \left(Z_{j_i-1} + \frac{\delta_i}{2}\right)^2 \text{ and } \left|\frac{y}{x}\right| \geq 4(k + n/2)\Bigg\},$$

while $B'_{i,k'}$ is defined by

$$B'_{i,k'} = \Bigg\{(x, y) \in \mathscr{H} : \left(Z_{j_i} - \frac{\delta_i}{2}\right)^2 \leq x^2 + y^2 \leq \left(Z_{j_i} + \frac{\delta'_i}{2}\right)^2 \text{ and } \left|\frac{y}{x}\right| \geq 4\big(k' + n/2\big)\Bigg\}.$$

In expanding the neighborhood B_i of the n_i zeros along the Bessel ray $\mathscr{H}_\rho$ to the larger neighborhood B_{i,k_i} within the Heisenberg fan $\mathscr{H}_n$, we use the form just given. First select $k = k_i \in \mathbf{Z}_+$ with the property that, for each $k \geq k_i$, exactly n_i of Laguerre zeros on the ray $\mathscr{H}_{k,\pm}$ are inside of $B_{i,k} \cap \mathscr{H}_{k,\pm}$. We also choose k_i to minimize all possible k_i satisfying the previous property and use this to form the neighborhood B_{i,k_i}. Similarly, B_i' is expanded to form $B'_{i,k_i'}$ by choosing $k_i' \in \mathbf{Z}_+$ with the property that, for each $k' \geq k_{i'}$, exactly one of the Laguerre zeros on the ray $\mathscr{H}_{k',\pm}$ is inside of $B'_{i,k_i'} \cap \mathscr{H}_{k',\pm}$. Using these same values of k_i and k_i', for each $i \in \mathbf{Z}_+$, we also expand from $\{C_i\}$ to $\{C_{i,k_i}\}$ and from $\{C_i'\}$ to $\{C'_{i,k_i'}\}$. In the same manner we expand from $\{V_i\}$ to $\{V_{i,k_i}\}$ and from $\{V_i'\}$ to $\{V'_{i,k_i'}\}$. Finally pick up remainder of Laguerre zeros and cover the rest of $\mathscr{H}$ by adding neighborhoods $\{D_i\}$ and $\{D_i'\}$

as well as neighborhoods $\{W_i\}$ and $\{W_i'\}$ that separate the previous neighborhoods, using the same method as given in the proof of Theorem 1.

Now that all the neighborhood systems are in place, the formation of the deconvolving sequences is identical to that given in Theorem 1. Sequences of "local identities" $\{\rho_{1,i}\}$ and $\{\rho_{0,i}\}$ are formed, such that

$$\widetilde{\rho}_{1,i} = \begin{cases} 1 & \text{on } C'_{i,j_i'}, \\ 0 & \text{on } F'_{i,j_i'}, \end{cases}$$

and

$$\widetilde{\rho}_{0,i} = \begin{cases} 1 & \text{on } C_{i,j_i}, \\ 0 & \text{on } F_{i,j_i}. \end{cases}$$

Similarly we form sequences of "local identities" $\{\pi_{1,i}\}$ and $\{\pi_{0,i}\}$ such that

$$\widetilde{\pi}_{1,i} = \begin{cases} 1 & \text{on } D_i', \\ 0 & \text{on } (W_i')^c, \end{cases}$$

and

$$\widetilde{\pi}_{0,i} = \begin{cases} 1 & \text{on } D_i, \\ 0 & \text{on } (W_i)^c. \end{cases}$$

These are put together in the same manner as in the proof of Theorem 1, forming ρ_m^* which is an "identity" for K_m. This further splits into the two "identities" ρ_1^{m*} and ρ_2^{m*} such that $\rho_m^* = \rho_1^{m*} + \rho_2^{m*}$, while ρ_1^{m*} and ρ_2^{m*} are "identities" on the neighborhoods of the zeros of $\widetilde{T}_1$ and $\widetilde{T}_2$, respectively.

All that remains is the "local inversion" of $\widetilde{T}_1$ and $\widetilde{T}_2$ away from their zeros, found by applying the inverses $\phi_j \in C^\infty$ satisfying $\phi_j(t) = 1/t$ for $|t| \geq M_j$ while $\phi_j(t) = 0$ for $|t| \leq M_j/2$, where the M_j are determined according to the size of $\widetilde{T}_1$ and $\widetilde{T}_2$ on the appropriate neighborhood systems within K_j. Note that $\widetilde{T}_1$ does not vanish on $\bar{V}' = (\cup_{i=1}^\infty V'_{i,j_i'}) \cup (\cup_{i=1}^\infty D_i')$, and similarly $\widetilde{T}_2$ does not vanish on $\bar{V} = (\cup_{i=1}^\infty V_{i,j_i}) \cup (\cup_{i=1}^\infty D_i')$. For inversion on K_j, we divide into neighborhoods of the zero sets $\bar{V}_j' = K_j \cap \bar{V}'$ and $\bar{V}_j = K_j \cap \bar{V}$. Then $M_j = \min\{M_{j,1}, M_{j,2}\}$, where $M_{j,1} = \min_{x \in \bar{V}_j'} |\widetilde{T}_1(x)|$ and $M_{j,2} = \min_{x \in \bar{V}_j} |\widetilde{T}_2(x)|$. Note that $(\phi_j \circ T_1)$ and $(\phi_j \circ T_2)$ invert $\widetilde{T}_1$ and $\widetilde{T}_2$ on $\bar{V}_j'$ and $\bar{V}_j$, respectively, away from their zero sets, as follows:

$$(\phi_j \circ T_1)^{\sim}|_{\bar{V}_j'} = \phi_j(\widetilde{T}_1)|_{\bar{V}_j'} = 1/\widetilde{T}_1,$$

and

$$(\phi_j \circ T_2)^{\sim}|_{\bar{V}_j} = \phi_j(\widetilde{T}_2)|_{\bar{V}_j} = 1/\widetilde{T}_2.$$

Note that the main difference between this case and that of Theorem 1 is the existence of coalescing zeros in this case. To deal with the locations these zeros coalesce, we have more variation in the width of the neighborhoods B'_i, extending to the neighborhoods B'_{i,j'_i}, which separate the zeros of $\widetilde{T}_2$ from those of $\widetilde{T}_1$. The neighborhood systems hold up so that the deconvolving sequences can be formed. However, the rate at which the zeros coalesce can effect the rate at which M_j approaches 0 as well as the corresponding rate at which the $(\phi_j \circ T_1)^{\sim}$ grows near the zero set of $\widetilde{T}_2$. These issues and other related issues will be considered in Sect. 5 dealing with issues of convergence related to the sequences of deconvolvers.

The deconvolving sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ are now formed by defining

$$\nu_{1,j} = \rho_{1,j} * (\phi_j \circ T_1) \text{ and } \nu_{2,j} = \rho_{2,j} * (\phi_j \circ T_2).$$

As previously, this gives the transforms

$$\begin{aligned}(T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim}|_{K_j} &= \widetilde{\rho}_{1,j}|_{K_j} \cdot \widetilde{T}_1 \cdot 1/\widetilde{T}_1 + \widetilde{\rho}_{2,j}|_{K_j} \cdot \widetilde{T}_2 \cdot 1/\widetilde{T}_2 \\ &= \widetilde{\rho}_{1,j}|_{K_j} + \widetilde{\rho}_{2,j}|_{K_j} \equiv 1.\end{aligned}$$

We can also easily see that $(T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim}|_{(V_j)^c} = 0$. This verifies the claims in the theorem, and the proof is complete. □

As noted above, the difference in these cases of the sphere and ball in Sect. 3.1 and the two balls of appropriate radii in Sect. 3.2 is the issue of distance of separation of the zeros for $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$, which carries over to the rates of growth of the sequences $\{\widetilde{\nu}_{1,j}\}$ and $\{\widetilde{\nu}_{2,j}\}$ near these zero sets as j becomes infinite. This issue and related issues of convergence will be revisited in Sect. 5, where we also address how the deconvolving sequences $\{\widetilde{\nu}_{1,j}\}$ and $\{\widetilde{\nu}_{2,j}\}$ may be applied to recover the function f.

4 Using Weyl Calculus and the Group Fourier Transform

Observe how in the proof of the results Theorem 1 and Theorem 2 above, the construction of the "local inverses" ρ_1 and ρ_2 was based on analysis of the Bessel zeros, sets U_1 and U_2. The neighborhoods of each of these zeros naturally extended to Laguerre zeros along an infinite number of Laguerre rays, based on the subspace topology. The remaining Laguerre zeros are finite along each of the Laguerre rays and are also locally finite. Since these zeros are easily incorporated using a finite number of appropriate neighborhoods, they do not affect the process. Thus, the construction of the "local inverses" in Sect. 3 above is based upon the construction for the Bessel ray. It appears that the distribution of the zeros along the Bessel ray determines the potential for deconvolution, provided that there are no common Laguerre zeros. In the context of the Weyl calculus, the Laguerre spectrum is a quantization of the Bessel part, and the behavior of the Laguerre part, in the limit as $k \to \infty$ and $\lambda \to 0$, determines the behavior of the Bessel part.

Recall the suggestion [10] that use of the Weyl calculus can establish connections between Pompeiu type results in Euclidean space and the Heisenberg group. Furthermore in [11] there was also the suggestion that the cases of the Pompeiu problem for the Heisenberg group and Euclidean space are actually very close, and in particular deconvolution should extend to the Heisenberg case. In this section we investigate specific points regarding how the deconvolution results of Theorem 1 and Theorem 2 can be viewed primarily from the perspective of the "Euclidean part" of the zero set, along the Bessel ray.

In this section we discuss the relations between these two avenues of investigation. The Weyl representation reduces to this central Bessel ray for the representations $\pi_{(\xi,\eta)}$. The issue of extending the deconvolution process from the Bessel ray $\mathscr{H}_\rho$ to the Laguerre rays $\cup_{k\in\mathbf{Z}_+}\mathscr{H}_{k,\pm}$ has parallels to the passage between the operator-valued Weyl representation $\pi_{\pm\lambda}$ on the entire spectrum and the Euclidean part $\pi_{(\xi,\eta)}$ on the central Bessel ray. We focus on the operator-valued Weyl representations $\pi_{\pm\lambda}$ and $\pi_{(\xi,\eta)}$ of the distributions T_1 and T_2, representing the sets over which the average is taken. Through use of the Weyl calculus we relate $\widetilde{T}_j(0;\rho)$, the Bessel part of the Gelfand transform, to $\pi_{(\xi,\eta)}T_j$, the Euclidean part of the Weyl representation. This corresponds to analysis of the Bessel zeros U_j for $j = 1, 2$, mentioned above. In fact the conclusion of the existence of deconvolving sequences in Theorem 1 and Theorem 2 can be based on two points, the behavior of the Bessel zeros and nonoverlapping of the zeros of the Laguerre part.

The Weyl calculus for the Pompeiu problem on $\mathbf{H}^n$ allows a unification of both Laguerre and Bessel parts of the zero sets U_1 and U_2 into the kernels of two operator-valued functions. Furthermore, in the case where $\lambda \to 0$, this carries over to the representation $\pi_{(\xi,\eta)}$, the Euclidean transform on $\mathbf{C}^n$. In this limit, the operator-valued functions become identical to functions used in the analysis of the Pompeiu problem on $\mathbf{C}^n$. This is a nice bridge from Heisenberg to Euclidean and from Euclidean to Heisenberg, and we will utilize it. In the case of Theorem 1, there are no conditions needed for the common radius of the sphere and ball. In fact, these two sets were selected to provide a representative example for the nice cases where there exists a uniform separation between the zero sets. As such, these sets are in a category comparable to the moment type results of [6, 9, 18] and do not require sets of exceptional radii. This corresponds to the fact that

$$\ker\Big\{j_{n-1}\Big(r\sqrt{|\lambda|\big(P^2+Q^2\big)}\Big)\Big\}\cap\ker\Big\{j_n\Big(r\sqrt{|\lambda|\big(P^2+Q^2\big)}\Big)\Big\}=\{0\},$$

a fact which is true for any radius r. However, in many cases exceptional radii are required, such as Theorem 2, corresponding to results such as those in [1, 5, 10]. In the context of the two-radius theorem for the Pompeiu problem on $\mathbf{H}^n$, the conditions for existence of the deconvolving sequences of Theorem 2 are expressed as follows.

Theorem 3 *Let r_1 and r_2 be two radii satisfying the condition*

$$\ker\Big\{j_n\Big(r_1\sqrt{|\lambda|\big(P^2+Q^2\big)}\Big)\Big\}\cap\ker\Big\{j_n\Big(r_2\sqrt{|\lambda|\big(P^2+Q^2\big)}\Big)\Big\}=\{0\},$$

so that $\widetilde{T}_1$ *and* $\widetilde{T}_2$ *do not have any common zeros. Then there exist sequences of functions* $\{\nu_{1,j}\}$ *and* $\{\nu_{2,j}\}$ *with the property that*

$$\widetilde{T}_1\widetilde{\nu}_{1,j} + \widetilde{T}_2\widetilde{\nu}_{2,j} \equiv 1 \quad \textit{on } K_j$$

and

$$\widetilde{T}_1\widetilde{\nu}_{1,j} + \widetilde{T}_2\widetilde{\nu}_{2,j} \equiv 0 \quad \textit{on } V_j^c,$$

where each V_j *is an open set such that* $K_j \subset V_j \subset K_{j+1}$.

Note that in the case of the n-ball and $(n-1)$-sphere of Theorem 1, the condition

$$\ker\Big\{ j_n\Big(r\sqrt{|\lambda|\big(P^2+Q^2\big)}\Big)\Big\} \cap \ker\Big\{ j_{n-1}\Big(r\sqrt{|\lambda|\big(P^2+Q^2\big)}\Big)\Big\} = \{0\}$$

is automatic due to results of zeros of Bessel functions of consecutive indices and does not relate to the radius r. Similarly there is no condition for exceptional radii needed in Theorem 1.

Here we describe the transforms associated to the measures used in Theorem 1 and Theorem 2 using the operator-valued Fourier transform on $\mathbf{H}^n$. As described in Sect. 2, the group Fourier transform for the measure μ with respect to the representation $\pi_{\pm\lambda}$ can be determined from the standard Euclidean Fourier transform on $\mathbf{R}^{2n+1}$, $\mathscr{F}_{2n+1}(\mu)(\mathbf{x},\mathbf{y},t)$, by substituting the operators P and Q to give

$$\pi_{\pm\lambda}(\mu_r) = \mathscr{F}_{2n+1}(\mu_r)\big(\mp\lambda^{1/2}P, -\lambda^{1/2}, \mp\lambda\big) \quad \text{for } \lambda \in \mathbf{R}_+ \setminus \{0\}.$$

The case of the one-dimensional measures $\pi_{(\xi,\eta)}$, corresponding directly to the Euclidean case, can be attained as a limit as $\lambda \to 0$ or by substitution of (ξ,η) in the form

$$\pi_{(\xi,\eta)}(\mu_r) = \mathscr{F}_{2n+1}(\mu_r)(-\xi,-\eta,0) \quad \text{for } (\xi,\eta) \in \mathbf{R}^n \times \mathbf{R}^n.$$

Since the issue of deconvolution on the associated Euclidean spaces is already settled, we may look here first. We first address the issue of common zeros required for injectivity of the Pompeiu transform and inherent to Hörmander's strongly coprime condition. In the case of the sphere and ball of Theorem 1, we have

$$\pi_{(\xi,\eta)}(S_r) = c\,\frac{J_{n-1}(r\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(r\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^{n-1}}$$

and

$$\pi_{(\xi,\eta)}(T_r) = c\frac{J_n(r\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(r\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^n}$$
$$= c'\int_0^r \frac{J_{n-1}(\rho\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(\rho\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^{n-1}}\rho^{2n-1}\,d\rho.$$

These have no common zeros as a consequence of the well-known result for Bessel functions of separate integer indices. In fact, there is a uniform separation among the zeros, which are interlaced. Note that in moving to the infinite-dimensional representations and using the series [15]

$$j_{n-1}(ts) = \frac{J_{n-1}(ts)}{(ts)^{n-1}} = 2\sum_{j=0}^{\infty}(-1)^j e^{-t^2/2}L_j^{(n-1)}\big(t^2\big)e^{-s^2/2}L_j^{(n-1)}\big(s^2\big) \tag{5}$$

to express the Laguerre part of the spectrum for these operators, we may write $\pi_{\pm\lambda}(\mu_r)$ in the form

$$\pi_{\pm\lambda}(S_r) = c\sum_{j=0}^{\infty}F_\lambda(2j+1)(-1)^j e^{-|\lambda|(P^2+Q^2)}L_j^{(n-1)}\big(2|\lambda|\big(P^2+Q^2\big)\big),$$

which can be represented using $H = P^2+Q^2$, the harmonic oscillator Hamiltonian. Thus we have

$$\pi_{\pm\lambda}(S_r) = c\sum_{j=0}^{\infty}F_\lambda(2j+1)(-1)^j e^{-|\lambda|H}L_j^{(n-1)}\big(2|\lambda|H\big).$$

Noting that H has the Hermite functions E_α as eigenfunctions with eigenvalues $2\alpha+1$, we see that E_j is also an eigenfunction for the operator $e^{-H}L_j^{(n-1)}(2H)$. This implies that each $(\lambda,k)\in\mathbf{R}^*\times\mathbf{Z}_+$ such that $L_k^{(n-1)}(|\lambda|r^2/2)=0$ yields a function E_k in the kernel of $\pi_{\pm\lambda}(S_r)$. The kernel of the operator-valued function $j_{n-1}(r\sqrt{|\lambda|(P^2+Q^2)})$ is then given by $(\xi,\eta)\in\mathbf{R}^n\times\mathbf{R}^n$ such that

$$J_{n-1}\big(r\sqrt{|\xi|^2+|\eta|^2}\big)=0$$

and $(\lambda,k)\in\mathbf{R}^*\times\mathbf{Z}_+$ such that $L_k^{(n-1)}(|\lambda|r^2/2)=0$. We observed above that the zeros of J_n and J_{n-1} have a uniform separation, due to the indices. To consider the other part of the common kernel, we consider also $\pi_{\pm\lambda}(T_r)$, similarly computed through the use of (5) to be

$$\pi_{\pm\lambda}(T_r) = c\sum_{j=0}^{\infty}(-1)^j\left(\int_0^r e^{-|\lambda|\rho^2/4}L_j^{(n-1)}\big(|\lambda|\rho^2/2\big)\rho^{2n-1}\,d\rho\right)$$
$$\times e^{-(P^2+Q^2)}L_j^{(n-1)}\big(2\big(P^2+Q^2\big)\big),$$

which similarly has the Hermite functions E_α as eigenfunctions. In this case the operator $\pi_{\pm\lambda}(T_r)$ has a the eigenfunction E_k in its kernel for any $(\lambda, k) \in \mathbf{R}^* \times \mathbf{Z}_+$ such that

$$\int_0^r e^{-|\lambda|\rho^2/4} L_k^{(n-1)}\big(|\lambda|\rho^2/2\big)\rho^{2n-1}\, d\rho = 0.$$

Then recognize that $\int_0^{r^2|\lambda|/2} e^{-x/2} x^n L_j^{(n-1)}(x)\, dx$ and $e^{-r^2|\lambda|/4} L_j^{(n-1)}(|\lambda| r^2/2)$ will also have a uniform separation between their zero sets, as was observed previously, in Sect. 3.1. Thus in this case the entire issue of injectivity, required as a prerequisite for deconvolution, is automatic. Furthermore the uniform separation within the Bessel part of the transform extends to the whole spectrum $\mathscr{H}$. This uniform separation makes the larger problem of deconvolution easier, and this case allows for the most direct application of these methods. We will have more to say later about the role of the zero sets in the larger deconvolution problem.

The more general case is sometimes more akin to the case of the balls of separate radii r_1 and r_2, as found in Theorem 2. In this case the zero sets may coalesce, and furthermore there may be issues related to the size of the functions at the zero sets. We investigate this case by forming the operator-valued transforms

$$\begin{aligned}\pi_{(\xi,\eta)}(T_{r_1}) &= c\frac{J_n(r_1\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(r_1\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^n}\\ &= \int_0^{r_1} \frac{J_{n-1}(\rho\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(\rho\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^{n-1}}\rho^{2n-1}\, d\rho\end{aligned}$$

and

$$\begin{aligned}\pi_{(\xi,\eta)}(T_{r_2}) &= c\frac{J_n(r_2\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(r_2\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^n}\\ &= \int_0^{r_2} \frac{J_{n-1}(\rho\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})}{(\rho\sqrt{\xi_1^2+\cdots+\xi_n^2+\eta_1^2+\cdots+\eta_n^2})^{n-1}}\rho^{2n-1}\, d\rho.\end{aligned}$$

Condition 1 of Theorem 2 is equivalent to the lack of a common kernel for these two representations. For the operator-valued functions $j_n(r_i\sqrt{|\lambda|(P^2+Q^2)})$, we still need to expand to the representations $\pi_{\pm\lambda}(T_{r_i})$ using the series (5) above. As in the above case of the group Fourier transform of T_r, we compute

$$\pi_{\pm\lambda}(T_{r_i}) = c\sum_{j=0}^{\infty}(-1)^j\left(\int_0^{r_i} e^{-\frac{|\lambda|\rho^2}{4}} L_j^{(n-1)}\left(\frac{|\lambda|\rho^2}{2}\right)\rho^{2n-1}\,d\rho\right)$$
$$\times\ e^{-(P^2+Q^2)} L_j^{(n-1)}\big(2(P^2+Q^2)\big)$$
$$= c\sum_{j=0}^{\infty}(-1)^j\left(\frac{2^{n-1}}{|\lambda|^n}\Psi_j^{(n-1)}\left(\frac{|\lambda|r_i^2}{2}\right)\right)e^{-(P^2+Q^2)} L_j^{(n-1)}\big(2(P^2+Q^2)\big),$$

which similarly has the Hermite functions E_α as eigenfunctions. Thus the operator $\pi_{\pm\lambda}(T_{r_i})$ has the eigenfunction E_k in its kernel for any $(\lambda, k) \in \mathbf{R}^* \times \mathbf{Z}_+$ such that

$$\frac{2^{n-1}}{|\lambda|^n}\Psi_k^{(n-1)}\big(|\lambda|r_i^2/2\big) = \int_0^{r_i} e^{-|\lambda|\rho^2/4} L_k^{(n-1)}\big(|\lambda|\rho^2/2\big)\rho^{2n-1}\,d\rho = 0.$$

Thus we see that condition 2 of Theorem 2 corresponds to no common kernel of $\pi_{\pm\lambda}(T_{r_1})$ and $\pi_{\pm\lambda}(T_{r_2})$ for $\lambda \in \mathbf{R}^*$. Now

$$\ker\left\{j_n\left(r_1\sqrt{|\lambda|\big(P^2+Q^2\big)}\right)\right\} \cap \ker\left\{j_n\left(r_2\sqrt{|\lambda|\big(P^2+Q^2\big)}\right)\right\} = \{0\}$$

implies that conditions 1 and 2 of Theorem 2 are met, which in turn implies the existence of the sequences of deconvolvers. This completes the proof of the theorem.

Note, however, that the lack of common zeros provided by the unified condition on the operator-valued Bessel functions of Theorem 3 address only part of the larger problem of deconvolution. This point can be seen from the analogous results for Euclidean spaces where the issue for deconvolution of the Pompeiu problem for two balls of appropriate radii divides into two separate cases, based on arithmetic conditions associated to the radii. As observed in [1], the separate conditions for the Bessel and Laguerre parts of the spectrum are unified by addressing the common kernel of the operator-valued transforms. However, at the level of the transform of the associated Euclidean space, the results of [4, 7, 8] demonstrate that for the problem of deconvolution and application of Hörmander's strongly coprime condition, it is necessary to divide into two cases related to the separation of these zeros and how rapidly they can coalesce. The condition dividing these cases is how well the quotient of radii $\frac{r_1}{r_2}$ can be approximated by quotients of zeros p/q of the Bessel function B_n found in these transforms. We require the following definition for N-well approximation. First, let E_n be an infinite set with elements ordered by

$$E_n = \{\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots\}, \quad \text{where } \lambda_j < \lambda_{j+1}.$$

Definition 1 For $N > 0$, a positive number α is called N-well approximated by ratios of E_n if, for every $\ell > 1$, there exist indices j, k such that

$$|\alpha - \lambda_k/\lambda_j| \le 1/\big(\ell j^N\big).$$

If for every $N > 0$, the number α is not N-well approximated by ratios of E_n, then α is called poorly approximated by ratios of E_n. We mention a result of Berenstein and Gay [3, Proposition 6] demonstrating that when $\frac{r_1}{r_2}$ is not N-well approximated by ratios of E_n, then $\widehat{\mu}_{r_2}$ satisfies the estimate

$$\left|\widehat{\mu}_{r_2}(\lambda_k)\right| \geq C/k^{N+(n-1)/2}$$

at λ_k, the zeros of $\widehat{\mu}_{r_1}$. From this result it is easy to see that if radii r_1 and r_2 are poorly approximated by zeros of J_n, it can be shown that Hörmander's strongly coprime condition holds, implying the existence of compactly supported deconvolvers. When the radii r_1 and r_2 are N-well approximated, Hörmander's strongly coprime condition is not met, implying that any deconvolvers cannot be compactly supported. Thus the distribution of these zero sets is integrally related to the issues of Hörmander's strongly coprime condition and the issue of deconvolution.

Note also that the procedure we have used in the above two results strongly suggests that the Bessel zeros are where the issue lies. When we can describe these zeros and find suitable neighborhoods to separate them, then it appears we can extend to a full neighborhood of the Bessel ray. What remains would then be only a finite number of Laguerre zeros, and a finite number of zeros should not introduce difficulties. Note that this assumes no common zeros, which assumption must be made in order to have injectivity. However, as we observe in the next section, this set of requirements refers to the type of deconvolution in the sense of limits as given in Sect. 3. Somewhat more will be required for the stronger form of deconvolution, in the sense of [4, 7], where a set of deconvolvers $\nu_1, \ldots, \nu_n$ that are compactly supported distributions are shown to exist. We address these issues in the next section.

5 Convergence of Deconvolving Sequences

In this section we address the issue of applying the sequences of deconvolvers formed in Sect. 3 to perform the deconvolution. The problem of deconvolution for the Pompeiu problem can be interpreted to mean the reconstruction of a given function f from the integral information given in the Pompeiu problem, in this case representable using the convolutions $f * T_1$ and $f * T_2$. We observe in this section how the sequences $\{\widehat{\nu}_{1,j}\}$ and $\{\widehat{\nu}_{2,j}\}$ can be used to reconstruct the function f. We also address related issues including the appropriate spaces for the functions and distributions we are working with as well as issues of convergence. Finally we will relate these issues to fundamental issues used in understanding deconvolution for Euclidean space, the Paley–Weiner theorem and Hörmander's strongly coprime condition.

We will utilize a theorem of Benson, Jenkins, and Ratcliff [2, Theorem 6.1] to describe the range of $\mathscr{S}(\mathbf{H}^n)$ under the spherical function transform. In the notation of this theorem, the Heisenberg fan $\mathscr{H}$ is represented by $\triangle(K, \mathbf{H}^n)$, where $K = U(n)$. The space $\widehat{\mathscr{S}}(K, \mathbf{H}^n)$ consists of functions that are rapidly decreasing on $\triangle(K, \mathbf{H}^n)$. The space $\mathscr{S}_K$ corresponds to radial functions in Schwartz space.

Theorem 4 ([2]) *If $f \in \mathscr{S}(\mathbf{H}^n)$, then $\widehat{f} \in \widehat{\mathscr{S}}(K, \mathscr{H}^n)$. Conversely, if $F \in \widehat{\mathscr{S}}(K, \mathscr{H}_n)$, then $F = \widehat{f}$ for some $f \in \mathscr{S}(\mathbf{H}^n)$. Moreover, the map $\widehat{}: \mathscr{S}(\mathbf{H}^n) \to \widehat{\mathscr{S}}(K, \mathscr{H}^n)$ is a bijection.*

First notice that in the limit, our deconvolving sequences $\{\widetilde{\nu}_{1,j}\}$ and $\{\widetilde{\nu}_{2,j}\}$ allow us to construct the sequence of functions $\{\widetilde{f}_j\}$ defined by

$$\begin{aligned}\widetilde{f}_j &\equiv (f * T_1)^{\sim} \cdot \widetilde{\nu}_{1,j} + (f * T_2)^{\sim} \cdot \widetilde{\nu}_{2,j} \\ &\equiv \widetilde{f} \cdot \widetilde{T}_1 \cdot \widetilde{\nu}_{1,j} + \widetilde{f} \cdot \widetilde{T}_2 \cdot \widetilde{\nu}_{2,j}.\end{aligned}$$

Noting that each $\widetilde{f}_j$ has the property that $\widetilde{f}_j|_{K_j} = \widetilde{f}|_{K_j}$ and $\widetilde{f}_j|_{F_j} = 0$, as constructed above in Sect. 3, we easily pass to the limit to attain the deconvolution

$$\widetilde{f} \equiv \lim_{j \to \infty} (f * T_1)^{\sim} \cdot \widetilde{\nu}_{1,j} + (f * T_2)^{\sim} \cdot \widetilde{\nu}_{2,j}.$$

This solves the problem of deconvolution, as f can be reconstructed from its “averages” $f * \mu_1$ and $f * \mu_2$. It is not as strong as the usual method of deconvolution, as we have here used a limiting procedure. It now remains to discuss the convergence and to address the appropriate function spaces for both the deconvolving sequences and the associated sequences of functions. We further discuss the existence of limits $\nu_1 = \lim \nu_{1,j}$ and $\nu_2 = \lim \nu_{2,j}$, forming individual deconvolvers from the sequences.

Before passing to the limit, we consider the issue of the spherical function transform, showing the existence of sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ whose spherical transforms yield the deconvolving sequences we have constructed. The above theorem of [2] characterizing the image of $\mathscr{S}(\mathbf{H}^n)$ under the spherical function transform will be used. At the level of the sequence of functions $\{\widetilde{f}_j\}$, as constructed above, we recognize that since $\widetilde{f}_j|_{K_j} \equiv 1$ while $\widetilde{f}_j|_{F_j} \equiv 0$, $\widetilde{f}_j$ is rapidly decreasing on $\triangle(K, \mathbf{H}^n)$, and thus $\widetilde{f}_j \in \widehat{\mathscr{S}}(K, \mathscr{H}_n)$. By the theorem there exists $F_j \in \mathscr{S}(\mathbf{H}^n)$ such that $\widetilde{F}_j \equiv \widetilde{f}_j$. Furthermore note that these F_j in the sequence can be explicitly constructed from the convolutions $f * T_1$ and $f * T_2$ using sequences of Schwartz functions $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ as follows. We observe that $\widetilde{\nu}_{1,j}|_{F_j} \equiv 0$ and $\widetilde{\nu}_{2,j}|_{F_j} \equiv 0$ imply that $\widetilde{\nu}_{1,j}, \widetilde{\nu}_{2,j} \in \widehat{\mathscr{S}}(K, \mathscr{H}_n)$. This in turn implies the existence of the desired sequences of $\nu_{1,j}, \nu_{2,j} \in \mathscr{S}(\mathbf{H}^n)$. Forming $(f * T_1) * \nu_{1,j} + (f * T_2) * \nu_{2,j}$, we see that

$$\begin{aligned}\left[(f * T_1) * \nu_{1,j} + (f * T_2) * \nu_{2,j}\right]^{\sim} &= \widetilde{f} \cdot \widetilde{T}_1 \cdot \widetilde{\nu}_{1,j} + \widetilde{f} \cdot \widetilde{T}_2 \cdot \widetilde{\nu}_{2,j} \\ &= \widetilde{f}_j.\end{aligned}$$

Thus $(f * T_1) * \nu_{1,j} + (f * T_2) * \nu_{2,j} = F_j \in \mathscr{S}(\mathbf{H}^n)$. We will demonstrate below that the sequence $\{F_j\}$ approaches f in the limit.

Noting that $F_j = f * (T_1 * \nu_{1,j} + T_2 * \nu_{2,j})$, we now describe the appropriate space for the elements $\widetilde{\nu}_{1,j}, \widetilde{\nu}_{2,j}$ as well as $T_1 * \nu_{1,j}, T_2 * \nu_{2,j}$. Each

$$(T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim} \in \widehat{\mathscr{S}}(K, \mathscr{H}_n)$$

since these were constructed to satisfy

$$(T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim} \equiv \big(\widetilde{T}_1 \cdot \widetilde{\nu}_{1,j} + \widetilde{T}_2 \cdot \widetilde{\nu}_{2,j}\big)|_{K_j} \equiv \begin{cases} 1 & \text{on } K_j, \\ 0 & \text{on } F_j, \end{cases}$$

and thus clearly are rapidly decreasing on $\triangle(K, \mathbf{H}^n)$. It follows from the theorem of [2] that $T_1 * \nu_{1,j} + T_2 * \nu_{2,j} \in \mathscr{S}(\mathbf{H}^n)$. A similar argument shows that each $T_1 * \nu_{1,j}, T_2 * \nu_{2,j} \in \mathscr{S}(\mathbf{H}^n)$.

On the side of the spherical transform we have constructed the $\widetilde{\nu}_{1,j}, \widetilde{\nu}_{2,j}$ in order to deconvolve $f * T_1$ and $f * T_2$ on K_j. Letting $\phi_j = T_1 * \nu_{1,j} + T_2 * \nu_{2,j}$, for each j, we have

$$\widetilde{\phi}_j|_{K_j} = (T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim}|_{K_j} = \big(\widetilde{T}_1 \cdot \widetilde{\nu}_{1,j} + \widetilde{T}_2 \cdot \widetilde{\nu}_{2,j}\big)|_{K_j} \equiv 1,$$

so that in the limit

$$\widetilde{\phi}_j = (T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim} \quad \to \quad \mathbf{1}_{\triangle(K,\mathscr{H}_n)}.$$

We then recognize that $\mathbf{1}_{\triangle(K,\mathscr{H}_n)}$ is a tempered distribution on $\triangle(K, \mathscr{H}_n)$. Furthermore, this is the spherical function transform of the Dirac delta function $\delta \in \mathscr{S}'(\mathbf{H}^n)$. Since we know that $\widetilde{\delta}(\lambda, k) = \psi_{\mathbf{k}}^{\lambda}(0) = 1$ and $\widetilde{\delta}(0; \rho) = \mathscr{J}_\rho(0) = 1$, we may write $\widetilde{\delta} = \mathbf{1}_{\triangle(K,\mathscr{H}_n)}$. Thus the limit

$$(T_1 * \nu_{1,j} + T_2 * \nu_{2,j})^{\sim} \quad \to \quad \mathbf{1}_{\triangle(K,\mathscr{H}_n)} = \widetilde{\delta}$$

tells us that $\lim_{j\to\infty} \phi_j = \lim_{j\to\infty} T_1 * \nu_{1,j} + T_2 * \nu_{2,j} = \delta$, by uniqueness of the Gelfand transform. Also not that each $\phi_j \in \mathscr{S}$ and that these converge to the tempered distribution $\delta \in \mathscr{S}'$.

In the approach to deconvolution outlined in [8] the Hörmander strongly coprime condition is used to demonstrate existence of deconvolvers $\nu_1, \dots, \nu_n$ as compactly supported distributions. The arithmetic condition for the radii of the disks not to be N-well approximated by the zeros of the Bessel function J_1 is used to give the required estimates for the strongly coprime condition (2) near the Bessel zeros. The explicit construction of deconvolvers given in [4, 7] utilize a different set of conditions, closely related to these. Our approach is different from these in that we have produced sequences of deconvolvers $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$ in $\mathscr{S}(\mathbf{H}^n)$. Since the transformations $\widetilde{\nu}_{1,j}$ and $\widetilde{\nu}_{2,j}$ are compactly supported for each j, we were able to use results on Schwartz space rather than Hörmander's strongly coprime condition. In this section we have been considering how to utilize these deconvolving sequences to form the deconvolution through a limiting process. When we also address the issue of the limits of the sequences of deconvolvers themselves, $\nu_1 = \lim \nu_{1,j}$ and $\nu_2 = \lim \nu_{2,j}$, then the Paley–Weiner theorem and the strongly coprime condition of Hörmander again become relevant. Considering the limit in the sense of distributions, we need to describe $\lim\langle f, \nu_{1,j}\rangle$ for every $f \in \mathscr{S}(\mathbf{H}^n)$. Due to the definition of $\nu_{1,j}$ in terms of its Gelfand transform $\widetilde{\nu}_{1,j}$, we consider the limits $\langle \widetilde{f}, \widetilde{\nu}_{1,j}\rangle$ for $\widetilde{f} \in \widehat{\mathscr{S}}$, where the inner products $\langle \widetilde{f}, \widetilde{\nu}_{1,j}\rangle$ on the space $\mathscr{H}_n$ are to be interpreted using Godement's Plancherel measure, as given in [2]. For convergence, it is necessary that $\widetilde{\nu}_{1,j}$ does not grow too rapidly. Due to the manner in which these were constructed, this issue is directly related to the proximity of the zero sets of $\widetilde{T}_1$ and $\widetilde{T}_2$, and their rates of decay near these zero sets. After setting up the limits and the issue of their convergence, we will briefly address three separate cases.

To discuss the limits of the deconvolving sequences $\{\nu_{1,j}\}$ and $\{\nu_{2,j}\}$, we must consider the limits in the sense of distributions. Since the deconvolving sequences have been defined in terms of the transforms $\widetilde{\nu}_{1,j}$ and $\widetilde{\nu}_{2,j}$, we consider the limits of these sequences as tempered distributions. Tempered distributions use the space of Schwartz functions as test functions, and this allows results to be transferred to $\nu_{1,j}$ and $\nu_{2,j}$ through the above result of [2]. Since $\widehat{\nu}_{1,j} \in \widehat{\mathscr{S}} \subset \widehat{\mathscr{S}'}$, we have that

$$\langle f, \nu_{1,j}\rangle = \langle \widetilde{f}, \widetilde{\nu}_{1,j}\rangle \quad \text{for all } \widetilde{f} \in \widehat{\mathscr{S}},$$

and we want to investigate the behavior in the limit as $j \to \infty$. If we can show that the limit converges for each $\widehat{f} \in \widehat{\mathscr{S}}$, this will imply the existence of $\widehat{\nu}_1 = \lim \widehat{\nu}_{1,j}$ and $\nu_1 = \lim \nu_{1,j}$ as tempered distributions, in $\widehat{\mathscr{S}'}$ and $\mathscr{S}'$, respectively. However, this condition depends on the rate of decay, or growth, of $\widetilde{\nu}_{1,j}$. Recall that the definition of $\widetilde{\nu}_{1,j}$ and $\widetilde{\nu}_{2,j}$ requires inversion of $\widetilde{T}_1$ and $\widetilde{T}_2$ away from their zeros. Although $\widetilde{\nu}_{1,j}, \widetilde{\nu}_{2,j} \in \widehat{\mathscr{S}}$ for each j, depending on the proximity of the zeros of $\widetilde{T}_1$ and $\widetilde{T}_2$ and the growth of $\widetilde{\nu}_{1,j}$ and $\widetilde{\nu}_{2,j}$ near these zeros, the limit of $\widetilde{\nu}_{1,j}$ and $\widetilde{\nu}_{2,j}$ may not remain in Schwartz space. Even if the $\widetilde{\nu}_{1,j}$ grow rapidly near the zero sets as j increases, the limit still exists in the space $\mathscr{D}'$ since we know that for each $f \in \mathscr{D}$, there exists k such that

$$\lim\langle f, \widetilde{\nu}_{1,j}\rangle = \langle f, \widetilde{\nu}_{1,k}\rangle,$$

where $\operatorname{supp}(f) \subset K_k$. Thus the rate of growth of the $\nu_{1,j}$ and the space in which this convergence occurs are directly related to the distribution of the zeros of $\widetilde{T}_1$ and $\widetilde{T}_2$, and these issues also relate directly to the Paley–Weiner theorem. In Euclidean space $\mathbf{C}^n$ the strongly coprime condition of Hörmander gives a condition for the existence of $\widehat{\nu}_1, \widehat{\nu}_2$ as Fourier transforms of compactly supported distributions, or equivalently for the existence of ν_1, ν_2 as compactly supported distributions. In the case of Sect. 3.1 the zeros have uniform separation, and it is easy to show Hörmander's strongly condition is satisfied. However in the case of Sect. 3.2, whether or not Hörmander's condition is satisfied is determined by whether the ratio of radii $\frac{r_1}{r_2}$ is N-well approximated or is poorly approximated by ratios of Bessel zeros E_n, where

$$E_n = \left\{x \in \mathbf{R} : J_n(x) = 0\right\} = \{\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots\}, \quad \text{where } \lambda_j < \lambda_{j+1}.$$

In the first two cases, where the zeros have a uniform separation for the ball and sphere of Sect. 3.1 or the radii of the two balls of Sect. 3.2 where the radii are poorly approximated by ratios of E_n, it is possible to show that the convergence exists in the space of tempered distributions. However in the third case in which the ratio of the radii of the two balls of Sect. 3.2 are N-well approximated, the convergence is more delicate, and we can only guarantee $\lim \widetilde{\nu}_{1,j} = \widetilde{\nu}_1 \in \mathscr{D}'$. Nevertheless the results of the deconvolving sequences are still valid in the space of tempered distributions, as discussed above. Note that the above conclusions were based on application of the results of [2] for Schwartz space and did not require stronger Paley–Weiner-type results. Ideally we would like to be able find methods to show the existence of deconvolvers as compactly supported distributions, or even to extend methods of [4, 7] to make an explicit construction of such compactly supported deconvolvers using methods of summation, differentiation, integration, and convolution.

In the sequel to this paper we plan to revisit this issue and to deal more directly with the issue of Hörmander's strongly coprime condition in the Heisenberg group setting.

6 Extending Deconvolution from the Bessel Ray

Consider the case of radial distributions $T_1, \ldots, T_n$ compactly supported satisfying Hörmander's strongly coprime condition for $\mathbf{C}^n$, implying the existence of distributions $\nu_1, \ldots, \nu_n$ radial and compactly supported such that

$$\widehat{T}_1(\xi)\widehat{\nu}_1(\xi) + \cdots + \widehat{T}_n(\xi)\widehat{\nu}_n(\xi) \equiv 1,$$

which can be written as

$$\widehat{T}_1(r)\widehat{\nu}_1(r) + \cdots + \widehat{T}_n(r)\widehat{\nu}_n(r) \equiv 1,$$

where $r = |\xi|$. Noting that $\widehat{T}_j(|\xi|) = \widehat{T}_j(r) = \widetilde{T}_j(0;\rho)$ and likewise $\widehat{\nu}_j(|\xi|) = \widehat{\nu}_j(r) = \widetilde{\nu}_j(0;\rho)$, this is equivalent to a deconvolution of $T_1, \ldots, T_n$ on the Bessel ray $\mathscr{H}_\rho$,

$$\widetilde{T}_1(0;\rho)\widetilde{\nu}_1(0;\rho) + \cdots + \widetilde{T}_n(0;\rho)\widetilde{\nu}_n(0;\rho) \equiv 1. \tag{6}$$

It is important to ask whether such a deconvolution can be extended to the Gelfand transforms $\widetilde{T}_1, \ldots, \widetilde{T}_n$ on all of the Heisenberg brush $\mathscr{H}$. It is not guaranteed that the deconvolvers on Euclidean space will extend to work for the Heisenberg group. Considering the Gelfand transform of the same sum, $\widetilde{T}_1\widetilde{\nu}_1 + \cdots + \widetilde{T}_n\widetilde{\nu}_n$, the goal is to make this expression uniformly equal to 1 for all $(\lambda, k) \in \mathbf{R}^* \times \mathbf{Z}_+$,

$$\widetilde{T}_1(\lambda,k)\widetilde{\nu}_1(\lambda,k) + \cdots + \widetilde{T}_n(\lambda,k)\widetilde{\nu}_n(\lambda,k) \equiv 1. \tag{7}$$

The existence of $\nu_1, \ldots, \nu_n$ satisfying (6) and (7) would solve the problem of deconvolution for $\mathbf{H}^n$; however a solution of (6) has not been shown to extend to (7).

Our general goal for a set of radial distributions $T_1, \ldots, T_n$ is to extend the Euclidean deconvolution

$$\widehat{T}_1(\xi)\widehat{\nu}_1(\xi) + \cdots + \widehat{T}_n(\xi)\widehat{\nu}_n(\xi) \equiv 1$$

to all of the Heisenberg brush $\mathscr{H}$ by expanding upon the method developed in Sect. 3. In this section, we develop a method for extending the deconvolution from the central Bessel ray $\mathscr{H}_\beta$ to all of the Heisenberg fan $\mathscr{H}$ for any T_r and S_r satisfying Hörmander's strongly coprime condition, or equivalently, satisfying condition (6). That is to say, given ν_1 and ν_2 satisfying

$$\widehat{T_r}(\xi)\cdot\widehat{\nu_1}(\xi)+\widehat{S_r}(\xi)\cdot\widehat{\nu_2}(\xi)\equiv 1 \quad \text{for all } \xi\in\mathbf{C}^n,$$

we want to find μ_1 and μ_2 satisfying

$$\widetilde{T_r}(p)\cdot\widetilde{\mu_1}(p)+\widetilde{S_r}(p)\cdot\widetilde{\mu_2}(p)\equiv 1 \quad \text{for all } p\in\mathscr{H},$$

equivalent to conditions (6) and (7), and furthermore $\widetilde{\mu}_j(0;\rho)=\widehat{\mu}_j(r)$, where $r=|\xi|$. Note that the constructions in Sect. 3 do not quite solve this problem, since the ν_1 and ν_2 produced do not necessarily agree on $\mathscr{H}_\rho$ with a given Euclidean deconvolution. However, in this section we will demonstrate the existence of the desired μ_1 and μ_2, and part of the construction is based on the methods of Sect. 3.

The idea is to use the existing Euclidean deconvolution $\widehat{T_r}\widehat{\nu_1}+\widehat{S_r}\widehat{\nu_2}\equiv 1$ on the Bessel ray $\mathscr{H}_\rho$ while using the construction of Sect. 3 on a neighborhood R away from the Bessel ray. Note that this methods of Sect. 3 essentially amounts to pasting together the inverse $\frac{1}{\widetilde{T_1}}$ away from the zeros set of $\widetilde{T_1}$ with the inverse $\frac{1}{\widetilde{T_2}}$ away from the zeros set of $\widetilde{T_2}$. The two main points are then first to extend the Euclidean deconvolution $\equiv 1$ to a neighborhood V_ρ of the Bessel ray $\mathscr{H}_\rho$ and second to orchestrate the overlapping of these two methods on the space in between the two neighborhoods V_ρ and R, through a construction comparable to a partition of unity. We begin with the determination of the open set V_ρ for which we extend the deconvolution to $\equiv 1$ for all $p\in V_\rho$. Here we will rely on the continuity of the transforms $\widetilde{T_1},\widetilde{T_2},\widetilde{\nu_1},\widetilde{\nu_2}$ in the relative subspace topology to yield existence of a neighborhood V_ρ where

$$\left|\widetilde{T_1}(p)\widetilde{\nu_1}(p)+\widetilde{T_2}(p)\widetilde{\nu_2}(p)-1\right|<\varepsilon \quad \text{for all } p\in V_\rho,$$

where $\varepsilon>0$ is a sufficiently small number. For the neighborhood R, we want to be sure that all the zeros of $\phi(p)=\widetilde{\nu_1}(p)\widetilde{T_r}(p)+\widetilde{\nu_2}(p)\widetilde{S_r}(p)$ are contained inside of R. Also, R should be of the form $R=\cup_{k=1}^{\alpha}\mathscr{H}_{k,\pm}$. We simply choose the minimum value of α that will satisfy the property of R containing all the zeros of ϕ. Then auxiliary distributions Ψ_1 and Ψ_2 will be constructed such that $\widetilde{\Psi_1}|_R\equiv 1$ and $\widetilde{\Psi_2}|_R\equiv 0$.

We also choose to make V_ε of a similar form, i.e., $V_\varepsilon=\cup_{k=\beta}^{\infty}\mathscr{H}_{k,\pm}$, where β is chosen to be a minimum so that V_ε still satisfies the inequality $|\phi(p)-1|<\varepsilon$ on V_ε. This will also provide the condition $R\cap V_\varepsilon=\emptyset$.

Theorem 5 *Consider S_r and T_r radial distributions satisfying Hörmander's strongly coprime condition on $\mathbf{C}^n$, i.e., such that there exist ν_1 and ν_2, radial, compactly supported distributions satisfying $\widehat{T_r}(\xi)\widehat{\nu_1}(\xi)+\widehat{S_r}(\xi)\widehat{\nu_2}(\xi)\equiv 1$ for all $\xi\in\mathbf{C}^n$. Also assume that for all $(\lambda,k)\in\mathbf{R}^*\times\mathbf{Z}_+$, either $\widetilde{S_r}(\lambda,k)\neq 0$ or $\widetilde{T_r}(\lambda,k)\neq 0$. Then there exist μ_1,μ_2 such that*

$$\widetilde{S_r}\widetilde{\mu_1}(p)+\widetilde{T_r}\widetilde{\mu_2}(p)\equiv 1 \quad \textit{for all } p\in\mathscr{H}$$

and such that $\widetilde{\mu}_1(0;\rho)=\widetilde{\nu}_1(\xi)$ *and* $\widetilde{\mu}_2(0;\rho)=\widetilde{\nu}_2(\xi)$ *for all* $\rho\in\mathbf{R}_+$ *and all* $\xi\in\mathbf{C}^n$, *where* $|\xi|=\rho$.

Proof First construct the regions R and V_ε, as described above, according to the magnitude of $|\widetilde{T}_r(\lambda,k)\widetilde{\nu}_1(\lambda,k)+\widetilde{S}_r(\lambda,k)\widetilde{\nu}_2(\lambda,k)|$, and the location of any zeros. If this expression has no zeros, it is possible to then define R by minimizing α to instead satisfy the property that R contains all p such that $\phi(p)=\gamma$, where γ is a real number near 0 appropriately chosen so the set of such p is nonempty.

Then construct Ψ_1 as follows:

$$\widetilde{\Psi}_1=\begin{cases}\frac{1}{k-\alpha} & k>\alpha,\\ 1 & k\le\alpha,\end{cases}$$

yielding $\lim_{k\to\infty}\widetilde{\Psi}_1(p_k)=0$ for $p_k\in\mathscr{H}_{k,\pm}$, and furthermore $\widetilde{\Psi}_1|_{\mathscr{H}_\rho}\equiv 0$.

Then we can construct Ψ_2 as follows:

$$\widetilde{\Psi}_2=\begin{cases}\frac{1-\widetilde{\Psi}_1}{\widetilde{T}_r\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\nu}_2} & \text{on } \mathscr{H}\setminus R,\\ 0 & \text{on } R.\end{cases}$$

It is possible to make this inversion since $\widetilde{T}_r\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\nu}_2$ is bounded away from 0 on $\mathscr{H}\setminus R$.

Then using $\widetilde{\mu}_1=\widetilde{\Psi}_1(\frac{1}{\widetilde{T}_r})\widetilde{\rho}_1+\widetilde{\Psi}_2\widetilde{\nu}_1$ and $\widetilde{\mu}_2=\widetilde{\Psi}_1(\frac{1}{\widetilde{S}_r})\widetilde{\rho}_2+\widetilde{\Psi}_2\widetilde{\nu}_2$, we have the desired relation $\widetilde{T}_r\widetilde{\mu}_1+\widetilde{S}_r\widetilde{\mu}_2$ for all of $\mathscr{H}$, as follows:

$$\begin{aligned}\widetilde{T}_r\widetilde{\mu}_1+\widetilde{S}_r\widetilde{\mu}_2&=\widetilde{T}_r\widetilde{\Psi}_1\left(\frac{1}{\widetilde{T}_r}\right)\widetilde{\rho}_1+\widetilde{T}_r\widetilde{\Psi}_2\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\Psi}_1\left(\frac{1}{\widetilde{S}_r}\right)\widetilde{\rho}_2+\widetilde{S}_r\widetilde{\Psi}_2\widetilde{\nu}_2\\&=\widetilde{\Psi}_1\widetilde{\rho}_1+\widetilde{T}_r\widetilde{\Psi}_2\widetilde{\nu}_1+\widetilde{\Psi}_1\widetilde{\rho}_2+\widetilde{S}_r\widetilde{\Psi}_2\widetilde{\nu}_2\\&=\widetilde{\Psi}_1(\widetilde{\rho}_1+\widetilde{\rho}_2)+\widetilde{\Psi}_2\big(\widetilde{T}_r\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\nu}_2\big)\\&=\widetilde{\Psi}_1+\widetilde{\Psi}_2\big(\widetilde{T}_r\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\nu}_2\big).\end{aligned}$$

Clearly on R this reduces to $\widetilde{\Psi}_1=1$, while on $\mathscr{H}\setminus R$ it reduces to

$$\widetilde{\Psi}_1+\frac{1-\widetilde{\Psi}_1}{\widetilde{T}_r\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\nu}_2}\big(\widetilde{T}_r\widetilde{\nu}_1+\widetilde{S}_r\widetilde{\nu}_2\big)=\widetilde{\Psi}_1+\big(1-\widetilde{\Psi}_1\big)=1.$$

Thus we have constructed μ_1 and μ_2 with the property that $\widetilde{S}_r\widetilde{\mu}_1+\widetilde{T}_r\widetilde{\mu}_2\equiv 1$ on all of $\mathscr{H}$, as required. This completes the proof of this theorem. □

In summary, we have constructed μ_1 and μ_2 such that $\widetilde{\mu}_1|_{\mathscr{H}_\rho}\equiv\widehat{\nu}_1(|\xi|)$ and $\widetilde{\mu}_2|_{\mathscr{H}_\rho}\equiv\widehat{\nu}_2(|\xi|)$, thus extending the deconvolvers ν_1,ν_2, compactly supported in $\mathbf{C}^n$, as guaranteed by Hörmander's result, to μ_1 and μ_2 defined on $\mathbf{H}^n$ with transforms $\widehat{\mu}_1,\widehat{\mu}_2$ defined on all of $\mathscr{H}$. Using the additional assumption of no common zeros for $\widetilde{S}_r$ and $\widetilde{T}_r$ on all of $\mathscr{H}$, it is possible to form a deconvolving sequence by the method of Sect. 3.2. The $\rho_{1,j}$ and $\rho_{2,j}$ used as the "platform" for the deconvolving sequences were used above to interpolate between the "local inverses" $1/\widetilde{S}_r$ and $1/\widetilde{T}_r$, away from the zero sets, and the deconvolvers $\widetilde{\nu}_1$ and $\widetilde{\nu}_2$ along the central Bessel ray, corresponding to $\widehat{\mathbf{C}}^n$.

Note that for radial distributions, Theorem 5 guarantees the existence of deconvolvers Ψ_1 and Ψ_2 for $\mathbf{H}^n$ when Hörmander's strongly coprime condition is satisfied in the Euclidean setting and in addition the conditions for no common Laguerre zeros are met. This essentially validates the claim in Sect. 4 that deconvolution for $\mathbf{H}^n$ depends essentially on what happens for the central Bessel ray, provided that there are no common Laguerre zeros. The use of the type of deconvolving sequences constructed in Sect. 3 further illustrates the importance of the methods developed therein. These results also suggest that there should be a suitable version of Hörmander's strongly coprime condition that will also apply to $\mathbf{H}^n$. We plan to explore this and the related issues discussed in Sects. 4 and 5 in the sequel.

Acknowledgements Part of the research work for this paper was completed during the visits of the three authors at the National Center for Theoretical Sciences in Hsinchu, Taiwan. They would like to thank NCTS (Hsinchu) for partial support of this research. They take great pleasure in expressing their thanks to Professor Winnie Li, the director of NCTS, for the invitation and the warm hospitality during their visits in Taiwan. The second author would also like to use this opportunity to thank the Academia Sinica in Taipei for the invitation during the summer of 2007. He would also like to thank the colleagues at the Sinica for the warm hospitality during his visit in Taipei.

References

1. Agranovsky, M., Berenstein, C., Chang, D.C., Pascuas, D.: Injectivity of the Pompeiu transform in the Heisenberg group. J. Anal. Math. **63**, 131–173 (1994)
2. Benson, C., Jenkins, J., Ratcliffe, G.: The spherical transform of a Schwartz function on the Heisenberg group. J. Funct. Anal. **154**, 379–423 (1998)
3. Berenstein, C.A., Gay, R.: A local version of the two circles theorem. Isr. J. Math. **55**, 267–288 (1986)
4. Berenstein, C.A., Taylor, B.A., Yger, A.: Sur quelques formules explicites de déconvolution. J. Opt. (Paris) **14**, 75–82 (1983)
5. Berenstein, C., Chang, D.C., Tie, J.: Laguerre Calculus and Its Applications on the Heisenberg Group. AMS/IP Series in Advanced Mathematics, vol. 22. International Press, Cambridge (2001)
6. Berenstein, C., Chang, D.C., Eby, W.: L^p results for the Pompeiu problem with moments on the Heisenberg group. J. Fourier Anal. Appl. **10**, 545–571 (2004)
7. Berenstein, C., Yger, A.: Le problème de déconvolution. J. Funct. Anal. **54**, 113–160 (1983)
8. Casey, S.D., Berenstein, C.A., Walnut, D.F.: Exact multichannel deconvolution on radial domains. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97), vol. 3, pp. 1865–1868 (1997)
9. Chang, D.C., Eby, W.: Tauberian type theorem for the **m**-spherical transform on the Heisenberg group. Math. Nachr. **280**, 815–837 (2007)
10. Eby, W.: Moment results for the Heisenberg group interpreted using the Weyl calculus. In: Recent Progress on Some Problems in Several Complex Variables and Partial Differential Equations. Contemp. Math., vol. 400, pp. 95–105 (2006)

11. Eby, W.: Aspects of the Pompeiu problem in the setting of the Heisenberg group. In: Birman, J., Catto, S., Nicolescu, B. (eds.) Group Theoretical Methods in Physics, pp. 220–225. Canopus publishing, Bristol (2006)
12. Ehrenpreis, L.: Solution of some problems of division, I, Division by a polynomial of derivation. Am. J. Math. **76**, 883–903 (1954)
13. Ehrenpreis, L.: Solution of some problems of division, II, Division by a punctual distribution. Am. J. Math. **77**, 286–292 (1955)
14. Ehrenpreis, L.: Solution of some problems of division, III, Division in the spaces $\mathscr{D}'$, $\mathscr{H}$, $\mathscr{Q}_A$, $\mathscr{O}$. Am. J. Math. **78**, 685–715 (1956)
15. Erdelyi, A., Magnus, W., Oberhettinger, F., Tricomi, F.G.: Higher Transcendental Functions, vol. 2. McGraw-Hill, New York (1953)
16. Folland, G.: Harmonic Analysis in Phase Space. Annals of Mathematical Studies, vol. 122. Princeton University Press, Princeton (1989)
17. Hörmander, L.: Generators for some rings of analytic functions. Bull. Am. Math. Soc. **73**, 943–949 (1967)
18. Zalcman, L.: Mean values and differential equations. Isr. J. Math. **14**, 339–352 (1973)

Theta Functions Wronskians and Weierstrass Points for Linear Spaces of Meromorphic Functions

Hershel M. Farkas

Abstract In this note we consider the Weierstrass points for the linear space of meromorphic functions on a compact Riemann surface whose divisors are multiples of $\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}$, where P_i are points of the surface, and α is a positive integer for which there is no holomorphic differential on the surface whose divisor is a multiple of $P_0^\alpha P_1 \cdots P_{g-1}$. Thus the dimension of our linear space is precisely α.

The Weierstrass points for our space are those points $Q \neq P_i$ for which there is a function in the space which vanishes to order at least α at the point Q. Thus the Weierstrass points are all zeros of the Wronskian determinant of a basis for our space, and the weight of the Weierstrass point is the order of the zero.

We show that all the Weierstrass points are zeros of the Riemann theta function $\theta(\alpha \Phi_{P_0}(P) - e)$ on the surface where $e = \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$. The question we investigate is whether the order of the zero of the theta function agrees with the order of the zero of the Wronskian. We prove that this is so at least in the case of zeros of order $k = 1, 2$.

1 Introduction

The work exposed here should be viewed as a continuation of the material presented in [1–3], and the reader should look at those papers and the references therein cited for motivation and background. This was a topic that I often discussed with Leon Ehrenpreis and therefore feel that it is appropriate for this volume dedicated to his memory.

The motivation for the present discussion is the following: We consider a compact Riemann surface of genus g at least 2 with canonical homology basis $(\gamma_1, \ldots, \gamma_g; \delta_1, \ldots, \delta_g)$. We let $(\theta_1, \ldots, \theta_g)$ be the basis, dual to the canonical homology basis, of the linear space of holomorphic differentials on the surface. This

Dedicated to the memory of Leon Ehrenpreis.

H.M. Farkas (✉)
Institute of Mathematics, The Hebrew University of Jerusalem, E. Safra Campus, Givat Ram, Jerusalem, Israel
e-mail: farkas@math.huji.ac.il

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_7, © Springer-Verlag Italia 2012

gives rise to a complex symmetric matrix $\Pi = \pi_{ij}$, $i, j = 1, \ldots, g$, with $Im(\Pi)$ positive definite. We can then construct the associated Riemann theta function

$$\theta(Z, \Pi) = \sum_{N \in Z^g} \exp(2\pi i)[1/2N\Pi N + NZ]$$

where $N = (n_1, \ldots, n_g)$ and $Z = (z_1, \ldots, z_g)$. It is easy to see that the theta function is an even function of Z.

Following Riemann, choosing a base point P_0 on the surface, we define a map Φ_{P_0} of the surface into its Jacobi variety, J(S), where J(S) is the quotient space C^g modulo the group of translations

$$Z \mapsto Z + e^i \quad i = 1, \ldots, g, \qquad Z \mapsto Z + \Pi^i \quad i = 1, \ldots, g,$$

where e^i, Π^i, are the ith columns of the identity matrix I and the matrix Π, respectively. We then consider the locally defined holomorphic function

$$f(P) = \theta\big(\Phi_{P_0}(P), \Pi\big).$$

$f(p)$ is not a single-valued function on the surface, but its zeroes are well defined. More generally, we can consider for each $\alpha \in Z^+$ and $e \in C^g$ the multivalued function

$$f(p) = \theta\big(\alpha\Phi_{P_0}(P) - e, \Pi\big).$$

The properties of this function are given in [4], Chap. 6, and are summarized in the following proposition.

Proposition 1 *The multivalued function $f(P)$ either vanishes identically on the Riemann surface (in which case it is single valued), or it vanishes at $\alpha^2 g$ points on the surface (counting multiplicities). In the latter case the zeroes $R_1, \ldots, R_{\alpha^2 g}$ satisfy*

$$\alpha e = \Phi_{P_0}(R_1 \cdots R_{\alpha^2 g}) + \alpha^2 K_{P_0},$$

where K_{P_0} is the vector of Riemann constants with base point P_0. If $\theta(e) \neq 0$, then surely $f(P_0) \neq 0$. If $\theta(e) = 0$, then by Riemann's vanishing theorem

$$e = \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}.$$

In the latter case, f(P) vanishes identically if and only if $i(P_0^\alpha P_1 \cdots P_{g-1}) \geq 1$.

This proposition is proven in [4], p. 312.

Let $e \equiv \Phi_{P_0}(P_1 \cdots P_g) + K_{P_0}$, assume that $P_0 \neq P_i$ for all i and that $i(P_0^\alpha P_1 \cdots P_g) = 0$. This is of course automatic whenever $i(P_1 \cdots P_g) = 0$. If $i(P_1 \cdots P_g) > 0$, then we have the equivalence

$$\Phi_{P_0}(P_1 \cdots P_g) \equiv \Phi_{P_0}\big(P_0 P_1' \cdots P_{g-1}'\big)$$

for a divisor $P_0 P_1' \cdots P_{g-1}'$ equivalent to $P_1 \cdots P_g$. Hence in the latter case we shall consider $e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$. The function $f(P) = \theta(\alpha\Phi_{P_0}(P) - e)$ in both cases is a nonidentically vanishing multivalued function on the associated

Riemann surface with well-defined zeros. The difference between the two cases is whether $Q = P_0$ is a zero of $f(P)$ or not. We shall also always assume that $\alpha \geq 2$.

Recalling the Riemann vanishing theorem [4], a point Q is a zero of $f(P)$ iff there exists an integral divisor of degree $g-1$, $T_1 \cdots T_{g-1}$, with the property that

$$\alpha \Phi_{P_0}(Q) - e \equiv -\Phi_{P_0}(T_1 \cdots T_{g-1}) - K_{P_0}.$$

In the case that $e \equiv \Phi_{P_0}(P_1 \cdots P_g) + K_{P_0}$ with $i(P_1 \cdots P_g) = 0$, this translates to the condition that

$$\phi_{P_0}\left(\frac{Q^\alpha T_1 \cdots T_{g-1}}{P_0^{\alpha-1} P_1 \cdots P_g}\right) \equiv 0,$$

while in the case $e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$ the meaning is that

$$\phi_{P_0}\left(\frac{Q^\alpha T_1 \cdots T_{g-1}}{P_0^{\alpha} P_1 \cdots P_{g-1}}\right) \equiv 0.$$

In either case however we have $r[\frac{1}{P_0^{\alpha-1} P_1 \cdots P_g}] = r[\frac{1}{P_0^{\alpha} P_1 \cdots P_{g-1}}] = \alpha$, so that in both cases Q is a zero of the Wronskian of a basis for the linear space $L[\frac{1}{P_0^{\alpha-1} P_1 \cdots P_g}]$ or the linear space $L[\frac{1}{P_0^{\alpha} P_1 \cdots P_{g-1}}]$. For what we are going to show here, there is no difference between the cases; so rather than having to specify which space we are dealing with, we will choose to use only the latter space where $e \equiv \Phi_{P_0}(P_1..P_{g-1}) + K_{P_0}$. Let us further observe that the zeroes of the function $f(P)$, $R_1, \ldots, R_{\alpha^2 g}$, satisfy

Lemma 1

$$\Phi_{P_0}\left(\frac{R_1 \cdots R_{\alpha^2 g}}{P_1^\alpha \cdots P_{g-1}^\alpha}\right) \equiv \frac{\alpha^2 - \alpha}{2}(-2k_{P_0}).$$

Proof The nonidentical vanishing of $\theta(\alpha \Phi_{P_0}(P) - e)$ gives that

$$\alpha e \equiv \Phi_{P_0}(R_1 \cdots R_{\alpha^2 g}) + \alpha^2 K_{P_0}.$$

Since we chose e such that $e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$, it follows that

$$\Phi_{P_0}\left(\frac{R_1 \cdots R_{\alpha^2 g}}{P_1^\alpha \cdots P_{g-1}^\alpha}\right) \equiv \frac{\alpha^2 - \alpha}{2}(-2K_{P_0}).$$

□

We now recall that Δ_q is the divisor of a meromorphic q-differential if and only if degree $\Delta_q = q(2g-2)$ and $\phi_{P_0}(\Delta_q) \equiv q(-2K_{P_0})$. Let $q = \frac{\alpha^2 - \alpha}{2}$. Then the divisor $\frac{R_1 \cdots R_{\alpha^2 g}}{P_0^{\alpha^2} P_1^\alpha \cdots P_{g-1}^\alpha}$ has degree $\alpha^2 g - \alpha g - \alpha^2 + \alpha$ and thus has the degree of a q-canonical divisor. Moreover our lemma has shown that its image in $J(S)$ is $\frac{\alpha^2-\alpha}{2}(-2K_{P_0})$, so that the divisor is the divisor of a meromorphic q-differential with $q = \frac{\alpha^2-\alpha}{2}$. In particular this shows us that the divisor of zeroes of $f(P)$ is related strongly to the zeroes of a meromorphic q-differential. In fact if we assume that $r[\frac{1}{P_0^j P_1 \cdots P_{g-1}}] = j+1$ for $j = 0, \dots, \alpha - 1$, we find that $P_0^{\alpha^2-\alpha} P_1^\alpha \cdots P_{g-1}^\alpha$ is the divisor of poles of the Wronskian determinant of a basis for $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$.

The issue we wish to raise here is the following: The above discussion has shown us that $Q \neq P_i$ is a zero of $f(P)$ if and only if Q is a zero of the Wronskian of a basis for the linear space $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$. The issue is whether the multiplicities of the zeros are also equal.

We do not give a complete answer to this question but will at least prove the following theorem:

Theorem 1 *For $k = 1, 2$, $Q \neq P_i$ is a kth-order zero of the Wronskian of a basis for $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ if and only if Q is a kth-order zero of $f(P) = \theta(\alpha \Phi_{P_0}(P) - e)$.*

2 Weierstrass Points for $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$

Definition 1 For $Q \neq P_0$, consider the sequence

$$r\left[\frac{Q^m}{P_0^\alpha P_1 \cdots P_{g-1}}\right], \quad m = 0, \dots, \alpha + g.$$

$m = n$ will be called a "drop" for the linear space $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$, provided that

$$r\left[\frac{Q^{n-1}}{P_0^\alpha P_1 \cdots P_{g-1}}\right] = r\left[\frac{Q^n}{P_0^\alpha P_1 \cdots P_{g-1}}\right] + 1.$$

It is clear that there are precisely α "drops" at any point Q. If we denote the drops by d_i, then the order of the Wronskian at the point Q is equal to $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2}$. The reason this is true is the fact that if the drops are d_i, then there is always a function in $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with a zero of order $d_i - 1$ at the point Q. We shall denote the Wronskian determinant by W.

Definition 2 For $Q \neq P_0$, a basis $f_1, \dots, f_\alpha$ for $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ will be called a basis adapted to Q, provided that $ord_Q f_i = d_i - 1$.

Definition 3 A point $Q \neq P_i$ will be called a Weierstrass point for the linear space $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$, provided that there is a function in the space with a zero of order at least α at Q.

Lemma 2 *If $Q \neq P_i$, then* 1 *is always a "drop" at Q.*

Proof Q is a non-"drop" if and only if $r[\frac{Q}{P_0^\alpha P_1 \cdots P_{g-1}}] = \alpha$, which is true if and only if $i(\frac{P_0^\alpha P_1 \cdots P_{g-1}}{Q}) = 1$. This contradicts $i(P_0^\alpha P_1 \cdots P_{g-1}) = 0$. □

Remark If we assume that for each point P_i, $i = 1, \ldots, g$, there is at least one function in the space $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with a pole at P_i, then the assumption $Q \neq P_0$ would suffice in the above lemma.

It is clear that a point Q is a Weierstrass point if and only if there is a "drop" at Q of size at least $\alpha + 1$, or stated otherwise that there is a non-"drop" of size less than or equal to α. Thus if the "drops" are the positive integers $1, 2, \ldots, \alpha$, the point is not a Weierstrass point. The point is a simple Weierstrass point (a simple zero of the Wronskian), provided that the sequence of "drops" is $1, 2, \ldots, \alpha - 1, \alpha + 1$. In fact, the following lemma is fairly obvious.

Lemma 3 *Assume that there are m non-"drops" in the set* $1, 2, \ldots, \alpha$. *Denote these non-"drops" by* $\alpha - l_i$, $i = 1, \ldots, m$, *with* $l_1 \geq 0$ *so that there are also m "drops" which are at least* $\alpha + 1$. *Denote these "drops" by* $\alpha + k_i$, $i = 1, \ldots, m$, *with* $k_1 \geq 1$. *Then* $ord_Q W = \sum_i k_i + \sum_i l_i$, *where W is the Wronskian of a basis for* $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$.

Proof It is clear that

$$ord_Q W = \sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2} = \sum_{i=1}^{m} \alpha + k_i - \sum_{i=1}^{m} \alpha - l_i = \sum_i k_i + \sum_i l_i. \quad \square$$

We now will require the following (well-known) properties of theta functions on Riemann surfaces. These all follow from the Riemann vanishing theorem [4, 5].

Proposition 2 *If* $f \equiv \Phi_{P_0}(R_1 \cdots R_{g-1}) + K_{P_0}$ *and if* $i(R_1 \cdots R_{g-1}) = 1$, *then*

$$\sum_{i}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i(P)$$

vanishes at the points $R_1, \ldots, R_{g-1}, S_1, \ldots, S_{g-1}$, *where the divisor* $R_1 \cdots R_{g-1} S_1 \cdots S_{g-1}$ *is a canonical divisor.*

Proposition 3 *If* $f \equiv \Phi_{P_0}(R_1 \cdots R_{g-1}) + K_{P_0}$, *then* $\theta(\Phi_Q(P) - f)$ *vanishes identically on the surface whenever* $i(QR_1 \cdots R_{g-1}) > 0$.

Proposition 4 *If* $f \equiv \Phi_{P_0}(R_1 \cdots R_{g-1}) + K_{P_0}$ *and* $i(R_1 \cdots R_{g-1}) \geq 2$, *then* $\theta(\Phi_Q(P) - f)$ *vanishes identically on the surface for any choice of base point Q. Furthermore,* $\frac{\partial \theta(\pm f)}{\partial z_i} = 0$ *for each* $i = 1, \ldots, g$.

3 Multiple Zeros of $\theta(\alpha \Phi_{P_0}(P) - e)$

Theorem 2 *Let* $e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$ *with* $P_i \neq P_0$. *Assume that* $i(P_0^\alpha P_1 \cdots P_{g-1}) = 0$. *Then* $\theta(\alpha \Phi_{P_0}(P) - e)$ *is nonidentically vanishing. Let*

$Q \neq P_0$ be a Weierstrass point for the linear space $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ and assume that Q is a simple Weierstrass point. Then Q is also a simple zero of $\theta(\alpha \Phi_{P_0}(P) - e)$. Conversely, if Q is a simple zero of $\theta(\alpha \Phi_{P_0}(P) - e)$, then Q is also a simple Weierstrass point of $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$.

Proof Recall that P_0 is always a zero of $\theta(\alpha \Phi_{P_0}(P) - e)$ independently of being a Weierstrass point of $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$. $Q \neq P_0$ a simple Weierstrass point means that the "drops" at Q are $1, 2, \ldots, \alpha - 1, \alpha + 1$, so that the basis adapted to Q, $f_1, \ldots, f_\alpha$ has the property that $ord_Q f_i = i - 1$ for $i = 1, \ldots, \alpha - 1$ and $ord_Q f_\alpha = \alpha$. In other words, we have the divisor of f_i,

$$(f_i) = \frac{Q^{i-1} \Delta_i}{P_0^\alpha P_1 \cdots P_{g-1}}, \quad i = 1, \ldots, \alpha - 1,$$

and

$$(f_\alpha) = \frac{Q^\alpha \Delta_\alpha}{P_0^\alpha P_1 \cdots P_{g-1}}.$$

The divisor Δ_α will play a special role, so we write

$$\Delta_\alpha = T_1 \cdots T_{g-1}$$

and observe that by construction Q is not in the support of any Δ_i. In particular, $Q \neq T_i$ for any i. Consider now $g_i = \frac{f_i}{f_\alpha}$. This is clearly a basis for the space $L[\frac{1}{Q^\alpha \Delta_\alpha}]$. Moreover we have

$$(g_i) = \frac{\Delta_i}{Q^{\alpha - i + 1} \Delta_\alpha}, \quad i = 1, \ldots, \alpha - 1,$$

and $g_\alpha \equiv 1$.

Our objective now is to show that $i(QT_1 \cdots T_{g-1}) = i(Q\Delta_\alpha) = 0$. Recall that we already know that $Q \neq T_i$ for any i. To this end, let us assume that $i(QT_1 \cdots T_{g-1}) > 0$. Then we would have a holomorphic differential whose divisor is a multiple of $Q\Delta_\alpha$. Call this differential ω and consider the meromorphic differentials ωg_i. Clearly the divisor of ωg_α is just the divisor of the holomorphic differential ω, but note that the meromorphic differential ωg_i for $i = 1, \ldots, \alpha - 1$ has at most only a singularity at Q, and it is a pole of order precisely $\alpha - i$ for $i = 1, \ldots, \alpha - 1$. In particular, when $i = \alpha - 1$, we get a simple pole. This is impossible, so it means that, in fact, the holomorphic differential ω has a higher-order zero at Q. Recall however that $Q^\alpha \Delta_\alpha \equiv P_0^\alpha P_1 \cdots P_{g-1}$ so that $i(Q^\alpha \Delta_\alpha) = i(P_0^\alpha P_1 \cdots P_{g-1}) = 0$. Thus the highest power possible is $\alpha - 1$, which will give a contradiction from ωg_1. It thus follows that $i(Q\Delta_\alpha) = 0$.

The above is the key ingredient needed to show that Q is only a simple zero of $\theta(\alpha \Phi_{P_0}(P) - e)$. We repeat the argument showing that Q is a zero of $\theta(\alpha \Phi_{P_0}(P) - e)$. Since Q is a Weierstrass point for our space, we know that there is an integral divisor of degree $g - 1$, $T_1 \cdots T_{g-1}$, such that $\frac{Q^\alpha T_1 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}}$ is a principal divisor. It follows that

$$\alpha \Phi_{P_0}(Q) + \Phi_{P_0}(T_1 \cdots T_{g-1}) - \Phi_{P_0}\big(P_0^\alpha P_1 \cdots P_{g-1}\big) \equiv 0.$$

Thus we have

$$\alpha\Phi_{P_0}(Q) - e \equiv \alpha\Phi_{P_0}(Q) - \Phi_{P_0}\big(P_0^{\alpha}P_1\cdots P_{g-1}\big) - K_{P_0}$$
$$\equiv -\Phi_{P_0}(T_1\cdots T_{g-1}) - K_{P_0}.$$

Denoting $\Phi_{P_0}(T_1\cdots T_{g-1}) + K_{P_0}$ by f, we have

$$\theta\big(\alpha\Phi(Q) - e\big) = \theta(-f),$$

and this vanishes by the Riemann vanishing theorem and the fact that the theta function is even. The condition that Q be a simple zero is now the condition that

$$\frac{d}{dz}\theta\big(\alpha\Phi_{P_0}(Q) - e\big) \neq 0,$$

where z is a local coordinate at Q. A simple computation shows that

$$\frac{d}{dz}\theta\big(\alpha\Phi_{P_0}(Q) - e\big) = \alpha\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i(Q),$$

where $\theta_i(P)$ is the normalized basis of the holomorphic differentials dual to the canonical homology basis which gave rise to the theta function.

Our propositions above tell us that

$$\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i(P)$$

is either identically zero in the case that $i(T_1\cdots T_{g-1}) \geq 2$, which is not the case here since it would imply that one could find a representative of $T_1\cdots T_{g-1}$ with a Q in its support which is ruled out by the fact that the Weierstrass point was simple, or that it vanishes only at the points $T_1,\dots,T_{g-1}, R_1,\dots,R_{g-1}$ with $T_1\cdots T_{g-1}R_1\cdots R_{g-1}$ canonical. We showed however that $i(QT_1\cdots T_{g-1}) = 0$, so this is also ruled out. This shows that the zero of at the point Q of $\theta(\alpha\Phi_{P_0}(P) - e)$ is simple.

We now treat the converse and shall assume that Q is a simple zero of $\theta(\alpha\Phi_{P_0}(P) - e)$. This already implies that $\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i(Q) \neq 0$, so that it is clear that $Q \neq T_i,\ i(QT_1\cdots T_{g-1}) = 0$, and $i(T_1\cdots T_{g-1}) = 1$. It thus follows that the largest "drop" is $\alpha + 1$ and thus that there is precisely one non-"drop" in $1, 2, \dots, \alpha$. If the non-"drop" is α, then the Weierstrass point is simple, and we are done. If the non-"drop" is any positive integer less than α, then we have that α and $\alpha + 1$ are both "drops". This implies that there are functions in the space with zero divisors $Q^{\alpha-1}S_1\cdots S_g$ and $Q^{\alpha}T_1\cdots T_{g-1}$. The equivalence of these divisors yield that $S_1\cdots S_g \equiv QT_1\cdots T_{g-1}$, and since $Q \neq S_i$, this yields that $i(QT_1\cdots T_{g-1}) \geq 1$, which is a contradiction. It thus follows that indeed Q is a simple Weierstrass point for $L[\frac{1}{P_0^{\alpha}P_1\cdots P_{g-1}}]$. This concludes the proof of our result for $k = 1$. □

We now turn to the case $k = 2$ and divide it up into a series of lemmas.

Lemma 4 *Let* $Q \neq P_0$ *be a Weierstrass point for* $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ *such that* $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2} = 2$. *Then there are only two possibilities for the sequence of "drops":*

$$(1) \quad 1, 2, \ldots, \alpha - 1, \alpha + 2;$$

$$(2) \quad 1, 2, \ldots, \alpha - 2, \alpha, \alpha + 1.$$

These correspond to the choices of the largest "drop" being $\alpha + 2$ *or* $\alpha + 1$.

Proof This follows from the formula written for $ord_Q W$ previously. The only possible solutions are $k_1 = 2, l_1 = 0$ or $k_1 = 1, l_1 = 1$. □

Lemma 5 *If the drops are* $1, 2, \ldots, \alpha - 1, \alpha + 2$, *then* $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) = 2$.

Proof We begin by constructing the basis for the space adapted to Q, $f_1, \ldots, f_\alpha$ and observe that

$$(f_\alpha) = \frac{Q^{\alpha+1} T_2 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}},$$

which we shall write as $\frac{Q^\alpha Q T_2 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}}$, thus having $Q = T_1$ in our previous notation. This already shows that $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) \geq 2$, something we already knew from the $k = 1$ case that we have already proven. This also shows that no other $T_i = Q$. We wish to show in addition that if we write $\Delta_\alpha = Q T_2 \cdots T_{g-1}$, then $i(Q^j \Delta_\alpha) = 0$ for all j. We already know this for $j \geq \alpha$. The purpose of this is to show that while indeed in this case

$$\sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i(Q) = 0$$

since in this case $Q = T_1$ or stated otherwise Q is in the support of Δ_α, it is also the case that when $i(Q T_2 \cdots T_{g-1}) = 1$ (which we continue assuming now), then

$$\sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i'(Q) \neq 0,$$

which says that Q is a simple zero of the differential $\sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i(P)$.

We now look again at the basis adapted to Q for our space $f_1, \ldots, f_\alpha$ and observe that just as before we have

$$ord_Q f_i = i - 1, \quad i = 1, \ldots, \alpha - 1,$$

but now we have $ord_Q f_\alpha = \alpha + 1$.

We, as before, construct $g_i = \frac{f_i}{f_\alpha}$ and observe as before that

$$(g_i) = \frac{\Delta_i}{Q^{\alpha-i+1} Q T_2 \cdots T_{g-1}}, \quad i = 1, \ldots, \alpha - 1,$$

so that if $i(Q^j(QT_2\cdots T_{g-1}))$ were positive for any $j \leq \alpha - 1$, we would find a meromorphic differential with a simple pole at Q, which is of course impossible. Hence the conclusion is

$$i(Q\Delta_\alpha) = i\big(Q(QT_2\cdots T_{g-1})\big) = 0,$$

which means that

$$\sum_{i=1}^{g} \frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q) \neq 0,$$

as we wished to show. The conclusion $i(Q\Delta_\alpha) = i(Q(QT_2\cdots T_{g-1})) = 0$ also shows that $i(QT_2\cdots T_{g-1})$ cannot exceed 1. Hence the assumption above that $i(QT_2\cdots T_{g-1}) = 1$ was the only possibility.

The next ingredient that we require is that $\theta(\Phi_Q(P) + f)$ vanishes identically on the surface. This is so because

$$f \equiv \Phi_{P_0}(QT_2\cdots T_{g-1}) + K_{P_0},$$

so that

$$-f \equiv \Phi_{P_0}(R_1\cdots R_{g-1}) + K_{P_0},$$

where $QT_2\cdots T_{g-1}R_1\cdots R_{g-1}$ is a canonical divisor. The result therefore is a consequence of the propositions above.

The identical vanishing of $\theta(\Phi_Q(P) + f)$ on the surface gives us that

$$\theta(f) = 0, \quad \sum_{i=1}^{g} \frac{\partial\theta(f)}{\partial z_i}\theta_i(Q) = 0,$$

$$\sum_{i,j=1}^{g} \frac{\partial^2\theta(f)}{\partial z_i\partial z_j}\theta_i(Q)\theta_j(Q) + \sum_{i=1}^{g} \frac{\partial\theta(f)}{\partial z_i}\theta_i'(Q) = 0.$$

The first equality follows from the vanishing at Q, and the second and third equalities from the vanishing of the first and second derivatives at Q.

In particular, we have

$$\sum_{i,j=1}^{g} \frac{\partial^2\theta(f)}{\partial z_i\partial z_j}\theta_i(Q)\theta_j(Q) = -\sum_{i=1}^{g} \frac{\partial\theta(f)}{\partial z_i}\theta_i'(Q) = \sum_{i=1}^{g} \frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q).$$

Suppose now that $ord_Q\theta(\alpha\Phi_{P_0}(P) - e) > 2$. Then the second derivative with respect to a local coordinate at Q of the function would have to vanish. This would give that

$$\alpha^2\sum_{i,j=1}^{g} \frac{\partial^2\theta(-f)}{\partial z_i\partial z_j}\theta_i(Q)\theta_j(Q) + \alpha\sum_{i=1}^{g} \frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q) = 0.$$

These equalities, together with the fact that θ is an even function of z, give that

$$\sum_{i=1}^{g} \frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q) = 0,$$

which we have seen is a contradiction to our initial assumption on the sequence of “drops”. Hence we have shown that $ord_Q\theta(\alpha\Phi_{P_0}(P)-e)=2$ as we wished to show. This completes the proof of the lemma. □

The second case with the drops being $1,2,\dots,\alpha-2,\alpha,\alpha+1$ is much easier. Here it is automatically the case that $i(\Delta_\alpha)=1$ since there can be no representative with a Q in its support. We begin again constructing the basis for the space adapted to Q, and we see from the functions $f_{\alpha-1}$ and f_α that $Q^{\alpha-1}R_1\cdots R_g\equiv Q^\alpha T_1\cdots T_{g-1}$ obtaining the result that $i(QT_1\cdots T_{g-1})\geq 1$. The conclusion is that

$$\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i(Q)=0,\qquad \theta\bigl(\Phi_Q(P)-f\bigr)\equiv 0.$$

Furthermore we see from the associated functions $g_{\alpha-1},g_{\alpha-2}$ that $i(Q^j\Delta_\alpha)=0$ for all $j\geq 2$. This means that the zero of the differential $\sum_{i=1}^{g}\frac{\partial\theta(f)}{\partial z_i}\theta_i(P)$ at Q is necessarily simple. Using the argument given above, we see that the second condition, together with $ord_Q\theta(\alpha\Phi_{P_0}(P)-e)>2$, will lead to

$$\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q)=0,$$

which is a contradiction. Hence we have in this case as well the conclusion that $ord_Q\theta(\alpha\Phi_{P_0}(P)-e)=2$.

We have therefore shown that if the order of the zero of the Wronskian is two, then the order of the associated theta function at the point is two as well.

We now turn to the converse; so we assume that $ord_Q\theta(\alpha\Phi_{P_0}(P)-e)=2$. The question is what can we say about $\sum_{i=1}^{\alpha}d_i-\frac{\alpha(\alpha+1)}{2}$. The condition we are assuming here yields that

$$\theta(-f)=0,\quad \sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i(Q)=0,$$

but that

$$\alpha^2\sum_{i,j=1}^{g}\frac{\partial^2\theta(-f)}{\partial z_i\partial z_j}\theta_i(Q)\theta_j(Q)+\alpha\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q)\neq 0.$$

We must show that if $\sum_{i=1}^{\alpha}d_i-\frac{\alpha(\alpha+1)}{2}\geq 3$, then

$$\alpha^2\sum_{i,j=1}^{g}\frac{\partial^2\theta(-f)}{\partial z_i\partial z_j}\theta_i(Q)\theta_j(Q)+\alpha\sum_{i=1}^{g}\frac{\partial\theta(-f)}{\partial z_i}\theta_i'(Q)=0.$$

This contradiction will yield the result.

We shall first establish the contradiction when $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2} = 3$. In this case there are three possibilities for the "drops":

$$(1) \quad 1, 2, \ldots, \alpha-1, \alpha+3;$$

$$(2) \quad 1, 2, \ldots, \alpha-2, \alpha, \alpha+2;$$

$$(3) \quad 1, 2, \ldots, \alpha-3, \alpha-1, \alpha, \alpha+1.$$

These correspond to the cases $k_1 = 3,\ l_1 = 0;\ k_1 = 2,\ l_1 = 1;\ k_1 = 1,\ l_1 = 2$.

Case (1) is the easiest one to deal with. Constructing the basis of $L[\frac{1}{P_0^{\alpha} P_1 \cdots P_{g-1}}]$ adapted to Q, we find that the divisor of zeros of f_α is given by

$$(f_\alpha) = \frac{Q^\alpha T_1 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}} = \frac{Q^\alpha Q^2 T_3 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}}.$$

Hence, if we let $f \equiv \Phi_{P_0}(Q^2 T_3 \cdots T_{g-1}) + K_{P_0}$, we find that

$$\sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i(Q) = 0, \qquad \sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i'(Q) = 0.$$

This is true (trivially) also when $i(Q^2 T_3 \cdots T_{g-1}) \geq 2$. Letting now $R_1 \cdots R_{g-1}$ be the complement of $T_1 \cdots T_{g-1} = Q^2 T_3 \cdots T_{g-1}$ with respect to the canonical class, we have $-f \equiv \Phi_{P_0}(R_1 \cdots R_{g-1}) + K_{P_0}$, and therefore we obtain $i(QR_1 \cdots R_{g-1}) > 0$. This implies that $\theta(\Phi_Q(P) + f) \equiv 0$ on the surface, and therefore, as we have seen above, the vanishing of the second derivative, together with the condition $\sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i'(Q) = 0$, leads to the conclusion that $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) \geq 3$, which is a contradiction.

Case (2) is similar, but here we need look at the divisors of zero of the two functions $f_{\alpha-1}, f_\alpha$. Here we find

$$(f_{\alpha-1}) = \frac{Q^{\alpha-1} S_1 \cdots S_g}{P_0^\alpha P_1 \cdots P_{g-1}}, \qquad (f_\alpha) = \frac{Q^\alpha T_1 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}} = \frac{Q^\alpha (Q T_2 \cdots T_{g-1})}{P_0^\alpha P_1 \cdots P_{g-1}}.$$

It thus follows from the fact that these divisors are equivalent that

$$S_1 \cdots S_g \equiv Q^2 T_2 \cdots T_{g-1},$$

from which we can conclude that $i(Q^2 T_2 \cdots T_{g-1}) \geq 1$ and can proceed as we did in case (1).

Case (3) will require a bit more in that not only will we have to look at the divisors of the functions f_j but will also have to look at the divisors of the functions $g_j = \frac{f_j}{f_\alpha}$. We have in this case

$$(f_{\alpha-3}) = \frac{Q^{\alpha-4} U_1 \cdots U_{g+3}}{P_0^\alpha P_1 \cdots P_{g-1}}, \qquad (f_{\alpha-2}) = \frac{Q^{\alpha-2} S_1 \cdots S_{g+1}}{P_0^\alpha P_1 \cdots P_{g-1}},$$

$$(f_{\alpha-1}) = \frac{Q^{\alpha-1} R_1 \cdots R_g}{P_0^\alpha P_1 \cdots P_{g-1}}, \qquad (f_\alpha) = \frac{Q^\alpha T_1 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}},$$

from which we can easily see that

$$R_1 \cdots R_g \equiv Q T_1 \cdots T_{g-1},$$

so that $i(QT_1 \cdots T_{g-1}) \geq 1$. This will not be enough to obtain a contradiction, so we look and find that

$$(g_{\alpha-3}) = \frac{U_1 \cdots U_{g+3}}{Q^4 T_1 \cdots T_{g-1}}, \qquad (g_{\alpha-2}) = \frac{S_1 \cdots S_{g+1}}{Q^2 T_1 \cdots T_{g-1}},$$
$$(g_{\alpha-1}) = \frac{R_1 \cdots R_g}{Q T_1 \cdots T_{g-1}}, \qquad g_\alpha \equiv 1.$$

Let ω be a holomorphic differential in $\Omega(QT_1 \cdots T_{g-1})$ whose existence is guaranteed by $i(QT_1 \cdots T_{g-1}) \geq 1$. Let us assume however that $i(Q^2 T_1 \cdots T_{g-1}) = 0$. If we now multiply the functions g_i by the holomorphic differential ω, we obtain meromorphic differentials with the property that

$$ord_Q \omega g_{\alpha-3} = -3, \qquad ord_Q \omega g_{\alpha-2} = -1$$

and that $\omega g_{\alpha-1}, \omega g_\alpha$ are holomorphic. This is however impossible since one cannot have a meromorphic differential with a simple pole at Q. Hence it must be the case that $i(Q^2 T_1 \cdots T_{g-1}) \geq 1$. This will once again give $\sum_{i=1}^{g} \frac{\partial \theta(-f)}{\partial z_i} \theta_i'(Q) = 0$, which again gives us a contradiction to $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) = 2$.

We have actually shown here in the proof that if $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2} = 3$, then $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) \geq 3$. We believe that this true in general, but have not written down a proof. It is however clear from the above that the same proof will work whenever there is only one "drop" greater than α. However, when $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2} \geq 4$, there can be sequences when there will be more than one "drop" greater than α. For example, when we have equality, we have the sequence of "drops"

$$1, 2, \ldots, \alpha - 2, \alpha + 1, \alpha + 2,$$

where in our previous notation we have $k_1 = 1,\ k_2 = 2,\ l_1 = 0,\ l_2 = 1$. In such a situation we have

$$(f_{\alpha-1}) = \frac{Q^\alpha R_1 \cdots R_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}}, \qquad (f_\alpha) = \frac{Q^\alpha Q T_2 \cdots T_{g-1}}{P_0^\alpha P_1 \cdots P_{g-1}},$$

so that $i(T_1 \cdots T_{g-1}) = i(QT_2 \cdots T_{g-1}) \geq 2$. In this case the Riemann vanishing theorem says that $\frac{\partial \theta(f)}{\partial z_i} = 0$ for all i, so that one gets the usual contradiction to $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) = 2$. Another way of saying this perhaps is that since $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2}$ is simply the order of the zero of the Wronskian at Q, raising the order simply puts on "more" conditions and does not remove any prior ones.

We have therefore concluded the proof of the following:

Theorem 3 *For $k = 1, 2$, we have that $\sum_{i=1}^{\alpha} d_i - \frac{\alpha(\alpha+1)}{2} = k$ is equivalent to $ord_Q \theta(\alpha \Phi_{P_0}(P) - e) = k$.*

4 Classical Weierstrass Points

We would like to show that our ideas above are indeed generalizations of the classical situation or, in other words, that an appropriate choice of e and α gives rise to the classical Weierstrass points on the surface. Let us consider

$$e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0},$$

where the points $P_1, \ldots, P_{g-1}$ satisfy the condition that $i(P_0^{g-1} P_1 \cdots P_{g-1}) = 1$. In this case $\alpha \geq g$, and let us consider $\theta(g\Phi_{P_0}(P) - e)$. Here the space $L[\frac{1}{P_0^g P_1 \cdots P_{g-1}}]$ is g-dimensional, and $Q \neq P_0$ is a zero of $\theta(g\Phi_{P_0}(P) - e)$ and a Weierstrass point for the space if and only if there is a function $f \in L[\frac{1}{P_0^g P_1 \cdots P_{g-1}}]$ with $(f) = \frac{Q^g T_1 \cdots T_{g-1}}{P_0^g P_1 \cdots P_{g-1}}$.

We claim that in this case, Q must be a classical Weierstrass point. The proof is immediate. By hypothesis $i(P_0^{g-1} P_1 \cdots P_{g-1}) = 1$, so that there is a holomorphic differential ω with $(\omega) = P_0^{g-1} P_1 \cdots P_{g-1}$. Hence, ωf is a (meromorphic) differential with $(\omega f) = \frac{Q^g T_1 \cdots T_{g-1}}{P_0}$. This is of course not possible, so it is necessarily the case that some point $T_i = P_0$. Without loss of generality, let $T_1 = P_0$. It then follows that $(\omega f) = Q^g T_2 \cdots T_{g-1}$ and indeed Q is a classical Weierstrass point. If we were to write a basis for $L[\frac{1}{P_0^g P_1 \cdots P_{g-1}}]$ adapted to Q, we would find that what we called a "drop" in this case is in fact what classically one calls a "gap". The astute reader will of course see that in fact the function we have constructed here is really $\theta(g\Phi_{P_0}(P) + K_{P_0})$, from which one can readily see that the zeros are classical Weierstrass points.

5 Multiplicity at P_0

In this section we say something about the order of vanishing at the base point P_0. To this end, let $e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$, and let α_0 be the least value of α for which $i(P_0^\alpha P_1 \cdots P_{g-1}) = 0$. Furthermore, let us assume that $P_i \neq P_j$ for $i \neq j$ and that for each point P_i, $i = 1, \ldots, g-1$, there is a function in $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with a simple pole at P_i. Finally, let us also assume that $i(P_1 \cdots P_{g-1}) = 1$. It follows that for any $\alpha \geq \alpha_0$, we have that $\theta(\alpha\Phi_{P_0}(P) - e)$ is nonidentically vanishing on the surface, and we would like to know its order of vanishing at $P = P_0$.

If $\alpha_0 = 1$, then we clearly have $r[\frac{1}{P_0^j P_1 \cdots P_{g-1}}] = j$ for all positive j, which implies that we can find a basis $f_1, \ldots, f_\alpha$ of $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with the property that $f_1 \equiv 1$, $-ord_{P_0} f_i = i$ for $i = 2, \ldots, \alpha$, so that

$$-ord_{P_0} W(f_1, \ldots, f_\alpha) = -ord_{P_0} W(df_2, \ldots, df_\alpha) = (\alpha + 1)(\alpha - 1) = \alpha^2 - 1.$$

If $\alpha_0 \geq 2$, then we have $r[\frac{1}{P_0^j P_1 \cdots P_{g-1}}] = j+1$ for $j = 1, \ldots, \alpha_0 - 1$, $r[\frac{1}{P_0^j P_1 \cdots P_{g-1}}] = j$ for $j = \alpha_0, \ldots, \alpha$. This implies that we can find a basis $f_1, \ldots, f_\alpha$ of $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$

with the property that $f_1 \equiv 1$, $-ord_{P_0} f_i = i-1$ for $i = 2, \ldots, \alpha_0$, $-ord_{P_0} f_i = i$ for $i = \alpha_0 + 1, \ldots, \alpha$. From this it follows that

$$\begin{aligned} -ord_{P_0} W(f_1, \ldots, f_\alpha) &= -ord_{P_0} W(df_2, \ldots, df_\alpha) \\ &= (\alpha_0 - 1)\alpha + (\alpha + 1)(\alpha - \alpha_0) = \alpha^2 - \alpha_0. \end{aligned}$$

The condition assumed above that for each point P_i, $i = 1, \ldots, g-1$, there is a function in $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with a simple pole at P_i implies that we can find a basis for the space $f_1, \ldots, f_\alpha$ with the property that f_1 has a simple pole at P_i and that f_i, $i = 2, \ldots, \alpha$, are all holomorphic at P_i. From here it is clear that the order of the Wronskian determinant of this basis for the space, $W(f_1, \ldots, f_\alpha)$, has a pole of order at most α at each such point P_i. If we assume that the order at each point P_i is actually equal to α, then the polar divisor of this Wronskian is $P_0^{\alpha^2-\alpha_0} P_1^\alpha \cdots P_{g-1}^\alpha$. This Wronskian is a q-differential with $q = \frac{\alpha(\alpha-1)}{2}$, so that its degree is $\alpha(\alpha-1)(g-1)$. Its divisor of zeros must have degree $\alpha^2 g - \alpha_0$. Each zero of this Wronskian is, as we have already seen, a zero of $f(P) = \theta(\alpha \Phi_{P_0}(P) - e)$, and hence we have proven the following:

Theorem 4 *Let $f(P) = \theta(\Phi_{P_0}(P) - e)$, where $e \equiv \Phi_{P_0}(P_1 \cdots P_{g-1}) + K_{P_0}$, $P_i \neq P_O$ for any i, and $i(P_1 \cdots P_{g-1}) = 1$. Let α_0 be the least positive integer such that $i(P_0^{\alpha_0} P_1 \cdots P_{g-1}) = 0$ and assume that for each P_i in the support of $P_1 \cdots P_{g-1}$, there is a function in $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with a simple pole at P_i, and that the order of the pole of the Wronskian determinant at P_i is α. Then the $\alpha^2 g$ zeroes of $f(P)$ are the zeroes of the Wronskian determinant of a basis for $L[\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}]$ with an additional zero of order α_0 at $P = P_0$.*

Let us now take an example with $\alpha = 2$ and $\alpha_0 = 1$. In this case there is a nonconstant function f in $L[\frac{1}{P_0^2 P_1 \cdots P_{g-1}}]$ with a double pole at P_0. There are thus $g+1$ poles and hence a total of $4g$ branch points. There is one branch point over the point at ∞, so that the Wronskian (differential in this case) has $4g-1$ zeros in agreement with our result. If however $\alpha_0 = 2$, then there is a nonconstant function f in $L(\frac{1}{P_0^2 P_1 \cdots P_{g-1}})$ with a simple pole at P_0. Hence the total number of branch points is $2(2g-1) = 4g-2$, again in agreement with our result.

In the discussion above we assumed that $i(P_1 \cdots P_{g-1}) = 1$. Now we shall remove that assumption and assume that $i(P_1 \cdots P_{g-1}) = r \geq 2$. This implies that $\alpha_0 \geq r$. We shall assume that $\alpha_0 = r$. This clearly implies that

$$r\left(\frac{1}{P_1 \cdots P_{g-1}}\right) = r, \qquad r\left(\frac{1}{P_0 P_1 \cdots P_{g-1}}\right) = r, \ldots, \ r\left(\frac{1}{P_0^r P_1 \cdots P_{g-1}}\right) = r,$$
$$r\left(\frac{1}{P_0^{r+1} P_1 \cdots P_{g-1}}\right) = r+1, \ldots, \ r\left(\frac{1}{P_0^\alpha P_1 \cdots P_{g-1}}\right) = \alpha.$$

It thus follows that we can find a basis for this space $f_0, f_1, \ldots, f_{r-1}, f_r, \ldots, f_{\alpha-1}$ with the property that $f_0 \equiv 1$, $f_2, \ldots, f_{r-1}$ are all holomorphic at P_0 and such that for all j, $r \leq j \leq \alpha - 1$, we have $ord_{P_0} f_j = j + 1$. It is then clear that

$$-ord_{P_0}W(f_0,\dots,f_{\alpha-1}) = -ord_{P_0}W(df_1,\dots,df_{\alpha-1}) = (\alpha+r)(\alpha-r)$$
$$= \alpha^2 - r^2.$$

At each point P_i as before assume that $-ord_{P_i}W(df_1,\dots,df_{\alpha-1}) = \alpha$. It thus follows as before that the divisor of zeros of the Wronskian must be of degree

$$\alpha(\alpha-1)(g-1) + \alpha(g-1) + \alpha^2 - r^2 = \alpha^2 g - r^2.$$

It thus follows that we have proven the following theorem.

Theorem 5 *Let $f(P) = \theta(\alpha\Phi_{P_0}(P) - e)$, where $e \equiv \Phi_{P_0}(P_1\cdots P_{g-1}) + K_{P_0}$, $P_i \neq P_O$ for any i, and $i(P_1\cdots P_{g-1}) = r$. Let α_0 be the least positive integer such that $i(P_0^{\alpha_0}P_1\cdots P_{g-1}) = 0$ and assume that for each P_i in the support of $P_1\cdots P_{g-1}$, there is a function in $L[\frac{1}{P_0^\alpha P_1\cdots P_{g-1}}]$ with a simple pole at P_i. Assume that $\alpha_0 = r$. Then the $\alpha^2 g$ zeroes of $f(P)$ are the zeroes of the Wronskian determinant of a basis for $L[\frac{1}{P_0^\alpha P_1\cdots P_{g-1}}]$ and an additional zero of order r^2 at P_0.*

There is an additional interesting fact that we can obtain in the case where $e \equiv \Phi_{P_0}(P_0^{\alpha_0}P_1\cdots P_{g-1}) + K_{P_0}$ and $i(P_1\cdots P_{g-1}) = r = \alpha_0$. Since by the definition of α_0,

$$i\big(P_0^{\alpha_0-1}P_1\cdots P_{g-1}\big) = 1, \qquad i\big(P_0^{\alpha_0}P_1\cdots P_{g-1}\big) = 0,$$

there is no function in $L[\frac{1}{P_0^\alpha P_1\cdots P_{g-1}}]$ with a pole of order α_0 at P_0. This allows us to do the following: For each function $f \in L[\frac{1}{P_0^{\alpha-1}P_1\cdots P_{g-1}}]$, ωf is a holomorphic differential where div(ω) is a multiple of $P_0^{\alpha_0-1}P_1\cdots P_{g-1}$. This means that if $f_0 \equiv 1, f_1,\dots,f_{\alpha_0-1}$ is a basis of $L[\frac{1}{P_0^\alpha P_1\cdots P_{g-1}}]$, then $\omega, \omega f_1,\dots,\omega f_{\alpha_0-1}$ is a basis for a linear space of holomorphic differentials. Recalling that

$$W(\omega,\omega f_1,\dots,\omega f_{\alpha_0-1}) = \omega^{\alpha_0}W(1,f_1,\dots,f_{\alpha_0-1}) = \omega^{\alpha_0}W(df_1,\dots,df_{\alpha_0-1}),$$

we see that when $\alpha = \alpha_0$, the Wronskian of a basis for the space of meromorphic functions can be computed in terms of the Wronskian of a basis of a linear space of holomorphic differentials.

Explicitly, the space of holomorphic differentials is the space $\Omega(Q_1\cdots Q_{g-\alpha_0})$, i.e., the space of holomorphic differentials whose divisors are multiples of the divisor $\zeta = Q_1\cdots Q_{g-\alpha_0}$, where the integral divisor ζ is the complement of the integral divisor $P_0^{\alpha_0-1}P_1\cdots P_{g-1}$ with respect to the canonical class. The Weierstrass points for this space are the points P such that $i(P^{\alpha_0}Q_1\cdots Q_{g-\alpha_0}) > 0$. This also explains the phenomenon of the previous section, where we showed how the classical Weierstrass points enter the discussion.

References

1. Farkas, H.M.: Identities on Compact Riemann Surfaces, Geometric and Algebraical Aspects in Several Complex Variables, Cetraro (Italy), pp. 97–106 (1989)

2. Farkas, H.M.: Weierstrass Points and the Theta Divisor. Israel Mathematical Conference Proceedings, vol. 14, pp. 65–79 (2000)
3. Farkas, H.M., Kra, I.: Special sets of points on compact Riemann surfaces. Contemp. Math. **256**, 75–94 (2000)
4. Farkas, H.M., Kra, I.: Riemann Surfaces, 2nd edn. Springer, Berlin (1991)
5. Lewittes, J.: Riemann surfaces and the theta function. Acta Math. **111**, 37–61 (1964)

The Admissibility Theorem for the Spatial X-Ray Transform over the Two-Element Field

Eric L. Grinberg

Abstract We consider the Radon transform along lines in an n-dimensional vector space over the two-element field. It is well known that this transform is injective and highly overdetermined. We classify the minimal collections of lines for which the restricted Radon transform is also injective. This is an instance of I.M. Gelfand's *admissibility problem*. The solution is in stark contrast to the more uniform cases of the affine hyperplane transform and the projective line transform, which are addressed in other papers (Feldman and Grinberg in Admissible Complexes for the Projective X-Ray Transform over a Finite Field, preprint, 2012; Grinberg in J. Comb. Theory, Ser. A 53:316–320, 1990). The presentation here is intended to be widely accessible, requiring minimum background.

1 Introduction

Dedication and two Mathematical Moments

This paper is dedicated to the memory of Leon Ehrenpreis. His colleagues were fortunate to have countlessly many discussions with him, after seminars (and during), in offices, hallways, and at the lounge blackboard. These served to inspire, energize, and generate many new ideas. The subject of this essay may well have come up in one of these chats.

During graduate school, I learned about the role of *spreads* in integral geometry from Ethan Bolker, via an early, handwritten version of [2], and when I joined Temple University, the concept of linear spreads followed and came up in early conversations with Leon. He found spreads to be useful in his approach to integral geometry, and he formulated a nonlinear variant which he employed in framing his notion of the *nonlinear Radon transform*. See the major work [5] and the review [1] by Carlos Berenstein.

I recall vividly a two-panel chalk board with the level sets of a homogeneous polynomial drawn on one panel, and the heat equation displayed on the other. I also

E.L. Grinberg (✉)
Department of Mathematics, University of Massachusetts Boston, Boston, MA 02125, USA
e-mail: eric.grinberg@umb.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_8, © Springer-Verlag Italia 2012

recall sharing a car ride with Ethan and Leon, from San Francisco to Arcata, CA, on the way to the 1989 AMS summer conference on integral geometry and tomography, which led to [10] and [3]. It is safe to say that the majority of the travel time was devoted to an intensive discussion of Radon transforms (and I hope that this did not impair the safety of the ride). The beautiful California coastline was superceded by admissible line complexes.

The structure of spreads (discussed concretely below) is particularly simple in the case of the hyperplane Radon transform over finite fields, and this can be used to solve the admissibility problem in that context. In contrast, the structure of spreads is not as simple for transforms that integrate over planes of larger codimension, and thus we expect the admissibility problem to have a more complicated solution. Here we investigate the simplest higher-codimension case.

2 The Radon Transform in a Finite Geometry

The theme of integral geometry, in the style introduced by P. Funk and J. Radon and prominent in the work of Leon Ehrenpreis, involves the recovery of functions (or data) from integrals. In applications one might imagine recovering the density distribution of biological tissue from X-ray data. If "all" integrals (X-ray) measurements are available, then the problem is overdetermined. It is natural to look for minimal sets of data (X-rays) with which complete recovery is still possible (even though in applications such minimal measurements may present stability problems). Finding and classifying such minimal families is an instance of I.M. Gelfand's *admissibility* problem [8], which initially occurred in the context of the Plancherel formula for semi-simple Lie groups. In the continuous category, the problem depends in part on the choice of function spaces, mapping properties of integral operators, and smoothness properties of collections of lines. Here we focus on a finite model of integral geometry in which analytic considerations are removed and sets of lines take center stage. In the admissibility theory work of Gelfand and collaborators within the continuous category ($\mathbb{R}^3$ or $\mathbb{C}^3$ and their projective and higher-dimensional analogs) the family of lines meeting a curve (the Chow variety) and the family of lines tangent to a surface occur as admissible complexes [7, 12]. Here we will search for finite analogs of these. For discussions of Radon transforms in finite geometries, see, e.g., [13, 15]. Recent results on admissibility in the context of finite projective spaces may be found in [6].

Starting with the q-element field $\mathbb{F}_q$, one can build lines, planes, vector spaces of dimension n, projective spaces, Grassmannians, and more. It is easy to use counting measure to define the Radon transform taking functions on $\mathbb{F}_q^n$ to functions on the set of k-planes in $\mathbb{F}_q^n$:

$$R_k f(H) = \sum_{\{x \in H\}} f(x),$$

where H is a k-dimensional affine plane in $\mathbb{F}_q^n$. Informally we write

$$R_k : \{\text{point functions}\} \longrightarrow \{k \text{ plane functions}\}.$$

It is natural to ask: *is the transform R_k invertible?* Rather than answer the question in this specific case, we consider a more general context, informally borrowing from S.S. Chern's (1942) formulation of integral geometry [4]. Consider the following *double fibration diagram*:

$$\begin{array}{ccc} & Z & \\ {\scriptstyle\pi}\swarrow & & \searrow{\scriptstyle\rho} \\ X & & Y \end{array}$$

Chern's formulation was presented in the continuous category; here X, Y, and Z are finite sets. We think of X as our space of points, Y as the family of lines or, more generally, submanifolds of X, and we think of Z as the *incidence manifold* of point-line (or point-generalized line) pairs, so that the point belongs to the line

$$\big\{(x, y)|x \in y\big\}.$$

The maps π and ρ are projection functions, e.g., $\pi(x, y) = x$, so that $\pi \times \rho$ is one to one. Thus thinking of X as a set of points and Y as a family of subsets of X is manifested by [11]:

$$F_y = \pi \circ \rho^{-1}\{y\}.$$

When y is a line, F_y is the set of points on the line. Dually, for every point x, we have the set of all subsets y passing through x:

$$G_x = \rho \circ \pi^{-1}\{x\}.$$

With the definitions of F_y and G_x, it is possible to relax the condition that π, ρ be projections and consider more general diagrams, though we will not need these here. The double fibration diagram has been used extensively as a paradigm for Radon transforms and their generalizations by Gelfand and collaborators, S. Helgason, V. Guillemin and S. Sternberg, and many many others.

A double fibration diagram together with a choice of measures leads to an integral transform. In the finite category we will use counting measure and define the notion of Radon transform without making any additional choices.

Let $C(X), C(Y)$ denote ($\mathbb{R}$- or $\mathbb{C}$-valued) functions on sets X, Y, respectively, and let $f(x), g(y)$ be functions in the appropriate spaces; then we define the Radon transform from point functions to line functions by "integrating" over points in a line and the dual Radon transform by reversing the role of points and lines:

$$R : C(X) \longrightarrow C(Y), \qquad R^t : C(Y) \longrightarrow C(X),$$

$$Rf(y) \equiv \sum_{\{x|x\in y\}} f(x), \qquad R^t g(x) \equiv \sum_{\{y|x\in y\}} g(y).$$

With X, Y, Z and the double fibration diagram so general, can anything be said about invertibility of the induced Radon transform? Surprisingly, the answer is affirmative:

Theorem (Bolker [2]) *Assume that the double fibration diagram satisfies the following two conditions*:

- $\#G_x = \alpha \;\; \forall x \in X$ (*uniform count of lines through each point*),
- $\#G_{x_1} \cap G_{x_2} = \beta \; \forall x_1 \neq x_2$ (*uniform count of lines through each point pair*),

for constants α, β *with* $0 \neq \alpha \neq \beta$. *Then the Radon transform associated with the diagram is invertible*, *with an explicit inversion formula*.

The conditions above, bundled together, are now known as the *Bolker condition*, which is used extensively. The proof of the theorem is straightforward.

Proof We first construct a basis for $C(X)$. Let δ_p be the function on X with value 1 at $p \in X$ and 0 elsewhere. Let n be the cardinality of X. Then $\{\delta_x\}_{x\in X}$ is a basis for $C(X)$, which has dimension n. There is a similar basis for $C(Y)$. The matrix of the composed transform $R^t \circ R$ in this basis is

$$\begin{pmatrix} \alpha & \cdots & \beta \\ \vdots & \ddots & \vdots \\ \beta & \cdots & \alpha \end{pmatrix} = (\alpha - \beta)\mathbf{I} + \beta\mathbf{1},$$

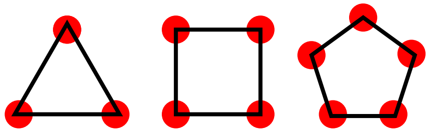
where $\mathbf{I}$, $\mathbf{1}$ denote the $n \times n$ identity matrix, and the $n \times n$ matrix with 1s in every entry, respectively. To invert, note that $(a\mathbf{I} + b\mathbf{1}) \cdot (c\mathbf{I} + d\mathbf{1}) = (ac)\mathbf{I} + (ad + bc + nbd)\mathbf{1}$. □

The Bolker condition is satisfied by many geometric double fibrations but does not hold for many others, even when the Radon transform is injective. The Radon transform on a triangle, rectangle, and pentagon can be represented by the graphs and matrices in Fig. 1. It is easy to verify the properties in Table 1 for the Radon transform on these geometries.

The k-plane transform in $\mathbb{F}_q^n$ satisfies the Bolker condition, since given points x_1, x_2, there is an affine map T that carries x_1 to x_2 and the set of lines through x_1 to the set of lines through x_2, and given two pairs of points, there is an affine map

Table 1 Bolker conditions for finite geometries

# sides	Bolker C. Satisfied?	R injective?
3	Yes	Yes
4	No	No
5	No	Yes



$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Fig. 1 Some finite geometries with and without the Bolker condition, and corresponding matrices

that carries one pair to the other and the lines through one pair to the lines through the other pair. More generally, the Bolker condition holds whenever there is a doubly transitive group action that preserves the appropriate incidence relations. When group symmetry is available, it is natural consider the use of group representations. Interestingly, representation theory can be used to understand Radon transforms on the one hand, e.g., [16], and Radon transforms can be used to understand representation theory, e.g., [11, 14].

We may also inquire about a range characterization: when is a function of k-planes the Radon transform of a function of points? We first look at the hyperplane case, $k = n - 1$.

Definition A *spread* of hyperplanes in $\mathbb{F}_q^n$ is a presentation of $\mathbb{F}_q^n$ as a disjoint union of hyperplanes.

Fact A function $g(H)$ of hyperplanes H in $\mathbb{F}_q^n$ is the Radon transform of a function of points $x \in \mathbb{F}_q^n$ **only if** the average of $g(H)$ over any spread is the same as the average over any other spread:

$$\sum_{\{H \in \Omega_1\}} g(H) = \sum_{\{H \in \Omega_2\}} g(H) \quad \text{(for any two spreads } \Omega_1, \Omega_2).$$

These are called the *Cavalieri conditions*. By way of illustration, in the diagram in Fig. 2, they state that the sum over lines with positive slope equals the sum over lines with negative slope.

Theorem (Bolker) *The Cavalieri conditions characterize the range of the hyperplane Radon transform over a finite field.*

The proof is based on a counting argument. This range condition yields an *admissibility* theorem.

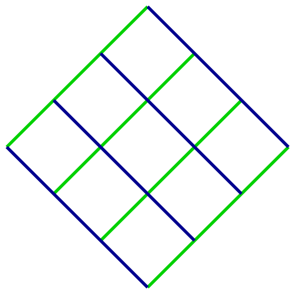


Fig. 2 Two spreads leading to a Cavalieri condition

3 Admissible Complexes

Definition Recall that a **complex** of hyperplanes $\mathscr{C}$ is a collection of hyperplanes $\{H | H \in \mathscr{C}\}$ so that $\#\mathscr{C} = \#\mathbb{F}_q^n = q^n$ (there are as many hyperplanes as points). We shall also use "complex" to denote the appropriate number of lines, curves, etc.

Definition The complex $\mathscr{C}$ is said to be **admissible** if the Radon transform operation, restricted to planes belonging to $\mathscr{C}$ is still injective:

$$R_{\mathscr{C}} : C(\mathbb{F}_q^n) \longrightarrow C(\mathscr{C}).$$

Theorem ([9]) *A complex $\mathscr{C}$ of hyperplanes in $\mathbb{F}_q^n$ is admissible if and only if it omits precisely one plane from each spread, except for one spread, which belongs to $\mathscr{C}$ in its entirety.*

To prove "if", it suffices to show that $R_{\mathscr{C}} f$ determines Rf. A counting argument shows that every complex contains an entire spread. To evaluate Rf on a plane H which $\mathscr{C}$ omits, simply use the total mass of f encoded in a spread that belongs to $\mathscr{C}$ in its entirety. To prove "only if", take two parallel hyperplanes and construct a "capacitor" charge distribution: $+1$ on plane, -1 on the other, and zero elsewhere. Only the two chosen planes can "see" this distribution via the Radon transform. The rest have vanishing Radon transform because of cancellation.

Thus the hyperplane case turns out to be the easy case. We now explore the next simplest, the line transform in $\mathbb{Z}_2^3$. The three-dimensional vector space over $\mathbb{Z}_2$ has 8 points, 7 lines through a given point, 28 lines in all (see Fig. 3).

Here are some ways to construct admissible complexes:

- Write $\mathbb{Z}_2^3$ as a union of two parallel planes (a *spread* of planes) and choose an admissible set of lines on each plane (four lines chosen in each plane).
- Choose one plane $\mathscr{P} \subset \mathbb{Z}_2^3$, choose an admissible set of (four) lines within $\mathscr{P}$, then extend four "*legs*" perpendicular to $\mathscr{P}$.
- Construct, if possible, admissible complexes in $\mathbb{Z}_2^3$ without using planar relatively admissible complexes.

The first two methods are illustrated in Fig. 4.

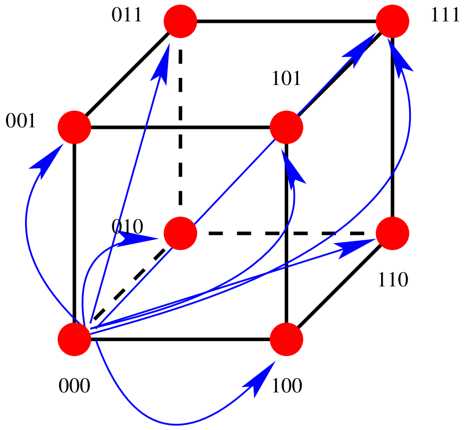


Fig. 3 Lines in $\mathbb{Z}_2^3$

The Radon transform for lines in $\mathbb{Z}_2^3$ can be represented by the following 28×8 matrix:

$$\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}.$$

Fig. 4 Some ways to construct admissible complexes

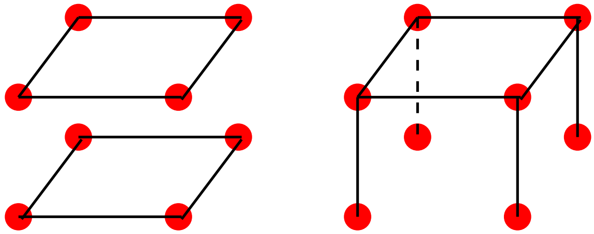


Fig. 5 A brute force program to count admissible complexes

A brute force program to list all admissible complexes:
There are $\binom{28}{8} = 3,108,105$ complexes in $\mathbb{Z}_2^3$.
How many are admissible?

```
a=0
mat= zeros( 28,8)
acc =0
for i=1:7
 for j=i+1:8
  acc=acc+1 ;
  mat(acc,i)=1 ;
  mat(acc,j)=1 ;
 endfor
endfor

for a1=1:21
 for a2=a1+1:22
  for a3=a2+1:23
   for a4=a3+1:24
    for a5=a4+1:25
     for a6=a5+1:26
      for a7=a6+1:27
       for a8=a7+1:28
        minor=zeros(8,8);
        minor(1,1:8) = mat(a1,1:8);
        minor(2,1:8) = mat(a2,1:8);
        minor(3,1:8) = mat(a3,1:8);
        minor(4,1:8) = mat(a4,1:8);
        minor(5,1:8) = mat(a5,1:8);
        minor(6,1:8) = mat(a6,1:8);
        minor(7,1:8) = mat(a7,1:8);
        minor(8,1:8) = mat(a8,1:8);

        a = det(minor)

        if a != 0
 printf("%d,%d,%d,%d,%d,%d,%d,%d",a1,a2,a3,a4,a5,a6,a7,a8)
        endif

       endfor
      endfor
     endfor
    endfor
   endfor
  endfor
 endfor
endfor
```

The admissibility problems asks:

What are the nonsingular 8×8 *minors of this matrix*?

We would like an answer that is geometrically motivated. The linear algebra computer environment `Octave` can be used to locate all admissible complexes for this transform, as illustrated in Fig. 5.

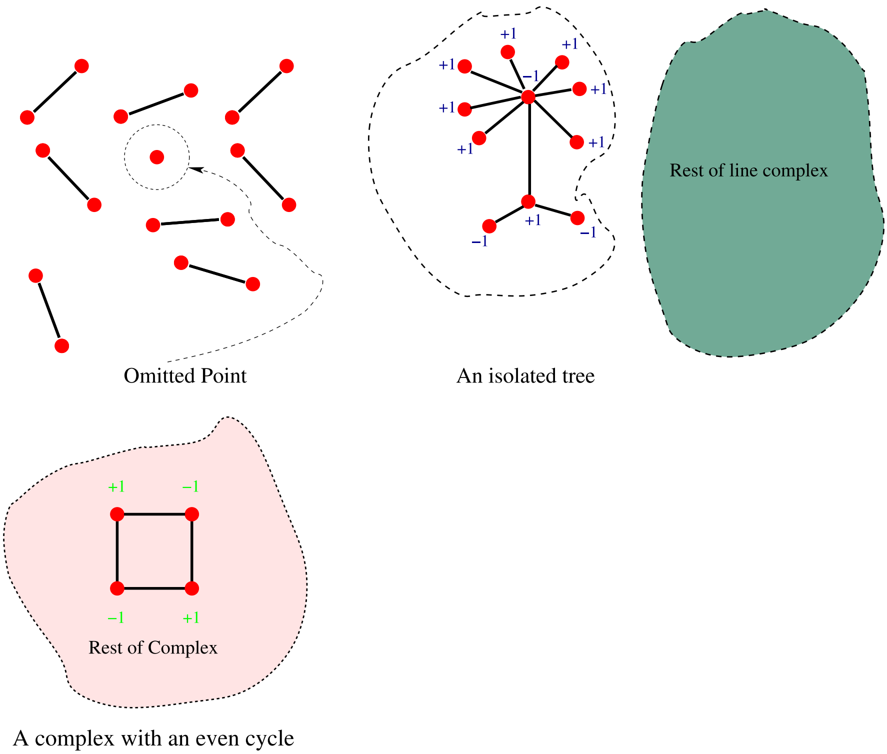


Fig. 6 Some inadmissible configurations

The program results give:

- 3,108,105 line complexes,
- 2,170,667 inadmissible complexes,
- 937,438 admissible complexes.

Can we describe the *moduli space* of admissible complexes? Can we enumerate them without using brute force?

There are some clear obstructions to admissibility. In particular, a complex $\mathscr{C}$ is inadmissible if it has any of the following features.

- An omitted point.
- An isolated tree.
- An even cycle.

Clearly, a line complex that does not pass through a particular point cannot recover data at that point. Similarly, complexes with even cycles or with isolated trees are rank-deficient, as manifested by a $+1, -1$ data pattern. These contexts are illustrated in Fig. 6.

It turns out that these are *the only* obstructions to admissibility.

Theorem (Admissibility for $\mathbb{Z}_2^n$) *Let $\mathscr{C}$ be a line complex in $\mathbb{Z}_2^n$. Assume that $\mathscr{C}$ omits no point, has no isolated trees, and does not contain an even cycle. Then $\mathscr{C}$ is admissible.*

Proof Take a point $p \in \mathbb{Z}_2^n$. There is a line $\ell \in \mathscr{C}$ containing P. Expand ℓ to a maximal connected set of lines, $\mathscr{M}$. Then $\mathscr{M}$ cannot be a tree, so $\mathscr{M}$ contains cycles, hence odd cycles. Each odd cycle is "self inverting". Every point in $\mathscr{M}$ is linked to an odd cycle by a contiguous path of lines, hence is solvable. □

Note Added in Proof The program was mildly corrected, yielding 937,440 admissibles; these have been explicitly enumerated by Mehmet Orhon and the author in a forthcoming preprint.

Acknowledgement The author thanks the referee for helpful comments and suggestions.

Appendix: Counting a Majority of Inadmissible Complexes

Here we will count two basic archetypes of inadmissible complexes, along with their intersection. This will serve to illustrate the combinatorics of the complete count.

A.1 Complexes that Omit One or More Points

First we enumerate complexes that are "missing points", that is, complexes $\mathscr{C}$ so that there exist points $p \in \mathbb{F}_2^3$ so that no line $\ell \in \mathscr{C}$ passes through p. It turns out that there are many of these. There are seven lines through p, so the complexes that miss p have 8 lines chosen from the $28 - 7 = 21$. Now $\binom{21}{8} = 203,440$. Multiplying this by the number of points, 8, and accounting for double counting (because there are complexes that omit more than one point) we obtain:

Lemma 1 *There are* $\binom{21}{8} \times 8 = 1{,}627{,}920$ *complexes that omit points. Here, each complex is counted with multiplicity equal to the number of points in $\mathbb{F}_2^3$ which it misses.*

A.1.1 Complexes that Omit Two or More Points

How many complexes miss two points? There are $7 + 7 - 1 = 13$ lines through one or the other or both points. So a complex that misses both points has 8 lines chosen from among $28 - 13 = 15$ lines. There are 28 pairs of points, so we have double counted $28 \times \binom{15}{8} = 28 \times 6{,}435 = 180{,}180$ complexes. (Note that we have double counted the double counting, because there are complexes that miss three points.)

Lemma 2 *The number of complexes that omit a pair of points is* $28 \times \binom{15}{8} = 28 \times 6{,}435 = 180{,}180$. *Here each complex is counted with multiplicity equal to the number of pairs of points that it misses.*

A.1.2 Complexes that Omit Three or More Points

How many lines pass through one or more of three given points? All but the 10 that form the complete graph on the remaining 5 points. Thus, to exhibit all complexes omitting three or more points, choose three points from 8 and then choose 8 lines from among 10. Thus we have:

Lemma 3 *The number of complexes that omit precisely three points is* $\binom{10}{8} \times \binom{8}{3} = 2,520$. *There are no line complexes that miss four or more points.*

Putting the above lemmas together we have:

Lemma 4 *The number of complexes that avoid one or more points is* $1{,}627{,}920 - 180{,}180 + 2{,}520 = 1{,}450{,}260$. *This count is without multiplicity.*

A.2 Complexes with Isolated Lines

A.2.1 Complexes with One or More Isolated Lines

Another type of nonadmissible complex is one where a single line ℓ is "isolated", i.e., meets no other line in the complex. (This is the simplest case of an isolated tree.) How many of these are there? Well, how many lines meet ℓ? $7 + 7 - 1 = 13 = 28 - 15$. So the number of complexes having ℓ as an isolated line is $\binom{15}{7} = 6{,}435$. Accounting for each of 28 lines, with the usual double counting reminder, we have:

Lemma 5 *There are* $6{,}435 \times 28 = 180{,}180$ *complexes with one or more isolated lines. Each complex is counted with multiplicity equal to the number of isolated lines it has.*

A.2.2 Complexes with Two or More Disjoint Isolated Lines

If ℓ is a line, there are 13 lines meeting ℓ and 15 lines disjoint from l. Thus there are $(28)(15)/2 = 210$ pairs of disjoint lines. Given a complex with a pair of disjoint lines, the other 6 lines of the complex must form the complete graph on the remaining four points. Thus there are 210 complexes with precisely two disjoint isolated lines. Clearly a complex cannot have three disjoint isolated lines.

Lemma 6 *There are* $(28)(15)/2 = 210$ *complexes with precisely two isolated lines, and there are no complexes with three or more isolated lines.*

Lemma 7 *There are* $180{,}180 - 210 = 179{,}970$ *complexes with one or more isolated lines. These complexes are counted without multiplicity.*

A.3 Complexes with Both Omitted Points and Isolated Lines

A.3.1 Complexes with One or More Isolated Lines and One or More Omitted Points

There are five points disjoint from the designated omitted point and the isolated line, hence there are $\binom{5}{2} = 10$ permissibile lines. We must choose 7 lines among these to form a complex, and there are 8×28 point, line pairs.

Lemma 8 *There are no complexes with one isolated line and two omitted points.*

Proof The complement of the union of the omitted points and the isolated line has 4 points, and these form 6 lines, not enough to form a line complex. □

Lemma 9 *There are no complexes with two disjoint isolated lines and an omitted point.*

Proof There are five points in the union of the two lines and point, hence three points left, not enough to span a line complex. □

Lemma 10 *The number of complexes with one isolated line and one omitted point is* $(8 \times 21)\binom{10}{7} = 20{,}160$. *The count is multiplicity free.*

Proof There are $8 \times 21 = 168$ disjoint point-line pairs (or $28 \times 6 = 168$ disjoint line-point pairs). Given a disjoint point-line pair, there are 5 remaining points and $\binom{5}{2} = 10$ lines in their complete graph. Of these we must choose 7 to obtain a line complex. Because of the preceding lemmas, there are no multiplicities. Hence the claimed count is verified. □

We have counted a majority of inadmissible complexes and illustrated the combinatorics of intersections of archetypes. If sufficient interest develops, we will post a completion of this analysis.

References

1. Berenstein, C.A.: Review of The universality of the Radon transform by Leon Ehrenpreis. Math. Reviews, MR2019604 (2007)
2. Bolker, E.D.: The Finite Radon Transform. Integral Geometry (Brunswick, Maine, 1984). Contemp. Math., vol. 63, pp. 27–50. Am. Math. Soc., Providence (1987)
3. Bolker, E.D., Grinberg, E.L., Kung, J.P.: Admissible complexes for the combinatorial Radon transform. A progress report. In: Integral Geometry and Tomography, Arcata, CA, 1989. Contemp. Math., vol. 113, pp. 1–3. Am. Math. Soc., Providence (1990)
4. Chern, S.S.: Integral geometry in Klein spaces. Ann. Math. **43**(2), 403–422 (1942)
5. Ehrenpreis, L.: The Universality of the Radon Transform, with Appendix by P. Kuchment and E.T. Quinto. Oxford Mathematical Monographs. Clarendon/Oxford University Press, New York (2003)

6. Feldman, D.V., Grinberg, E.L.: Admissible Complexes for the Projective X-Ray Transform over a Finite Field, preprint (2012)
7. Gelfand, I.M., Graev, M.I.: Integral transformations connected with straight line complexes in complex affine space. Dokl. Akad. Nauk SSSR **2**, 809–812 (1961)
8. Gelfand, I.M., Graev, M.I., Vilenkin, N.: Generalized Functions, vol. 5. Academic Press, New York (1966)
9. Grinberg, E.L.: The admissibility theorem for the hyperplane transform over a finite field. J. Comb. Theory, Ser. A **53**, 316–320 (1990)
10. Grinberg, E.L., Quinto, E.T. (eds.): Integral geometry and tomography. In: Proc. AMS-IMS-SIAM Joint Summer Research Conf., Humboldt State U., Arcata, CA, 1989. Contemp. Math., vol. 113. American Mathematical Society, Providence (1990)
11. Guillemin, V., Sternberg, S.: Geometric Asymptotics, Chap. VI and Appendix. Mathematical Surveys, No. 14. American Mathematical Society, Providence (1977)
12. Kirillov, A.A.: On a problem of I.M. Gelfand. Dokl. Akad. Nauk SSSR **137**, 276–277 (2000) (Russian); translated as Sov. Math. Dokl. **2**, 268–269 (1961)
13. Kung, J.P.: The Radon transforms of a combinatorial geometry I. J. Comb. Theory, Ser. A **26**(2), 97–102 (1979)
14. Sternberg, S.: Group Theory and Physics. Cambridge University Press, Cambridge (1994). Appendix C
15. Strichartz, R.S.: Radon inversion—variations on a theme. Am. Math. Mon. **89**(6), 377–384 (1982) 420–423
16. Zelevinskii, A.V.: Generalized Radon transforms in spaces of functions on Grassmann manifolds over a finite field (Russian). Usp. Mat. Nauk **28**(5173), 243–244 (1973)

Microlocal Analysis of Fixed Singularities of WKB Solutions of a Schrödinger Equation with a Merging Triplet of Two Simple Poles and a Simple Turning Point

Shingo Kamimoto, Takahiro Kawai, and Yoshitsugu Takei

Abstract We first show that the WKB-theoretic canonical form of an M2P1T (merging two poles and one turning point) Schrödinger equation is given by the algebraic Mathieu equation. We further show that, in analyzing the structure of WKB solutions of a Mathieu equation near fixed singular points relevant to simple poles of the equation, we can focus our attention on the pole part of the equation so that we may reduce it to the Legendre equation. The Borel transformation of WKB-theoretic transformations thus obtained gives rise to microdifferential relations, which lead to the microlocal analysis of the Borel transformed WKB solutions of an M2P1T equation near their fixed singular points. The fully detailed account of the results will be given in Kamimoto et al. (Exact WKB analysis of a Schrödinger equation with a merging triplet of two simple poles and one simple turning point—its relevance to the Mathieu equation and the Legendre equation, 2011).

1 Introduction

The purpose of this article is to announce the main results of [9] emphasizing the atypical points in its reasoning which cannot be found in earlier papers dealing with seemingly related problems, such as [3] and [8]. As the logical structure of the argument in [9] is intricate, we try to explain the ideas that underlie its formulation of the problem. The target of [9] is the exact WKB analysis of the Schrödinger equation

$$\left(\frac{d^2}{dx^2} - \eta^2 Q(x,a)\right)\psi = 0, \tag{1}$$

In memory of the late Professor Leon Ehrenpreis.

S. Kamimoto (✉)
Graduate School of Mathematical Sciences, University of Tokyo, Tokyo, 153-8914 Japan
e-mail: kamimoto@ms.u-tokyo.ac.jp

T. Kawai · Y. Takei
Research Institute for Mathematical Sciences, Kyoto University, Kyoto, 606-8502 Japan

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_9, © Springer-Verlag Italia 2012

where η is a large parameter, and the potential Q contains a triplet of two simple poles and one simple turning point that merge as the parameter a tends to 0. Here "exact WKB analysis" means WKB analysis based on the Borel transformation with respect to the large parameter η; thus our principal aim is to analyze the singularity structure of the Borel transformed WKB solution $\psi_B(x, a, y)$, which solves the Borel transformed Schrödinger equation

$$\left(\frac{\partial^2}{\partial x^2} - Q(x, a)\frac{\partial^2}{\partial y^2}\right)\psi_B(x, a, y) = 0. \tag{2}$$

Hence the exact WKB analysis belongs to the most favorite field of the late Professor Ehrenpreis, Fourier analysis in the complex domain (see [6]). Our interest in the class of Schrödinger equations with a merging triplet of poles and a turning point originates from our desire to understand the semi-global structure of a Schrödinger equation with two simple poles in its potential. As is now well known (see [12, 13]), a simple pole gives an effect to the Borel transformed WKB solutions that is similar to the effect which a turning point gives. Thus the analysis of the class of Schrödinger equations with two simple poles in their potentials is a natural counterpart of the classes of equations studied in [3] (Schrödinger equations with a merging pair of simple turning points) and in [8] (Schrödinger equations with a merging pair of a simple pole and a simple turning point). One can then easily guess that a WKB-theoretic canonical form of such a Schrödinger equation is the Legendre equation with a large parameter, that is,

$$\left(\frac{d^2}{dx^2} - \eta^2 Q_{\mathrm{Leg}}(x, a)\right)\psi = 0, \tag{3}$$

where

$$Q_{\mathrm{Leg}} = \frac{\lambda}{x^2 - a^2} + \eta^{-2}\left(\frac{\gamma_+}{(x-a)^2} + \frac{\gamma_-}{(x+a)^2}\right) \tag{4}$$

with $\gamma_\pm$ being complex numbers and with λ being an infinite series in η^{-1} with constant coefficients that satisfies an appropriate growth order condition to be discussed later. To emphasize the fact that λ is not a genuine constant but an infinite series, we sometimes call such an equation the ∞-Legendre equation. Parenthetically we note that, in what follows, we basically concentrate our attention to the core part of the potential, that is, $\lambda/(x^2 - a^2)$ by mainly considering the situation where γ_+ and γ_- are 0; this limitation is helpful in clarifying the logical structure of our reasoning by avoiding technical complexities. By the way, in the exact WKB analysis, an important subject is the analytic structure of the Borel transformed WKB solutions near their fixed singularities (see [11, pp. 112–113]; see also [4, 5], and [17]), that is, singularities located at

$$y = -\int_\alpha^x \sqrt{Q(x, a)}\,dx + 2l\int_\alpha^{\tilde{\alpha}} \sqrt{Q(x, a)}\,dx \quad (l = \pm 1, \pm 2, \dots), \tag{5}$$

where α and $\tilde{\alpha}$ are turning points (with a simple pole being regarded as a turning point) of the equation. An important point in [3] and [8] is that the period integral

$$2\int_{\alpha}^{\tilde{\alpha}} \sqrt{Q(x,a)}\,dx \tag{6}$$

tends to 0 when we let a tend to 0; hence by showing that the domain of definition of the transformation operator to the canonical form can be chosen to be independent of a, we can analyze the analytic structure of the Borel transformed WKB solution near a fixed singularity with $|l| \gg 1$. But this time we find that

$$\int_{-a}^{a} \frac{dx}{\sqrt{x^2-a^2}} = \pi i \tag{7}$$

does not change even as a tends to 0. Thus the strategy in [3] and [8] is not effective in this case. To circumvent the problem, we dismantle the potential of its homogeneity and seek for the class of Schrödinger equations which can be transformed to

$$\left(\frac{d^2}{dx^2} - \eta^2 \frac{aA + xB}{x^2 - a^2}\right)\psi = 0 \tag{8}$$

with A and B being infinite series in η^{-1} that are independent of x, that is, the algebraic ∞-Mathieu equation (if we follow the usage of the terminology of [7, p. 98]), which we call the ∞-Mathieu equation for short. In view of the explicit form of the potential in (8), we imagine that the class which we now try to analyze would consist of Schrödinger equations with two simple poles and one simple turning point. Fortunately, this guess turns out to be correct, as is explained in Sect. 2 below. Thus, widening the target class gives a clean result, but the problem is the fact that the Mathieu equation is a notoriously difficult object to analyze. Hence we next contrive to deduce the analytic properties of Borel transformed WKB solutions near the fixed singularities relevant to the pair of simple poles, which was our original target, by "driving off" the simple turning point. This contrivance will be explained in Sect. 4, but here we note the following geometric fact that explains why we introduce an auxiliary parameter ρ into our formulation (see Definition 1 below).

To describe the geometric situation, let A_0 (resp., B_0) denote the degree 0 (in η) part of A (resp., B). Then we can confirm that

$$A_0|_{a=0} \neq 0 \quad \text{(see (50) and (15))} \tag{9}$$

and

$$B_0|_{a=0} = Z_0\rho \quad \text{with } Z_0 = \pm 1 \quad \text{(see (51))}. \tag{10}$$

Now, keeping $a/\rho =: \kappa (\neq 0)$ fixed sufficiently small, we let ρ tend to 0. Then, since the turning point t_0 of (8) is given by

$$-\frac{aA_0}{B_0} = -\frac{\kappa A_0(0)}{Z_0 + \kappa\beta} + O(\rho) \tag{11}$$

with some constant β, it stays away from 0. On the other hand, the simple poles $t = \pm a$ tend to 0. Thus one may expect that the singularity structure of Borel transformed WKB solutions near the fixed singularities relevant to the simple poles can be deduced from that of Borel transformed WKB solutions of the Schrödinger equation whose potential contains two simple poles only, i.e., without a turning point. And this expectation is realized in Sect. 4. In ending Introduction we note that in deducing the results in the final section (Sect. 5) from those in Sect. 2 and Sect. 4, we make full use of microdifferential relations among objects on the Borel plane which are discussed in Sect. 3.

2 Definition of an M2P1T Equation and Its Reduction to the Mathieu Equation

In what follows, U (resp., V and O) denotes a sufficiently small open neighborhood of the origin $\{t \in \mathbb{C}; t = 0\}$ (resp., $\{a \in \mathbb{C}; a = 0\}$ and $\{\rho \in \mathbb{C}; \rho = 0\}$), and $f(t, a, \rho)$ denotes a holomorphic function that has the following form (12) on $U \times V \times O$:

$$f(t, a, \rho) = t\rho g(t, \rho) + \sum_{j \geq 1} a^j f^{(j)}(t, \rho) \tag{12}$$

with

$$g(t, \rho) \text{ and } f^{(j)}(t, \rho) \text{ being holomorphic on } U \times O, \tag{13}$$

$$g(0, \rho) = 1, \tag{14}$$

$$f^{(1)}(0, 0) \neq 0, \tag{15}$$

$$\rho^2 \neq \left(f^{(1)}(0, \rho)\right)^2 \quad \text{for } \rho \text{ in } O. \tag{16}$$

In what follows we use the symbols $f^{(0)}(t, \rho)$ and $\tilde{f}^{(0)}(t, \rho)$ respectively, to denote $t\rho g(t, \rho)$ and $\rho g(t, \rho)$.

Definition 1 Let $f(t, a, \rho)$ be as above, let $g_{\pm}(t)$ be holomorphic functions on U, and let Q denote the following potential:

$$\frac{f(t, a, \rho)}{t^2 - a^2} + \eta^{-2}\left(\frac{g_+(t)}{(t-a)^2} + \frac{g_-(t)}{(t+a)^2}\right) \quad (\eta\text{: a large parameter}). \tag{17}$$

Then the Schrödinger operator

$$\frac{d^2}{dt^2} - \eta^2 Q(t, a, \rho, \eta) \tag{18}$$

is called an M2P1T (merging two poles and one turning point) operator.

Remark 1 For the sake of simplicity, we assume the following condition (19) in Sect. 2:

$$g_+ = g_- = 0. \tag{19}$$

Remark 2 It immediately follows from (14) that (18) for $\rho \neq 0$ has a simple turning point when V is chosen sufficiently small.

Remark 3 It follows from the trivial relation

$$\frac{t\tilde{f}^{(0)} + af^{(1)}}{t^2 - a^2} = \frac{\tilde{f}^{(0)} + f^{(1)}}{2(t-a)} + \frac{\tilde{f}^{(0)} - f^{(1)}}{2(t+a)} \tag{20}$$

that we obtain a sum of simple poles at $a = 0$, not a double pole. Parenthetically we note that assumption (16) guarantees that their residues are different from 0.

Remark 4 The reader might wonder why the assumption about the structure of $\tilde{f}^{(0)}(t, \rho)$ is so restrictive. But, since we want to uniformly deal with the problem for an arbitrarily small parameter $\rho (\neq 0)$, some strict restriction on the structure of $\tilde{f}^{(0)}(t, \rho)$ is inevitable. Actually one will be able to find that the function $x_0^{(0)}(t, \rho)$ given by (38) below cannot be holomorphic on a fixed neighborhood of the origin $\{t = 0\}$ if we choose, for example,

$$\tilde{f}^{(0)}(t, \rho) = t + \rho, \tag{21}$$

although it satisfies

$$\tilde{f}^{(0)}(0, \rho) = \rho, \tag{22}$$

the condition we frequently use in our computation.

The purpose of this section is to show that an M2P1T equation is WKB-theoretically transformed to an ∞-Mathieu equation. We refer the reader to [11, Sect. 2] for the basic properties of "WKB-theoretic transformations," but we note their heuristic explanation as follows: in an intuitive description its core is a formal coordinate transformation from t to $x = x(t, a, \rho, \eta)$ defined by an infinite series

$$x(t, a, \rho, \eta) = \sum_{k \geq 0} x_{2k}(t, a, \rho)\eta^{-2k} \tag{23}$$

which satisfies

$$Q(t, a, \rho, \eta) = \left(\frac{\partial x}{\partial t}\right)^2 \left(\frac{aA + xB}{x^2 - a^2}\right) - \frac{1}{2}\eta^{-2}\{x; t\} \tag{24}$$

for some infinite series

$$A = \sum_{k \geq 0} A_{2k}(a, \rho)\eta^{-2k} \tag{25}$$

and

$$B = \sum_{k \geq 0} B_{2k}(a, \rho)\eta^{-2k}, \tag{26}$$

where $\{x; t\}$ stands for the Schwarzian derivative

$$-2\left(\frac{\partial x}{\partial t}\right)^{1/2} \frac{\partial^2}{\partial t^2}\left(\frac{\partial x}{\partial t}\right)^{-1/2}. \tag{27}$$

In what follows we call the Schrödinger operator

$$\frac{d^2}{dx^2} - \eta^2 \frac{aA + xB}{x^2 - a^2} \tag{28}$$

an ∞-Mathieu operator. Using appropriate growth order conditions that $x_{2k}(t, a, \rho)$, $A_{2k}(a, \rho)$, and $B_{2k}(a, \rho)$ satisfy, we can construct microdifferential operators $\mathscr{X}$ and $\mathscr{Y}$ so that they "intertwine" the Borel transformed M2P1T operator and the Borel transformed ∞-Mathieu operator; we have (Theorem 2)

$$N\mathscr{X} = \mathscr{Y} M_\infty, \tag{29}$$

where M_∞ denotes the Borel transformed ∞-Mathieu operator, and N denotes the Borel transformed M2P1T operator written in (x, y)-variable with the effect of the coordinate change appropriately taken into account (cf. (122) for the concrete form of N). See Sect. 3 for the explicit description of $\mathscr{X}$ in terms of the infinite series x.

In constructing the infinite series x, A, and B, we further expand $x_{2k}(t, a, \rho)$ etc. in powers of a; that is, we will seek for x, A, and B in the form of double series as follows:

$$x = \sum_{j,k \geq 0} x_{2k}^{(j)}(t, \rho) a^j \eta^{-2k}, \tag{30}$$

$$A = \sum_{j,k \geq 0} A_{2k}^{(j)}(\rho) a^j \eta^{-2k}, \tag{31}$$

$$B = \sum_{j,k \geq 0} B_{2k}^{(j)}(\rho) a^j \eta^{-2k}. \tag{32}$$

Substituting these series into (24) and comparing the coefficient of η^0, we find

$$\frac{f(t, a, \rho)}{t^2 - a^2} = \left(\frac{\partial x_0}{\partial t}\right)^2 \frac{aA_0 + x_0 B_0}{x_0^2 - a^2}, \tag{33}$$

where

$$x_0(t, a, \rho) = \sum_{j \geq 0} x_0^{(j)}(t, \rho) a^j, \tag{34}$$

$$A_0(a,\rho) = \sum_{j\geq 0} A_0^{(j)}(\rho)a^j, \tag{35}$$

$$B_0(a,\rho) = \sum_{j\geq 0} B_0^{(j)}(\rho)a^j. \tag{36}$$

After multiplying (33) by $(t^2-a^2)(x_0^2-a^2)$ we compare the coefficient of a^p to find

$$\begin{aligned}
&- f^{(p-2)} + \sum_{j+k+l=p} x_0^{(j)} x_0^{(k)} f^{(l)} \\
&\quad = t^2\Bigg(\sum_{j+k+l=p} \frac{\partial x_0^{(j)}}{\partial t}\frac{\partial x_0^{(k)}}{\partial t} A_0^{(l-1)} + \sum_{j+k+l+m=p} \frac{\partial x_0^{(j)}}{\partial t}\frac{\partial x_0^{(k)}}{\partial t} x_0^{(l)} B_0^{(m)}\Bigg) \\
&\qquad - \Bigg(\sum_{j+k+l=p-2} \frac{\partial x_0^{(j)}}{\partial t}\frac{\partial x_0^{(k)}}{\partial t} A_0^{(l-1)} + \sum_{j+k+l+m=p-2} \frac{\partial x_0^{(j)}}{\partial t}\frac{\partial x_0^{(k)}}{\partial t} x_0^{(l)} B_0^{(m)}\Bigg).
\end{aligned} \tag{37.p}$$

In (37.p) terms whose indices do not meet the requirements should be ignored, as usual. With this convention, (37.p) with $p=0$ or 1 is of a peculiar form. For example, we find

$$t x_0^{(0)2} \tilde{f}^{(0)} = t^2 x_0^{(0)\prime 2} x_0^{(0)} B_0^{(0)}. \tag{37.0}$$

Here, and in what follows, x' stands for $\partial x/\partial t$. Hence we find

$$x_0^{(0)}(t,\rho) = \frac{1}{4B_0^{(0)}}\left(\int_0^t \sqrt{\frac{\tilde{f}^{(0)}(t,\rho)}{t}}\,dt\right)^2, \tag{38}$$

where $B_0^{(0)}$ is a nonzero constant to be fixed later. Then it follows from assumptions (13) and (14) that there exists a holomorphic function $\tilde{x}_0^{(0)}(t,\rho)$ that satisfies

$$x_0^{(0)}(t,\rho) = t\tilde{x}_0^{(0)}(t,\rho) \tag{39}$$

with

$$\tilde{x}_0^{(0)}(0,\rho) = \frac{\rho}{B_0^{(0)}}. \tag{40}$$

Next we consider the case $p=1$. Then, by using (39) we find

$$\begin{aligned}
&2t\tilde{x}_0^{(0)} x_0^{(1)} t \tilde{f}^{(0)} + t^2 \tilde{x}_0^{(0)2} f^{(1)} \\
&\quad = t^2\big(x_0^{(0)\prime 2} A_0^{(0)} + 2x_0^{(0)\prime} x_0^{(1)\prime} x_0^{(0)} B_0^{(0)} + x_0^{(0)\prime 2} x_0^{(1)} B_0^{(0)} + x_0^{(0)\prime 2} x_0^{(0)} B_0^{(1)}\big).
\end{aligned} \tag{37.1}$$

Hence it suffices to solve

$$2x_0^{(0)\prime}x_0^{(1)\prime}x_0^{(0)}B_0^{(0)} + x_0^{(0)\prime 2}x_0^{(1)}B_0^{(0)} - 2\tilde{x}_0^{(0)}x_0^{(1)}\tilde{f}^{(0)} \\ = -x_0^{(0)\prime 2}A_0^{(0)} - x_0^{(0)\prime 2}x_0^{(0)}B_0^{(1)} + \tilde{x}_0^{(0)2}f^{(1)}. \tag{41}$$

Here and in what follows, we use a new variable s given by

$$s = x_0^{(0)}(t, \rho). \tag{42}$$

Using the symbol $\dot{x}$ to denote dx/ds, we then find the following equation (43) with the help of (38):

$$B_0^{(0)}\left(2s\frac{d}{ds} - 1\right)x_0^{(1)}(s, \rho) \\ = -A_0^{(0)} - sB_0^{(1)} + \left[\left(x_0^{(0)\prime}\right)^{-2}\tilde{x}_0^{(0)2}f^{(1)}\right]\big(t(s, \rho), \rho\big), \tag{43}$$

where $t(s, \rho)$ denotes the inverse function of $s = x_0^{(0)}(t, \rho)$. It is clear that (43) admits a solution $x_0^{(1)}(s, \rho)$ that is holomorphic near $s = 0$ for arbitrary constants $A_0^{(0)}$ and $B_0^{(1)}$, which are to be fixed later. Furthermore we can immediately see that

$$x_0^{(1)}(0, \rho) = \frac{1}{B_0^{(0)}}\big(A_0^{(0)} - f^{(1)}(0, \rho)\big), \tag{44}$$

$$\dot{x}_0^{(1)}(0, \rho) = \frac{1}{B_0^{(0)}}\Big(-B_0^{(1)} + Z_0^{-1}\big(z'(0, \rho)f^{(1)}(0, \rho) + f^{(1)\prime}(0, \rho)\big)\Big), \tag{45}$$

where

$$Z_0 = x_0^{(0)\prime}(0, \rho) \tag{46}$$

and

$$z(t, \rho) = \big(x_0^{(0)\prime}(t, \rho)\big)^{-2}\tilde{x}_0^{(0)}(t, \rho)^2. \tag{47}$$

For $p \geq 2$, (37.p) assumes the following form:

$$C_0^{(p)}(\rho) + D_0^{(p)}(\rho)t + t^2\mathscr{E}_0^{(p)} = 0, \tag{48.p}$$

where $C_0^{(p)}$ and $D_0^{(p)}$ are free from t, and $\mathscr{E}_0^{(p)}$ contains in it at least

$$\sum_{j+k+l=p} x_0^{(j)\prime}x_0^{(k)\prime}A_0^{(l-1)} + \sum_{j+k+l+m=p} x_0^{(j)\prime}x_0^{(k)\prime}x_0^{(l)}B_0^{(m)}. \tag{49}$$

One can readily find that $C_0^{(2)}$ is absent in (48.2) and that $D_0^{(2)} = 0$ gives a quadratic constraint on $(A_0^{(0)}, B_0^{(0)})$ (see [9, (1.1.1.17)]). Hence, by assuming $D_0^{(2)} = 0$, we can solve the equation $\mathscr{E}_0^{(2)} = 0$ to find $x_0^{(2)}(t, \rho)$ that is holomorphic near $t = 0$. As one of the most exciting points in our computation becomes visible at the next stage, we hasten to study the situation where $p = 3$; we will come back to the explicit computation of $x_0^{(2)}(t, \rho)$ after the study of the case. For this purpose, we assume that $C_0^{(3)} = 0$. Then a straightforward computation shows that this gives another quadratic constraint on $(A_0^{(0)}, B_0^{(0)})$ (see [9, (1.1.1.19)]). The equations $D_0^{(2)} = C_0^{(3)} = 0$ lead to

$$A_0^{(0)} = f^{(1)}(0, \rho) \tag{50}$$

and

$$B_0^{(0)2} = \rho^2. \tag{51}$$

Thus it follows from (40) and (46) that

$$Z_0^2 = 1. \tag{52}$$

And, by (44) we find the following amazing result:

$$x_0^{(1)}(0, \rho) = 0. \tag{53}$$

Relation (53), together with (52), plays a crucially important role at several points in the reasoning of [9]. As a typical example of such points, we show here how (52) and (53) effect the computation of $x_0^{(2)}(0, \rho)$. To begin with, we rewrite (37.2) explicitly in $s(= x_0^{(0)}(t, \rho))$-variable:

$$\begin{aligned} B_0^{(0)}&\left(2s\frac{d}{ds} - 1\right)x_0^{(2)}(s, \rho) \\ &= -A_0^{(1)} - B_0^{(2)}s - 2\dot{x}_0^{(1)}(s, \rho)A_0^{(0)} \\ &\quad - 2\dot{x}_0^{(1)}(s, \rho)x_0^{(1)}(s, \rho)B_0^{(0)} - 2\dot{x}_0^{(1)}(s, \rho)sB_0^{(1)} \\ &\quad - x_0^{(1)}(s, \rho)B_0^{(1)} - \dot{x}_0^{(1)}(s, \rho)^2 sB_0^{(0)} + \big(z(t, \rho)f^{(2)}(t, \rho) \\ &\quad - \big[t^{-1}\big(x_0^{(0)\prime}(t, \rho)\big)^{-2}\big(\mathscr{B}^{(1)}(t, \rho) - \mathscr{B}^{(1)}(0, \rho)\big)\big]\big)\big|_{t=t(s,\rho)}, \end{aligned} \tag{37$'$.2}$$

where $z(t, \rho)$ is the function given by (47), and

$$\mathscr{B}^{(1)}(t, \rho) = \tilde{f}^{(0)} - x_0^{(1)2}\tilde{f}^{(0)} - 2\tilde{x}_0^{(0)}x_0^{(1)}f^{(1)} - x_0^{(0)\prime 2}\tilde{x}_0^{(0)}B_0^{(0)}. \tag{54}$$

By way of parenthesis, we note that the condition $D_0^{(2)} = 0$ is given by $\mathscr{B}^{(1)}(0, \rho) = 0$. To evaluate the term in the brackets in (37′.2) at $s = 0$, we compute $\partial\mathscr{B}^{(1)}/\partial t|_{t=0}$ to find

$$\rho g'(0, \rho) - 2Z_0 f^{(1)}(0, \rho)x_0^{(1)\prime}(0, \rho) - 2B_0^{(0)}x_0^{(0)\prime\prime}(0, \rho) - B_0^{(0)}\tilde{x}_0^{(0)\prime}(0, \rho). \tag{55}$$

In this computation we have repeatedly used (52) and (53); for example, we have used (53) to claim

$$\big(x_0^{(1)2}\tilde{f}^{(0)}\big)'\big|_{t=0} = x_0^{(1)}\big(\tilde{x}_0^{(0)} f^{(1)}\big)'\big|_{t=0} = 0. \tag{56}$$

Using (52), we further notice a remarkable cancellation of terms in the right-hand side of (37′.2) when they are evaluated at $s = 0$; it follows from (50) that $-2\dot{x}_0^{(1)}(0,\rho)A_0^{(0)}$ is canceled by $-(x_0^{(0)\prime}(0,\rho))^{-2}(-2Z_0 f^{(1)}(0,\rho)x_0^{(1)\prime}(0,\rho))$ in (55), i.e.,

$$-2\dot{x}_0^{(1)}(0,\rho)A_0^{(0)} + 2Z_0^{-2}\big(Z_0^2 A_0^{(0)}\dot{x}_0^{(1)}(0,\rho)\big) = 0. \tag{57}$$

An important implication of (57) is that the canceling terms originally depended on $B_0^{(1)}$ through $\dot{x}_0^{(1)}(0,\rho)$ (see (45)). Furthermore, other $B_0^{(1)}$-dependent terms in the right-hand side of (37′.2), i.e.,

$$-2\dot{x}_0^{(1)}(s,\rho)x_0^{(1)}(s,\rho)B_0^{(0)} - 2\dot{x}_0^{(1)}(s,\rho)sB_0^{(1)} - x_0^{(1)}(s,\rho)B_0^{(1)} - \dot{x}_0^{(1)}(s,\rho)^2 sB_0^{(0)}, \tag{58}$$

also vanish when evaluated at $s = 0$, thanks to (53). It then follows from (37′.2) that

$$B_0^{(0)}x_0^{(2)}(0,\rho) = A_0^{(1)} - f^{(2)}(0,\rho) + \chi_0^{(0)}B_0^{(0)}, \tag{59}$$

where $\chi_0^{(0)}$ is a constant fixed by $g(t,\rho)$ (and $Z_0 = \pm 1$). Thus $x_0^{(2)}(0,\rho)$ is free from $B_0^{(1)}$, and this fact, together with the explicit form of $\dot{x}_0^{(1)}(0,\rho)$ given by (45), enables us to explicitly describe $D_0^{(3)}$ and $C_0^{(4)}$. An important point is that these "cancellations and vanishings" occur for every $p \geq 2$ and that they make the concrete expression of the core parts of $D_0^{(p+1)}$ and $C_0^{(p+2)}$ to be "uniform," as is shown below:

$$C_0^{(p+2)} - 2\left(A_0^{(p-1)} - \frac{A_0^{(0)}}{B_0^{(0)}}B_0^{(p-1)}\right) \quad \text{depends only on } \big(A_0^{(q)}, B_0^{(q)}\big)$$
$$(q \leq p-2) \text{ and given data such as } f^{(q)}(0,\rho)(q \leq p-1), \tag{60}$$

and

$$D_0^{(p+1)} - 2Z_0\left(\frac{A_0^{(0)}}{B_0^{(0)}}A_0^{(p-1)} - B_0^{(p-1)}\right) \quad \text{depends only on } \big(A_0^{(q)}, B_0^{(q)}\big)$$
$$(q \leq p-2) \text{ and given data.} \tag{61}$$

As is clear from (60) and (61) we can determine $(A_0^{(p-1)}, B_0^{(p-1)})$ $(p \geq 2)$ recursively by solving linear equations. (The solvability of the equations is guaranteed by the assumption (16) together with the explicit computations (50) and (51) of $A_0^{(0)}$ and $B_0^{(0)}$.) Here we emphasize the importance of the point that the main parts "$2(A_0^{(p-1)} - A_0^{(0)}B_0^{(p-1)}/B_0^{(0)})$" and "$2Z_0(A_0^{(0)}A_0^{(p-1)}/B_0^{(0)} - B_0^{(p-1)})$" are of the same form for every p. Parenthetically we note that $C_0^{(p+2)}$ (resp., $D_0^{(p+1)}$) read off from (37.$p+2$) (resp., (37.$p+1$)) at first contains $x_0^{(p)}(0,\rho)$ and $x_0^{(p-1)\prime}(0,\rho)$; their "principal parts," the parts which may be dependent on $A_0^{(p-1)}$ and $B_0^{(p-1)}$, are at first respectively given as follows (see [9, Lemma 1.1.2.1]):

$$\big[\big(x_0^{(0)\prime}\big)^2 A_0^{(p-1)} + 2x_0^{(0)\prime} x_0^{(p-1)\prime} A_0^{(0)} + x_0^{(0)\prime 2} x_0^{(p)} B_0^{(0)}\big]\big|_{t=0}, \tag{62}$$

$$\begin{aligned}&\big[2\tilde{x}_0^{(0)} x_0^{(0)\prime} x_0^{(p-1)\prime} B_0^{(0)} + \tilde{x}_0^{(0)} x_0^{(0)\prime 2} B_0^{(p-1)} + 2\tilde{x}_0^{(0)} f^{(1)} x_0^{(p)} \\ &\quad + \big(x_0^{(0)\prime}\big)^2 x_0^{(p-1)\prime} B_0^{(0)}\big]\big|_{t=0}.\end{aligned} \tag{63}$$

Thus the clean and uniform results (60) and (61) are almost miraculous, and at the same time we believe that, without such uniform expressions, it should be impossible to find conditions that would guarantee the recursive solvability of equations $C_0^{(p+2)} = D_0^{(p+1)} = 0$.

Thus a naive way of inductively determining $(x_0^{(p)}, A_0^{(p)}, B_0^{(p)})$ $(p \geq 1)$ is as follows:

In order to find a holomorphic (in t) solution $x_0^{(p)}(t,\rho)$ of (37.p), one first requires $C_0^{(p)} = D_0^{(p)} = 0$; then by rewriting (37.$p$) in $s(= x_0^{(0)}(t,\rho))$-variable we find

$$B_0^{(0)}\left(2s\frac{d}{ds} - 1\right) x_0^{(p)}(s,\rho) = -A_0^{(p-1)} - B_0^{(p)} s + B_0^{(0)} R_0^{(p)}(s,\rho), \tag{37$'$.p}$$

where

$$\begin{aligned} &B_0^{(0)} R_0^{(p)}(s,\rho) \\ &\quad = -\sum_{\substack{i+j+k=p-1\\ k\leq p-2}} \dot{x}_0^{(i)} \dot{x}_0^{(j)} A_0^{(k)} - \sum_{\substack{i+j+k+l=p\\ i,j,k,l\leq p-1}} \dot{x}_0^{(i)} \dot{x}_0^{(j)} x_0^{(k)} B_0^{(l)} \\ &\qquad + \Bigg[\big(x_0^{(0)\prime}(t,\rho)\big)^{-2} t^{-2} \Bigg(\sum_{i+j+k=p-3} x_0^{(i)\prime} x_0^{(j)\prime} A_0^{(k)} \\ &\qquad + \sum_{i+j+k+l=p-2} x_0^{(i)\prime} x_0^{(j)\prime} x_0^{(k)} B_0^{(l)} \\ &\qquad + \sum_{\substack{i+j+k=p\\ k\geq 1}} x_0^{(i)} x_0^{(j)} f^{(k)} + \sum_{\substack{i+j=p\\ i,j\geq 1}} x_0^{(i)} x_0^{(j)} f^{(0)} - f^{(p-2)} \Bigg)\Bigg]\Bigg|_{t=t(s,\rho)}. \end{aligned} \tag{64.p}$$

It is then clear that (37$'$.p) admits a holomorphic solution $x_0^{(p)}(s,\rho)$ for any complex numbers $A_0^{(p-1)}$ and $B_0^{(p)}$, as we have assumed $C_0^{(p)} = D_0^{(p)} = 0$. On the other hand, if we admit (60) and (61), the equation $C_0^{(p)} = 0$ combined with $D_0^{(p-1)} = 0$, a relation required in the preceding stage, will fix $A_0^{(p-3)}$ and $B_0^{(p-3)}$ (for $p \geq 4$), which have not yet been completely fixed so far. At the same time, the condition $D_0^{(p)} = 0$ will be used at the next stage to fix $A_0^{(p-2)}$ and $B_0^{(p-2)}$. Thus the reader might find the reasoning to be somewhat clumsy, particularly because of the unevenness of the indices in question. Hence we present here the core of the more refined induction procedure with some comments on its background. We note that the induction scheme we present below is also suited for the growth order estimation of the functions constructed. See [9, Sects. 1.1.3 and 1.2] for the details.

Let us first prepare some notation. We denote a triplet $\{x_0^{(r)}(s,\rho), A_0^{(r)}, B_0^{(r)}\}$ by $T_0^{(r)}$ and use the symbol $\mathfrak{A}_0(p)$ to mean the assertion that $T_0^{(r)}$ is given for $0 \le r \le p$ so that each of them satisfies the following conditions (65.r)–(69.r):

$$x_0^{(r)}(s,\rho) \text{ is a holomorphic solution of } (37'.r) \text{ near } s=0, \tag{65.r}$$

$$x_0^{(r)}(s,\rho) \text{ depends on } \left(\overrightarrow{A}_0[r-1], \overrightarrow{B}_0[r]\right) \underset{\text{def}}{=} \left(A_0^{(0)}, A_0^{(1)}, \ldots, A_0^{(r-1)}, B_0^{(0)}, B_0^{(1)}, \ldots, B_0^{(r)}\right), \tag{66.r}$$

$$C_0^{(r+3)} \text{ and } D_0^{(r+2)} \text{ depend on } \left(\overrightarrow{A}_0[r], \overrightarrow{B}_0[r]\right), \text{ and } \left(\overrightarrow{A}_0[r], \overrightarrow{B}_0[r]\right) \text{ annihilates them}, \tag{67.r}$$

$$C_0^{(r+3)} - 2\left(A_0^{(r)} - \frac{A_0^{(0)}}{B_0^{(0)}} B_0^{(r)}\right) \text{ is independent of } \left(A_0^{(r)}, B_0^{(r)}\right), \tag{68.r}$$

$$D_0^{(r+2)} - 2Z_0\left(\frac{A_0^{(0)}}{B_0^{(0)}} A_0^{(r)} - B_0^{(r)}\right) \text{ is independent of } \left(A_0^{(r)}, B_0^{(r)}\right). \tag{69.r}$$

Then we obtain the following:

Proposition 1 *The assertion $\mathfrak{A}_0(p)$ is valid for every $p \ge 1$.*

The proof of this proposition is done in an inductive manner (cf. [9, Sect. 1.1.3]). But we imagine that the first reactions to this proposition of the reader might be the following:

[A] Is the claim logically self-contained? For example, the concrete expression (62) (resp., (63)) of $C_0^{(p+2)}$ (resp., $D_0^{(p+1)}$) indicates that we need $x_0^{(p_0+1)}(0,\rho)$ for the description of $C_0^{(p_0+3)}$ and $D_0^{(p_0+2)}$, but $\mathfrak{A}_0(p_0)$ refers to $T_0^{(r)}$ ($r \le p_0$) only.

[B] Well, this may not be a logical question but a rather psychological one. Still, I wonder why (67.p_0) is valid despite the presence of $x_0^{(p_0+1)}$ in $C_0^{(p_0+3)}$; in view of (66.p_0+1), I think that $\overrightarrow{B}_0[r]$ in (67.r) might be $\overrightarrow{B}_0[r+1]$.

So let us first dispel potential sources of such uneasiness. Actually both [A] and [B] are reasonable concerns, and the core of the proof of Proposition 1 is closely related to them. The answer to [A] is rather easy: although $x_0^{(p_0+1)}(s,\rho)$ is not referred to in $\mathfrak{A}_0(p_0)$, the assertion $\mathfrak{A}_0(p_0)$ trivially entails the vanishing of $C_0^{(p_0+1)}$ and $D_0^{(p_0+1)}$, and hence the existence of a holomorphic solution $x_0^{(p_0+1)}(s,\rho)$ of $(37'.p_0+1)$ is guaranteed. Then it follows from $(37'.p_0+1)$ that $x_0^{(p_0+1)}(0,\rho)$ is given by

$$x_0^{(p_0+1)}(0,\rho) = \left(B_0^{(0)}\right)^{-1} A_0^{(p_0)} - R_0^{(p_0+1)}(0,\rho). \tag{70}$$

Thus $x_0^{(p_0+1)}(0,\rho)$ is described by $T_0^{(r)}$ $(r \le p_0)$. Note that $R_0^{(p_0+1)}(s,\rho)$ is determined by $T_0^{(r)}$ $(r \le p_0)$ (cf. $(64.p)$). This concrete expression of $x_0^{(p_0+1)}(0,\rho)$ will also alleviate the anxiety [B]. Still, the reader might wonder:

[B′] How can we proceed with a seemingly rather vague expression like (70)? For example, how can we find $(68.p_0+1)$ and $(69.p_0+1)$, which are needed to proceed one step further, that is, to confirm $\mathfrak{A}_0(p_0+1)$ using the data in $\mathfrak{A}_0(p_0)$?

Well, then, we present the core of the proof of Proposition 1, which will clarify all these.

Remark 5 Here we have tried to follow the late Professor Ehrenpreis in his style of lecturing—how do you find it, Professor Ehrenpreis?

To perform the induction procedure, let us suppose that $\mathfrak{A}_0(p_0)$ is validated. Then, as we noted, to see (70), we have

$$C_0^{(p_0+1)} = D_0^{(p_0+1)} = 0, \tag{71}$$

and hence we can find a holomorphic solution $x_0^{(p_0+1)}(s,\rho)$ of $(37'.p_0+1)$ for any complex number $B_0^{(p_0+1)}$, which meets the requirements $(65.p_0+1)$ and $(66.p_0+1)$. Now, the intriguing part of the proof begins here. Since $\mathfrak{A}_0(p_0)$ entails

$$C_0^{(p_0+2)} = D_0^{(p_0+2)} = 0, \tag{72}$$

we can further find a holomorphic solution $x_0^{(p_0+2)}(s,\rho)$ of $(37'.p_0+2)$ for any complex numbers $A_0^{(p_0+1)}$ and $B_0^{(p_0+2)}$. To confirm $\mathfrak{A}_0(p_0+1)$, we do not make full use of $x_0^{(p_0+2)}(s,\rho)$ but use only $x_0^{(p_0+2)}(0,\rho)$ for the computation of $C_0^{(p_0+4)}$ and $D_0^{(p_0+3)}$. Since it follows from $(37'.p_0+2)$ that

$$B_0^{(0)} x_0^{(p_0+2)}(0,\rho) = A_0^{(p_0+1)} - B_0^{(0)} R_0^{(p_0+2)}(0,\rho), \tag{73}$$

the following Lemma 1 is the key to the proof.

Lemma 1 *Let us suppose that $\mathfrak{A}_0(p_0)$ is validated. Then we find*

$$B_0^{(0)} R_0^{(p_0+2)}(0,\rho) \text{ is free from } B_0^{(p_0+1)}. \tag{74}$$

Before giving the proof of this lemma, we note the following three facts: first, once the lemma is proved, the confirmation of $\mathfrak{A}_0(p_0+1)$ is an easy task as we will note later. Second, although this is a rather obvious comment, the complex number $B_0^{(p_0+2)}$ introduced to define $x_0^{(p_0+2)}(s,\rho)$ is actually irrelevant to $x_0^{(p_0+2)}(0,\rho)$ and has no relevance to the later argument; in validating $\mathfrak{A}_0(p_0+2)$ we may use another complex number $\tilde{B}_0^{(p_0+2)}$ to construct $\tilde{x}_0^{(p_0+2)}(s,\rho)$ needed there, which may be different from $x_0^{(p_0+2)}(s,\rho)$ constructed above for the auxiliary purpose of finding the constant $x_0^{(p_0+2)}(0,\rho)$, which is irrelevant to $B_0^{(p_0+2)}$. Third, the cancellation among several terms to be observed in the proof of Lemma 1 also plays crucially important roles in the estimation of growth orders of $T_0^{(p)}$ etc. (see [C1] and [C2] after Remark 7).

Now we give:

Proof of Lemma 1 In view of (66.r) ($r \le p_0$) we find that the terms in $B_0^{(0)}R_0^{(p_0+2)}(0,\rho)$ which may contain $B_0^{(p_0+1)}$ are those which contain $x_0^{(p_0+1)}$, $\dot{x}_0^{(p_0+1)}$, and $B_0^{(p_0+1)}$ itself. Furthermore we note that $x_0^{(p_0+1)}(0,\rho)$ is seen to be free from $B_0^{(p_0+1)}$ by (70) together with the fact that $R_0^{(p_0+1)}(s,\rho)$ is determined by $T_0^{(r)}$ ($r \le p_0$). Thus we do not worry about $-(\sum_{i+j+k=1}\dot{x}_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)B_0^{(k)}) \times x_0^{(p_0+1)}(0,\rho)$ in our computation. Hence it is enough to examine the contribution from the following terms:

$$-\Bigg(\sum_{i+j+k=1}\dot{x}_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)x_0^{(k)}(0,\rho)\Bigg)B_0^{(p_0+1)}, \tag{75}$$

$$-\Bigg(\sum_{\substack{i+j=p_0+2\\ i,j\le p_0+1}}\dot{x}_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)\Bigg)x_0^{(0)}(0,\rho)B_0^{(0)}$$
$$-\Bigg(\sum_{i+j=p_0+1}\dot{x}_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)\Bigg)\Bigg(\sum_{k+l=1}x_0^{(k)}(0,\rho)B_0^{(l)}\Bigg), \tag{76}$$

$$-2\dot{x}_0^{(0)}(0,\rho)\dot{x}_0^{(p_0+1)}(0,\rho)A_0^{(0)}, \tag{77}$$

and

the terms that appear in the coefficients of the Taylor expansion in s of
$$\big[\big(x_0^{(0)\prime}\big)^{-2}t^{-2}\big(2x_0^{(0)}x_0^{(p_0+1)}f^{(1)}+2x_0^{(1)}x_0^{(p_0+1)}f^{(0)}\big)\big]\big|_{t=t(s,\rho)}. \tag{78}$$

Here we observe the following two facts:

any term that may contain $B_0^{(p_0+1)}$ in (75) and (76) vanishes because of the vanishing of $x_0^{(i)}(0,\rho)$ $(i=0,1)$, (79)

and

$$-2\dot{x}_0^{(0)}(0,\rho)\dot{x}_0^{(p_0+1)}(0,\rho)A_0^{(0)} + 2\big(x_0^{(0)\prime}\big)^{-2}\tilde{x}_0^{(0)}x_0^{(p_0+1)\prime}f^{(1)}\big|_{t=t(0,\rho)} = 0, \tag{80}$$

where the second term in (80) is the unique relevant term in (78). (Cf. Remark 6 below.) It is then evident that (79) (resp., (80)) is a counterpart of (56) (resp., (57)), which we encountered in the computation of $x_0^{(2)}(0,\rho)$. In any event, (79) and (80) clearly prove the lemma. □

Remark 6 Since $x_0^{(p_0+1)}(0,\rho)$ is free from $B_0^{(p_0+1)}$ as noted above, $B_0^{(p_0+1)}$ is not contained in

$$2\big(x_0^{(0)\prime}(0,\rho)\big)^{-2}\tilde{x}_0^{(1)}(0,\rho)\tilde{f}^{(0)}(0,\rho)x_0^{(p_0+1)}(0,\rho), \tag{81}$$

despite the fact that (81) is resembling to the second term in (80) in the sense that (81) originates from

$$\big[\big(x_0^{(0)\prime}\big)^{-2}t^{-2}\big(2x_0^{(1)}x_0^{(p_0+1)}f^{(0)}\big)\big]\big|_{t=t(s,\rho)}, \tag{82}$$

which forms the pair to

$$\big[\big(x_0^{(0)\prime}\big)^{-2}t^{-2}\big(2x_0^{(0)}x_0^{(p_0+1)}f^{(1)}\big)\big]\big|_{t=t(s,\rho)} \tag{83}$$

in (78), the term which generates the second term in (80).

Now Lemma 1 and (73) imply

$$x_0^{(p_0+2)} - A_0^{(p_0+1)}/B_0^{(0)} \text{ depends on only } \big(\overrightarrow{A}_0[p_0], \overrightarrow{B}_0[p_0]\big). \tag{84}$$

On the other hand, $(37'.p_0+1)$ entails

$$B_0^{(0)}\dot{x}_0^{(p_0+1)}(0,\rho) + B_0^{(p_0+1)} = B_0^{(0)}\dot{R}_0^{(p_0+1)}(0,\rho), \tag{85}$$

which also depends on only $(\overrightarrow{A}_0[p_0], \overrightarrow{B}_0[p_0])$.

Substituting those into (62) and (63) with $p = p_0 + 2$, we can validate $(68.p_0+1)$ and $(69.p_0+1)$. Then we can readily choose $(A_0^{(p_0+1)}, B_0^{(p_0+1)})$ so that they satisfy

$$C_0^{(p_0+4)} = D_0^{(p_0+3)} = 0. \tag{86}$$

Thus the induction proceeds. This completes the proof of Proposition 1.

Remark 7 As (84) and (85) show, expressions like (70) nicely fit in with our induction scheme. This is the answer to the query [B′], and the important point in the answer is Lemma 1.

Thus we have formally constructed $T_0^{(p)} = \{x_0^{(p)}, A_0^{(p)}, B_0^{(p)}\}$ for every $p \geq 0$. We can further confirm (see [9, Lemma 1.2.3]) that they actually define a function

$$x_0(t, a, \rho) = \sum_{p \geq 0} x_0^{(p)}(t, \rho) a^p, \tag{87}$$

which is holomorphic on

$$\left\{(t, a, \rho) \in \mathbb{C}^3; |t| < r_0, \rho \neq 0, |a|, |\rho| < M_0, |a/\rho| < N_0\right\}, \tag{88}$$

and constants

$$A_0(a, \rho) = \sum_{p \geq 0} A_0^{(p)}(\rho) a^p \tag{89}$$

and

$$B_0(a, \rho) = \sum_{p \geq 0} B_0^{(p)}(\rho) a^p, \tag{90}$$

which are convergent on

$$\left\{(a, \rho) \in \mathbb{C}^2; \rho \neq 0, |a|, |\rho| < M_0, |a/\rho| < N_0\right\} \tag{91}$$

for some positive constants r_0, M_0, and N_0. Although we do not give the details of the proof here, we note the following core facts [C1] and [C2]. Here we use the symbol $(\sigma.j)$ ($j =$ i, ii, and iii) to denote the following sums in $R_0^{(p_0+1)}(s, \rho)$ (cf. (64.p) with $p = p_0 + 1$):

$$(\sigma.\mathrm{i}) \underset{\mathrm{def}}{=} - \sum_{i+j=p_0} \dot{x}_0^{(i)}(s, \rho) \dot{x}_0^{(j)}(s, \rho) A_0^{(0)} / B_0^{(0)}$$

$$\left(\text{cf. the first sum in } (64.p_0 + 1)\right), \tag{92}$$

$$(\sigma.\mathrm{ii}) \underset{\mathrm{def}}{=} \left[\left(x_0^{(0)\prime}(t, \rho)\right)^{-2} t^{-2} \left(\sum_{i+j=p_0} x_0^{(i)}(s, \rho) x_0^{(j)}(s, \rho) f^{(1)}(t, \rho) / B_0^{(0)} \right) \right] \Bigg|_{t=t(s,\rho)}$$

$$\left(\text{cf. the fifth sum in } (64.p_0 + 1)\right), \tag{93}$$

$$(\sigma.\mathrm{iii}) \underset{\mathrm{def}}{=} \left[\left(x_0^{(0)\prime}(t, \rho)\right)^{-2} t^{-1} \tilde{f}^{(0)}(t) \left(\sum_{\substack{i+j=p_0+1 \\ i,j \geq 1}} x_0^{(i)}(s, \rho) x_0^{(j)}(s, \rho) / B_0^{(0)} \right) \right] \Bigg|_{t=t(s,\rho)}$$

$$\left(\text{cf. the sixth sum in } (64.p_0 + 1)\right). \tag{94}$$

Now in inductively showing the domination of $\{x_0^{(p)}, A_0^{(p)}, B_0^{(p)}\}$ which guarantees the domains of convergence (88) and (91) we at first find that each of these three terms might block the induction reasoning from proceeding. But, fortunately we observe:

[C1] What we encounter in the induction process is the estimation of the integral of the form, say,

$$I(\mathrm{iii}) = \frac{1}{2\pi i}\oint \frac{(\sigma.\mathrm{iii})}{s}\,ds; \tag{95}$$

then by the Taylor expansion of

$$\sum_{\substack{i+j=p_0+1 \\ i,j\geq 1}} x_0^{(i)}(s,\rho)x_0^{(j)}(s,\rho), \tag{96}$$

we find the following from the relation $\tilde{f}^{(0)} = \rho g$:

$$\begin{aligned} \left|I(\mathrm{iii})\right| = \Bigg|\frac{1}{2\pi i}\oint \left(\frac{dt}{ds}\right)^2\left(\frac{s}{t}\right) Z_0\, g(t,\rho)\Bigg\{ &\sum_{\substack{i+j=p_0+1 \\ i,j\geq 1}} x_0^{(i)}(0,\rho)x_0^{(j)}(0,\rho) \\ &+ 2s\Bigg(\sum_{\substack{i+j=p_0+1 \\ i,j\geq 1}} x_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)\Bigg) + O\big(s^2\big)\Bigg\}\frac{ds}{s^2}\Bigg|. \end{aligned} \tag{97}$$

Then in order to make the induction reasoning run smoothly, we use (53); the second sum in the integrand of the right-hand side gives the contribution of the form

$$\frac{1}{2\pi i}\oint 2\Bigg(\sum_{\substack{i+j=p_0+1 \\ i\geq 2,\, j\geq 1}} x_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)\Bigg)\frac{ds}{s}. \tag{98}$$

See [9] for the details which show how this gain in the margin of indices is important in the induction procedures.

[C2] The integral

$$I(\mathrm{i}) = \frac{1}{2\pi i}\oint \frac{(\sigma.\mathrm{i})}{s}\,ds \tag{99}$$

is, notably enough, canceled by the contribution

$$I_0 = \frac{1}{2\pi i}\frac{1}{B_0^{(0)}}\oint \frac{s^2}{t^2}\left(\frac{dt}{ds}\right)^2\Bigg(\sum_{i+j=p_0} \dot{x}_0^{(i)}(0,\rho)\dot{x}_0^{(j)}(0,\rho)\Bigg) f^{(1)}(t,\rho)\frac{ds}{s}, \tag{100}$$

which originates from

$$I(\mathrm{ii}) = \frac{1}{2\pi i}\oint \frac{(\sigma.\mathrm{ii})}{s}\,ds, \tag{101}$$

and, furthermore, $I(\mathrm{ii}) - I_0$ is amenable to the induction procedure, as is shown in [9].

We readily find [C1] and [C2] are reasonable counterparts of (79) and (80), respectively.

Thus we have succeeded in constructing $\{x_0(t,a,\rho),\ A_0(a.\rho),\ B_0(a,\rho)\}$ which satisfies the highest degree (i.e., degree 0) part in η of the required relation (24); hence the reasonable approach to the proof of (24) is to try to construct the perturbation series $\{x=\sum_{k\geq 0} x_{2k}\eta^{-2k},\ A=\sum_{k\geq 0} A_{2k}\eta^{-2k},\ B=\sum_{k\geq 0} B_{2k}\eta^{-2k}\}$ so that they satisfy (24). As we mentioned earlier, we further expand $\{x_{2k}, A_{2k}, B_{2k}\}$ into the power series of a (cf. (30), (31), and (32)), and by comparing the coefficients of a^p in the coefficients of η^{-2n} $(n\geq 1)$ of (24) multiplied by $(t^2-a^2)(x^2-a^2)$ we obtain

$$\begin{aligned}\sum_{\substack{q+r+u=p\\ i+j=n}} x_{2i}^{(q)}x_{2j}^{(r)}f^{(u)} = t^2\Bigg[&\sum_{\substack{q+r+u=p-1\\ i+j+k=n}} x_{2i}^{(q)\prime}x_{2j}^{(r)\prime}A_{2k}^{(u)} + \sum_{\substack{q+r+u+v=p\\ i+j+k+l=n}} x_{2i}^{(q)\prime}x_{2j}^{(r)\prime}x_{2k}^{(u)}B_{2l}^{(v)}\\ &-\frac{1}{2}\sum_{\substack{q+r+u=p\\ i+j+k=n-1}} x_{2i}^{(q)}x_{2j}^{(r)}\{x;t\}_{2k}^{(u)} + \frac{1}{2}\{x;t\}_{2(n-1)}^{(p-2)}\Bigg]\\ &-\Bigg[\sum_{\substack{q+r+u=p-3\\ i+j+k=n}} x_{2i}^{(q)\prime}x_{2j}^{(r)\prime}A_{2k}^{(u)} + \sum_{\substack{q+r+u+v=p-2\\ i+j+k+l=n}} x_{2i}^{(q)\prime}x_{2j}^{(r)\prime}x_{2k}^{(u)}B_{2l}^{(v)}\\ &-\frac{1}{2}\sum_{\substack{q+r+u=p-2\\ i+j+k=n-1}} x_{2i}^{(q)}x_{2j}^{(r)}\{x;t\}_{2k}^{(u)} + \frac{1}{2}\{x;t\}_{2(n-1)}^{(p-4)}\Bigg],\end{aligned} \tag{102}$$

where $\{x;t\}_{2k}^{(q)}$ designates the coefficient of $a^q\eta^{-2k}$ of $\{x;t\}$, that is,

$$\{x;t\} = \sum_{q,k\geq 0}\{x;t\}_{2k}^{(q)}a^q\eta^{-2k}. \tag{103}$$

In view of the resemblance between (37.p) and (102), one expects that the construction and domination of the triplet $T_{2n}^{(r)} = \{x_{2n}^{(r)}(s,\rho),\ A_{2n}^{(r)}(\rho),\ B_{2n}^{(r)}(\rho)\}$ $(n\geq 1)$ may be performed in parallel with the construction and domination of $T_0^{(r)}$, and actually this is really the case. We only note the following facts:

in the recursive construction of $x_{2n}^{(p)}(s,\rho)$ $(p=0,1,2,\ldots)$ the relation

$$x_{2n}^{(0)}(0,\rho)=0 \text{ plays an important role,} \tag{104}$$

assertions similar to [C1] and [C2] (with the appropriate shift of indices)

$$\text{also play important roles,} \tag{105}$$

and

in dominating the growth order of $T_{2n}^{(p)}$ we first dominate $\{x; t\}_{2(n-1)}^{(p)}$ using the induction hypothesis and then employ the similar argument used in dominating $T_0^{(p)}$. (106)

We refer the reader to [9, Sect. 1.2] for the details. Here we content ourselves by quoting the final result which will be used later.

Theorem 1 *Let $Q(t, a, \rho, \eta)$ be a potential of an M2P1T operator given by (17). Then there exist positive constants r_0, M_0, N_0, R_0 and holomorphic functions $A_{2n}(a, \rho)$, $B_{2n}(a, \rho)$, and $x_{2n}(t, a, \rho)$ $(n \geq 0)$ on*

$$\left\{(t, a, \rho) \in \mathbb{C}^3; |t| < r_0, \rho \neq 0, |a|, |\rho| < M_0, |a/\rho| < N_0\right\} \tag{107}$$

for which the following conditions are satisfied there:

$$A(a, \rho, \eta),\ B(a, \rho, \eta),\ \textit{and}\ x(t, a, \rho, \eta)\ \textit{satisfy}\ (24), \tag{108}$$

$$\frac{1}{2}\left|f^{(1)}(0,0)\right| \leq \left|A_0(a, \rho)\right| \leq 2\left|f^{(1)}(0,0)\right|, \tag{109}$$

$$\left|B_0(a, \rho)\right| \leq 2|\rho|, \tag{110}$$

$$\frac{\partial x_0}{\partial t}(t, a, \rho) \neq 0, \tag{111}$$

$$x_0^2(\pm a, a, \rho) = a^2, \tag{112}$$

$$\textit{if}\ t = t_0(a, \rho)\ \textit{satisfies}\ f(t_0, a, \rho) = 0,\ \textit{then}\ a A_0(a, \rho) + x_0(t_0, a, \rho) B_0(a, \rho) = 0\ \textit{holds}, \tag{113}$$

the following estimates hold for $n \geq 1$:

$$\left|A_{2n}(a, \rho)\right| \leq (2n)! R_0^n |\rho|^{1-n}, \tag{114}$$

$$\left|B_{2n}(a, \rho)\right| \leq (2n)! R_0^n |\rho|^{1-n}, \tag{115}$$

$$\left|x_{2n}(t, a, \rho)\right| \leq (2n)! R_0^n |\rho|^{-n}, \tag{116}$$

$$\left|\frac{dx_{2n}}{dt}(t, a, \rho)\right| \leq (2n)! R_0^n |\rho|^{-n}. \tag{117}$$

Remark 8 Although we have presented the results assuming (19), the construction and the domination of $\{x = \sum_{k\geq 0} x_{2k}\eta^{-2k},\ A = \sum_{k\geq 0} A_{2k}\eta^{-2k},\ B = \sum_{k\geq 0} B_{2k}\eta^{-2k}\}$ can be done without the assumption. In this case the potential of the canonical form of an M2P1T equation is

$$\frac{aA + xB}{x^2 - a^2} + \eta^{-2}\left(\frac{g_+(a)}{(x-a)^2} + \frac{g_-(-a)}{(x+a)^2}\right). \tag{118}$$

3 Intertwining the Borel Transformed Schrödinger Operators

As was first observed in [2], the analytic meaning of the formal coordinate transformation becomes most transparent with the help of the Borel transformation. To describe the situation concretely, let us first introduce the inverse function $h(x, a, \rho)$ of $x = x_0(t, a, \rho)$, that is,

$$x = x_0\big(h(x, a, \rho), a, \rho\big),\ t = h\big(x_0(t, a, \rho), a, \rho\big). \tag{119}$$

Since we formally find

$$\psi\big(x_0 + \eta^{-2}x_2 + \eta^{-4}x_4 + \cdots, \eta\big) = \sum_{n\geq 0} \frac{1}{n!}\Big(\sum_{k\geq 1} x_{2k}\eta^{-2k}\Big)^n \frac{\partial^n}{\partial x^n}\psi(x, \eta)\Big|_{x=x_0}, \tag{120}$$

its Borel transform has the form

$$\begin{aligned}&\left(\sum_{n\geq 0}\frac{1}{n!}\left(\sum_{k\geq 1} x_{2k}\big(h(x, a, \rho), a, \rho\big)\left(\frac{\partial}{\partial y}\right)^{-2k}\right)^n \frac{\partial^n}{\partial x^n}\right)\psi_B(x, y)\\ &\quad =: \exp\left(\left(\sum_{k\geq 1} x_{2k}\big(h(x, a, \rho), a, \rho\big)\eta^{-2k}\right)\xi\right) : \psi_B(x, y).\end{aligned} \tag{121}$$

In the right-hand side of (121), and also in what follows, we denote by ξ the symbol of $\partial/\partial x$ and use the ideograms in the symbol calculus of microdifferential operators; in particular the ideogram $:\sigma:$ designates the normal ordered product determined by a symbol σ. We note that $:\sigma:$ makes sense as a microdifferential operator when the formal series σ satisfies some growth order conditions like those we discussed in Theorem 1. (Cf. Theorem 2 below.) See [1] for the details of the symbol calculus. Relation (121) indicates that the structure of Schrödinger equations should be most clearly understood when they are Borel transformed. Actually we find Theorem 2 below by making use of the formal series constructed in Sect. 2.

To state the theorem let us prepare some notations.

Let N denote the Borel transform of an M2P1T operator written in (x, y)-coordinates, that is,

$$N = \left(\frac{\partial h}{\partial x}\right)^{-2}\frac{\partial^2}{\partial x^2} - \frac{\partial^2 h}{\partial x^2}\left(\frac{\partial h}{\partial x}\right)^{-3}\frac{\partial}{\partial x} - Q\left(h(x, a, \rho), a, \rho, \frac{\partial}{\partial y}\right)\frac{\partial^2}{\partial y^2}. \tag{122}$$

We also denote the Borel transform of the ∞-Mathieu equation by M_∞. Using $\{x_{2n}\}_{n\geq 0}$ and the function h in (119), we define

$$r_{2k} = x_{2k}\big(h(x, a, \rho), a, \rho\big), \tag{123}$$

$$r = \sum_{k\geq 1} r_{2k}\eta^{-2k}, \tag{124}$$

$$s = x + r, \tag{125}$$

$$\mathscr{X} =: \left(\frac{\partial h}{\partial x}\right)^{1/2} \left(\frac{\partial s}{\partial x}\right)^{-1/2} \exp(r\xi) :, \tag{126}$$

$$\mathscr{Y} =: \left(\frac{\partial h}{\partial x}\right)^{-3/2} \left(\frac{\partial s}{\partial x}\right)^{3/2} \exp(r\xi) : . \tag{127}$$

To describe the geometric situation we introduce the following set W where C_0, δ_0, and δ_1 are some positive constants:

$$W = \left\{(a, \rho) \in \mathbb{C}^2;\ |a| \leq C_0|\rho|,\ 0 < |\rho| < \delta_0,\ |a| < \delta_1\right\}. \tag{128}$$

With these notations, we can deduce Theorem 2 below from the results in Sect. 2 by using the same reasoning as in the proof of Theorem 2.6 of [8].

Theorem 2 *Let U be a sufficiently small open neighborhood of the closed interval $[-a, a]$. Then, for sufficiently small constants C_0, δ_0, and δ_1, the microdifferential operators $\mathscr{X}$ and $\mathscr{Y}$ intertwine N and M_∞ on $U \times W_0$ with the exception of $(x^2 - a^2)\eta = 0$, that is, we have*

$$N\mathscr{X} = \mathscr{Y} M_\infty \tag{129}$$

with $\mathscr{X}$ and $\mathscr{Y}$ being invertible there.

Although the ∞-Mathieu equation contains infinite series A and B, they satisfy the growth order conditions stated in Theorem 1. The growth order conditions enable us to relate, by microdifferential operators, the Borel transformed ∞-Mathieu operator and the Borel transformed Mathieu operator $M = M(A, B, c_+, c_-)$, that is,

$$M(A, B, c_+, c_-) = \frac{\partial^2}{\partial x^2} - \frac{aA + xB}{x^2 - a^2}\frac{\partial^2}{\partial y^2} - \frac{c_+}{(x-a)^2} - \frac{c_-}{(x+a)^2} \tag{130}$$

with A, B, and $c_\pm$ being genuine constants, as the following Theorem 3 shows.

Theorem 3 *There exist microdifferential operators $\mathscr{A}$ and $\mathscr{B}$ for which the following relation holds*:

$$\mathscr{A}\mathscr{B}M = M_\infty \mathscr{A}\mathscr{B}. \tag{131}$$

The proof is essentially the same as the proof of Theorem 4.1 of [10]; it suffices to define

$$\mathscr{A} =: \exp\left(\sum_{k\geq 1} A_{2k}\eta^{-2k}\right) a\alpha_0 : \tag{132}$$

and

$$\mathscr{B} =: \exp\left(\sum_{d\geq 1} B_{2k}\eta^{-2k}\right) \beta_0 :, \tag{133}$$

where α_0 (resp., β_0) stands for the symbol of $\partial/\partial(aA_0)$ (resp., $\partial/\partial B_0$).

These theorems assert that the microlocal structure of Borel transformed WKB solutions of an M2P1T equation coincides with that of the Mathieu equation. By appropriately representing the action of the microdifferential operator in question as an integro-differential operator acting on multivalued analytic functions, we can deduce informations on the alien derivatives of WKB solutions of an M2P1T equation from those of its canonical equation. To attain this goal, we first show the following:

Theorem 4 *The action of the microdifferential operator $\mathscr{X}$ (given by (126)) upon the Borel transformed WKB solution $\psi_{+,B}$ of the ∞-Mathieu equation is expressed as an integro-differential operator of the form*

$$\mathscr{X}\psi_{+,B} = \int_{-y_+}^{y} K\big(x, a, \rho, y - y', \partial/\partial x\big)\psi_{+,B}\big(x, a, \rho, y'\big)\, dy', \tag{134}$$

where

$$y_+(x, a, \rho) = \int_a^x \sqrt{\frac{aA_0(a, \rho) + xB_0(a, \rho)}{x^2 - a^2}}\, dx, \tag{135}$$

and $K(x, a, \rho, y, \partial/\partial x)$ is a differential operator of infinite order (in the sense of [15]) which is defined on $\{(x, a, \rho, y) \in \mathbb{C}^4;\ (x, a, \rho) \in U \times W,\ |y| < C|\rho|^{1/2}\}$ for some positive constant C. Similar expressions are also available for the action of $\mathscr{A}$ and $\mathscr{B}$ on the Borel transformed WKB solutions of the Mathieu equation.

4 Can We Focus Our Attention on the Simple Poles of the Mathieu Equation?

As we emphasized in Introduction, our original problem was to analyze the singularity structure of Borel transformed WKB solutions near fixed singularities determined by a pair of simple poles contained in the potential. But the canonical equation of an M2P1T equation, i.e., the Mathieu equation, contains a simple turning point besides two simple poles. Unfortunately no effective WKB-theoretic results are known for the Mathieu equation, but T. Koike has succeeded in computing the Voros coefficient for the Legendre equation (private communication; see also [14]). Hence, if we can somehow focus our attention on the simple poles of the Mathieu equation, we will be able to make use of the results of Koike. Actually this expectation is realized in Sect. 5. The problem is what we mean by saying "focus our attention on the pole part." The answer is given by Theorem 5 below. In what follows, $Q_L(z, C, \gamma_+, \gamma_-)$ denotes

$$\frac{aC}{z^2 - a^2} + \eta^{-2}\left(\frac{\gamma_+}{(z - a)^2} + \frac{\gamma_-}{(z + a)^2}\right), \tag{136}$$

and $Q_M(x, A, B, c_+, c_-)$ denotes

$$\frac{aA + xB}{x^2 - a^2} + \eta^{-2}\left(\frac{c_+}{(x-a)^2} + \frac{c_-}{(x+a)^2}\right). \tag{137}$$

Theorem 5 *Let $r_1(>1)$ and r_2 be positive constants with r_2 sufficiently small and denote by Ω_{r_1,r_2} the following set*:

$$\left\{(x, a, A, B) \in \mathbb{C}^4; |x| < r_1|a|, a \neq 0, A \neq 0, |B| < r_2|A|\right\}. \tag{138}$$

Then we can construct infinite series

$$z(x, a, A, B, \eta) = \sum_{k \geq 0} z_{2k}(x, a, A, B)\eta^{-2k} \tag{139}$$

and

$$C(a, A, B, \eta) = \sum_{k \geq 0} C_{2k}(a, A, B)\eta^{-2k} \tag{140}$$

so that they satisfy the following conditions (141)–(145):

$$z_{2k} \text{ and } C_{2k} \text{ are holomorphic on } \Omega_{r_1,r_2}, \tag{141}$$

$$\text{for each fixed constants } a, A, \text{ and } B, \text{ the function } z_0(x, a, A, B) \text{ of } x \text{ is injective on } \left\{x \in \mathbb{C}; |x| < r_1|a|\right\}, \tag{142}$$

$$\left(z_0(\pm a, a, A, B)\right)^2 = a^2, \tag{143}$$

$$\frac{\partial z_0}{\partial x}(x, a, A, B) \neq 0 \quad \text{on } \Omega_{r_1,r_2}, \tag{144}$$

$$Q_M(x, A, B, c_+, c_-) = \left(\frac{\partial z}{\partial x}\right)^2 Q_L\big(z(x, a, A, B, \eta), C, c_+, c_-\big) - \frac{1}{2}\eta^{-2}\{z; x\}. \tag{145}$$

Further the constructed series z and C satisfy the following estimates:

$$\text{for any } \varepsilon > 0, \text{ we can find sufficiently small } r_2 \text{ for which} \tag{146}$$

$$\left|z_{2k}(x, a, A, B)\right| \leq (2k)!\varepsilon^k |aA|^{-k} \tag{146.i}$$

and

$$\left|C_{2k}(a, A, B)\right| \leq (2k)!\varepsilon^k |aA|^{-k} \tag{146.ii}$$

hold on Ω_{r_1,r_2} for every $k \geq 1$.

In parallel with the reasoning in Sect. 3, relation (145), together with estimates (146.i) and (146.ii), entails that the Borel transformed Mathieu operator and the

Borel transformed Legendre operator are intertwined on Ω_{r_1,r_2} by microdifferential operators and that the microdifferential operators enjoy the integral representation similar to (134). The point is that the simple turning point of the Mathieu equation, i.e., $-aA/B$, is necessitated to be outside Ω_{r_1,r_2} for sufficiently small r_2. We refer the reader to [9] for the proof of Theorem 5; the formal construction of the series z and C is rather straightforward, but their estimation is quite intricate.

As Koike has explicitly written down the Voros coefficient for the Legendre-type equation with a large parameter that has the form

$$\left(\frac{d^2}{dz^2} - \eta^2\left(\frac{a\Lambda^2}{z^2-a^2} + \eta^{-1}\frac{\sqrt{a}\Lambda}{z^2-a^2} + \eta^{-2}\frac{az\nu + a^2(\mu^2-1)}{(z^2-a^2)^2}\right)\right)\phi = 0, \quad (147)$$

we prepare Lemma 2 below so that we may make use of Koike's results in Sect. 5.

Lemma 2 *We can rewrite*

$$\left(\frac{d^2}{dz^2} - \eta^2 Q_L(z, C, c_+, c_-)\right)\psi = 0 \quad (148)$$

in the form (147) *if we choose* μ, ν, *and*

$$\Lambda(a, C, \eta) = \sum_{k\geq 0} \Lambda_k(a, C)\eta^{-k} \quad (149)$$

as

$$\mu^2 = 1 + 2(c_+ + c_-), \quad (150)$$

$$\nu = 2(c_+ - c_-), \quad (151)$$

$$\Lambda = \sqrt{C - (\sqrt{a}\eta)^{-2}\left(c_+ + c_- - \frac{1}{4}\right)} - \frac{(\sqrt{a}\eta)^{-1}}{2}. \quad (152)$$

The proof is straightforward.

5 Singularity Structure of the Borel Transformed WKB Solutions of an M2P1T Equation

As stated in Sect. 4, we can focus our attention on the pole part of the Mathieu equation so that the part may be analyzed with the help of the results for the Legendre equation. Hence by the same reasoning as in [8, Sect. 5] (see [4] and [16] for the basic properties of the alien derivative) we obtain the following:

Theorem 6 *Let* $\tilde{\psi}_+(t, a, \rho, \eta)$ *be a WKB solution of a generic* (*i.e.*, $a \neq 0$, $\rho \neq 0$) *M2P1T equation that is normalized at a simple pole* $\{t = a\}$. *Then for every positive* *integer l, we can find positive constants δ_1 and δ_2 so that the following relation (153) holds, where $\Delta_{y=-y_+(t,a,\rho)+l\varpi}$ designates the alien derivative at the fixed singularity $-y_+(t,a,\rho)+l\varpi$, and the suffix B indicates the Borel transform in the parentheses*:

$$
\begin{aligned}
&(\Delta_{y=-y_+(t,a,\rho)+l\varpi}\tilde{\psi}_+)_B(t,a,\rho,y)\\
&\quad=\frac{(-1)^l}{l}\left\{1+(-1)^l-\cosh\left(2\pi i l\sqrt{\frac{\mu^2+\sqrt{\mu^4-\nu^2}}{2}}\right)\right.\\
&\qquad\left.-\cosh\left(2\pi i l\sqrt{\frac{\mu^2-\sqrt{\mu^4-\nu^2}}{2}}\right)\right\}\\
&\qquad\times\left(\exp\left(-l\oint_\gamma \tilde{S}_{\mathrm{odd}}\,dt\right)\tilde{\psi}_+\right)_B(t,a,\rho,y),
\end{aligned}
\tag{153}
$$

where $\tilde{S}_{\mathrm{odd}}$ denotes the odd part of the solution $\tilde{S}$ of the Riccati equation associated with the $M2P1T$ equation, and γ is a closed curve that encircles two simple poles counterclockwise, and

$$\mu^2=1+2\big(g_+(a)+g_-(-a)\big), \tag{154}$$

$$\nu=2\big(g_+(a)-g_-(-a)\big), \tag{155}$$

$$y_+(t,a,\rho)=\int_a^t\sqrt{\frac{f(t,a,\rho)}{t^2-a^2}}\,dt, \tag{156}$$

$$\varpi(a,\rho)=\oint_\gamma\sqrt{\frac{f(t,a,\rho)}{t^2-a^2}}\,dt. \tag{157}$$

Remark 9 The highest degree part in η of $\oint_\gamma \tilde{S}_{\mathrm{odd}}\,dt$ is $\eta\varpi(a,\rho)$.

Acknowledgements The research of the authors has been supported in part by JSPS grants-in-aid No. 22-6971, No. 20340028 and No. 21340029.

References

1. Aoki, T.: Symbols and Formal Symbols of Pseudodifferential Operators. Advanced Studies in Pure Math., vol. 4, pp. 181–208. North-Holland, Amsterdam (1984)
2. Aoki, T., Kawai, T., Takei, Y.: The Bender–Wu analysis and the Voros theory. In: Special Functions, pp. 1–29. Springer, Berlin (1991)
3. Aoki, T., Kawai, T., Takei, Y.: The Bender–Wu Analysis and the Voros Theory. II. Advanced Studies in Pure Math., vol. 54, pp. 19–94. Math. Soc. Japan, Tokyo (2009)
4. Delabaere, E., Pham, F.: Resurgent methods in semi-classical asymptotics. Ann. Inst. Henri Poincaré **71**, 1–94 (1999)
5. Ecalle, J.: Les fonctions résurgentes, I, II, III, Publ. Math. d'Orsay, Univ. Paris-Sud, 1981 (Tome I, II), 1985 (Tome III)

6. Ehrenpreis, L.: The Borel transform. In: Algebraic Analysis of Differential Equations, pp. 119–131. Springer, Berlin (2008)
7. Erdély, A.: Higher Transcendental Functions, III, McGraw-Hill, 1955; reprinted in 1981 by Robert E. Krieger Publishing Company, Malabar, Florida
8. Kamimoto, S., Kawai, T., Koike, T., Takei, Y.: On the WKB theoretic structure of the Schrödinger operators with a merging pair of a simple pole and a simple turning point. Kyoto J. Math. **50**, 101–164 (2010)
9. Kamimoto, S., Kawai, T., Takei, Y.: Exact WKB analysis of a Schrödinger equation with a merging triplet of two simple poles and one simple turning point—its relevance to the Mathieu equation and the Legendre equation, preprint (RIMS-1735), 2011
10. Kawai, T., Koike, T., Takei, Y.: On the exact WKB analysis of higher order simple-pole type operators. Adv. Math. **228**, 63–96 (2011)
11. Kawai, T., Takei, Y.: Algebraic Analysis of Singular Perturbation Theory. Am. Math. Soc., Providence (2005)
12. Koike, T.: On a regular singular point in the exact WKB analysis. In: Toward the Exact WKB Analysis of Differential Equations, Linear or Non-Linear, pp. 39–54. Kyoto Univ. Press, Kyoto (2000)
13. Koike, T.: On the exact WKB analysis of second order linear ordinary differential equations with simple poles. Publ. RIMS, Kyoto Univ. **36**, 297–319 (2000)
14. Koike, T., Takei, Y.: On the Voros coefficient for the Whittaker equation with a large parameter—some progress around Sato's conjecture in exact WKB analysis. Publ. RIMS, Kyoto Univ. **47**, 375–395 (2011)
15. Sato, M., Kawai, T., Kashiwara, M.: Microfunctions and Pseudo-differential Equations. Lect. Notes in Math., vol. 287, pp. 265–529. Springer, Berlin (1973)
16. Sauzin, D.: Resurgent functions and splitting problems. RIMS Kôkyûroku, 1493, RIMS, pp. 48–117 (2006)
17. Voros, A.: The return of the quartic oscillator—the complex WKB method. Ann. Inst. Henri Poincaré **39**, 211–338 (1983)

Geometric Properties of Boundary Orbit Accumulation Points

Steven G. Krantz

Abstract We study the automorphism group action on a bounded domain in $\mathbb{C}^n$. In particular, we consider boundary orbit accumulation points, and what geometric properties they must have. These properties are formulated in the language of Levi geometry.

1 Introduction

In this paper a *domain* $\Omega \subseteq \mathbb{C}^n$ is a connected open set. We let $\mathscr{O}(\Omega)$ denote the algebra of holomorphic functions on Ω. Also we let $\mathrm{Aut}(\Omega)$ be the group (under composition of mappings) of biholomorphic self-maps of Ω. The standard topology on $\mathrm{Aut}(\Omega)$ is that of uniform convergence on compact sets (equivalently, the compact-open topology).

We shall use the following notation: D denotes the unit disc in the complex plane. We let $D^2 = D \times D$ denote the bidisc, and $D^n = D \times D \times \cdots \times D$ the polydisc in $\mathbb{C}^n$. The symbol $B = B^n$ is the unit ball in $\mathbb{C}^n$.

Certainly domains with transitive automorphism group are of some interest. But they are relatively few in number (see the classification theory of Cartan, as described in [11]). A very natural and compelling alternative is to study domains with *noncompact automorphism group*. A bounded domain Ω has noncompact automorphism group if there is a sequence $\varphi_j \in \mathrm{Aut}(\Omega)$ such that no subsequence converges uniformly on compact sets to another automorphism. Obversely, the automorphism group is compact if every sequence $\{\varphi_j\}$ in $\mathrm{Aut}(\Omega)$ has a subsequence that converges uniformly on compact sets to another automorphism. In this regard, the following result of H. Cartan is central and useful (see [23]):

Theorem 1 *Let $\Omega \subseteq \mathbb{C}^n$ be a bounded domain. Then Ω has a noncompact automorphism group if and only if there are a point $P \in \partial\Omega$ and a point $X \in \Omega$ and automorphisms $\varphi_j \in \mathrm{Aut}(\Omega)$ such that $\lim_{j\to\infty} \varphi_j(X) = P$.*

Dedicated to Leon Ehrenpreis, a fine mathematician and a wonderful human being.

S.G. Krantz (✉)
Department of Mathematics, Washington University in St. Louis, St. Louis, MO 63130, USA
e-mail: sk@math.wustl.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_10, © Springer-Verlag Italia 2012

A point $P \in \partial\Omega$ is called a *boundary orbit accumulation point* if there are a point $X \in \Omega$ and a sequence $\varphi_j \in \mathrm{Aut}(\Omega)$ such that $\lim_{j\to\infty} \varphi_j(X) = P$. Of special interest is the case where there is a *single* automorphism ψ such that $\lim_{j\to\infty} \psi^j(X) = P$. (Here ψ^j denotes the composition of ψ with itself j times when $j = 0, 1, 2, \ldots$; also, if $j < 0$, then ψ^j denotes the composition of ψ^{-1} with itself $|j|$ times.) In this latter circumstance we call P a *special boundary orbit accumulation point*. It is not clear when an arbitrary boundary orbit accumulation point is a special boundary orbit accumulation point.

In this paper, a point P in the boundary of a domain

$$\Omega = \left\{ z \in \mathbb{C}^n : \rho(z) < 0 \right\}$$

(where ρ is a *defining function* for Ω) is *Levi pseudoconvex* if the Levi form

$$\mathscr{L}_\rho(z,\lambda) \equiv \sum_{j,k=1}^{n} \frac{\partial^2 \rho}{\partial z_j \partial \overline{z}_k}(P)\lambda_j \overline{\lambda}_k$$

is positive semi-definite for all complex tangent n-vectors λ such that $\sum_j \lambda_j (\partial\rho/\partial z_j)(P) = 0$. The point P is *strictly* or *strongly Levi pseudoconvex* if the Levi form is strictly positive definite on complex tangent vectors. Finally, the point P is *Levi pseudoconcave* if the Levi form has a negative eigenvalue on some complex tangent vector λ at P. It is *strictly* or *strongly Levi pseudoconcave* if all eigenvalues are negative. The book [18] is a good reference for all these matters. The notion of Hartogs pseudoconvexity is useful on domains with nonsmooth boundary. It is equivalent to Levi pseudoconvexity on domains with C^2 boundary. See [18] for the details.

In this paper we concentrate on bounded domains, but add a few remarks about unbounded domains.

We shall use Sect. 1 to collect some simple, preliminary results that have independent interest. It is a pleasure to thank the referee for many helpful suggestions.

2 Background Results

The result that inspires the present work comes from [10]:

Theorem 2 *Let $\Omega \subseteq \mathbb{C}^n$ be a smoothly bounded domain. Suppose that $P \in \partial\Omega$ is a boundary orbit accumulation point. Then P is a point of Levi pseudoconvexity.*

There is an analogous result for domains without smooth boundary (and in which the conclusion involves *Hartogs* pseudoconvexity). But we shall have no use for it in the present paper. See [10] for the details.

For interest's sake, we provide here an alternative formulation of Theorem 2:

Theorem: *If $P \in \partial\Omega$ is a boundary orbit accumulation point, then P does not belong to the holomorphic envelope $\widetilde{\Omega}$ of Ω in the following sense: There is no neighborhood U of P with a holomorphic mapping $\psi : U \to \widetilde{\Omega}$ such that $\psi = (\mathrm{id})^{-1}$ on $U \cap \Omega$.*

We first give an example to emphasize that, even though the boundary orbit accumulation point is pseudoconvex, nearby points need not be.

Example 1 Let $B \subseteq \mathbb{C}^n$ be the unit ball with defining function $\rho(z) = |z|^2 - 1$ (see [18] for the concept of defining function). Let ϕ be a C_c^∞ function on $\mathbb{C}^n$ with these properties:

(a) ϕ is real-valued, and $0 \le \phi(z) \le 1/10$ for all $z \in \mathbb{C}^n$.
(b) ϕ is radial about the point $(i, 0)$.
(c) $\operatorname{supp}\phi \subseteq B((i, 0), 1/10)$.
(d) $\phi(z) = 1/10$ for $|z - (i, 0)| < 1/20$.

Set

$$\Omega' = \left\{z \in \mathbb{C}^2 : -1 + |z|^2 + \phi(z) < 0\right\}$$

and

$$\Omega = \bigcap_{j=-\infty}^{\infty} \Phi_{1/2}^{j}\left(\Omega'\right),$$

where

$$\Phi_a(z_1, z_2) = \left(\frac{z_1 - a}{1 - \overline{a}z_1}, \frac{\sqrt{1 - |a|^2}z_2}{1 - \overline{a}z_1}\right) \tag{1}$$

for $a \in \mathbb{C}$, $|a| < 1$. It is easy to check, by direct calculation, that Φ_a is an automorphism of the unit ball $B \subseteq \mathbb{C}^n$, and the domain Ω will be the unit ball with countably many strongly pseudoconcave dents that accumulate at the points $(1, 0)$ and $(-1, 0)$.

Now it is plain that the point $(1, 0) \in \partial\Omega$ is a boundary orbit accumulation point. In fact we may let $X = (0, 0)$ and $\varphi_j(z) = \Phi_{1/2}^{-j}(z)$ for $j = 1, 2, \ldots$. So $(1, 0) \in \partial\Omega$ is certainly pseudoconvex. Notice that, at points in B along the normal line through $(i, 0)$, $-1 + |z|^2 + \phi(z)$ is negative when z is at least $1/5$ units from the boundary of the ball B, and $-1 + |z|^2 + \phi(z)$ is positive at $(i, 0)$. So there must be an intermediate point $\widetilde{z}$ on this line segment—a point *inside* the unit ball B—where $-1 + |z|^2 + \phi(z)$ vanishes. It follows that $\widetilde{z}$ is a boundary point of Ω'. Hence $\Phi_{1/2}^{j}(\widetilde{z})$ is a boundary point for each j. The boundary points $\Phi_{1/2}^{j}(\widetilde{z})$ will be strictly pseudoconcave.

It must be noted that, in Example 1, Ω does *not* have smooth boundary. In fact, at the boundary points $(1, 0)$, $(-1, 0)$, the boundary is only Lipschitz.

The following example is reasonably well known and gives a nice way to generate example of non-pseudoconvex domains with noncompact automorphism group.

Example 2 Define

$$\Omega = \{(w, z_1, z_2) \in \mathbb{C}^3 : |w|^2 + (|z_1|^2 - |z_2|^2)^2 + \lambda |z_1|^4 < 1\}.$$

This domain is not pseudoconvex. Indeed, at the boundary point (0, 1, 0) the vector $\langle 1, 0, 0\rangle$ is a tangent vector and is a negative direction for the Levi form. One can see that the domain has noncompact automorphism group by mapping it to the unbounded domain $\Omega^* \equiv \{4\text{Im}\,(W) + (|Z_1|^2 - |Z_2|^2)^2 + \lambda |Z_1|^4 < 0\}$ by way of the mapping

$$w = \frac{1 - iW}{1 + iW},$$

$$z_1 = \frac{Z_1}{(1 + iW)^{1/2}},$$

$$z_2 = \frac{Z_2}{(1 + iW)^{1/2}}.$$

The domain Ω^* clearly has noncompact automorphism group because translations in the Re W direction are automorphisms. Hence so does Ω.

More generally, one can consider a domain of the form

$$E = \{(w, z_1, z_2) : |w|^2 + p(z_1, z_2) < 1\}.$$

Here $M_1, M_2, \ldots, M_n$ are positive integer weights, and we take p to be a polynomial of the form

$$p(z, \bar{z}) = \sum{}' a_{i_1, i_2, \ldots, i_n, j_1, j_2, \ldots, j_n} z_1^{i_1} \cdot z_2^{i_2} \cdots z_n^{i_n} \cdot \bar{z}_1^{j_1} \cdot \bar{z}_2^{j_2} \cdots \bar{z}_n^{j_n}, \tag{2}$$

where $\sum'$ indicates that the summation is taken over all multi-indices $(i_1, i_2, \ldots, i_n)$ and $(j_1, j_2, \ldots, j_n)$ such that $i_1/M_1 + i_2/M_2 + \cdots + i_n/M_n = 1$ and $j_1/M_1 + j_2/M_2 + \cdots j_n/M_n = 1$.

Then, as above, the mapping

$$w = \frac{1 - iW}{w + iW},$$

$$z_1 = \frac{Z_1}{(1 + iW)^{1/M}},$$

$$z_2 = \frac{Z_2}{(1 + iW)^{1/M}}$$

sends E to an unbounded domain on which translations in the Re W variable act as automorphisms. Whenever p is not plurisubharmonic, we obtain a non-pseudconvex example.

Of course the disc $D \subseteq \mathbb{C}$ has noncompact automorphism group. Let

$$\varphi_a(\zeta) = \frac{\zeta - a}{1 - \overline{a}\zeta}$$

for $a \in \mathbb{C}$, $|a| < 1$. Then the automorphisms

$$\left\{ \frac{\zeta - (1 - 1/j)}{1 - (1 - 1/j)\zeta} : j \in \mathbb{Z} \right\}$$

are a sequence of automorphisms of D that have no subsequence converging to an automorphism. Indeed, any subsequence either converges to the constant function 1 or the constant function -1. It is a fact—see [19]—that any domain in $\mathbb{C}$ having C^1 boundary and noncompact automorphism group must be conformally equivalent to the disc. This is true without any topological hypotheses on the domain! In $\mathbb{C}^n$, the first result of this nature—due to Bun Wong [26] and Rosay [24]—is that any C^2 bounded domain in $\mathbb{C}^n$ with a boundary orbit accumulation point that is strongly pseudoconvex must be biholomorphic to the unit ball B. It is not known in general which smoothly bounded domains have noncompact automorphism group. Certainly there are finite type domains with noncompact automorphism group—see [2, 3]. More on this matter in what follows.

Example 3 In the complex plane $\mathbb{C}$, there are unbounded domains with noncompact automorphism group that are not the disc. The simplest example is when the domain Ω is the entire complex plane $\mathbb{C}$. The punctured plane also has this property.

3 New Results

The statement of Theorem 2 makes it desirable to have a formulation purely in terms of the intrinsic, invariant geometry of the domain. For instance, one might hope to be able to say something about the completeness of the Kobayashi metric at a boundary orbit accumulation point. Unfortunately, the following example dashes that hope:

Example 4 Let $B \subseteq \mathbb{C}^2$ be the unit ball with defining function $\rho(z) = |z|^2 - 1$. Let ϕ be a C_c^∞ function on $\mathbb{C}^n$ with these properties:

(a) ϕ is real-valued, and $0 \le \phi(z) \le 1/10$ for all $z \in \mathbb{C}^n$.
(b) ϕ is radial about the point $(i, 0)$.
(c) $\operatorname{supp} \phi \subseteq B((i, 0), 1/10)$.
(d) $\phi(z) = 1/10$ for $|z - (i, 0)| < 1/20$.

Set

$$\Omega' = \left\{ z \in \mathbb{C}^2 : -1 + |z|^2 + \phi(z) < 0 \right\}$$

and

$$\Omega = \bigcap_{-1<a<1} \Phi_a\big(\Omega'\big),$$

where

$$\Phi_a(z_1, z_2) = \left(\frac{z_1 - a}{1 - \bar{a} z_1}, \frac{\sqrt{1-|a|^2} z_2}{1 - \bar{a} z_1} \right) \tag{3}$$

for $a \in \mathbb{C}$, $|a| < 1$. Then Ω is the unit ball with a groove stretching from $(-1, 0)$ to $(1, 0)$. This new domain is strongly pseudoconcave along an entire curve from $(-1, 0)$ to $(1, 0)$. Of course the point $(1, 0)$ is still a boundary orbit accumulation point. Indeed the automorphisms Φ_a, $-1 < a < 0$, send $(0, 0)$ to $(1, 0)$. And, along the curve $\gamma(t) = (t, 0)$, $0 < t < 1$, the Kobayashi distance to the boundary point is infinite.

Now write $\Phi_a(z) = (\varphi_a^1(z), \varphi_a^2(z))$. Let the point $\widetilde{z}$ be as at the end of Example 1. We consider, in the same domain Ω as above, the curve

$$\lambda : t \longmapsto t \cdot \Phi_t(\widetilde{z}), \quad 0 \le t \le 1.$$

Then this curve terminates at $(1, 0)$ and is tangent to the boundary of Ω to order $1/2$ at that point (as a calculation with the automorphisms in (3) shows). This last means that $|\rho(\lambda(t))| \approx C \cdot t^{1/2}$, where ρ is a defining function for Ω. For $Q \in \partial\Omega$, let ν_Q denote the outward Euclidean unit normal vector at Q. Note that Ref. [20] shows that, on the domain $W = B(0, 2) \setminus \overline{B}(0, 1)$, near the strongly pseudoconcave boundary point $Q = (-1, 0)$, a point $Q^* = Q - \delta\nu_Q$ satisfies the estimate

$$F_K^{\Omega}\big(Q^*, \nu\big) \approx \delta^{-3/4}.$$

By scaling this estimate we see that, on the domain $\widetilde{W} = B(0, 2\alpha) \setminus \overline{B}(0, \alpha)$, with boundary point $\widetilde{Q} = (-\alpha, 0)$ and interior point $\widetilde{Q}^* = \widetilde{Q} - \delta\nu_{\widetilde{Q}}$, we have the estimate

$$F_K^{\widetilde{W}}\big(\widetilde{Q}^*, \nu\big) \approx \alpha^{-1/4} (\alpha\delta)^{-3/4}. \tag{4}$$

Moreover, in the tangential direction, we have

$$F_K^{\widetilde{\Omega}}\big(\widetilde{Q}^*, \tau\big) \approx C\,. \tag{5}$$

We may utilize these estimates as follows. At a point on the curve λ that is distance t from the point $(1, 0)$ in the normal direction z_1, the curve λ is laterally distant about $\sqrt{t}$ from $\partial\Omega$. Hence, by estimate (4), we have that the Kobayashi metric in the normal direction has size

$$\sqrt{t}^{-1/4} \cdot (\sqrt{t} \cdot \sqrt{t})^{-3/4} = t^{-7/8}.$$

This, combined with (5), tells us that the curve λ has finite length as $t \to 1^-$.

Therefore, at least on a domain with Lipschitz boundary, it is *not* the case that a boundary orbit accumulation point will be a point at which the Kobayashi metric is complete.

On the positive side, we can prove the following result:

Proposition 1 *Let $\Omega \subseteq \mathbb{C}^n$ be a smoothly bounded domain and $P \in \partial\Omega$ a boundary orbit accumulation point that is holomorphically simple* (*i.e*., *there is no complex variety through P that lies in the boundary*). *Assume that a boundary neighborhood of P is pseudoconvex. Then it is not possible for Ω to have Levi pseudoconcave boundary points in any part of $\partial\Omega$.*

Proof Let $\mathscr{W}$ denote the set of boundary points where the Levi form has a strictly negative eigenvalue. Thus $\mathscr{W} \subseteq \partial\Omega$ is a relatively open set. Letting $\widehat{\Omega}$ denote the envelope of holomorphy of Ω, we see that the identity map id extends holomorphically to a neighborhood of $\mathscr{W}$. Thus $\mathscr{W}$ is naturally identified with a subset $\mathrm{id}(\mathscr{W}) \equiv \widehat{\mathscr{W}} \subseteq \widehat{\Omega}$. If the boundary $\partial\Omega$ is holomorphically simple near P then, for every compact $K \subseteq \Omega$, there is a sequence of automorphisms φ_j such that $\varphi_j(K) \to P$ uniformly as $j \to \infty$ (this is a standard result, but see [8]). These automorphisms extend to mappings $\pi \circ \widetilde{\varphi}_j : \widetilde{\Omega} \to \mathbb{C}^n$, where π is the projection of the envelope into $\mathbb{C}^n$. As $j \to \infty$, we see that $\pi \circ \widetilde{\varphi}_j \to P$ uniformly on compact subsets of $\widetilde{\Omega}$. In particular, for any point $Q \in \mathscr{W}$, a sequence $\varphi_j(Q)$ will approach P as $j \to \infty$. Since $\mathscr{W}$ is invariant under the φ_j, this gives the contradiction that $\partial\Omega$ is not pseudoconvex in a neighborhood of P. □

Remark 1 In [5], Bedford and Pinchuk proved the following elegant theorem:

> **Theorem:** *Let $\Omega \subseteq \mathbb{C}^2$ be a domain with real analytic boundary. If there exists a boundary orbit accumulation point for Ω, then Ω must be biholomorphic to a domain of the form*
>
> $$E_m = \left\{(z_1, z_2) : |z_1|^2 + |z_2|^{2m} < 1\right\}$$
>
> *for some positive integer m.*

It follows from the Bedford–Pinchuk theorem that the domain must be globally pseudoconvex. Clearly our Proposition 1 is philosophically related to this result. [We note that David Catlin has observed—unpublished—that the last theorem actually holds in the generality of domains of finite type in $\mathbb{C}^2$.] In later papers Bedford and Pinchuk produced analogous results in $\mathbb{C}^n$.

A well-known conjecture in the subject says this:

> **The Greene–Krantz Conjecture:** Let Ω be a smoothly bounded domain in $\mathbb{C}^n$. If $P \in \partial\Omega$ is a boundary orbit accumulation point, then P is a point of finite type in the sense of Kohn–D'Angelo–Catlin.

This conjecture has not been established in full generality. But results in [13] and [14] support the conjecture. Now we have:

Proposition 2 *Let $\Omega \subseteq \mathbb{C}^2$ be a smoothly bounded*, *pseudoconvex domain. Let $P \in \partial\Omega$ be a boundary orbit accumulation point. Assume that the Greene–Krantz conjecture is true. Then any path ending at P will have infinite length in the Kobayashi metric.*

Proof This result is almost obvious for the hypothesis implies that P is of finite type. And now the estimates on the Kobayashi metric in [6], together with the calculations in [1], give the result about infinite length of paths. □

4 A Boundary Orbit Accumulation Point Characterization of Domains

In [24] and [26], for instance, it is shown that, if a bounded domain has a strongly pseudoconvex boundary orbit accumulation point, then that domain must be biholomorphic to the unit ball in $\mathbb{C}^n$. Put in other words, if two distinct bounded domains have boundary orbit accumulation points, and if those boundary orbit accumulation points are both strongly pseudoconvex, then the two domains must be biholomorphic (since they are both biholomorphic to the ball).

One might more generally formulate this question as follows:

> Suppose that Ω_1 and Ω_2 are two bounded domains in $\mathbb{C}^n$. Assume that Ω_1 has boundary orbit accumulation point P_1 and Ω_2 has boundary orbit accumulation point P_2. If P_1 and P_2 have the same Levi geometry, may we conclude that Ω_1 is biholomorphic to Ω_2?

I do not know the full answer to this question at this time. However, the following partial answer may be proved using known techniques:

Proposition 3 *Let Ω_1, Ω_2 be bounded domains in $\mathbb{C}^n$. Let $P_1 \in \partial\Omega_1$ and $P_2 \in \partial\Omega_2$ each be boundary orbit accumulation points. Assume that $\partial\Omega_j$ is smooth near P_j, $j = 1, 2$. Suppose that each P_j is of finite type in the sense of Kohn–Catlin–D'Angelo and is also a peak point. Finally assume that there is a neighborhood U_1 of P_1 and a neighborhood U_2 of P_2 and a biholomorphic mapping*

$$\Phi : U_1 \cap \Omega_1 \to U_2 \cap \Omega_2$$

such that (i) *Φ continues to a diffeomorphism of $\partial\Omega_1 \cap U_1$ to $\partial\Omega_2 \cap U_2$,* (ii) *$\Phi(P_1) = P_2$.*

Then Ω_1 is biholomorphic to Ω_2.

Proof Choose a point $X_1 \in \Omega_1$ and automorphisms φ_j of Ω_1 so that $\varphi_j(X_1) \to P_1$. Likewise choose a point $X_2 \in \Omega_2$ and automorphisms ψ_j of Ω_2 such that $\psi_j(X_2) \to P_2$. A standard argument (see [18], Chap. 11) shows that, for any compact set $K \subseteq \Omega_1$, $\varphi_j(z)$ converges to P_1 uniformly for $z \in K$. A similar statement holds for Ω_2.

Let K be a large compact set inside Ω_1. Choose j so large that $\varphi_j(K) \subseteq U_1 \cap \Omega_1$. Likewise let L be a large compact set inside Ω_2. Choose k so large that $\psi_k(L) \subseteq U_2 \cap \Omega_2$. Let $\epsilon > 0$ be small and set $U_1^\epsilon = \{z \in U_1 : \text{dist}(z, {}^cU_1 > \epsilon)\}$. Similarly set $U_2^\epsilon = \{z \in U_2 : \text{dist}(z, {}^cU_2 > \epsilon)\}$. By shrinking ϵ if necessary, we may assume that $\Phi(\varphi_j(K)) \subseteq U_2^\epsilon$. By enlarging L if necessary, we may suppose that $\varphi_k(L) \supseteq U_2^\epsilon$.

Now consider $(\psi_k)^{-1} \circ \Phi \circ \varphi_j$. This will be a univalent holomorphic mapping that takes $K \subseteq \Omega_1$ to L, and the mapping is invertible. We may similarly assume that the inverse mapping takes L to K. The set of all such mappings, as K exhausts Ω_1 and L exhausts Ω_2, forms a normal family. And we may extract a convergent subsequence that converges to a biholomorphic mapping of Ω_1 to Ω_2. That is the result that we seek. □

There are a number of different approaches to the classical Bun Wong–Rosay theorem. Useful references are [8, 9, 15–17, 24, 26].

5 Concluding Remarks

Recall that, in Remark 1, we discussed a result of Bedford and Pinchuk that characterizes complex ellipsoids in terms of noncompact automorphism group actions.

But it must be pointed out that, in higher dimensions, we cannot hope for a conclusion as simple as "the domain must be a complex ellipsoid." For consider the domain

$$\Omega^* = \left\{(z_1, z_2, z_3) \in \mathbb{C}^3 : |z_1|^2 + \left(|z_2|^2 + |z_3|^2\right)^2 < 1\right\}.$$

It has automorphisms of the form

$$\Phi_a(z_1, z_2, z_3) = \left(\frac{z_1 - a}{1 - \overline{a}z_1}, \frac{\sqrt[4]{1-|a|^2}z_2}{\sqrt{1-\overline{a}z_1}}, \frac{\sqrt[4]{1-|a|^2}z_3}{\sqrt{1-\overline{a}z_1}}\right)$$

for $a \in \mathbb{C}$, $|a| < 1$. If we let a take the values $1 - 1/j$ for $j = 1, 2, \ldots$, then we see immediately that Ω^* has noncompact automorphism group. And Ω^* is *not* an ellipsoid, even biholomorphically (see [7]). It has been conjectured by Catlin and others (see [21] for the details) that the correct conclusion in higher dimensions is that the defining function of the domain should satisfy a certain homogeneity condition. Results along these lines have been obtained in [4].

It has been noted that the Greene–Krantz conjecture asserts that a boundary orbit accumulation point *must be* of finite type. There is some evidence to support the conjecture—see, for instance, [13, 14]. If it turns out to be true, then the Bedford–Pinchuk theorem cited above can probably be streamlined to say that a smoothly bounded domain in $\mathbb{C}^2$ with noncompact automorphism group must be an ellipsoid. We note also that the results of the paper [25]—we mention particularly Theorem 1.1—may be conceptually simplified with the Greene–Krantz conjecture.

It is certainly a matter of some interest to understand the nature of boundary orbit accumulation points. We know that they must be pseudoconvex, and the Greene–Krantz conjecture posits even more specific information about these points. Another subject of some study is boundary orbit accumulation *sets*—see, for instance, [12] and [22]. Much more can in principle be said about these sets.

Automorphism groups are in some sense an invariant that is a substitute for the lack in several complex variables of a uniformization theorem or a Riemann mapping theorem. It is in our best interest to develop their properties so that they can be used effectively to study and classify domains up to biholomorphic equivalence.

We hope to study these matters further in future papers.

Acknowledgements The author is supported in part by the National Science Foundation and by the Dean of the Graduate School at Washington University.

References

1. Aladro, G.: The comparability of the Kobayashi approach region and the admissible approach region. Ill. J. Math. **33**, 42–63 (1989)
2. Bedford, E., Pinchuk, S.: Domains in $\mathbb{C}^2$ with noncompact group of automorphisms. Math. Sb. Nov. Ser. **135**(2), 147–157 (1988)
3. Bedford, E., Pinchuk, S.: Domains in $\mathbb{C}^{n+1}$, with non-compact automorphism group. J. Geom. Anal. **3**, 165–191 (1991)
4. Bedford, E., Pinchuk, S.: Convex domains with noncompact groups of automorphisms (Russian). Mat. Sb. **185**, 3–26 (1994); translation in Russ. Acad. Sci. Sb. Math. **82**, 1–20 (1995)
5. Bedford, E., Pinchuk, S.: Domains in $\mathbb{C}^2$ with noncompact automorphism groups. Indiana Univ. Math. J. **47**, 199–222 (1998)
6. Catlin, D.: Estimates of invariant metrics on pseudoconvex domains of dimension two. Math. Z. **200**, 429–466 (1989)
7. Coupet, B., Pinchuk, S.: Holomorphic equivalence problem for weighted homogeneous rigid domains in $\mathbb{C}^{n+1}$. In: Complex Analysis in Modern Mathematics, FAZIS, Moscow, pp. 57–70 (2001) (Russian)
8. Gaussier, H., Kim, K.-T., Krantz, S.G.: A note on the Wong–Rosay theorem in complex manifolds. Complex Var. Theory Appl. **47**, 761–768 (2002)
9. Greene, R.E., Kim, K.-T., Krantz, S.G.: The Geometry of Complex Domains. Birkhäuser, Boston (2011, to appear)
10. Greene, R.E., Krantz, S.G.: Invariants of Bergman geometry and results concerning the automorphism groups of domains in $\mathbb{C}^n$. In: Geometrical and Algebraical Aspects in Several Complex Variables (Cetraro, 1989), pp. 107–136, Sem. Conf., 8, EditEl, Rende (1991)
11. Helgason, S.: Differential Geometry and Symmetric Spaces. Academic Press, New York (1962)
12. Isaev, A., Krantz, S.G.: On the boundary orbit accumulation set for a domain with noncompact automorphism group. Mich. Math. J. **43**, 611–617 (1996)
13. Kim, K.-T.: Domains in $\mathbb{C}^n$ with a piecewise Levi flat boundary which possess a noncompact automorphism group. Math. Ann. **292**, 575–586 (1992)
14. Kim, K.-T., Krantz, S.G.: Complex scaling and domains with non-compact automorphism group. Ill. J. Math. **45**, 1273–1299 (2001)
15. Kim, K.-T., Krantz, S.G.: Some new results on domains in complex space with non-compact automorphism group. J. Math. Anal. Appl. **281**, 417–424 (2003)
16. Kim, K.-T., Krantz, S.G.: A Kobayashi metric version of Bun Wong's theorem. Complex Var. Elliptic Equ. **54**, 355–369 (2009)
17. Kim, K.-T., Krantz, S.G.: Characterization of the Hilbert ball by its automorphism group. Trans. Am. Math. Soc. **354**, 2797–2818 (2002)
18. Krantz, S.G.: Function Theory of Several Complex Variables, 2nd edn. American Mathematical Society, Providence (2001)
19. Krantz, S.G.: Characterizations of smooth domains in $\mathbb{C}$ by their biholomorphic self maps. Am. Math. Mon. **90**, 555–557 (1983)

20. Krantz, S.G.: On the boundary behavior of the Kobayashi metric. Rocky Mt. J. Math. **22**, 227–233 (1992)
21. Krantz, S.G.: Recent progress and future directions in several complex variables. In: Complex Analysis Seminar. Lecture Notes, vol. 1268, pp. 1–23. Springer, Berlin (1987)
22. Krantz, S.G.: Topological/geometric properties of the orbit accumulation set. Complex Var. Elliptic Equ. (to appear)
23. Narasimhan, R.: Several Complex Variables. University of Chicago Press, Chicago (1971)
24. Rosay, J.-P.: Sur une characterization de la boule parmi les domains de $\mathbb{C}^n$ par son groupe d'automorphismes. Ann. Inst. Fourier **XXIX**, 91–97 (1979)
25. Verma, K.: A characterization of domains in $\mathbb{C}^2$ with noncompact automorphism group. Math. Ann. **344**, 645–701 (2009)
26. Wong, B.: Characterizations of the ball in $\mathbb{C}^n$ by its automorphism group. Invent. Math. **41**, 253–257 (1977)

Microlocal Analysis of Elliptical Radon Transforms with Foci on a Line

Venkateswaran P. Krishnan, Howard Levinson, and Eric Todd Quinto

Abstract In this paper, we take a microlocal approach to the study of an integral geometric problem involving integrals of a function on the plane over two-dimensional sets of ellipses on the plane. We focus on two cases: (a) the family of ellipses where one focus is fixed at the origin and the other moves along the x-axis, and (b) the family of ellipses having a common offset geometry.

For case (a), we characterize the Radon transform as a Fourier integral operator associated to a fold and blowdown. This has implications on how the operator adds singularities, how backprojection reconstructions will show those singularities, and in comparison of the strengths of the original and added singularities in a Sobolev sense.

For case (b), we show that this Radon transform has similar structure to case (a): it is a Fourier integral operator associated to a fold and blowdown. This case is related to previous results of authors one and three. We characterize singularities that are added by the reconstruction operator, and we present reconstructions from the authors' algorithm that illustrate the microlocal properties.

1 Introduction

In Synthetic Aperture Radar (SAR) imaging, a region of interest on the surface of the earth is illuminated by electromagnetic waves from a moving airborne platform.

We dedicate this article to the memory of Leon Ehrenpreis, a brilliant mathematician and a Mensch.

V.P. Krishnan
Tata Institute of Fundamental Research Centre for Applicable Mathematics, Bangalore 560065, India
e-mail: vkrishnan@math.tifrbng.res.in

H. Levinson
1759 Beechwood Blvd., Pittsburgh, PA 15217, USA
e-mail: howielevinson@gmail.com

E.T. Quinto (✉)
Tufts University, Medford, MA 02155, USA
e-mail: todd.quinto@tufts.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_11, © Springer-Verlag Italia 2012

The goal is to reconstruct an image of the region based on the measurement of scattered waves. For an in-depth treatment of SAR imaging, we refer the reader to [6, 8]. SAR imaging is similar to other imaging problems such as Sonar where acoustic waves are used to reconstruct the shape of objects on the ocean floor [3].

Depending on the acquisition geometry, the transmitter and the receiver can be located on the same platform (monostatic SAR imaging) or different airborne platforms (bistatic SAR imaging).

There are several advantages to considering bistatic data acquisition geometries. The receivers, compared to the transmitters, are passive and hence are more difficult to detect. Hence, by separating their locations, the receivers alone can be in an unsafe environment, while the transmitters are in a safe environment. Bistatic SAR acquisition geometry arises naturally when imaging using a stationary transmitter such as a television or radio broadcasting station. Finally, bistatic SAR systems are more resistant to electronic countermeasures such as target shaping to reduce scattering in the direction of incident waves [32].

Under certain simplifying assumptions, the scattered data can be viewed as integrals of a function over a family of ellipses in the case of bistatic SAR, compared to a family of circles for the case of monostatic SAR. Thus, imaging using a bistatic SAR system leads to the question of recovering a function given its integrals over a family of ellipses. With this as our motivation, we analyze two elliptical Radon transforms in this paper. In Sect. 2 we give microlocal properties of the transform that integrates over ellipses with one focus fixed at the origin and the other focus moving on a line. We show using microlocal analysis why there are added singularities in reconstructions. In Sect. 3 we consider the elliptical transform involving a common offset geometry, where the foci are on a line at a fixed positive distance apart and move along this line. In Sect. 4 we describe our algorithm and reconstructions from that algorithm. As before, we explain, using microlocal analysis, why there are added singularities in the reconstructions.

Radon transforms over circles and spheres have a rich theory starting from the early 1900s. In 1916, Funk inverted the transform integrating over great circles on the sphere [22]. Then researchers such as John [33], Courant and Hilbert [9], Helgason [30], and many others proved important results for spherical integrals in $\mathbb{R}^n$ and manifolds. The article [58] gives a very readable summary of the large number of themes in the field up to that point. In the article [1], microlocal and harmonic analysis are used to characterize invertibility for the circular Radon transform with centers on a curve.

Our elliptical transform in Sect. 2 integrates over ellipses that enclose the origin. Helgason [30] proved a support theorem for the transform integrating over spheres in $\mathbb{R}^n$ enclosing the origin under the assumption that the function is rapidly decaying at infinity. Globevnik [23, Theorem 1] characterizes the null space of the Radon transform integrating over circles enclosing the origin.

Leon Ehrenpreis considered spherical Radon transforms in several contexts. For example, [12] is a lovely article involving integrals over spheres tangent to a set, and he discussed spherical integrals in relation to Huygens Principle in his book *The Universality of the Radon Transform* [13, p. 132 ff]. In the book, he applied

Radon transforms to PDE, harmonic analysis, and Fourier analysis, as well as tomography and even topics related to number theory. He developed a theory of the nonparametric Radon transform [13, p. 4 ff], and our two elliptic transforms can be put in this framework. We work the details out for case (a) in Example 1.

Less is known about integrals over ellipses. Volchkov [56] and others considered convolution integrals over sets such as ellipsoids. Elliptical integrals come up in ultrasound [2, 54] as well. The sound source and receiver are at different locations, and the sound wavefronts are elliptical giving rise to elliptical Radon transforms.

Microlocal analysis has a long history in integral geometry starting with [27–29]. Then many other authors have applied microlocal analysis to integral geometric problems. A very partial listing of the themes and a few papers in those areas include microlocal properties of the operators and their compositions [25, 26, 46, 52, 53], applications to support theorems and uniqueness [1, 4, 5, 35, 37, 48], applications to SAR imaging [7, 16, 17, 36, 41, 43], and applications to other modalities in tomography including X-ray CT [21, 34, 47], SPECT [50], electron microscopy [51], and seismic imaging [10, 11, 18, 19, 40, 45].

2 Analysis of an Elliptical Radon Transform with One Fixed Focus

In this section, we study the microlocal analysis of an elliptical Radon transform integrating over ellipses in which one focus is fixed at the origin and the other is free to move along the horizontal axis. As explained in the introduction, this acquisition geometry is related to one in SAR imaging. The receiver is passive, often smaller and less expensive to replace than the transmitter. Therefore, in dangerous environments, it might be advantageous to let the transmitter and receiver move independently. One useful case to study is where the receiver can use a radio or cellphone transmitter that is already in the environment. Thus, the radar problem has a fixed transmitter location, and movable receiver becomes of interest. The transmitter becomes one fixed focus of the ellipsoidal wavefronts, and the receiver becomes the other focus.

The transform we now study is motivated by this SAR transform. It is an elliptical Radon transform with one focus fixed on the ground and the other moving along the horizontal axis. For the SAR transform, the transmitter and receiver would be above the ground. From now on, we will let $X = \mathbb{R}^2$ and denote points in X as (x_1, x_2). We let

$$Y_o = \{(s, L) : L > |s|\}, \tag{1}$$

where the subscript o refers to the fact that one focus is at the origin. We parameterize the ellipse with foci $(0, 0)$ and $(s, 0)$ and major diameter L by

$$E_o(s, L) = \{x \in \mathbb{R}^2 : |x| + |x - (s, 0)| = L\} \quad \text{for } (s, L) \in Y_o.$$

The restriction $L > |s|$ in the definition of Y_o is required because the major diameter must be longer than the distance between the foci.

The integral geometry problem that we are interested in is recovery of f from

$$\mathscr{R}_o f(s, L) = \int_{|x|+|x-(s,0)|=L} f(x)\,\mathrm{d}l(x) \quad \text{for } (s, L) \in Y_o.$$

Here dl is the arc-length measure. This transform is just the integral of f over the ellipse $E_o(s, L)$.

Example 1 Ehrenpreis's nonparametric Radon transform is defined as integrals over sets which are defined by *spreads* [13, p. 4ff]. Spreads are foliations of space that depend on a parameter. For each fixed value of the parameter, the leaves of the foliation define manifolds the Radon transform integrates over. For all parameters, all the leaves of all the foliations are diffeomorphic copies of one manifold, such as a line, plane, ellipse, or circle. The transform $\mathscr{R}_o$ is easily put into this framework. We fix s, and then, for $L > s$, the map $L \mapsto E_o(s, L)$ foliates the plane (except for the segment between the origin and s) by ellipses. For any s, the leaves of the foliations are ellipses, and so they are diffeomorphic.

Because of the nonuniqueness results for integrals over spheres enclosing the origin [23], we expect that the transform $\mathscr{R}_o$ is not invertible. However, we might still be able to reconstruct singularities, so we will now understand what this transform and its adjoint do to singularities by analyzing the microlocal properties of the transform $\mathscr{R}_o$ and the imaging operator $\mathscr{R}_o^*\mathscr{R}_o$ (see Remark 1).

Our first theorem is the following:

Theorem 1 *$\mathscr{R}_o$ is a Fourier integral operator of order $-1/2$ with canonical relation Λ_o defined by*

$$\begin{aligned}\Lambda_o = \Bigg\{\Bigg(&s, L, -\omega\frac{x_1 - s}{\sqrt{(x_1-s)^2 + x_2^2}}, -\omega;\\ &x_1, x_2, -\omega\left(\frac{x_1}{\sqrt{x_1^2+x_2^2}} + \frac{x_1 - s}{\sqrt{(x_1-s)^2+x_2^2}}\right),\\ &-\omega\left(\frac{x_2}{\sqrt{x_1^2+x_2^2}} + \frac{x_2}{\sqrt{(x_1-s)^2+x_2^2}}\right)\Bigg)\\ &: \omega \neq 0, (s, L) \in Y_o,\ x \in E_o(s, L)\Bigg\}\end{aligned}$$

and with global parameterization (s, x_1, x_2, ω). The left projection $\pi_L : \Lambda_o \to T^*Y_o \setminus \mathbf{0}$ has a fold singularity along $\Sigma = \{(s, x_1, 0, \omega)\}$. The right projection $\pi_R : \Lambda_o \to T^*X \setminus \mathbf{0}$ has a blowdown singularity along Σ.

For the definitions of fold and blowdown singularities, we refer the reader to [24] or [25]. While we do not show this here, knowing that π_L is a fold and π_R is a blowdown has implications for the comparison of the strengths (in a Sobolev sense) of the original and added singularities discussed in Theorem 2.

Proof We use the framework of [27–29] and introduce the *incidence relation of* $\mathscr{R}_o$. This is the set

$$Z_o = \{(s, L, x) : (s, L) \in Y_o,\ x \in E_o(s, L)\}.$$

Then by results in [27, 29] we know that $\mathscr{R}_o$ is an elliptic Fourier integral operator of order $-1/2$ associated to the Lagrangian manifold $N^*(Z_o) \setminus \mathbf{0}$ (since we will show that neither π_L nor π_R maps to the zero section). Computing $N^*Z_o \setminus \mathbf{0}$ and twisting it gives the canonical relation Λ_o above. It is easy to see that (s, x_1, x_2, ω) is a global parameterization of Λ_o.

We have

$$\pi_L(s, x, \omega) = \left(s, |x| + |x - (s, 0)|, -\omega \frac{x_1 - s}{\sqrt{(x_1 - s)^2 + x_2^2}}, -\omega\right).$$

Since $\omega \neq 0$, we have that $\pi_L : \Lambda_o \to T^*Y_o \setminus \mathbf{0}$. Now

$$(\pi_L)_* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ * & (\frac{x_1}{|x|} + \frac{x_1 - s}{|x-(s,0)|}) & (\frac{x_2}{|x|} + \frac{x_2}{|x-(s,0)|}) & * \\ * & -\omega \frac{x_2^2}{|x-(s,0)|^3} & \omega \frac{(x_1-s)x_2}{|x-(s,0)|^3} & * \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and

$$\det\big((\pi_L)_*\big) = \omega \frac{x_2}{|x - (s, 0)|^2} \left(1 + \frac{x_1(x_1 - s) + x_2^2}{|x||x - (s, 0)|}\right).$$

Lemma 1 *Under the hypothesis of* (1), $1 + \frac{x_1(x_1-s)+x_2^2}{|x||x-(s,0)|} > 0$.

Proof It is easy to see that $(x_1(x_1 - s) + x_2^2)^2 < |x|^2|x - (s, 0)|^2$ is equivalent to $x_2^2 s^2 > 0$. By the hypothesis that $L > |s|$, if $x_2 = 0$, the term $\frac{x_1(x_1-s)}{|x_1||x_1-s|} = 1$ for all x_1 and s, from which the lemma follows. □

Therefore $\det((\pi_L)_*) = 0$ if and only if $x_2 = 0$. Also since $\mathrm{d}(\det(\pi_L)_*)$ on Σ is nonvanishing, we have that π_L drops rank by one simply on Σ.

Now it remains to show that $T\Sigma \cap \mathrm{Kernel}(\pi_L)_* = \{0\}$. This follows from the fact that, above Σ, $\mathrm{Kernel}(\pi_L)_* = \mathrm{span}(\frac{\partial}{\partial x_2})$ and $T\Sigma = \mathrm{span}(\frac{\partial}{\partial s}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial \omega})$. This concludes the proof that $\pi_L : \Lambda_o \to T^*Y_o \setminus \mathbf{0}$ has a fold singularity along Σ.

Next we consider $\pi_R : \Lambda_o \to T^*X$:

$$\pi_R(s, x, \omega) = \left(x_1, x_2, -\omega\left(\frac{x_1}{|x|} + \frac{x_1 - s}{|x - (s, 0)|}\right), -\omega\left(\frac{x_2}{|x|} + \frac{x_2}{|x - (s, 0)|}\right)\right).$$

We now show that $\pi_R : \Lambda_o \to T^*X \setminus \mathbf{0}$. For suppose π_R maps to the zero section, then $x_2 = 0$. Now since $L > |s|$, we have that x_1 and $x_1 - s$ have the same sign. Therefore, $\frac{x_1}{|x_1} + \frac{x_1 - s}{|x_1 - s|}$ is never 0. Hence π_R never maps to the zero section.

Now

$$(\pi_R)_* = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \omega \frac{x_2^2}{|x-(s,0)|^3} & * & * & -(\frac{x_1}{|x|} + \frac{x_1 - s}{|x-(s,0)|}) \\ -\omega \frac{(x_1 - s)x_2}{|x-(s,0)|^3} & * & * & (\frac{x_2}{|x|} + \frac{x_2}{|x-(s,0)|}) \end{pmatrix}.$$

Since $\det((\pi_R)_*) = \det((\pi_L)_*)$, π_R drops ranks by one simply along Σ. Furthermore above Σ, since $\text{Kernel}(\pi_R)_* = \text{span}(\frac{\partial}{\partial s}) \subset T\Sigma$, π_R has a blowdown singularity along Σ. □

Next we analyze the wavefront set of the imaging operator $\mathscr{R}_o^* \mathscr{R}_o$.

Remark 1 For the composition of $\mathscr{R}_o$ with $\mathscr{R}_o^*$ to be well defined, we have to modify $\mathscr{R}_o$ by introducing an infinitely differentiable cut-off function χ_o defined on Y_o that is identically 1 on a compact subset of Y_o and 0 outside a bigger compact subset of Y_o. In the next theorem, we assume that $\mathscr{R}_o^*$ is modified using this cut-off function χ_o.

Theorem 2 *The wavefront set of the imaging operator satisfies the following*:

$$WF\big(\mathscr{R}_o^* \mathscr{R}_o\big) \subset \Delta \cup C_1,$$

where

$$\Delta := \big\{(x_1, x_2, \xi_1, \xi_2; x_1, x_2, \xi_1, \xi_2)\big\} \quad \textit{and}$$
$$C_1 := \big\{(x_1, x_2, \xi_1, \xi_2; x_1, -x_2, \xi_1, -\xi_2)\big\}.$$

Here over the point $x = (x_1, x_2)$, (ξ_1, ξ_2) *consists of all nonzero multiples of the vector*

$$-\nabla_x\big(|x| + |x - (s, 0)|\big)$$

for all $s \in \mathbb{R}$.

Remark 2 Given a point x and a focus location $(s, 0)$, a vector (ξ_1, ξ_2) as in the theorem above is a vector perpendicular to the ellipse $E_o(s, L)$ (where $L = |x| + |x - (s, 0)|$) at the point x.

Note that Remark 4 in Sect. 3 applies to this transform and there is the left–right ambiguity for $\mathscr{R}_o^* \mathscr{R}_o$ as in the common offset case discussed in that section. The implications for imaging are the same as for Theorem 4 in the common offset case; singularities of a function f on one side of the x_1 axis can be reflected to the other side in the reconstruction $\mathscr{R}_o^* \mathscr{R}_o f$.

Proof Using the Hörmander–Sato Lemma, we have that $WF(\mathscr{R}_o^*\mathscr{R}) \subset \Lambda_o^t \circ \Lambda_o$. The composition of these two canonical relations is given as follows:

$$\Lambda_o^t \circ \Lambda_o = \Bigg\{\Bigg(x_1, x_2, -\omega\left(\frac{x_1}{|x|} + \frac{x_1 - s}{|x-(s,0)|}\right), -\omega\left(\frac{x_2}{|x|} + \frac{x_2}{|x-(s,0)|}\right);$$
$$y_1, y_2, -\omega\left(\frac{y_1}{|y|} + \frac{y_1 - s}{|y-(s,0)|}\right), -\omega\left(\frac{y_2}{|y|} + \frac{y_2}{|y-(s,0)|}\right)\Bigg)$$
$$: |x| + |x-(s,0)| = |y| + |y-(s,0)| \frac{x_1 - s}{|x-(s,0)|} = \frac{y_1 - s}{|y-(s,0)|}\Bigg\}.$$

Lemma 2 *For all* $s > 0$, *the set of all* (x_1, x_2), (y_1, y_2) *that satisfy*

$$|x| + |x-(s,0)| = |y| + |y-(s,0)|, \tag{2}$$

$$\frac{x_1 - s}{|x-(s,0)|} = \frac{y_1 - s}{|y-(s,0)|} \tag{3}$$

necessarily satisfy the relations: $x_1 = y_1$ *and* $x_2 = \pm y_2$.

Proof It is straightforward to verify for the case $s = 0$. For $s \neq 0$, we use the following coordinate change to elliptical coordinates:

$$x_1 = \frac{s}{2} + \frac{s}{2}\cosh\rho\cos\theta, \qquad y_1 = \frac{s}{2} + \frac{s}{2}\cosh\rho'\cos\theta',$$
$$x_2 = \frac{s}{2}\sinh\rho\sin\theta, \qquad y_2 = \frac{s}{2}\sinh\rho'\sin\theta'.$$

From the first equation in (2), we have $s\cos\rho = s\cos\rho'$, which then gives $\rho = \rho'$. From the second equation in (2), we have

$$\frac{\cosh\rho\cos\theta - 1}{\cosh\rho - \cos\theta} = \frac{\cosh\rho'\cos\theta' - 1}{\cosh\rho' - \cos\theta'}.$$

Using the fact that $\cosh\rho = \cosh\rho'$ and simplifying this, we obtain, $\cos\theta = \cos\theta'$. Therefore, $\theta = 2n\pi \pm \theta'$. This then gives $\sin\theta = \pm\sin\theta'$. Now going back to (x_1, y_1) and (x_2, y_2), we have $x_1 = y_1$ and $x_2 = \pm y_2$. □

Now to finish the proof of the theorem, when $x = y$, $\Lambda_o^t \circ \Lambda_o \subset \Delta = \{(x, \xi; x, \xi)\}$ and when $x_1 = y_1$ and $x_2 = -y_2$, $\Lambda_o^t \circ \Lambda_o \subset C_1 = \{(x_1, x_2, \xi_1, \xi_2; x_1, -x_2, \xi_1, -\xi_2)\}$. □

3 Analysis of a Common Offset Elliptical Radon Transform

In this section, we consider an elliptical Radon transform over a family of ellipses in which the foci move along the x_1-axis and are spaced a constant distance apart.

We parameterize the right and left foci, respectively, by

$$\gamma_T(s) = (s + \alpha, 0) \quad \text{and} \quad \gamma_R(s) = (s - \alpha, 0),$$

where $\alpha > 0$ is fixed. If this were a radar problem, then γ_T would be the location of the transmitter and γ_R would be the location of the receiver. In radar imaging, the phrase "common offset" comes from the fact that the transmitter γ_T and receiver γ_R are *offset* a fixed distance from each other. In the case of common offset SAR, the transmitter and receiver (the foci of an ellipsoid) are on a line $h > 0$ units above the plane to be reconstructed, and they travel along a line with one behind the other.

The transform we now study is motivated by this SAR transform. It is an elliptical Radon transform in which the foci are a fixed distance apart as they move along the x_1 axis in the plane. Again, $X = \mathbb{R}^2$, and we let

$$Y_c = \big\{(s, L) : L > 2\alpha\big\}, \tag{4}$$

where the subscript c refers to *common* offset. The ellipse with foci $\gamma_T(s)$ and $\gamma_R(s)$ and major diameter L is denoted

$$E_c(s, L) = \big\{x \in \mathbb{R}^2 : \big|x - \gamma_T(s)\big| + \big|x - \gamma_R(s)\big| = L\big\} \quad \text{for } (s, L) \in Y_c.$$

The restriction $L > 2\alpha$ is needed because the major diameter of the ellipse must be longer than the distance between the foci.

In this section, we consider the integral geometry problem of recovery of f from

$$\mathscr{R}_c f(s, L) = \int_{x \in E_c(s,L)} f(x)\,\mathrm{d}l(x) \quad \text{for } (s, t) \in Y_c, \tag{5}$$

which is the integral of f over the ellipse $E_c(s, L)$ in arc-length measure. As we discussed for $\mathscr{R}_o$ in Example 1, $\mathscr{R}_c$ can be put into Ehrenpreis's framework of spreads.

This case is very closely related to the results on common offset SAR in [36], and we will state our theorems and then explain how they follow from the results in [36].

Similar to Theorem 1, our first theorem in this section shows that $\mathscr{R}_c$ is an FIO, gives its canonical relation, and the mapping properties of the left and right projections from this canonical relation.

Theorem 3 *The common offset elliptical transform $\mathscr{R}_c$ is a Fourier integral operator of order $-1/2$ with canonical relation Λ_c defined by*

$$\Lambda_c = \Bigg\{\Bigg(s, L, -\omega\bigg(\frac{x_1 - s - \alpha}{|x - \gamma_T(s)|} + \frac{x_1 - s + \alpha}{|x - \gamma_R(s)|}\bigg), -\omega\Bigg);$$
$$\Bigg(x_1, x_2, -\omega\bigg(\frac{x_1 - s - \alpha}{|x - \gamma_T(s)|} + \frac{x_1 - s + \alpha}{|x - \gamma_R(s)|}\bigg),$$

$$- \omega\left(\frac{x_2}{|x-\gamma_T(s)|}+\frac{x_2}{|x-\gamma_R(s)|}\right)\Bigg)$$
$$: L=\sqrt{(x_1-s-\alpha)^2+x_2^2}+\sqrt{(x_1-s+\alpha)^2+x_2^2},\ \omega\neq 0\Bigg\}. \tag{6}$$

Furthermore the map λ *taking* (s, x_1, x_2, ω) *to the point in* Λ *given above is a global parameterization for* Λ.

Finally, the projection $\pi_L : \Lambda_c \to T^*Y_c \setminus \mathbf{0}$ *has a fold along* $\Sigma = \{s, x_1, 0, \omega\}$ *and* $\pi_R : \Lambda_c \to T^*X \setminus \mathbf{0}$ *has a blowdown along* Σ.

Proof The assertion (6) can be proven as in [36], but here, as in Theorem 1, we outline another proof using the framework of [27–29]. The incidence relation of $\mathscr{R}_c$ is the set

$$Z_c = \big\{(x, s, L) : (s, L) \in Y_c,\ x \in E_c(s, L)\big\}.$$

Then by results in [27, 29] we know $\mathscr{R}_c$ is an elliptic Fourier integral operator of order $-1/2$ associated to Lagrangian manifold $N^*(Z_c) \setminus \mathbf{0}$ (since we will show in the course of the proof that neither π_L nor π_R maps to the zero section). Computing $N^*(Z_c)$ and twisting it gives the canonical relation (6). This is the same as the canonical relation in [36] for $h = 0$ where h is the elevation of the transmitter and receiver above the reconstruction plane.

In the parameterization λ given in the theorem, the projection $\pi_L : \Lambda_c \to T^*Y_c$ is given by

$$\pi_L(s, x_1, x_2, \omega) = \left(s, \big(|x-\gamma_T(s)|+|x-\gamma_R(s)|\big), -\omega\left(\frac{x_1-s-\alpha}{|x-\gamma_T(s)|}+\frac{x_1-s+\alpha}{|x-\gamma_R(s)|}\right), -\omega\right). \tag{7}$$

It is clear that π_L maps to $T^*Y_c \setminus \mathbf{0}$ since $\omega \neq 0$. Now from [36], by letting $h = 0$ there, we get $\det((\pi_L)_*) = \omega x_2(\frac{1}{|x-\gamma_T(s)|^2|}+\frac{1}{|x-\gamma_R(s)|^2})(1+\frac{(x_1-s)^2+x_2^2-\alpha^2}{|x-\gamma_T(s)||x-\gamma_R(s)|})$. It is easy to see that $((x_1-s)^2+x_2^2-\alpha^2)^2 < (|x-\gamma_T(s)||x-\gamma_R(s)|)^2$ is equivalent to $4x_2^2\alpha^2 > 0$. Since $L > 2\alpha$, if $x_2 = 0$, $\frac{(x_1-s)^2-\alpha^2}{|x_1-s-\alpha||x_1-s+\alpha|} = 1$. Therefore, $\det((\pi_L)_*) = 0$ if and only if $x_2 = 0$. Also since $\mathrm{d}(\det(\pi_L)_*)$ on Σ is nonvanishing, we have that π_L drops rank by one simply on Σ. Now as in the proof of Theorem 1, we have that $T\Sigma = \mathrm{span}(\frac{\partial}{\partial s}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial \omega})$ and $\mathrm{Kernel}((\pi_L)_*) = \mathrm{span}(\frac{\partial}{\partial x_2})$ above Σ. This shows that $\pi_L : \Lambda_c \to T^*Y_c \setminus \mathbf{0}$ has a fold along Σ.

Next we consider $\pi_R : \Lambda \to T^*X$. This is given by

$$\pi_R(s, x_1, x_2, \omega) = \left(x_1, x_2, -\omega\left(\frac{x_1-s-\alpha}{|x-\gamma_T(s)|}+\frac{x_1-s-\alpha}{|x-\gamma_R(s)|}\right), - \omega\left(\frac{x_2}{|x-\gamma_T(s)|}+\frac{x_2}{|x-\gamma_R(s)|}\right)\right). \tag{8}$$

We now show that π_R does not map to the zero section. For π_R to map to the zero section, we must have $x_2 = 0$ and

$$\frac{x_1 - s - \alpha}{|x - \gamma_T(s)|} + \frac{x_1 - s - \alpha}{|x - \gamma_R(s)|} = 0. \tag{9}$$

Using $x_2 = 0$ in (9), we see that

$$\frac{x_1 - s - \alpha}{|x_1 - s - \alpha|} + \frac{x_1 - s + \alpha}{|x_1 - s + \alpha|} = 0. \tag{10}$$

However, since $(x_1, 0)$ is on an ellipse with foci $(s - \alpha, 0)$ and $(s + \alpha, 0)$, either $x_1 < s - \alpha$ or $x_1 > s + \alpha$. Therefore, both terms in (10) are nonzero and have the same sign. This shows that π_R does not map to the zero section.

Now we show that π_R has a blowdown singularity along Σ. $(\pi_R)_*$ is the same as in [36], by letting $h = 0$ there. Then as in [36], we have that $\mathrm{Kernel}((\pi_R)_*) \subset T\Sigma$. Therefore, π_R has a blowdown singularity along Σ. □

Next we consider the imaging operators $\mathscr{R}_c^* \mathscr{R}_c$ and $\mathscr{R}_c^* D \mathscr{R}_c$ where D is a differential operator on Y_c. As in the last section (see Remark 1), we modify $\mathscr{R}_c$ first by multiplying it by an infinitely differentiable cutoff function χ_c that is identically 1 in a compact subset of Y_c and 0 outside a bigger compact subset.

Theorem 4 *The wavefront sets of $\mathscr{R}_c^* \mathscr{R}_c$ and $\mathscr{R}_c^* D \mathscr{R}_c$ satisfy the following*:

$$WF\big(\mathscr{R}_c^* \mathscr{R}_c\big) \subset \Delta \cup C_1, \tag{11}$$

$$WF\big(\mathscr{R}_c^* D \mathscr{R}_c\big) \subset \Delta \cup C_1, \tag{12}$$

where

$$\Delta := \big\{(x_1, x_2, \xi_1, \xi_2; x_1, x_2, \xi_1, \xi_2)\big\} \quad \textit{and}$$
$$C_1 := \big\{(x_1, x_2, \xi_1, \xi_2; x_1, -x_2, \xi_1, -\xi_2)\big\}.$$

Here, over the point $x = (x_1, x_2)$, (ξ_1, ξ_2) consists of all nonzero multiples of the vectors

$$-\nabla_x\big(\big|x - \gamma_T(s)\big| + \big|x - \gamma_R(s)\big|\big)$$

for all $s \in \mathbb{R}$.

We include the differential operator D in (12) because we will discuss a reconstruction algorithm using this type of operator in Sect. 4.

Remark 3 Similar to Remark 2, note that given a point x and foci locations $\gamma_T(s)$ and $\gamma_R(s)$, a vector (ξ_1, ξ_2) as in the theorem above is a vector perpendicular to the ellipse $E_c(s, L)$ where $(L = |x - \gamma_T(s)| + |x - \gamma_R(s)|)$ at the point x.

Remark 4 Theorem 4 describes the added singularities in any reconstruction algorithm $\mathscr{R}_c^* D \mathscr{R}_c f$. Let f be a function of compact support in X. Using (12), one may infer [31] that

$$WF\big(\mathscr{R}_c^* D \mathscr{R}_c\big)(f) \subset \big(\Delta \circ WF(f)\big) \cup \big(C_1 \circ WF(f)\big).$$

Now,

$$\Delta \circ WF(f) = WF(f) \tag{13}$$

and

$$C_1 \circ WF(f) = \big\{(x_1, -x_2, \xi_1, -\xi_2) : (x_1, x_2, \xi_1, \xi_2) \in WF(f)\big\}. \tag{14}$$

Therefore, the reconstruction operator $\mathscr{R}_c^* D \mathscr{R}_c f$ will show singularities of f by (13). However, the operator will also put singularities at the mirror points with respect to the x_1 axis. This is demonstrated by (14) because a singularity above the point (x_1, x_2) can cause a singularity above $(x_1, -x_2)$. We will observe this so called *left–right ambiguity* in our reconstructions in Sect. 4.2. Finally, we note that the vectors $(\xi_1; \xi_2)$ are all perpendicular to ellipses in the data set because of the condition on $(\xi_1; \xi_2)$ given at the end of Theorem 4.

Proof The proof is similar to the one give in [36]. Since we use a slightly different coordinate system, we will give it for completeness.

By the Hörmander–Sato Lemma, we have that $WF(\mathscr{R}_c^* \mathscr{R}_c) \subset \Lambda_c^t \circ \Lambda_c$, where

$$\begin{aligned}
&\Lambda_c^t \circ \Lambda_c \\
&= \Bigg\{\Bigg(x_1, x_2, -\omega\bigg(\frac{x_1 - s - \alpha}{|x - \gamma_T(s)|} + \frac{x_1 - s + \alpha}{|x - \gamma_R(s)|}\bigg), \\
&- \omega\bigg(\frac{x_2}{|x - \gamma_T(s)|} + \frac{x_2}{|x - \gamma_R(s)|}\bigg)\Bigg); \\
&\Bigg(y_1, y_2, -\omega\bigg(\frac{y_1 - s - \alpha}{|y - \gamma_T(s)|} + \frac{y_1 - s + \alpha}{|y - \gamma_R(s)|}\bigg), -\omega\bigg(\frac{y_2}{|y - \gamma_T(s)|} + \frac{y_2}{|y - \gamma_R(s)|}\bigg)\Bigg) \\
&: \big|x - \gamma_T(s)\big| + \big|x - \gamma_R(s)\big| = \big|y - \gamma_T(s)\big| + \big|y - \gamma_R(s)\big|, \\
&\frac{x_1 - s - \alpha}{|x - \gamma_T(s)|} + \frac{x_1 - s + \alpha}{|x - \gamma_R(s)|} = \frac{y_1 - s - \alpha}{|y - \gamma_T(s)|} + \frac{y_1 - s + \alpha}{|y - \gamma_R(s)|}, \ \omega \neq 0\Bigg\}.
\end{aligned}$$

We now obtain a relation between (x_1, x_2) and (y_1, y_2). This is given by the following lemma.

Lemma 3 *For all s, the set of all (x_1, x_2), (y_1, y_2) that satisfy*

$$\big|x - \gamma_T(s)\big| + \big|x - \gamma_R(s)\big| = \big|y - \gamma_T(s)\big| + \big|y - \gamma_R(s)\big|, \tag{15}$$

$$\frac{x_1 - s - \alpha}{|x - \gamma_T(s)|} + \frac{x_1 - s + \alpha}{|x - \gamma_R(s)|} = \frac{y_1 - s - \alpha}{|y - \gamma_T(s)|} + \frac{y_1 - s + \alpha}{|y - \gamma_R(s)|}, \tag{16}$$

necessarily satisfy the following relations: $x_1 = y_1$ *and* $x_2 = \pm y_2$.

Proof In order to show this, we use the following change of coordinates:

$$x_1 = s + \alpha \cosh\rho\cos\theta, \quad y_1 = s + \alpha\cosh\rho'\cos\theta',$$
$$x_2 = \alpha\sinh\rho\sin\theta, \qquad y_2 = \alpha\sinh\rho'\sin\theta'.$$

Using this change of coordinates, we have

$$\begin{aligned} &\left|x-\gamma_T(s)\right| = \alpha(\cosh\rho - \cos\theta), && \left|x-\gamma_R(s)\right| = \alpha(\cosh\rho+\cos\theta),\\ &\frac{x_1 - s - \alpha}{|x-\gamma_T(s)|} = \frac{\cosh\rho\cos\theta - 1}{\cosh\rho - \cos\theta}, && \frac{x_1 - s + \alpha}{|x-\gamma_R(s)|} = \frac{\cosh\rho\cos\theta+1}{\cosh\rho+\cos\theta}. \end{aligned} \tag{17}$$

The terms involving y are obtained similarly. Now (15) and (16) transform as follows:

$$2\cosh\rho = 2\cosh\rho',$$
$$\frac{\cosh\rho\cos\theta - 1}{\cosh\rho - \cos\theta} + \frac{\cosh\rho\cos\theta+1}{\cosh\rho+\cos\theta} = \frac{\cosh\rho'\cos\theta'-1}{\cosh\rho'-\cos\theta'} + \frac{\cosh\rho'\cos\theta'+1}{\cosh\rho'+\cos\theta'}.$$

Using the first equality in the second equation, we have

$$\frac{\cos\theta}{\cosh^2\rho - \cos^2\theta} = \frac{\cos\theta'}{\cosh^2\rho - \cos^2\theta'}.$$

This gives $\cos\theta = \cos\theta'$. Therefore, $\theta = 2n\pi \pm \theta'$, which then gives $\sin\theta = \pm\sin\theta'$. Therefore, in terms of (x_1, x_2) and (y_1, y_2), we have $x_1 = y_1$ and $x_2 = \pm y_2$. □

Now to finish the proof of the theorem, when $x_1 = y_1$ and $x_2 = y_2$, there is contribution to $WF(\mathscr{R}_c^*\mathscr{R}_c)$ contained in the diagonal set $\Delta := \{(x_1, x_2, \xi_1, \xi_2; x_1, x_2, \xi_1, \xi_2)\}$ and when $x_1 = y_1$ and $x_2 = -y_2$, we have a contribution to $WF(\mathscr{R}_c^*\mathscr{R}_c)$ contained in C_1, where $C_1 := \{(x_1, x_2, \xi_1, \xi_2; x_1, -x_2, \xi_1, -\xi_2)\}$. We restrict to vectors perpendicular to ellipses in the data set, vectors given by the condition at the end of Theorem 4, because these are the vectors in $\pi_R(\Lambda_c)$ in (8). Finally note that introducing a differential operator D does not add any new singularities, and so the same proof holds for the analysis of $WF(\mathscr{R}_c^* D\mathscr{R}_c)$. This completes the proof of the theorem. □

4 Our Algorithm and Reconstructions for the Common Offset Elliptical Radon Transform

In this section we describe the authors' algorithm and the refinements and implementation from [38] for the common-offset ellipse problem that was discussed in Sect. 3. Recall that the forward operator $\mathscr{R}_c$ and its dual $\mathscr{R}_c^*$ are both of order $-1/2$. Our reconstruction operator is

$$\Lambda(f) = \mathscr{R}_c^*\big(\chi_c D\big(\mathscr{R}_c(f)\big)\big), \tag{18}$$

where D is a well-chosen second-order differential operator, and χ_c is a compactly supported cut off in L. Therefore, Λ is an operator of order one, so it emphasizes boundaries and other singularities.

One includes the cutoff function χ_c because $\mathscr{R}_c(f)$ does not have compact support in general, even if f has compact support. Therefore, one cannot evaluate $\mathscr{R}_c^*$ on $\mathscr{R}_c(f)$ in general, without this cutoff. We will provide more details about χ_c and the differential operator D later in this section, but for the moment we will discuss this general type of algorithm.

An algorithm like (18) is called a derivative-backprojection operator because it takes a derivative and then takes some type of dual operator, a so-called backprojection operator. Such an algorithm will, typically, reconstruct singularities of the object, such as jumps at boundaries. It will image shapes and locations of objects rather than density values, and it is not an inversion method. Backprojection algorithms typically use other filters besides derivatives, and such algorithms have been considered in the context of bistatic SAR imaging in [57].

Therefore, researchers need to understand which singularities the algorithm reconstructs, which singularities are not imaged, and which singularities can be added to the reconstruction by the algorithm. This is one reason microlocal analysis and theorems in Sect. 3 are important.

Derivative-backprojection algorithms are useful in many problems, in particular when there is no inversion formula, when there is limited data, and when one is interested only in shapes, not density values.

The earliest modern tomography algorithm of this type is Lambda tomography, which was independently developed by Smith and Vainberg [55] (see [14, 15] for the state of the art). This algorithm is for planar X-ray tomography, and it is useful in medical [14] and industrial tomography (e.g., [49]). The planar Lambda operator is an elliptic pseudodifferential operator, so the reconstruction shows all singularities. However, in limited angle tomography some singularities can be invisible, as in electron microscopy [51].

In three-dimensional tomography problems, singularities can be spread, and artifacts can be created that are of the same strength as the original singularities. This occurs in local backprojection algorithms for cone beam 3-D CT (e.g., [34, 39]), and this was proven in [21] (see [25] for general admissible line complexes on manifolds). A derivative-backprojection reconstruction algorithm was developed for slant-hole SPECT in [50]. It was shown in [20] that if one chooses the right differential operator D, then the added singularities are suppressed in relation to the genuine singularities, and so they are less obvious in the reconstruction. Unfortunately, $\mathscr{R}_c$ spreads singularities in a more complicated way than the slant-hole SPECT transform, and it is an open problem to find a differential operator to globally decrease the strength of the added singularities.

4.1 Our Algorithm

The choices of the differential operator D and of the cutoff function χ_c in our reconstruction operator Λ (18) are important, and we describe them in this section.

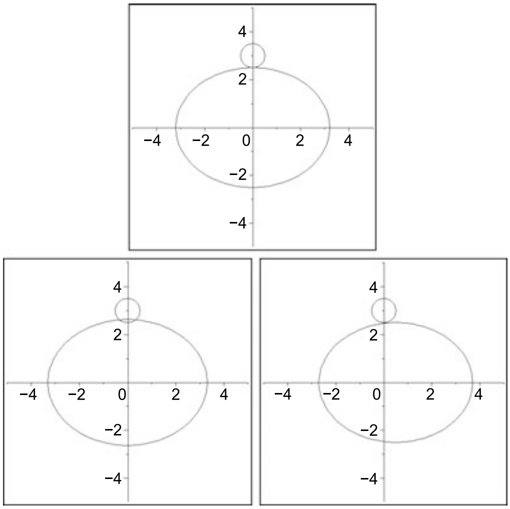


Fig. 1 The *top figure* shows the ellipse $E(s, L)$ tangent to a ball at the top point of the minor axis. The figure on the *lower left* shows the ellipse if L is increased slightly, and the ellipse intersects the ball. The figure on the *lower right* shows the ellipse if s is increased slightly. In this case, the ellipse remains outside the ball

It is shown in [38] that the operator

$$D = -\frac{\partial^2}{\partial L^2} \tag{19}$$

gives better reconstructions than the operator $-\frac{\partial^2}{\partial s^2}$. Boundaries are imaged more clearly as we will now explain using Fig. 1. Let f be the characteristic function of the ball in Fig. 1, and let x be the point of tangency of the ellipse in the top picture in Fig. 1. One can see from the lower left image in Fig. 1 that because the ellipse moves into the ball as L is increased, and the integral, $\mathscr{R}_c f(s, L)$, increases from zero like a square root function. Therefore $\frac{\partial^2}{\partial L^2}\mathscr{R}_c f$ will be unbounded at this ellipse. The reconstruction operator, $\mathscr{R}_c^* \frac{\partial^2}{\partial L^2}\mathscr{R}_c f$, averages $\frac{\partial^2}{\partial L^2}\mathscr{R}_c f$ all ellipses through x. Therefore, the reconstruction at x will be large.

However, movement in the s (horizontal) direction keeps the ellipse outside of the ball, so $\mathscr{R}_c f(s, L)$ remains zero, and $\frac{\partial^2}{\partial s^2}\mathscr{R}_c f$ will be zero at this ellipse. For ellipses nearby, the s derivative of the data will also be small. Therefore, the reconstruction operator, $\mathscr{R}_c^* \frac{\partial^2}{\partial s^2}\mathscr{R}_c f$, which averages this derivative on all ellipses through x will be small. These horizontal boundaries were almost invisible in the $\partial^2/\partial s^2$ reconstructions in [38]. If the ellipse was tangent at another point, then as s increased, the ellipse could intersect the ball, but $\mathscr{R}_c f$ would, in general, increase from zero more slowly than if L were increased and so the derivative in s would be smaller than the derivative in L.

4.2 Reconstructions

We now present reconstructions of the characteristic function of a ball of radius $1/2$ and centered at (0, 1): $B((0, 1), 1/2)$. The backprojection $\mathscr{R}_c^*$ is implemented using

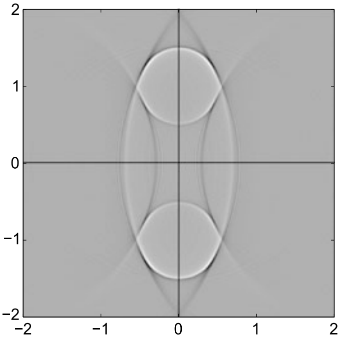


Fig. 2 Reconstruction of the ball $B((0,1),1/2)$ using the function χ_c supported on $[-3,3]$ and equal to 1 on $[-9/4,9/4]$

the trapezoidal rule, and the derivative D is implemented using a central second difference. The common offset is $d = 1/4$ ($\alpha = 1/8$). Details are in the second author's senior honors thesis [38].

We will analyze both types of artifacts in the reconstructions, those caused by the left–right ambiguity and those caused by the limited range on s.

As noted in Remark 4 after Theorem 4, the reconstruction operator (18) for the common offset elliptical transform has the *left–right ambiguity*: singularities on one side of the x_1 axis are reflected on the other side in the reconstruction. This explains why our reconstructions put copies of the circle on both the right and left of the flight path. This global spreading of singularities is more difficult to decrease than the local spreading in SPECT [20, 50] and electron microscopy [51].

The second type of added singularity is the "parentheses" surrounding and tangent to the circle, and they are explained by the limited values of s or, equivalently, the support of the cutoff χ_C. The choice of cutoff function χ_c makes an important difference to the reconstruction [38]. Two parameters, $M > m > 0$, are chosen, and the cutoff function $\chi_c(L)$ is supported in $[-M, M]$ and equal to one in $[-m, m]$. In this case χ_c does not need to be compactly supported on Y_c but only in L since the functions we reconstruct have compact support.

If one looks carefully at the reconstructions, one can show that the "parentheses" artifacts are parts of the boundaries of ellipses that are tangent to the circle and with $s = -M$ (for the ones "pointing" right) and with $s = M$ (for the ones "pointing" left). These are ellipses with foci at $(-3, 0)$ and $(-2.75, 0)$ tangent to the ball and two with foci at $(2.75, 0)$ and $(3, 0)$ tangent to the ball.

The authors believe that there are both microlocal reasons and practical grounds for these elliptical artifacts. If the integration had been over $[-3, 3]$ without a smooth cutoff χ_c, then the operator would not have smooth kernel, and that could cause the artifact because Λ would not be a smooth Fourier integral operator. However, the algorithm includes the smooth function χ_c and there is still an artifact. In order to reduce the effect of these artifacts, we changed the cutoff χ_c. In Fig. 3, the artifacts caused by these ellipses are decreased, but somewhat fewer singularities are visible than in the reconstruction in Fig. 2, which is with a sharper cutoff.

Smith's implementation of Lambda tomography includes a constant term in the derivative D. This shows contours of the object because it adds a multiple of the

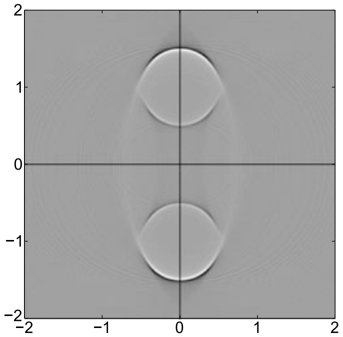


Fig. 3 Reconstruction of the ball $B((0,1),1/2)$ using the function χ_c supported on $[-3,3]$ and equal to 1 only at the origin, [0, 0]

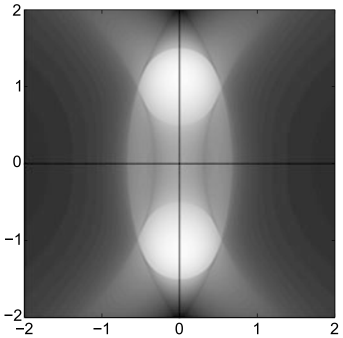


Fig. 4 Reconstruction with $D = 1 - \frac{\partial^2}{\partial L^2}$, which includes the simple backprojection as well as the derivative in L

simple backprojection; for our case, it would be $\mathscr{R}_c^* \chi_c \mathscr{R}_c$. The reconstruction in Fig. 4 illustrates this, and the inside of the ball has higher "density" than the outside.

5 Discussion

In this section, we will discuss the implications of our work for bistatic SAR, and we will suggest some open problems and conjectures.

The elliptical Radon transforms we consider in this article, while motivated by bistatic SAR imaging, are simplifications of the operators that appear in bistatic SAR. In our case, the transmitter and receiver are on the ground, and in general, in SAR, they are above the ground. The canonical relations in SAR are different from ours, but they become the same if the transmitter and receiver are on the ground. The SAR operators are also FIOs of a different order. For the common offset case with transmitter and receiver above the ground, the projections are a fold and blowdown [36], as in our case. It is easy to see that the same holds for the transform with one fixed focus above the origin and the other moving above the horizontal axis.

The appearance of ambiguities is a serious issue in SAR imaging. In the acquisition geometries we considered in this paper, we showed in Theorems 2 and 4 that there are only left-right ambiguities. We can decrease such ambiguities by focusing the antenna beam (known as beam forming) to the right or left of the flight path.

However, one then images only one half of the scene, and one must fly over the scene again to image the other side. In general, one needs to know the nature and structure of such ambiguities in order to decide if focusing the beam could decrease these ambiguities. For general bistatic acquisition geometries, this is an open problem. The structure of ambiguities could be very complicated in this case.

Monostatic SAR has colocated transmitter and receiver. For such SAR systems, more is known. For linear flight trajectory, a similar theorem to Theorem 3 is true, namely for that canonical relation, π_L is a fold along the set Σ at which it drops rank, and π_R is a blowdown along Σ, and Theorem 4 is also true in this case [16, 17, 42].

For linear flight paths and monostatic or bistatic SAR, it is conjectured that, without beam forming, the left–right ambiguity is intrinsic to the problem and cannot be eliminated.

However, for other flight paths, more can be done. Injectivity holds for the circular transform with centers on a curve as long as the curve is not a line or a Coxeter system of lines [1]. This suggests, but does not prove, that the general monostatic SAR transform is injective for such curves. For nonlinear flight tracks, there is a local left–right ambiguity as can be seen from reconstructions in [38]. However, these added singularities seem to be spread and look quantitatively weaker than for linear flight tracks. Felea [17] showed that for the monostatic SAR transform with circular flight tracks, one can displace added singularities far away from the image [17].

We conjecture that Felea's methods would work for the circular transform because it has the same canonical relation as the monostatic SAR transform. For the elliptical Radon transform with transmitter and receiver a fixed distance apart along a circle, the reconstruction operator is an elliptic pseudodifferential operator as long as the scene is sufficiently inside the circle [2]. This suggests that ideas in [17] might be helpful for the bistatic case with circular trajectories.

Nolan and Dowling [44] showed that if one takes monostatic data over two perpendicular linear flight paths, then the added singularities caused by the left–right ambiguity are quantitatively weaker than the image itself. This makes sense because, when one adds the images, only the real image reinforces itself.

The authors and their colleagues will continue investigating novel flight paths and reconstruction algorithms, evaluating them using microlocal analysis as we have done in this article for the elliptical transform with one fixed focus and for the common-offset case.

Acknowledgements All three authors were supported by Quinto's NSF grant DMS 0908015. Krishnan was supported by summer postdoctoral supplements to Quinto's NSF grant DMS 0908015 in 2010 (DMS 1028096) and 2011 (DMS 1129154), and by NSF grant DMS 1109417. Levinson received REU support.

References

1. Agranovsky, M., Quinto, E.T.: Injectivity sets for Radon transform over circles and complete systems of radial functions. J. Funct. Anal. **139**, 383–414 (1996)

2. Ambartsoumian, G., Krishnan, V., Quinto, E.T.: The microlocal analysis of the ultrasound operator with circular source and receiver trajectory. Tech. rep., University of Texas at Arlington, University of Bridgeport, Tufts University (2011, submitted)
3. Andersson, L.E.: On the determination of a function from spherical averages. SIAM J. Math. Anal. **19**, 214–232 (1988)
4. Boman, J.: Helgason's Support Theorem for Radon Transforms—A New Proof and a Generalization. Lecture Notes in Mathematics, vol. 1497, pp. 1–5. Springer, Berlin (1991)
5. Boman, J., Quinto, E.T.: Support theorems for real analytic Radon transforms. Duke Math. J. **55**, 943–948 (1987)
6. Cheney, M.: A mathematical tutorial on synthetic aperture radar. SIAM Rev. **43**(2), 301–312 (2001) (electronic). doi:10.1137/S0036144500368859
7. Cheney, M., Borden, B.: Microlocal structure of inverse synthetic aperture radar data. Inverse Probl. **19**(1), 173–193 (2003). doi:10.1088/0266-5611/19/1/310
8. Cheney, M., Borden, B.: Fundamentals of Radar imaging. In: CBMS-NSF Regional Conference. Series in Applied Mathematics, vol. 79. Society for Industrial and Applied Mathematics, Philadelphia (2009)
9. Courant, R., Hilbert, D.: Methods of Mathematical Physics, vol. II. Wiley-Interscience, New York (1962)
10. Miller, D.M.O., Beylkin, G.: A new slant on seismic imaging: Migration and integral geometry. Geophysics **52**(7), 943–964 (1987)
11. deHoop, M., Smith, H., Uhlmann, G., van der Hilst, R.: Seismic imaging with the generalized Radon transform: A curvelet transform perspective. Inverse Probl. **25** (2009). doi:10.1088/0266-5611/25/2/025005
12. Ehrenpreis, L.: Three Problems at Mount Holyoke. Contemporary Math., pp. 123–130 (2001)
13. Ehrenpreis, L.: The Universality of the Radon Transform. Oxford University Press, Oxford (2003)
14. Faridani, A., Finch, D., Ritman, E.L., Smith, K.T.: Local tomography, II. SIAM J. Appl. Math. **57**, 1095–1127 (1997)
15. Faridani, A., Ritman, E.L., Smith, K.T.: Local tomography. SIAM J. Appl. Math. **52**, 459–484 (1992)
16. Felea, R.: Composition of Fourier integral operators with fold and blowdown singularities. Commun. Partial Differ. Equ. **30**(10–12), 1717–1740 (2005). doi10.1080/03605300500299968
17. Felea, R.: Displacement of artefacts in inverse scattering. Inverse Probl. **23**(4), 1519–1531 (2007). doi:10.1088/0266-5611/23/4/009
18. Felea, R., Greenleaf, A.: An FIO calculus for marine seismic imaging: folds and cross caps. Commun. Partial Differ. Equ. **33**(1–3), 45–77 (2008). doi:10.1080/03605300701318716
19. Felea, R., Greenleaf, A., Pramanik, M.: An FIO calculus for marine seismic imaging, II: Sobolev estimates. Tech. rep., University of Rochester, Rochester Institute of Technology. Preprint (2010)
20. Felea, R., Quinto, E.T.: The microlocal properties of the local 3-D SPECT operator. SIAM J. Math. Anal. **43**(3), 1145–1157 (2011)
21. Finch, D.V., Lan, I.R., Uhlmann, G.: Microlocal analysis of the restricted X-ray transform with sources on a curve. In: Uhlmann, G. (ed.) Inside out, Inverse Problems and Applications. MSRI Publications, vol. 47, pp. 193–218. Cambridge University Press, Cambridge (2003)
22. Funk, P.: Über eine geometrische Anwendung der Abelschen Integralgleichung. Math. Ann. **77**, 129–135 (1916)
23. Globevnik, J.: Zero integrals on circles and characterizations of harmonic and analytic functions. Trans. Am. Math. Soc. **317**, 313–330 (1990)
24. Golubitsky, M., Guillemin, V.: Stable Mappings and Their Singularities. Graduate Texts in Mathematics, vol. 14. Springer, New York (1973)
25. Greenleaf, A., Uhlmann, G.: Nonlocal inversion formulas for the X-ray transform. Duke Math. J. **58**(1), 205–240 (1989). doi:10.1215/S0012-7094-89-05811-0

26. Greenleaf, A., Uhlmann, G.: Composition of some singular Fourier integral operators and estimates for restricted X-ray transforms. Ann. Inst. Fourier (Grenoble) **40**(2), 443–466 (1990)
27. Guillemin, V.: Some remarks on integral geometry. Tech. rep., MIT (1975)
28. Guillemin, V.: On some results of Gelfand in integral geometry. Proc. Symp. Pure Math. **43**, 149–155 (1985)
29. Guillemin, V., Sternberg, S.: Geometric Asymptotics. American Mathematical Society, Providence (1977). Mathematical Surveys, No. 14
30. Helgason, S.: The Radon transform on Euclidean spaces, compact two-point homogeneous spaces and Grassmann manifolds. Acta Math. **113**, 153–180 (1965)
31. Hörmander, L.: The Analysis of Linear Partial Differential Operators. I. Classics in Mathematics. Springer, Berlin (2003). Distribution theory and Fourier analysis, Reprint of the second (1990) edition [Springer, Berlin; MR1065993 (91m:35001a)]
32. Horne, A., Yates, G.: Bistatic synthetic aperture radar. In: Proceedings of IEEE Radar Conference, pp. 6–10 (2002)
33. John, F.: Plane Waves and Spherical Means Applied to Partial Differential Equations. Interscience, New York (1966)
34. Katsevich, A.: Improved cone beam local tomography. Inverse Probl. **22**, 627–643 (2006)
35. Krishnan, V.P.: A support theorem for the geodesic ray transform on functions. J. Fourier Anal. Appl. **15**(4), 515–520 (2009). doi:10.1007/s00041-009-9061-5
36. Krishnan, V.P., Quinto, E.T.: Microlocal aspects of bistatic synthetic aperture radar imaging. Inverse Probl. Theor. Imaging **5**, 659–674 (2011)
37. Krishnan, V.P., Stefanov, P.: A support theorem for the geodesic ray transform of symmetric tensor fields. Inverse Probl. Theor. Imaging **3**(3), 453–464 (2009)
38. Levinson, H.: Algorithms for Bistatic Radar and Ultrasound Imaging, Senior Honors Thesis with Highest Thesis Honors. Tufts University (2011)
39. Louis, A.K., Maaß, P.: Contour reconstruction in 3-D X-Ray CT. IEEE Trans. Med. Imaging **12**(4), 764–769 (1993)
40. Mensah, S., Franceschini, E.: Near-field ultrasound tomography. J. Acoust. Soc. Am. **121**(3), 1423–1433 (2007). doi:10.1109/TIP.2009.2039662
41. Nolan, C.J., Cheney, M.: Synthetic aperture inversion. Inverse Probl. **18**(1), 221–235 (2002). doi:10.1088/0266-5611/18/1/315
42. Nolan, C.J., Cheney, M.: Microlocal analysis of synthetic aperture radar imaging. J. Fourier Anal. Appl. **10**(2), 133–148 (2004). doi:10.1007/s00041-004-8008-0
43. Nolan, C.J., Cheney, M., Dowling, T., Gaburro, R.: Enhanced angular resolution from multiply scattered waves. Inverse Probl. **22**(5), 1817–1834 (2006). doi:10.1088/0266-5611/22/5/017
44. Nolan, C.J., Dowling, T.: Private communication
45. Nolan, C.J., Symes, W.W.: Global solution of a linearized inverse problem for the wave equation. Commun. Partial Differ. Equ. **22**(5–6), 919–952 (1997). doi:10.1080/03605309708821289
46. Quinto, E.T.: The dependence of the generalized Radon transform on defining measures. Trans. Am. Math. Soc. **257**, 331–346 (1980)
47. Quinto, E.T.: Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$. SIAM J. Math. Anal. **24**, 1215–1225 (1993)
48. Quinto, E.T.: Support theorems for the spherical Radon transform on manifolds. Int. Math. Res. Not. **2006**, 1–17 (2006). Article ID = 67205
49. Quinto, E.T.: Local algorithms in exterior tomography. J. Comput. Appl. Math. **199**, 141–148 (2007)
50. Quinto, E.T., Bakhos, T., Chung, S.: A local algorithm for Slant Hole SPECT. In: Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT), Centro De Georgi, Pisa. CRM Series, vol. 7, pp. 321–348 (2008)
51. Quinto, E.T., Öktem, O.: Local tomography in electron microscopy. SIAM J. Appl. Math. **68**, 1282–1303 (2008)
52. Stefanov, P.: Microlocal approach to tensor tomography and boundary and lens rigidity. Serdica Math. J. **34**(1), 67–112 (2008)

53. Stefanov, P., Uhlmann, G.: Boundary rigidity and stability for generic simple metrics. J. Am. Math. Soc. **18**(4), 975–1003 (2005) (electronic). doi:10.1090/S0894-0347-05-00494-7
54. Vaidyanathan, R.S., Lewis, M., Ambartsoumian, G., Aktosun, T.: Reconstruction algorithms for interior and exterior spherical Radon transform-based ultrasound imaging. In: Proceedings of SPIE, v. 7265, Medical Imaging 2009: Ultrasonic Imaging and Signal Processing, pp. 72,651I 1–8 (2009)
55. Vainberg, E., Kazak, I.A., Kurozaev, V.P.: Reconstruction of the internal three-dimensional structure of objects based on real-time integral projections. Sov. J. Nondestr. Test. **17**, 415–423 (1981)
56. Volchkov, V.V.: Integral Geometry and Convolution Equations. Kluwer Academic, Dordrecht (2003)
57. Yarman, C.E., Yazıcı, B., Cheney, M.: Bistatic synthetic aperture radar imaging with arbitrary trajectories. IEEE Trans. Image Process. **17**(1), 84–93 (2008)
58. Zalcman, L.: Offbeat integral geometry. Am. Math. Mon. **87**, 161–175 (1980)

Mathematics of Hybrid Imaging: A Brief Review

Peter Kuchment

Abstract The article provides a brief survey of the mathematics of newly being developed so-called "hybrid" (also called "multi-physics" or "multi-wave") imaging techniques.

1 Introduction

Leon Ehrenpreis was a mathematician of remarkable strength, famous accomplishments, and extremely wide interests. In the last couple of decades of his life, integral geometry was one of the main areas he was interested in. This had lead in particular to his involvement with problems of tomography [7, 53]. E.T. Quinto and the author had the honor of writing at Leon's request an Appendix [83] dedicated to tomography for his last book [52]. He was also interested in new developments in medical imaging, which in some instances turned out to be directly related to some integral geometric and PDE problems he considered in [52]. It is thus appropriate to address some of these new techniques in an article dedicated to Leon's memory.

It is natural to wonder, why do we need new methods of medical imaging in the first place, if we already have the whole bunch of well-developed ones [95, 104]? Indeed, one can mention the widely known "standard" X-ray CT (Computed Tomography) scanners, MRI, and ultrasound scanners, which can be found in most hospitals nowadays. There are also less known to the general public, but well developed by now PET (Positron Emission Tomography) and SPECT (Single Photon Emission Computed Tomography) techniques, Optical Tomography (OT), Electrical Impedance Tomography (EIT), Elastography, as well as quite a few others.

It seems that even the existing methods are way too many. Why would one need all of them? One of the main answers is that different imaging techniques "see" different things. One can say (in a crude approximation) that the CT scan can distinguish between the tissues of different density. What if two tissues (e.g., the cancer-

Dedicated to the memory of Professor Leon Ehrenpreis, a great mathematician, human being, and friend.

P. Kuchment (✉)
Texas A&M University, College Station, TX, 77843-3368, USA
e-mail: kuchment@math.tamu.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_12, © Springer-Verlag Italia 2012

ous and the healthy one) have essentially the same density but absorb significantly different amounts of light in certain part of the spectrum? One can think of using OT then rather than the X-ray CT. Some modalities can show the metabolism (e.g., the PET and SPECT), while some others cannot. Some new methods can show the level of oxygenation in blood, while those relying upon density would not be able to do so. This list can go on and on, and so the reader can see the point in having a variety of imaging techniques.

There are several other parameters that make a difference when using different types of scanners. Here are the most common ones:

1. *Safety* for the patient and practitioner. Indeed, X-ray scans are clearly not too safe, while, for instance, OT or ultrasound tomography are not harmful. If there is a choice, one surely would shoot for a safer method.
2. *Cost*, in the times of the high medical expenses, is clearly one of the major issues. Some imaging techniques (X-ray CT-scan, MRI, PET, and some others) require very expensive devices, with the price tags in millions of dollars. Some others, however, e.g., OT and EIT, are orders of magnitude cheaper.
3. *Contrast* is another important feature. For instance, if the parameter that a method can detect is, say, electric conductivity of the tissue, then one will be able to distinguish between the tissues that have a significant conductivity contrast and will not see any difference between tissues that happen to have very close electric conductivities. So, high contrast between the features that we would like to distinguish is crucial. One clearly would prefer the contrast that is on the order of hundreds (or at least dozens) of percents, while one percent contrast, albeit still usable, is much less desirable. For instance, some breast tumors on early stages might have almost no contrast with the healthy tissues with respect to ultrasound propagation, but a huge contrast (several hundreds of percents) in their optical and electric properties.
4. *Resolution*, which determines what is the smallest size of a feature that a method can "see," is another very important parameter. Indeed, resolution of several centimeters probably is not good for early breast tumor diagnostics, where submillimeter resolution is desirable.

These are the reasons, why the quest for new and "better" (at least in some of the parameters) imaging methods not only does not subside, but intensifies in the recent decades, involving new physics, engineering, and mathematics ideas. However, all attempts to find a "magic wand" method that would "see everything" and feature low cost, safety, high contrast and high resolution are clearly futile. Thus, having a variety of techniques at our disposal is apparently the way to go.

For a mathematician, however, the main motivation for working on these various tomographic modalities is that they bring about a large variety of challenging and beautiful mathematical problems that involve more or less all areas of mathematics.

Let us describe now the structure of this article. After a general discussion of hybrid methods in Sect. 2, we will concentrate in more details on three of the emerging hybrid methods. The largest Sect. 3 contains an overview of the most developed (both experimentally and mathematically) Thermo-/Photo-acoustic Tomography (TAT/PAT). The next, shorter, Sect. 4 addresses the much newer technique of Acousto-Electric Tomography.[1] Although AET was initially suggested (and its principle experimentally proven) by biomedical engineers [148], here mathematics is getting developed fast in the recent few years, while experimentalist still struggle with reaching good signal-to-noise ratios (SNR) in the measurements. This situation is reversed in the Ultrasound Modulated Optical Tomography (UMOT, also called UOT), which by now is significantly studied experimentally, while the mathematics of this modality is still in its infancy (and even mathematical models are still being agreed upon). This topic is addressed in ever shorter Sect. 6.

Both AET and UMOT rely upon an assumption of a "perfect focusing" of ultrasound at a given location, which is a crude approximation to the reality (see, e.g., [66]). The synthetic focusing technique, discussed in Sect. 5, allows one to use more realistic sets of ultrasound waves.

The common feature of AET and UMOT (as well as some other hybrid imaging techniques, such as CDI and MREIT, which are just mentioned in this text) is that the measurements provide the researcher with some interior information, i.e., a function of an interior location x. There has been a rather common feeling that such an interior information might stabilize the notoriously unstable modalities such as EIT or OT. This issue is briefly discussed in Sect. 7.

The topics surveyed in this article are highly technical and involve a wide range of interesting mathematics. Due to space limitations, the author was forced to show very few technical details and instead to try to give a hand-waving heuristic description. The literature references (as well as the references therein) will provide the interested reader with more details.

2 Hybrid Imaging Methods

As we have indicated in the introduction, each of the available imaging methods has its advantages and deficiencies. For instance, in breast imaging ultrasound provides a high (sub-millimeter) resolution while suffers from a low contrast. On the other hand, many tumors absorb much more energy of electromagnetic waves (in some specific energy bands) than healthy cells. Thus, using such electromagnetic waves offers very high contrast. Alas, the resolution in this case is very low. One can go on and on with such examples.

Since both the advantages and disadvantages of various modalities come from the underlying physics, can one do anything to combine the advantages and simultaneously alleviate the problems associated with different type of physical waves/radiation involved? The natural idea is to try to this end to combine different imaging modalities into some kind of "hybrid" ones. This is how the *hybrid* (also called *multi-physics* or *multi-wave*) *techniques* have been appearing in the last decade–decade and a half [10, 16, 19, 121, 131, 134, 138, 139, 141].

[1] Besides the name AET, suggested in the original paper [148], other names are also used: Ultrasound Modulated EIT (UMEIT) [17], Impediography [10], and some others.

The "hybridization" can occur at different stages of the imaging process. Let us recall the crude scheme of all CT methods: on the first step, some wave(s) are sent through the body and the outgoing (transmitted or reflected) waves are measured; on the second stage, mathematical processing of the measured data is done; finally, the third stage provides a picture (tomogram). Correspondingly, one can combine different techniques at different stages. The simplest, and already industrially implemented (e.g., in PET/CT scanners) idea is to run two types of scans of the same patient and then somehow "combine" the images so that they hopefully complement each other. Here reconstruction of both individual images requires neither new types of scanners nor new mathematical reconstruction algorithms. Certainly, some (often nontrivial) processing is needed to correctly overlap the two images (the so-called *image registration*).

Another option is to combine the techniques on the second stage. Namely, after collecting data from two independent types of scans, a reconstruction algorithm is applied that uses both data sets. The additional information can significantly improve the quality (stability, resolution, etc.) of the resulting picture (tomogram). Such combined techniques might not require any new physics or engineering, but do demand new mathematical processing. There are several recent imaging procedures that successfully implement this idea. Some of them (CDI, current density imaging [99–101], and MREIT [122, 142]) combine MRI and EIT measurements. Having the extra MRI data makes the mathematical problem of EIT reconstruction (the so-called Calderón problem, or inverse conductivity problem [15, 27, 31, 37, 41, 73, 95, 98, 129, 132, 133, 135]), known for its severe ill-posedness, significantly less ill-posed and thus allows for good quality reconstructions of the internal electrical conductivity maps. Another actively developing method of this kind, called MRE [96, 97], combines MRI with elastography: MRI allows one to observe propagation of elastic waves through the tissue, which then leads to mathematical reconstruction of mechanical properties (e.g., stiffness) that carry important medical diagnostic information. Probably the oldest such combination is of CT and SPECT. The CT scan provides the reconstruction of the attenuation map, which is used then to recover the distribution of a radio-pharmaceutical inside the patient's body [33].

Due to the author's limited expertise and lack of space, these types of hybrid methods will not be discussed in the article. The interested readers are referred to the literature cited above.

We reserve in this text the name "hybrid methods" only for the techniques that combine different types of waves already on the first, scanning stage. In the examples that we will discuss this will lead to one physical type of irradiation triggering or modulating the other one and thus producing new types of measurements, which hopefully allow one to improve images in comparison with the two techniques done separately.[2]

[2]The reader should be aware that this classification of hybrid modalities into three classes, although being reasonable, is not commonly accepted and is used in this text only for the author's convenience.

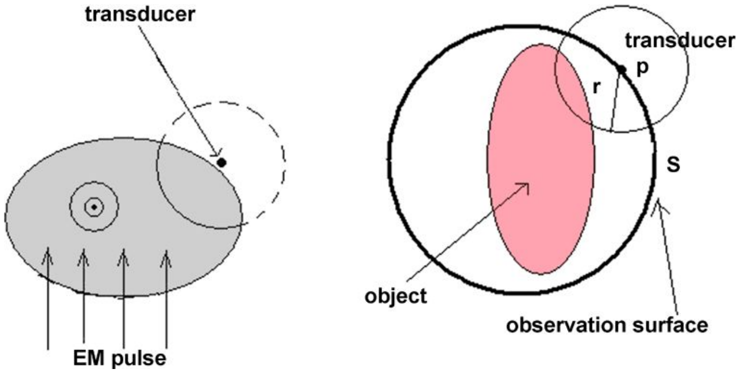


Fig. 1 The TAT procedure

We now move to considerations of some of the hybrid techniques in more detail. We start in the next section with the probably best developed (both experimentally and mathematically) among the hybrid techniques Thermo-/Photo-acoustic tomography (TAT/PAT) and then move to the less studied ultrasound modulated electrical and optical tomography.

3 Thermo-/Photo-Acoustic Tomography (TAT/PAT)

As we have already mentioned, in many medical diagnostics situations, ultrasound displays low contrast (and thus sees the tissues as almost homogeneous), while providing fine resolution. Optical or radio-frequency EM illumination, on the other hand, gives an enormous contrast between the cancerous and healthy cells, while both are known to suffer from low resolution. How can one combine their strengths? The answer is in the photo-acoustic effect, which was discovered by Alexander Graham Bell [28] but had to wait for another century for its applications to follow [32].

Imagine that a biological object is irradiated by a wide, homogeneous, but extremely short electro-magnetic pulse in radio-frequency range (Fig. 1). Some part of the electromagnetic (EM) energy will be absorbed throughout the tissues. Let us denote by $f(x)$ the density of energy absorbed at a location x. It is known that the values of this absorption function will be several times higher at the cancerous locations than in the surrounding healthy tissues [50, 89, 90, 108, 109, 130, 131, 138–141, 143]. Thus, if this function were known, it would provide an extremely valuable diagnostic information. However, the radio-frequency waves are too long to lead to any reasonable resolution. This is where the photo-acoustic effect kicks in: the EM energy absorption leads to heating, and higher levels of absorption lead to more heating. In turn, the resulting thermoelastic expansion creates a pressure wave $p(x,t)$ (acoustic wave), whose initial value is essentially (proportional to) the function of interest $f(x)$. Placing an ultrasonic transducer (a microphone) at a location y at the boundary of the object, one can measure the value of $p(y,t)$ at this point for any time $t \geq 0$. If we now surround the object by transducers located on an *observation surface* S (this can be done using optical interferometers, which do not obstruct the acoustic wave propagation), we can collect the values of $p(y,t)$ for all $(y,t) \in S \times \mathbb{R}^+$ (Fig. 1). Now the problem reduces to finding the initial values $p(x,0)$ inside the volume surrounded by the surface S. This is the idea of Thermoacoustic Tomography (TAT), which found its implementation in the middle of 1990s [75, 108, 109], in particular in the work of R. Kruger [75], who started a company OptoSonics manufacturing TAT devices. The Photo-acoustic version (PAT) differs by the choice of heating radiation, a laser beam instead of radio waves. A large part of the corresponding mathematics is parallel in TAT and PAT cases, and thus we will only mention TAT here.

The mathematics of TAT/PAT reconstruction happens to be fascinating and occupied attention of a large group of mathematicians throughout the 1st decade of this century. The reader can consult with the recent surveys, collections, and books [3, 12, 19, 56, 58, 77, 80, 84, 111, 114, 121, 124, 131, 134, 138] for details and further references.

We will now describe the mathematical model of TAT. Let us denote by $c(x)$ the sound speed in the interior of the body. Then the pressure p satisfies the following wave equation:

$$\begin{cases} \frac{\partial^2 p}{\partial t^2} = c^2(x)\Delta p \quad \text{in } \mathbb{R}^3 \times \mathbb{R}_+, \\ p(0,x) = f(x), \\ \frac{\partial p}{\partial t}(0,x) = 0. \end{cases} \tag{1}$$

The data g measured by a TAT machine provides the values of the pressure p on the observation surface S:

$$f(x) \mapsto g(y,t) := p\,|_{S\times\mathbb{R}_+}. \tag{2}$$

The goal of TAT thus is inverting this forward operator.

3.1 TAT/PAT and (Restricted) Spherical Mean Operators

In the case of an acoustically homogeneous medium (i.e., when $c(x) = \text{const}$), one can reduce the TAT problem to an equivalent integral geometry question. Namely, using the standard Kirchhoff–Poisson formulas [46, 71] for the solution of (1), the TAT inversion can be reduced to the equivalent (see [6]) problem of recovery the function $f(x)$ from its spherical averages over spheres centered at the transducers' locations, i.e., on the observation surface S:

$$f \mapsto g(y,t) := (R_S f)(y,t) = \int_{|x-y|=t} f(x)\,d\sigma(x), \quad y \in S,\ t \geq 0. \tag{3}$$

Here R_S denotes the operator of taking spherical averages over all spheres centered on S.

One can notice that such transforms were considered in more general situation and without any relation to tomography by Leon Ehrenpreis in his book [52] and by V. Lin and A. Pinkus in [92, 93] for the needs of approximation theory and neural networks. The restricted spherical mean transforms also play important role in the Radar and Sonar studies [40, 94]. One can find a list of other applications in [6].

3.2 Main Mathematical Problems in TAT

As in all tomographic methods, the following questions play the central role:

Uniqueness of reconstruction: Is the collected data g sufficient for the unique reconstruction of the tomogram (function $f(x)$)? When the author first looked at this problem in terms of system (1), he was confused for a second. Indeed, the measured data g seems to be the boundary value of the solution of the wave equation in a cylinder, and the function f to recover is the initial condition. This clearly is impossible, since g essentially does not carry any information about f, besides the standard junction conditions where these functions meet. However, the thing is that (1) is not a boundary-value problem in a cylinder, but rather a problem in the whole space, whose solution we observe on the surface S only. What is even more important, if the sound speed is nontrapping (e.g., constant) and S is a closed surface (i.e., the boundary of a bounded domain), then the local energy decay theorems [35, 51, 119, 136, 137] show that the solution decays in the interior of S. This turns out to be the main feature that brings about uniqueness and inversion [2, 68, 69, 124].
The question of sufficiency of the data collected on a nonclosed hyper-surface S is much more complicated and can be considered largely resolved only in $2D$, when the sound speed is constant and the function $f(x)$ is compactly supported (see [6],[3] as well as surveys [3, 77, 80] for the results and further discussion). This is a fascinating mathematical problem, which involves a variety of techniques from PDEs, integral, differential, and algebraic geometry, microlocal analysis, commutative algebra, etc. A more general problem (with spheres replaced by the level surfaces of a polynomial) is discussed in L. Ehrenpreis' book [52].

Reconstruction procedures: Having a uniqueness of reconstruction theorem is far from being sufficient for tomography, since one needs to be able to actually reconstruct the tomogram. So, the natural question, after proving uniqueness, is: how can one actually invert the mapping $f \mapsto g$? We will briefly describe the known algorithms in the next subsection.

Stability: Having uniqueness of reconstruction theorem, or even a reconstruction algorithm, is no guarantee that the reconstruction is practicably feasible (or

[3] Some microlocal arguments in [6], although correct, were incomplete. The missing arguments are provided in [123].

at least that it can provide sharp pictures). The word "stability" in this context means the effect that small errors in the measured data have on the reconstructed image. In other words, stability means well-posedness (in the Hadamard's sense [46]) of the inverse problem. Unfortunately, essentially all tomographic problems are ill-posed to some degree. Some of them, such as Radon transform inversion, are very mildly unstable, which allows for wonderful CT-scan images. Some others, like EIT and OT, are known to be severely ill-posed, and thus reconstruction of sharp images is practically impossible.

The stability of the TAT/PAT reconstruction with full data (i.e., for the observation surface surrounding the object completely) is known to be very good, the same as for the Radon transform inversion, which leads to excellent reconstructions. This applies both to the case of an acoustically homogeneous object (when $c(x) = \text{const}$), as well as to the case of a variable nontrapping sound speed $c(x)$ [2, 69, 107, 124, 126, 127]. The proofs are based upon inversion procedures and/or a microlocal analysis of the problem.

Range: As it is common in integral geometry and tomography [60, 61, 63, 64, 104, 105], the range of the forward operator $f \mapsto g$ has infinite codimension in the natural scales of function spaces. The description of this range is an important part of integral geometric and tomographic studies [ibid]. In the case of a spherical observation surface S and constant sound speed, the complete range descriptions are known [1, 4, 5, 9, 57]. Some of these range conditions are known to be necessary for more general observation surfaces and sound speeds, but in these cases complete range description is not known (and might be impossible).

Incomplete data: In the TAT/PAT case, one usually mentions an *incomplete data situation*, when the observation surface S does not surround the object completely [3, 77, 80, 145, 146]. This is a common, albeit somewhat misleading description, since depending on the location of the object, the "incomplete" data might be sufficient for unique, and sometimes even stable, reconstruction. It is quite common that even a small observation surface S can provide enough data for proving uniqueness of reconstruction. For example, in the case of a constant sound speed, it is known that any set S that is not a part of an algebraic hyper-surface leads to the injectivity of the spherical mean operator R_S and thus to unique TAT reconstruction (see [6] and references therein). Note that such S can be geometrically very small. It is clear that for all practical purposes, in spite of an uniqueness theorem, something should go wrong with the actual reconstruction in this case. And indeed, microlocal arguments similar to those of X-ray CT [118] show that most of the singularities (i.e., wave front set directions) of $f(x)$ will be "invisible" ("not audible") [3, 69, 77, 80, 107, 111, 112, 124, 145, 146]. This implies, in particular, the high instability of the reconstruction (all "invisible" interfaces will be blurred, no matter how sophisticated the algorithms are) [69, 107, 124, 145, 146]. This is also well known in the more standard X-ray CT and SPECT [82, 118]. The wonderful feature of the microlocal analysis is that it not only explains, but also allows one to predict these blurring effects [ibid].

On the other hand, if one is in the situation where there is a uniqueness result and all singularities of the object are "visible," one indeed can reconstruct the object stably [13, 14, 87]. Conditions of unique reconstruction and "visibility" have been also figured out for the case of variable sound speeds and can be expressed simply in terms of geometric optics rays [3, 69, 77, 80, 107, 124, 145, 146].

While the questions above are common for all tomographic techniques, there is one issue that is specific for TAT (a similar, still not completely resolved, problem arises also in SPECT, see [76] and references therein, as well as the recent papers [20, 34]). This is the following:

Speed recovery problem: Problem (1) involves the unknown function $f(x)$, the tomogram, as well as the sound speed $c(x)$, which in all reconstruction methods is assumed to be known. It has been observed [69, 70] that errors in the values $c(x)$ introduce significant artifacts into reconstruction. In other words, one needs to know the speed $c(x)$ well, which is normally not the case. One of the options suggested to alleviate this difficulty is to run an ultrasound transmission scan beforehand, which would provide the speed map [70]. However, there is numerical evidence [147, 149] that it might be possible to determine both the speed $c(x)$ and then the tomogram $f(x)$ from the TAT data. Proving the uniqueness of reconstruction of $c(x)$ happens to be a difficult problem. Some very limited partial results have been obtained recently (and not published, except [67, 69]) by various mathematicians: M. Agranovsky, D. Finch, K. Hickmann, Y. Hristova, P. Kuchment, L. Nguyen, P. Stefanov, and G. Uhlmann. However, the problem (which, according to D. Finch's observation, is closely related to also not completely resolved well-known transmission eigenvalue problem [36, 43–45, 74, 110, 128]) is essentially open. Somewhat more understanding has been achieved concerning possible instability of the speed reconstruction [107, 124, 125].

3.3 Reconstruction Methods in TAT/PAT

What techniques are available for actual TAT reconstructions? There are several groups of approaches that have been used successfully:

Closed-form inversion formulas: Closed-form inversion formulas in tomography bring about better theoretical understanding and lead to efficient reconstruction algorithms. The best known example is the so-called filtered backprojection (FBP) algorithm in X-ray tomography, which is derived from one of the popular inversion formulas for the classical Radon transform (see, for example, [63–65, 104, 105]).

The first such formulas for TAT (the very existence of which had not been clear) were obtained in odd dimensions in [55], for the observation surface S being a sphere and the sound speed being constant. Let the function $f(x)$ be supported within a ball of radius R, and the detectors be located on the boundary $S = \partial B$ of this ball. Some of the formulas obtained in [55] are:

$$f(x) = -\frac{1}{8\pi^2 R} \Delta_x \int_{\partial B} \frac{g(y, |y-x|)}{|y-x|} dA(y), \tag{4}$$

$$f(x) = -\frac{1}{8\pi^2 R} \int_{\partial B} \left(\frac{1}{r} \frac{\partial^2}{\partial r^2} g(y, r) \right) \bigg|_{r=|y-x|} dA(y), \tag{5}$$

$$f(x) = -\frac{1}{8\pi^2 R} \int_{\partial B} \left(\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \frac{g(y, r)}{r} \right) \right) \bigg|_{r=|y-x|} dA(y), \tag{6}$$

where $dA(y)$ is the surface measure on ∂B, and g represents the values of the spherical integrals (3).

Here differentiation with respect to r in (5) and (6) and the Laplace operator in (4) represent the filtration step, while the (weighted) integrals correspond to the backprojection: integration over the set of spheres passing through the point x and centered on S.

A different family of inversion formulas valid in any dimension was found in [85]. Still another set of closed-form inversion formulas applicable in even dimensions was found in [54]. Finally, a unified family of inversion formulas was derived in [106].

Having closed-form inversion formulas has the advantage that they usually lead to fast and precise inversion algorithms. However, there are several disadvantages of these formulas in TAT. First of all, the FBP formulas are now available only for the observation surface being a sphere (see discussion above), with the only exception of [88], where such formulas are derived for a cube and some other crystallographic domains. Secondly, it is known (e.g., [77]) that if some part of the source function $f(x)$ is supported outside the observation surface S, then its reconstruction inside S using FBP formulas might be incorrect. Finally, there are no FBP formulas known for the case of a variable sound speed.

Eigenfunction expansions: This approach, which theoretically works for arbitrary closed surfaces, was proposed in [86] (and extended in [2] to the case of variable sound speeds). It is based on expansion into eigenfunctions of the Laplacian operator in the interior B of S with zero Dirichlet conditions on S. It is thus nicely implementable whenever the spectrum and eigenfunctions of the Dirichlet Laplacian are known explicitly, e.g., for spheres, half-spheres, cylinders, cubes, and parallelepipeds, as well as the surfaces of some crystallographic domains.

The function $f(x)$ can be reconstructed inside B from the data g in (1), as the following $L^2(B)$-convergent series:

$$f(x) = \sum_k f_k \psi_k(x), \tag{7}$$

where $\psi_k(x)$ are properly normalized eigenfunctions of the operator $(-c^2(x)\Delta)$ in B with zero Dirichlet conditions, and λ_k^2 are the corresponding eigenvalues.

The Fourier coefficients f_k can be recovered from the data g in (3) using one of the following formulas:

$$\begin{aligned} f_k &= \lambda_k^{-2} g_k(0) - \lambda_k^{-3} \int_0^\infty \sin(\lambda_k t) g_k''(t)\, dt, \\ f_k &= \lambda_k^{-2} g_k(0) + \lambda_k^{-2} \int_0^\infty \cos(\lambda_k t) g_k'(t)\, dt, \quad \text{or} \\ f_k &= -\lambda_k^{-1} \int_0^\infty \sin(\lambda_k t) g_k(t)\, dt \\ &= -\lambda_k^{-1} \int_0^\infty \int_S \sin(\lambda_k t) g(x,t) \overline{\frac{\partial \psi_k}{\partial n}(x)}\, dx\, dt, \end{aligned} \tag{8}$$

where

$$g_k(t) = \int_S g(x,t) \overline{\frac{\partial \psi_k}{\partial n}(x)}\, dx.$$

This method becomes computationally efficient when the eigenvalues and eigenfunctions are known explicitly and a fast summation formula for the series (7) is available, for instance, when the acquisition surface S is a surface of a cube, and thus the eigenfunctions are products of sine functions. The resulting $3D$ reconstruction algorithm is extremely fast and precise (see [86]).
This method applies in any dimension and is stable. It also does not have the deficiencies of the FBP formulas that we have mentioned above. Namely, presence of a part of the function $f(x)$ outside $f(x)$ does not hurt the reconstruction inside. The method, at least theoretically, works for arbitrary closed observation surface s and variable speed $c(x)$. However, its practicality in these circumstances is still questionable.

Time reversal: We now describe an inversion technique that has the same advantages as the eigenfunction expansion method above and in addition is easy to implement for any shape of the observation surface and acoustically inhomogeneous media. One can come up easily with this method if one notices the underlying assumption of TAT, which is often hidden, and which we have discussed explicitly before: local energy decay of the solution of (1). Then one can naturally come up with the idea of running the wave equation in (1) back in time, starting at the infinite time with zero initial condition (which reflects the local energy decay) and using the measured data g as the boundary value. Eventually, at time $t = 0$, one arrives to the tomogram $f(x)$. Certainly, the solution never vanishes exactly at any finite time (unless the sound speed is constant and the dimension is odd, where the Huygens' principle kicks in). Thus, one has to start from some sufficiently large time $t = T$ with zero conditions and go back in time to arrive at an approximation of $f(x)$, which one expects to get better as $T \to \infty$. This approach has been implemented by various researchers, its feasibility was shown, and error estimates were provided (see, e.g., [62, 68, 69, 144]). This works [62, 69] even in $2D$ (where decay is the slowest) and in inhomogeneous media. However, when trapping occurs, some parts become "invisible" and blur away [69].

A more sophisticated (than just zero) cut-off at time $t = T$ is used in the version of time reversal suggested in [117, 124, 125]. This leads to an equation with a contraction operator, which allows the use of the Neumann series (fixed-point iterations) to obtain high-quality images.

It is the author's belief that the time reversal is the most versatile and easy to implement method of TAT reconstructions.

Algebraic reconstruction techniques: ART is the well-used workhorse in approaching inverse problems (especially, when analysis of the problem is too complicated). To put it simply, in ART one discretizes the problem and uses one's favorite method (usually an iterative one) for solving the resulting linear algebraic system. Such techniques have been used in CT for quite a while [104]. ART algorithms frequently produce very good images. However, they are notoriously slow. In TAT, they have been used successfully for reconstructions with partial data [13, 113] and sound speed recovery [147, 149].

Parametrix approaches: Some of the earlier noniterative reconstruction techniques of approximate nature [75, 115, 116] were based (explicitly or implicitly) upon microlocal analysis. For example, in [75], by approximating the integration spheres by their tangent planes at the point of reconstruction and applying one of the inversion formulas for the classical Radon transform, one reconstructs a decent approximation to the image.

Such techniques are related to the general scheme proposed in [29] for the inversion of the "generalized" Radon transform that integrates over curved manifolds. One constructs a *parametrix* (usually an integral Fourier operator (FIO)) for the forward operator $F : f \to g$, i.e. such operator P that the operators $PF - I$ and $FP - I$, while not equal to zero, as in the case of a true inversion, are "smoothing." Thus, applying a parametrix P to the data g, one recovers the image f up to addition of a smooth function. This also often reduces the problem to a Fredholm integral equation of the second kind, which is well amenable to numerical solution. In other words, the parametrix method provides an efficient pre-conditioner for an iterative solver; the convergence of such iterations can be much faster than that of algebraic iterative methods. On the other hand, parametrix reconstructions can be often accepted as approximate images.

3.4 Examples of TAT/PAT Reconstructions

Now, the mathematics looks nice, but does the TAT/PAT procedure really work? The reader can find below (Fig. 2) the wonderful PAT image (courtesy of Wikipedia) of a blood vessel structure inside a human hand.

The next Fig. 3 (from [145]) shows several TAT reconstructions of a muscle+fat phantom shown on top left.

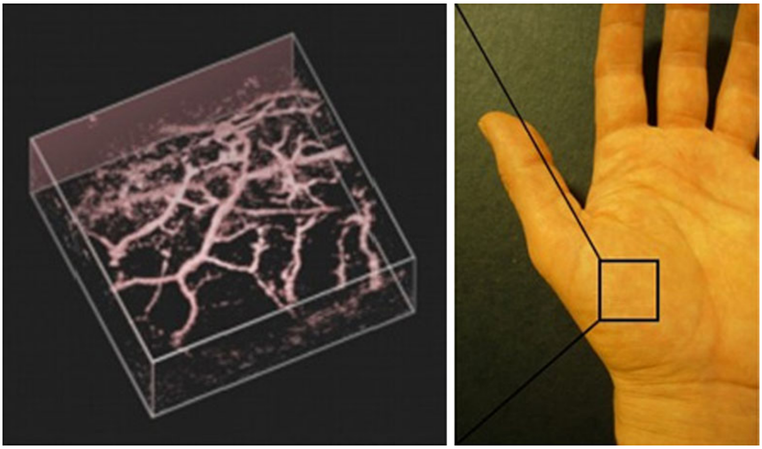


Fig. 2 A PAT reconstruction of the blood vessel system of a human hand. The picture is courtesy of Wikipedia

3.5 Quantitative PAT

The usual TAT/PAT procedure essentially recovers the initial pressure $f(x)$, which is proportional to the energy deposition function $H(x)$. How does $H(x)$ relate to the actual electric or optical parameters of the medium? It is less of a problem for TAT, where radio frequencies are used, and one can hope to obtain more or less homogeneous irradiation of all tissues. The situation is different in PAT. In the diffusion regime, one has

$$H(x) = \Gamma(x)\sigma(x)u(x),$$

where Γ is the so-called Grüneisen coefficient, $\sigma(x)$ is the EM energy absorption coefficient, and $u(x)$ is the radiation intensity. The following equation is satisfied by $u(x)$:

$$-\nabla \cdot D(x)\nabla u(x) + \sigma(x)u(x) = 0,$$

where D is the diffusion coefficient. The question arises whether, after doing the TAT/PAT reconstruction and recovering $H(x)$, one can go further and recover the actual optical parameters (D, σ, Γ) from $H(x)$? This is the goal of the so-called Quantitative PAT (QPAT), which has started developing only recently [21, 22, 24, 26, 47–49, 120, 147] and is in a very active stage now. In other words, QPAT takes of where PAT lands.

4 Acousto-Electric Tomography

Electrical impedance tomography (EIT) strives to recover the interior distribution of electric conductivity by measurements on the boundary [15, 27, 31, 37, 41, 42,

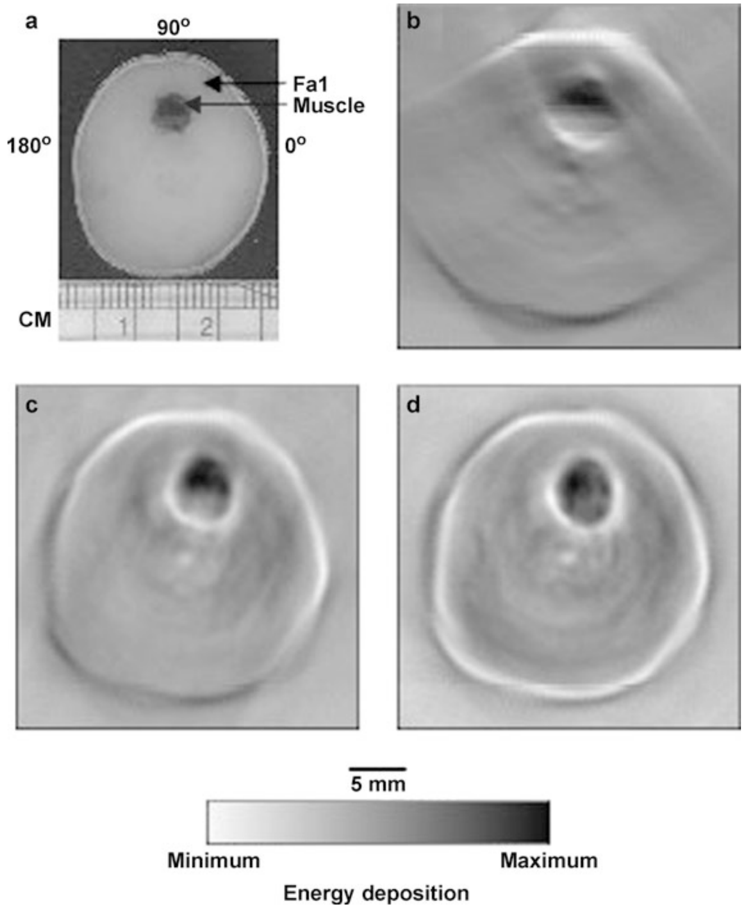


Fig. 3 **a**. The phantom. **b**. Reconstruction from partial data with "invisible" details blurred. **c**. Reconstruction from partial data with all features visible. **d**. Reconstruction from full data [145]

73, 95, 98, 129, 132, 133, 135]. Namely, one creates various boundary voltages and measures the resulting boundary currents (or vice versa) see Fig. 4. From these measurements one tries to recover the internal conductivity. The mathematical incarnation of EIT is the *inverse conductivity problem*, which was apparently suggested first by E. Calderón and has by now a glorious 30 years history. Efforts of many leading mathematicians were directed towards proving that the measured data is sufficient for the unique recovery of internal conductivity. This happens to be much more mathematically difficult topic than those arising in traditional tomography or in TAT. To large extend, this is due to the significant nonlinearity of the corresponding mapping and instability of its inversion. This is still an active area, for instance, since the optimal result in $3D$ is still not known. However, the general understanding is that one does have sufficient information for the reconstruction of the conductivity. We provide just a sample of references to this bursting with life

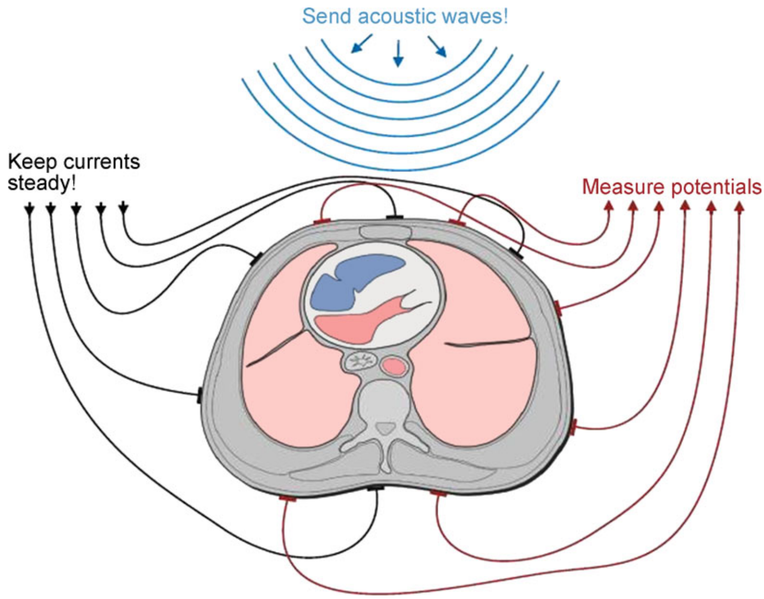


Fig. 4 The AET procedure: electrical boundary measurements are done concurrently with scanning the object with ultrasound. The picture is courtesy of L. Kunyansky

topic [15, 31, 37, 41, 73, 129, 132, 133, 135], where the reader can find plenty of information and further references.

In this text we will be interested in the issues related to the actual reconstruction in EIT, which has also attracted enormous attention of scientists, including mathematicians. Due to the already mentioned instability,[4] the pictures come out very much blurred and of low resolution. This is a rigorously proven fact of life, not the deficiency of the mathematics used. It is the author's opinion that by now experts have achieved as good EIT reconstructions as humanly possible … unless one changes the physical set-up of the measurements and/or uses some a priori information. Changing the experimental setup is what is suggested in Acousto-Electric Tomography (AET).

The main obstacle of EIT (and similarly of OT (optical tomography)) is that the signals measured at the boundary loose, exponentially fast with the depth, the information on where they came from. This is a simple-minded explanation of why reconstructions are blurred. So, if one could somehow get some type of "interior" information about where the signals came from, one could hope for stabler image reconstructions. AET is one of the implementations of this idea. It has been known for some time that ultrasound irradiation of soft tissues modifies the tissues' electric and optical properties (electro-acoustic effect [89, 90]). It was thus an easy step to decide to send an ultrasound beam that focuses on some internal location x and thus

[4] Here instability = ill-posedness = super-algebraic decay of singular values of the direct mapping.

modifies (by a multiplicative factor close to 1) the electric conductivity $\sigma(x)$ at this location. This would lead to a perturbation of the boundary EIT measurements, and what is crucial, the practitioner will know where the perturbation came from—from the point x. Then one could scan the focused beam throughout the whole object and get hopefully sufficient information for a stable reconstruction. This idea of AET was suggested and tried by a direct measurement in [148]. It was shown there that a detectable (albeit rather small) signal does exist. However, no reconstruction was done at that time. In a few years, the topic started developing fast [10, 11, 30, 38, 59, 78, 79, 148], sometimes with the researchers being unaware of the original work [148]. Let us describe briefly the current state of affairs in AET (although by the time of publication, the situation will definitely change).

The following observation was made experimentally and justified theoretically [89, 90, 148]: the acousto-electric effect, although detectable, is so small that one can safely linearize the problem.

Another smallness assumption was used in most works: the ability of sharp focusing at a given location (i.e., creating a delta-type ultrasound pulse). Such perfect focusing is clearly impossible (see the discussion in the book [66] devoted to this issue). Still, let us assume for the time being that sufficiently good focusing is possible (and return to this discussion in Sect. 5). This allows one to use the well-studied "small volume inclusion" asymptotics, as in [10, 11, 38, 39], where such asymptotics play the crucial role. On the other hand, in [78, 79], only smallness of the acousto-electric effect is needed, and no perfect focusing is required (see Sect. 5).

In all these works the authors studied what kind of interior quantities can be stably recovered from the measurements if perfect focusing (in particular, small volume asymptotics) is possible. For instance, if $u_1(x), u_2(x)$ are the (unknown) potentials created by some boundary current setup, then one can recover the values

$$\sigma(x)\nabla u_1(x) \cdot \nabla u_2(x)$$

for any interior point x, where $u \cdot v$ denotes the inner product of two vectors. Some other local functionals of the form $F(\sigma(x), u_i(x), \nabla u_i(x))$ could be recovered (this is also the case in the previously mentioned MREIT and CDI [99–101, 122, 142], which we do not discuss in this paper).

It was shown then that such values, if recovered from measurements, lead to locally unique and stable reconstruction of the conductivity σ [30, 38, 78, 79, 81]. Essentially, one can prove that the Fréchet derivative of the mapping

$$\sigma \mapsto \text{ values of } F(\sigma, u_i, \nabla u_i)$$

(in appropriate function spaces) is an injective semi-Fredholm operator.

A variety of inversion procedures has been suggested and mostly tested on numerical phantoms: those involving numerical optimization [11, 38], those reduced to solving well posed hyperbolic problems [16], or the ones that lead to solving transport equation or Poisson-type elliptic equations [78, 79].

In most cases one could achieve wonderful quality reconstructions; e.g., see Fig. 5, where the method of [79] is used, which reduces to solving a Poisson equation for determining the conductivity.

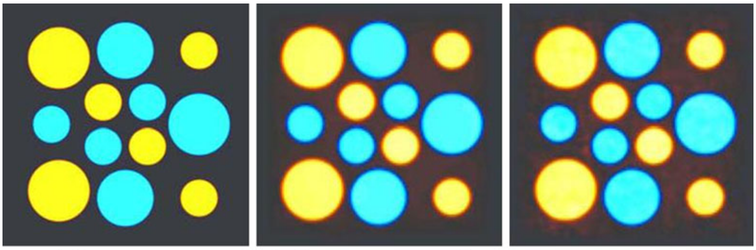


Fig. 5 An example of AET reconstruction. The picture is courtesy of L. Kunyansky

Looking at Fig. 5, one might (and should) feel cheated. Indeed, how can one get such a good reconstruction with the data contaminated by a 50% noise? The answer will be given in the next section.

By now, the mathematics of AET, albeit just a few years old, is already rather successful. The experimental implementation of the AET lags behind, due to the difficulty of acquiring good signal-to-noise ratios.

A different combination of ultrasound and EIT is suggested in [59]. Here one again creates currents through the interior of the body of interest. These currents lead to a small inhomogeneous heating of the tissues and thus to thermoelastic expansion. Then the TAT procedure, using the microphones surrounding the body, reconstructs a local functional $F(\sigma(x), u(x), \nabla u(x))$, after which one of the previously mentioned procedures of reconstruction can be applied.

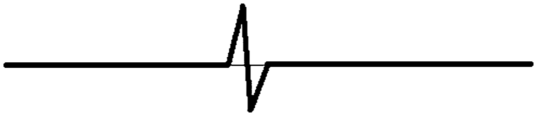
5 Synthetic Focusing in Hybrid Techniques

As it has been mentioned, the unfeasible [66] perfect focusing is assumed in most mathematical work on AET, in particular, when using the small volume asymptotics. Can this be avoided? The answer, as it was explained in [78] and then confirmed in [79], is a "yes." Indeed, the delta functions that are idealized focused beams, form a function "basis." Suppose that we can produce another, practically feasible set of ultrasound waves, which would also form such a basis. Then, using the smallness of the acousto-electric effect and thus linearity, one could mathematically process the data obtained from that basis of waves and "synthetically focus" them by changing the basis to delta-functions.

There are several examples of possible bases from [78]:

- Using large planar broad band transducers, one could generate a set of monochromatic planar waves with arbitrary wave vectors, and then the synthetic focusing would be just applying the inverse Fourier transform. This option was adopted in [16]. Its practical feasibility is not yet clear.
- Using point-like omni-directional transducers, one could generate thin spherical shell waves. Then, lo and behold, the synthetic focusing will boil down to inversion of a restricted spherical mean transform, and thus any of the standard TAT

Fig. 6 An N-shaped pulse



inversions would do it. This is the option of [79]. The problem with this is that it is much easier to create a short N-shaped (Fig. 6) spherical wave rather than δ-shaped such wave.

- One can create narrow "pencil beams," as it is done in [141], and then synthetic focusing would coincide with the inversion of the standard X-ray transform. This option has to struggle with impossibility of creating a truly homogeneous pencil ultrasound beam of sufficient length [66].

The sharp and amazingly stable with respect to the noise reconstruction, shown in Fig. 5, was done using the second option for the basis: thin spherical waves with consecutive TAT inversion for the synthetic focusing. After that an elliptic (Poisson-type) equation was solved to recover the conductivity. Now, what about having N-shaped rather than δ-shaped pulses? This "difficulty" turns out to be a blessing. Indeed, the TAT reconstruction includes a filtering portion, which increases the noise and decreases stability. However, with the N-shaped pulse, this filtration step is not needed, since it is already performed by the transducer. As the result, the synthetic focusing by the TAT inversion happens to be a *smoothing operator* and thus kills a lot of noise. If we could indeed produce and use δ-shaped pulses, the reconstruction would work, but would be very unlikely to survive a 50% noise.

6 Ultrasound-Modulated Optical Tomography (UMOT)

The idea of scanning an object with a focused ultrasound that we applied in AET can be tried with the optical tomography (OT) as well. The goal is the same: to improve drastically the resolution of OT (which is dismal on a centimeter depth and deeper). Since OT, like EIT, is a cheap, safe, and high-contrast modality, achieving this goal, and thus adding high resolution, would make it an invaluable diagnostic tool.

In comparison with the AET, the situation with UMOT is reversed: there is an extensive body of experimental research (see [141] and references therein), but the first glimpses on the mathematics of UMOT are just appearing [8, 25, 102, 103], and even the mathematical model is not settled down.

The set-up of OT is as follows: One sends a beam of (coherent or incoherent) laser light through the body of interest and observes the intensity and speckle patterns of the outgoing light. The features of interest are the internal distribution of the absorption and scattering coefficients. The contrast in optical properties of cancerous and healthy locations is often huge. However, diffused photons, when they reach a detector after multiple scattering, essentially loose any information about the locations they went through. This leads to dismal resolution at centimeters' depth (although good pictures can be obtained at skin depth). The idea of UMOT is to, concurrently with OT measurements, scan the body with a focused ultrasound and thus to acquire some interior (i.e., location-dependent) information, which hopefully would stabilize the problem. It has been argued in physics and biomedical engineering literature (e.g., [72, 91, 141]) that when using coherent light and measuring the ultrasound frequency Fourier component of the outgoing speckle pattern, one can recover the values of the following functional at an interior location x:

$$G(x,d)A^2(x)I(x).$$

Here A is the applied ultrasound power (assumed to be known), I is the light intensity, d is the detector position on the boundary, and $G(x,d)$ is the "probability of a photon emitted at the location x to reach the detector at the location d." In other words, $G(x,d)$ is a Green's function of the diffusion equation

$$-\nabla D(x)\nabla I(x) + \mu_a(x)I(x) = 0$$

inside the domain of the interest. The difficulty (at least for the author) is determining what the "probability of reaching the detector" means (e.g., does this mean the first time of reaching the point d?). Thus, it is not clear what boundary conditions the Green's function should correspond to.

It was assumed in [8] that the correct boundary conditions are those that correspond to the optical impedance at the boundary of the object. Under this condition and with the perfect focusing assumption, a reconstruction algorithm was applied that showed sufficiently sharp internal reconstructions of the absorption coefficient μ_a (although the quality was lower than in the AET case). It was also shown in [8] (see also the acknowledgments there) that (formally computed) Fréchet derivative of the forward mapping is a semi-Fredholm operator in natural function spaces. However, injectivity of this derivative was not shown. Thus there are so far no local injectivity results.

Some controversy surrounds the usage of coherent light. It is claimed in engineering literature [141] that the signal from ultrasound modulation in the case of incoherent light could not be detected so far. However, there already are some mathematical studies of UMOT using incoherent light [25].

Synthetic focusing in UMOT is possible. However, while the spherical waves should still do the job, the use of planar waves and the consequent inversion of the Fourier transform seems to be not an option here, due to the presence of the *square* of the acoustic power $A(x)$ in the measured functional.

7 Why the Improvement? Inverse Problems with Interior Information

The examples of AET and UMOT show how acquiring interior (i.e., attached to the internal locations) information stabilizes the utterly unstable inverse problems.

This is an example of a folklore meta-statement: "appropriate" internal information stabilizes the severely unstable problems like diffused OT or EIT.

This issue was studied in detail in [16, 18, 23], where several internal information functionals arising in applications (including those described above) were studied. Different functionals required sometimes different techniques. There is a feeling though that there might be an answer to a general question: What kind of a function $F(D(x), \sigma(x), u(x), \nabla u(x))$, if known, stabilizes the inverse boundary problems for

$$-\nabla \cdot D(x)\nabla u + \sigma u = 0?$$

A rather general answer was just given in [81]. Under some reasonable conditions on the functional, which cover all cases we have considered here, it is shown that the Fréchet derivative of the forward mapping is a semi-Fredholm operator. This explains why the observed improvement in stability occurs. Together with local uniqueness (which probably need to be proven individually in each case), this also gives local uniqueness and stability.

Acknowledgements This work was partially supported by the NSF DMS Grant 0604778, as well as by MSRI and IAMCS. The author expresses his gratitude to these institutions. Thanks also go to many colleagues with whom the author discussed the hybrid imaging methods (the author's attempt to write specific names produced a very long, while still incomplete, list).

References

1. Agranovsky, M., Finch, D., Kuchment, P.: Range conditions for a spherical mean transform. Inverse Probl. Imaging **3**, 373–382 (2009)
2. Agranovsky, M., Kuchment, P.: Uniqueness of reconstruction and an inversion procedure for thermoacoustic and photoacoustic tomography with variable sound speed. Inverse Probl. **23**, 2089–2102 (2007)
3. Agranovsky, M., Kuchment, P., Kunyansky, L.: On reconstruction formulas and algorithms for the TAT and PAT tomography. In: Wang, L.V. (ed.) Photoacoustic Imaging and Spectroscopy, pp. 89–101. CRC Press, Boca Raton (2009)
4. Agranovsky, M., Kuchment, P., Quinto, E.T.: Range descriptions for the spherical mean Radon transform. J. Funct. Anal. **248**, 344–386 (2007)
5. Agranovsky, M., Nguyen, L.: Range conditions for a spherical mean transform and global extension of solutions of Darboux equation. J. Anal. Math. **112**(1), 351–367 (2011). doi:10.1007/s11854-010-0033-0
6. Agranovsky, M., Quinto, E.T.: Injectivity sets for the Radon transform over circles and complete systems of radial functions. J. Funct. Anal. **139**, 383–414 (1996)
7. Aguilar, V., Ehrenpreis, L., Kuchment, P.: Range conditions for the exponential Radon transform. J. Anal. Math. **68**, 1–13 (1996)
8. Allmaras, M., Bangerth, W.: Reconstructions in ultrasound modulated optical tomography. Preprint arXiv:0910.2748v3
9. Ambartsoumian, G., Kuchment, P.: A range description for the planar circular Radon transform. SIAM J. Math. Anal. **38**, 681–692 (2006)
10. Ammari, H.: An Introduction to Mathematics of Emerging Biomedical Imaging. Springer, Berlin (2008)

11. Ammari, H., Bonnetier, E., Capdeboscq, Y., Tanter, M., Fink, M.: Electrical impedance tomography by elastic deformation. SIAM J. Appl. Math. **68**, 1557–1573 (2008)
12. Ammari, H., Bossy, E., Jugnon, V., Kang, H.: Mathematical modeling in photoacoustic imaging of small absorbers. SIAM Rev. **52**, 677–695 (2010)
13. Anastasio, M., Zhang, J., Pan, X., Zou, Y., Ku, G., Wang, L.V.: Half-time image reconstruction in thermoacoustic tomography. IEEE Trans. Med. Imaging **24**, 199–210 (2005)
14. Andreev, V., Popov, D., et al.: Image reconstruction in $3D$ optoacoustic tomography system with hemispherical transducer array. Proc. SPIE **4618**, 137–145 (2002)
15. Astala, K., Päivärinta, L.: Calderón's inverse conductivity problem in the plane. Ann. Math. **163**, 265–299 (2006)
16. Bal, G.: Hybrid inverse problems and internal information. In: G. Uhlmann, (ed.), Inside out: Inverse Problems and Applications. MSRI Publ., vol. 2 (to appear)
17. Bal, G.: Cauchy problem for Ultrasound Modulated EIT, preprint, April 2011
18. Bal, G., Bonnetier, E., Monard, F., Triki, F.: Inverse diffusion from knowledge of power densities, preprint, March 2011
19. Bal, G., Finch, D., Kuchment, P., Stefanov, P., Uhlmann, G. (eds.): Tomography and Inverse Transport Theory. Contemp. Math., vol. 559. AMS, Providence (2011)
20. Bal, G., Jollivet, A.: Combined source and attenuation reconstructions in SPECT. In: G. Bal, D. Finch, P. Kuchment, P. Stefanov, G. Uhlmann (eds.) Tomography and Inverse Transport Theory. Contemp. Math., vol. 559, pp. 13–28. AMS, Providence (2011)
21. Bal, G., Jollivet, A., Jugnon, V.: Inverse transport theory of photoacoustics. Inverse Probl. **26**(2), 025011 (2010)
22. Bal, G., Ren, K.: Multiple-source quantitative photoacoustic tomography (submitted)
23. Bal, G., Ren, K.: Non-uniqueness result for a hybrid inverse problem. In: G. Bal, D. Finch, P. Kuchment, P. Stefanov, G. Uhlmann (eds.) Tomography and Inverse Transport Theory. Contemp. Math., vol. 559, pp. 29–38. AMS, Providence (2011)
24. Bal, G., Ren, K., Uhlmann, G., Zhou, T.: Quantitative thermo-acoustics and related problems. Inverse Probl. **27**, 055007. doi:10.1088/0266-5611/27/5/055007
25. Bal, G., Schotland, J.C.: Inverse scattering and acousto-optic imaging. Phys. Rev. Lett. **104**, 043902 (2010)
26. Bal, G., Uhlmann, G.: Inverse diffusion theory of photoacoustics. Inverse Probl. **26**(8), 085010 (2010)
27. Barber, D.C., Brown, B.H.: Applied potential tomography. J. Phys. E., Sci. Instrum. **17**, 723–733 (1984)
28. Bell, A.G.: On the production and reproduction of sound by light. Am. J. Sci. **20**, 305–324 (1880)
29. Beylkin, G.: The inversion problem and applications of the generalized Radon transform. Commun. Pure Appl. Math. **37**, 579–599 (1984)
30. Bonnetier, E., Triki, F.: A stability result for electric impedance tomography by elastic perturbation. Presentation at the workshop "Inverse Problems: Theory and Applications", 12 November 2010. MSRI, Berkeley, CA
31. Borcea, L.: Electrical impedance tomography. Inverse Probl. **18**, R99–R136 (2002)
32. Bowen, T.: Radiation-induced thermoacoustic soft tissue imaging. Proc., IEEE Ultrason. Symp. **2**, 817–822 (1981)
33. Budinger, T.F., Gullberg, G.T., Huseman, R.H.: Emission computed tomography. In: Herman, G. (ed.) Image Reconstruction from Projections. Topics in Applied Physics, vol. 32, pp. 147–246. Springer, Berlin (1979)
34. Bukhgeim, A.L.: Inverse gravimetry approach to attenuated tomography. In: G. Bal, D. Finch, P. Kuchment, P. Stefanov, G. Uhlmann (eds.) Tomography and Inverse Transport Theory. Contemp. Math., vol. 559, pp. 49–64. AMS, Providence (2011)
35. Burq, N.: Décroissance de l'énergie locale de l'équation des ondes pour le problème extérieur et absence de résonance au voisinage du réel. Acta Math. **180**(1), 1–29 (1998)
36. Cakoni, F., Gintides, D., Haddar, H.: The existence of an infinite discrete set of transmission eigenvalues. SIAM J. Math. Anal. **42**(1), 237–255 (2010)

37. Calderón, A.: Selected Papers of Alberto P. Calderón. With commentary. Edited by A. Bellow, C.E. Kenig, P. Malliavin, American Mathematical Society, Providence (2008)
38. Capdeboscq, Y., Fehrenbach, J., de Gournay, F., Kavian, O.: Imaging by modification: numerical reconstruction of local conductivities from corresponding power density measurements. SIAM J. Imaging Sci. **2/4**, 1003–1030 (2009)
39. Capdeboscq, Y., Vogelius, M.: A general representation formula for boundary voltage perturbations caused by internal conductivity inhomogeneities of low volume fraction. Math. Model. Numer. Anal. **37**(1), 159–173 (2003)
40. Cheney, M.: A mathematical tutorial on synthetic aperture radar. SIAM Rev. **43**(2), 301–312 (2001)
41. Cheney, M., Isaacson, D., Newell, J.C.: Electrical impedance tomography. SIAM Rev. **41**(1), 85–101 (1999)
42. Cipra, B.: Shocking images from RPI. SIAM News, July 1994, 14–15
43. Colton, D.: Inverse acoustic and electromagnetic scattering theory. In: Uhlmann, G. (ed.) Inside out: Inverse Problems and Applications. MSRI Publ., pp. 67–110. Cambridge University Press, Cambridge (2003)
44. Colton, D., Kress, R.: Inverse Acoustic and Electromagnetic Scattering Theory. Applied Mathematical Sciences, vol. 93, 2nd edn. Springer, Berlin (1998)
45. Colton, D., Päivärinta, L., Sylvester, J.: The interior transmission problem. Inverse Probl. Imaging **1**, 13–28 (2007)
46. Courant, R., Hilbert, D.: Methods of Mathematical Physics. Partial Differential Equations, vol. II. Interscience, New York (1962)
47. Cox, B.T., Arridge, S.R., Beard, P.C.: Estimating chromophore distributions from multiwavelength photoacoustic images. J. Opt. Soc. Am. A **26**, 443–455 (2009)
48. Cox, B., Tarvainen, T., Arridge, S.: Multiple illumination quantitative photoacoustic tomography using transport and diffusion models. In: G. Bal, D. Finch, P. Kuchment, P. Stefanov, G. Uhlmann (eds.) Tomography and Inverse Transport Theory. Contemp. Math., vol. 559, pp. 1–12. AMS, Providence (2011)
49. Cox, B.T., Laufer, J.G., Beard, P.C.: The challenges for quantitative photoacoustic imaging. Proc. SPIE **7177**, 717713 (2009)
50. Diebold, G.J., Sun, T., Khan, M.I.: Photoacoustic monopole radiation in one, two, and three dimensions. Phys. Rev. Lett. **67**(24), 3384–3387 (1991)
51. Egorov, Y.V., Shubin, M.A.: Partial Differential Equations I. Encyclopaedia of Mathematical Sciences, vol. 30, pp. 1–259. Springer, Berlin (1992)
52. Ehrenpreis, L.: The Universality of the Radon Transform. Oxford Univ. Press, London (2003)
53. Ehrenpreis, L., Kuchment, P., Panchenko, A.: Attenuated Radon transform and F. John's equation. I. Range conditions. In: Grinberg, E.L., Berhanu, S., Knopp, M., Mendoza, G., Quinto, E.T. (eds.) Analysis, Geometry, Number Theory: The Mathematics of Leon Ehrenpreis. AMS, Providence (2000)
54. Finch, D., Haltmeier, M., Rakesh: Inversion of spherical means and the wave equation in even dimensions. SIAM J. Appl. Math. **68**(2), 392–412 (2007)
55. Finch, D., Patch, S., Rakesh: Determining a function from its mean values over a family of spheres. SIAM J. Math. Anal. **35**(5), 1213–1240 (2004)
56. Finch, D., Rakesh: The spherical mean value operator with centers on a sphere. Inverse Probl. **23**, S37–S50 (2007)
57. Finch, D., Rakesh: Range of the spherical mean value operator for functions supported in a ball. Inverse Probl. **22**, 923–938 (2006)
58. Finch, D., Rakesh: Recovering a function from its spherical mean values in two and three dimensions. In: Wang, L.V. (ed.) Photoacoustic Imaging and Spectroscopy, pp. 77–88. CRC Press, Boca Raton (2009)
59. Gebauer, B., Scherzer, O.: Impedance-acoustic tomography. SIAM J. Appl. Math. **69**, 565–576 (2009)

60. Gelfand, I., Gindikin, S., Graev, M.: Selected Topics in Integral Geometry. Transl. Math. Monogr., vol. 220. Am. Math. Soc., Providence (2003)
61. Gelfand, I., Graev, M., Vilenkin, N.: Generalized Functions. Integral Geometry and Representation Theory, vol. 5. Academic. Press, San Diego (1965)
62. Grün, H., Haltmeier, M., Paltauf, G., Burgholzer, P.: Photoacoustic tomography using a fiber based Fabry–Perot interferometer as an integrating line detector and image reconstruction by model-based time reversal method. Proc. SPIE **6631**, 663107 (2007)
63. Helgason, S.: The Radon Transform. Birkhäuser, Basel (1980)
64. Helgason, S.: Integral Geometry and Radon Transforms. Springer, New York (2010)
65. Herman, G. (ed.): Image Reconstruction from Projections. Topics in Applied Physics, vol. 32. Springer, Berlin (1979)
66. Hernandez-Figueroa, H.E., Zamboni-Rached, M., Recami, E. (eds.): Localized Waves. IEEE Press/Wiley, Hoboken (2008)
67. Hickmann, K.: Unique determination of acoustic properties from thermoacoustic data. Ph.D. Thesis, Oregon State University, Corvallis, OR (2010)
68. Hristova, Y.: Time reversal in thermoacoustic tomography: error estimate. Inverse Probl. **25**, 1–14 (2009)
69. Hristova, Y., Kuchment, P., Nguyen, L.: On reconstruction and time reversal in thermoacoustic tomography in homogeneous and non-homogeneous acoustic media. Inverse Probl. **24**, 055006 (2008)
70. Jin, X., Wang, L.V.: Thermoacoustic tomography with correction for acoustic speed variations. Phys. Med. Biol. **51**, 6437–6448 (2006)
71. John, F.: Plane Waves and Spherical Means Applied to Partial Differential Equations. Dover, New York (1971)
72. Kempe, M., Larionov, M., Zaslavsky, D., Genack, A.Z.: Acousto-optic tomography with multiply scattered light. J. Opt. Soc. Am. A **14**, 1151–1158 (1997)
73. Kenig, C.E., Sjöstrand, J., Uhlmann, G.: The Calderón problem with partial data. Ann. Math. **165**(2), 567–591 (2007)
74. Kirsch, A.: On the existence of transmission eigenvalues. Inverse Probl. Imaging **3**, 155–172 (2009)
75. Kruger, R.A., Liu, P., Fang, Y.R., Appledorn, C.R.: Photoacoustic ultrasound (PAUS) reconstruction tomography. Med. Phys. **22**, 1605–1609 (1995)
76. Kuchment, P.: Generalized transforms of Radon type and their applications. In: Olafsson, G., Quinto, E.T. (eds.) The Radon Transform, Inverse Problems, and Tomography. American Mathematical Society Short Course, January 3–4, 2005, Atlanta, Georgia. Proc. Symp. Appl. Math., vol. 63, pp. 67–91. AMS, Providence (2006)
77. Kuchment, P., Kunyansky, L.: Mathematics of thermoacoustic tomography. Eur. J. Appl. Math. **19**, 191–224 (2008)
78. Kuchment, P., Kunyansky, L.: Synthetic focusing in ultrasound modulated tomography. Inverse Probl. Imaging **4**(4), 665–673 (2010)
79. Kuchment, P., Kunyansky, L.: 2D and 3D reconstructions in acousto-electric tomography. Inverse Probl. **27**, 055013 (2011)
80. Kuchment, P., Kunyansky, L.: Mathematics of thermoacoustic and photoacoustic tomography. In: Scherzer, O. (ed.) Handbook of Mathematical Methods in Imaging, pp. 817–866. Springer, Berlin (2010)
81. Kuchment, P., Steinhauer, D.: Stabilizing Inverse Problems by Internal Data, preprint arXiv:1110.1819
82. Kuchment, P., Lancaster, K., Mogilevskaya, L.: On local tomography. Inverse Probl. **11**, 571–589 (1995)
83. Kuchment, P., Quinto, E.T.: Some problems of integral geometry arising in tomography. In: Ehrenpreis, L. (ed.) The Universality of the Radon Transform. Oxford Univ. Press, London (2003). Chap. XI
84. Kuchment, P., Scherzer, O.: Mathematical Methods in Photoacoustic Imaging to appear in Encyclopedia of Applied and Computational Mathematics. Springer, Berlin (2012)

85. Kunyansky, L.A.: Explicit inversion formulae for the spherical mean Radon transform. Inverse Probl. **23**, 373–383 (2007)
86. Kunyansky, L.: A series solution and a fast algorithm for the inversion of the spherical mean Radon transform. Inverse Probl. **23**, S11–S20 (2007)
87. Kunyansky, L.: Thermoacoustic tomography with detectors on an open curve: an efficient reconstruction algorithm. Inverse Probl. **24**(5), 055021 (2008)
88. Kunyansky, L.: Reconstruction of a function from its spherical (circular) means with the centers lying on the surface of certain polygons and polyhedra. Inverse Probl. **27**, 025012 (2011). doi:10.1088/0266-5611/27/2/025012
89. Lavandier, B., Jossinet, J., Cathignol, D.: Quantitative assessment of ultrasound-induced resistance change in saline solution. Med. Biol. Eng. Comput. **38**, 150–155 (2000)
90. Lavandier, B., Jossinet, J., Cathignol, D.: Experimental measurement of the acousto-electric interaction signal in saline solution. Ultrasonics **38**, 929–936 (2000)
91. Leutz, W., Maret, G.: Ultrasonic modulation of multiply scattered light. Physica B **204**, 14–19 (1995)
92. Lin, V., Pinkus, A.: Fundamentality of ridge functions. J. Approx. Theory **75**, 295–311 (1993)
93. Lin, V., Pinkus, A.: Approximation of multivariate functions. In: Dikshit, H.P., Micchelli, C.A. (eds.) Advances in Computational Mathematics, pp. 1–9. World Scientific, Singapore (1994)
94. Louis, A.K., Quinto, E.T.: Local tomographic methods in Sonar. In: Surveys on Solution Methods for Inverse Problems, pp. 147–154. Springer, Vienna (2000)
95. Mathematics and Physics of Emerging Biomedical Imaging. The National Academies Press (1996). Available online at http://www.nap.edu/catalog.php?record_id=5066#toc
96. McLaughlin, J.R., Yoon, J.-R.: Unique identifiability of elastic parameters from time-dependent interior displacement measurement. Inverse Probl. **20**, 25–45 (2004)
97. Muthupillai, R., Lomas, D.J., Rossman, P.J., Greenleaf, J.F., Manduca, A., Ehman, R.L.: Magnetic resonance elastography by direct visualization of propagating acoustic strain waves. Science **269**(5232), 1854–1857 (1995). doi:10.1126/science.7569924
98. Nachman, A.: Global uniqueness for a two-dimensional inverse boundary value problem. Ann. Math. **143**(1), 71–96 (1996)
99. Nachman, A., Tamasan, A., Timonov, A.: Conductivity imaging with a single measurement of boundary and interior data. Inverse Probl. **23**(6), 2551–2563 (2007)
100. Nachman, A., Tamasan, A., Timonov, A.: Recovering the conductivity from a single measurement of interior data. Inverse Probl. **25**(3), 035014 (2009)
101. Nachman, A., Tamasan, A., Timonov, A.: Current density impedance imaging. In: G. Bal, D. Finch, P. Kuchment, P. Stefanov, G. Uhlmann (eds.) Tomography and Inverse Transport Theory. Contemp. Math., vol. 559, pp. 135–150. AMS, Providence (2011)
102. Nam, H.: Ultrasound modulated optical tomography. Ph.D. thesis, Texas A&M University (2002)
103. Nam, H., Dobson, D.: Ultrasound modulated optical tomography, preprint (2004)
104. Natterer, F.: The Mathematics of Computerized Tomography. Wiley, New York (1986). Reprinted in 2001 by the SIAM
105. Natterer, F., Wübbeling, F.: Mathematical Methods in Image Reconstruction. Monographs on Mathematical Modeling and Computation, vol. 5. Philadelphia, SIAM (2001)
106. Nguyen, L.V.: A family of inversion formulas in thermoacoustic tomography. Inverse Probl. Imaging **3**(4), 649–675 (2009)
107. Nguyen, L.V.: On singularities and instability of reconstruction in thermoacoustic tomography, preprint arXiv:0911.5521v1
108. Oraevsky, A.A., Esenaliev, R.O., Jacques, S.L., Tittel, F.K.: Laser optoacoustic tomography for medical diagnostics principles. Proc. SPIE **2676**, 22 (1996)
109. Oraevsky, A.A., Jacques, S.L., Esenaliev, R.O., Tittel, F.K.: Laser-based optoacoustic imaging in biological tissues. Proc. SPIE **2134A**, 122–128 (1994)

110. Päivärinta, L., Sylvester, J.: Transmission Eigenvalues. SIAM J. Math. Anal. **40**, 738–753 (2008)
111. Palamodov, V.P.: Reconstructive Integral Geometry. Birkhäuser, Basel (2004)
112. Palamodov, V.: Remarks on the general Funk-Radon transform and thermoacoustic tomography. Inverse Probl. Imaging **4**(4), 693–702 (2010)
113. Paltauf, G., Viator, J.A., Prahl, S.A., Jacques, S.L.: Iterative reconstruction algorithm for optoacoustic imaging. J. Acoust. Soc. Am. **112**(4), 1536–1544 (2002)
114. Patch, S.K., Scherzer, O.: Photo- and thermo-acoustic imaging. Inverse Probl. **23**, S01–S10 (2007)
115. Popov, D.A., Sushko, D.V.: A parametrix for the problem of optical-acoustic tomography. Dokl. Math. **65**(1), 19–21 (2002)
116. Popov, D.A., Sushko, D.V.: Image restoration in optical-acoustic tomography. Probl. Inf. Transm. **40**(3), 254–278 (2004)
117. Qian, J., Stefanov, P., Uhlmann, G., Zhao, H.-K.: An efficient Neumann-series based algorithm for thermoacoustic tomography with variable sound speed. SIAM J. Imaging Sci. **4**, 850–883 (2011)
118. Quinto, E.T.: Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$. SIAM J. Math. Anal. **24**, 1215–1225 (1993)
119. Ralston, J.V.: Solutions of the wave equation with localized energy. Commun. Pure Appl. Math. **22**, 807–823 (1969)
120. Ren, K., Bal, G.: On multi-spectral quantitative photoacoustic tomography. Inverse Probl. **28**, 025010 (2012)
121. Scherzer, O. (ed.): Handbook of Mathematical Methods in Imaging. Springer, Berlin (2010)
122. Seo, J.K., Woo, E.Je.: Magnetic resonance electrical impedance tomography (MREIT). SIAM Rev. **53**(1), 40–68 (2011)
123. Stefanov, P., Uhlmann, G.: Integral geometry of tensor fields on a class of non-simple Riemannian manifolds. Am. J. Math. **130**(1), 239–268 (2008)
124. Stefanov, P., Uhlmann, G.: Thermoacoustic tomography with variable sound speed. Inverse Probl. **25**(7), 075011 (2009)
125. Stefanov, P., Uhlmann, G.: Thermoacoustic tomography arising in brain imaging. Inverse Probl. **27**, 045004 (2011)
126. Steinhauer, D.: A uniqueness theorem for thermoacoustic tomography in the case of limited boundary data, preprint arXiv:0902.2838
127. Steinhauer, D.: A reconstruction procedure for thermoacoustic tomography in the case of limited boundary data, preprint arXiv:0905.2954
128. Sylvester, J.: Discreteness of transmission eigenvalues via upper triangular compact operators, preprint arXiv:1104.4336
129. Sylvester, J., Uhlmann, G.: A global uniqueness theorem for an inverse boundary value problem. Ann. Math. **125**, 153–169 (1987)
130. Tam, A.C.: Applications of photoacoustic sensing techniques. Rev. Mod. Phys. **58**(2), 381–431 (1986)
131. Tuchin, V.V. (ed.): Handbook of Optical Biomedical Diagnostics. SPIE, Bellingham (2002)
132. Uhlmann, G.: Inverse boundary value problems and applications. Astérisque **207**, 153–211 (1992)
133. Uhlmann, G. (ed.): Inside Out: Inverse Problems and Applications. MSRI Publ. Cambridge University Press, Cambridge (2003)
134. Uhlmann, G.: (ed.) Inside Out: Inverse Problems and Applications. MSRI Publ., vol. 2 (to appear)
135. Uhlmann, G.: Commentary on Calderón's paper on an inverse boundary value problem. In: Bellow, A., Kenig, C.E., Malliavin, P. (eds.) Selected Papers of Alberto P. Calderón. With commentary, pp. 623–636. Am. Math. Soc., Providence (2008)
136. Vainberg, B.: The short-wave asymptotic behavior of the solutions of stationary problems, and the asymptotic behavior as $t \to \infty$ of the solutions of nonstationary problems. Russ. Math. Surv. **30**(2), 1–58 (1975)

137. Vainberg, B.: Asymptotics Methods in the Equations of Mathematical Physics. Gordon and Breach, New York (1982)
138. Vo-Dinh, T. (ed.): Biomedical Photonics Handbook. CRC Press, Boca Raton (2003)
139. Wang, L.V. (ed.): Photoacoustic Imaging and Spectroscopy. CRC Press, Boca Raton (2009)
140. Wang, K., Anastasio, M.A.: Photoacoustic and thermoacoustic tomography: image formation principles. In: Scherzer, O. (ed.) Handbook of Mathematical Methods in Imaging. Springer, Berlin (2010). Chap. 28
141. Wang, L., Wu, H.: Biomedical Optics. Wiley, New York (2007)
142. Woo, E.J., Seo, J.K.: Magnetic resonance electrical impedance tomography (MREIT) for high-resolution conductivity imaging. Physiol. Meas. **29**, R1–R26 (2008)
143. Xu, M., Wang, L.-H.V.: Photoacoustic imaging in biomedicine. Rev. Sci. Instrum. **77**, 041101 (2006)
144. Xu, Y., Wang, L.V.: Time reversal in photoacoustic tomography or thermoacoustic tomography. In: Wang, L.V. (ed.) Photoacoustic Imaging and Spectroscopy, pp. 117–120. CRC Press, Boca Raton (2009)
145. Xu, Y., Wang, L., Ambartsoumian, G., Kuchment, P.: Reconstructions in limited view thermoacoustic tomography. Med. Phys. **31**(4), 724–733 (2004)
146. Xu, Y., Wang, L., Ambartsoumian, G., Kuchment, P.: Limited view thermoacoustic tomography. In: Wang, L.H. (ed.) Photoacoustic Imaging and Spectroscopy, pp. 61–73. CRC Press, Boca Raton (2009). Chap. 6
147. Yuan, Z., Zhang, Q., Jiang, H.: Simultaneous reconstruction of acoustic and optical properties of heterogeneous media by quantitative photoacoustic tomography. Opt. Express **14**(15), 6749 (2006)
148. Zhang, H., Wang, L.: Acousto-electric tomography. Proc. SPIE **5320**, 145–149 (2004)
149. Zhang, J., Anastasio, M.A.: Reconstruction of speed-of-sound and electromagnetic absorption distributions in photoacoustic tomography. Proc. SPIE **6086**, 608619 (2006)

On Fermat-Type Functional and Partial Differential Equations

Bao Qin Li

Abstract This paper concerns entire and meromorphic solutions to functional and nonlinear partial differential equations of the form $a_1 f^m + a_2 g^n = a_3$ with function coefficients a_k, $k = 1, 2, 3$, where f and g are unknown functions or partial derivatives of an unknown function. We will discuss some recent results and also give, among other things, some new results on these equations.

1 Fermat-Type Equations

This paper concerns entire and meromorphic solutions to functional equations of the form

$$a_1 f^p + a_2 g^q = a_3 \tag{1}$$

and nonlinear partial differential equations

$$a_1 u_{z_1}^p + a_2 u_{z_2}^q = a_3, \tag{2}$$

where the coefficients a_k are given functions, and f, g, u are unknown. Complex solutions to such equations over some commonly studied function fields have been investigated by many authors, and there is an extensive literature on these equations and generalizations as well as connections to other problems (see [1, 4–6, 15–17, 19, 21, 24, 27], etc. and various references therein). These equations clearly contain the Fermat equations $f^p + g^p = 1$ as special cases. A classic result due to Montel [19] (see [13] for a simpler proof) states that entire solutions f, g of the Fermat equation $f^p + g^p = 1$ ($p \geq 3$) are necessarily constant. The Fermat theorem of this type for polynomial or rational function solutions to these equations and relations to the so-called ABC theorem can be found in [6, 14], etc. These results are also related to the geometric fact that the surface $x^p + y^p = 1$ is a Kobayashi

Dedicated to the memory of the late Professor Leon Ehrenpreis.

B.Q. Li (✉)
Department of Mathematics, Florida International University, Miami, FL 33157, USA
e-mail: libaoqin@fiu.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_13, © Springer-Verlag Italia 2012

hyperbolic manifold when $p \geq 3$, which implies the nonexistence of nonconstant entire holomorphic curves to the manifold (see, e.g., [21]). This also has a striking similarity with the Fermat last theorem in number theory. While the Fermat equation $x^p + y^p = 1$ when $p \geq 3$ does not admit nontrivial solutions in rational numbers by Fermat's last theorem [25, 26], the equation when $p = 2$, i.e., $x^2 + y^2 = 1$, does admit nontrivial rational solutions. While the equation $f^p + g^p = 1$ when $p \geq 3$ does not admit any nontrivial entire solutions, the equation when $p = 2$, i.e., $f^2 + g^2 = 1$, does admit nonconstant function solutions over some commonly studied function fields. Taking $f = u_{z_1}$ and $g = u_{z_2}$ in the functional equations $f^p + g^p = 1$, we then have the Fermat partial differential equations ($p \geq 3$) and the eikonal (eiconal) equation ($p = 2$). These partial differential equations in real variables, especially the eikonal equation and its generalizations arise in wave propagation, geometric optics, quantum mechanics, general relativity, etc. The nonexistence of nontrivial entire solutions to the Fermat equations $f^p + g^p = 1$ when $p \geq 3$ immediately implies that the Fermat partial differential equations have no nonlinear entire solutions. In spite of the failure of the above result for entire solutions of the functional equation when $p = 2$, the eikonal partial differential equation still has no nonlinear entire solutions in $\mathbf{C}^2$, and this result can be obtained from many results of different nature (see [8, 11, 15–17, 20], etc.). For instance, we obtained in [17] the following:

Theorem 1 *Meromorphic solutions f, g of $f^2 + g^2 = 1$ in $\mathbf{C}^2$ are constant if and only if f_{z_2} and g_{z_1} have the same zeros (counting multiplicities).*

If we let $f = u_{z_1}$ and $g = u_{z_2}$, then $f_{z_2} = g_{z_1} (= u_{z_1 z_2})$, and thus the condition in Theorem 1 is trivially satisfied, which immediately implies that entire or meromorphic solutions of the eikonal equation $u_{z_1}^2 + u_{z_2}^2 = 1$ must be linear.

We give another proof of this result using a theorem in our paper [16]:

Theorem 2 *Let u be a meromorphic function in $\mathbf{C}^2$. Then its partial derivatives u_{z_1} and u_{z_2} have a common right factor (in the sense of composition) if and only if u is a linear function, or $u = c_1 z_1 + f(z_2 + c_2 z_1)$, where f is a meromorphic function in the complex plane such that f' is nonlinear, and c_1 and $c_2 \neq 0$ are two constants.*

Now suppose that u is a meromorphic solution of $u_{z_1}^2 + u_{z_2}^2 = 1$. Note that $f^2 + g^2 = 1$ implies that $(f + ig)(f - ig) = 1$ and thus $f + ig = q$ and $f - ig = \frac{1}{q}$ for some meromorphic function q. That is,

$$\begin{aligned} f &= \frac{1}{2}\left(q + q^{-1}\right) = \frac{1}{2}\left(z + z^{-1}\right) \circ q, \\ g &= \frac{1}{2i}\left(q - q^{-1}\right) = \frac{1}{2i}\left(z - z^{-1}\right) \circ q. \end{aligned} \tag{3}$$

Thus, the partial derivatives u_{z_1} and u_{z_2} have a common right factor. By Theorem 2, if u is nonlinear, then $u = c_1 z_1 + h(z_2 + c_2 z_1)$, where h' is nonlinear, and c_1 and

$c_2 \neq 0$ are two constants. Thus, from the given equation it follows that

$$\begin{aligned} 0 &= u_{z_1}^2 + u_{z_2}^2 - 1 \\ &= \left\{c_1 + c_2 h'(z_2 + c_2 z_1)\right\}^2 + \left\{h'(z_2 + c_2 z_1)\right\}^2 - 1 \\ &= \left\{(c_1 + c_2 w)^2 + w^2 - 1\right\} \circ h'(z_2 + c_2 z_1). \end{aligned}$$

However, this composite function is clearly a nonzero function, a contradiction. Thus, u must be linear. This result for entire solutions of the eikonal equation was first proved in [11], and the same result for meromorphic solutions was given in [20] (the proof in [20], however, appears to contain an error in using the method of characteristics for partial differential equations).

These studies naturally extend to more general settings such as $f^p + g^q = 1$ for different positive integers p, q and $a_1 f^p + a_2 g^q = a_3$ over function fields. For instance, the nonexistence of nonconstant entire solutions is still true for $f^p + g^q = 1$ when $\frac{1}{p} + \frac{1}{q} < 1$ (due to a theorem of Cartan, see, e.g., [4]). It was shown in [5] that meromorphic solutions to the equation $f^p + g^p = 1$ reduce to constant when $p \geq 4$, and it was shown in [27] that the relation $a_1 f^p + a_2 g^q = 1$ with coefficients a_j, which grow more slowly than f, g, cannot hold for meromorphic solutions f, g when $p, q \geq 3$ unless $p = q = 3$, which was then improved for entire solutions [24] when $\frac{1}{p} + \frac{1}{q} < 1$. All these results will be contained and unified in a result (Lemma 1) presented below with a short proof, which will then be applied to prove a result (Theorem 4) on entire and meromorphic solutions of (1) and (2). In doing this, we also give a theorem (Theorem 3) on relationships between Nevanlinna characteristics of meromorphic functions and those of their partial derivatives, which is of independent interest.

2 Some Preliminaries

We will need some notation and facts from Nevanlinna theory (see, e.g., [22] and [23]) such as the counting function $N(r, 1/f)$ (resp. $\bar{N}(r, 1/f)$) for the zeros of a meromorphic function f in $\mathbf{C}^n$ counting multiplicities (resp. without counting multiplicities), the proximity function $m(r, f)$, the Nevanlinna characteristic function $T(r, f) := m(r, f) + N(r, f)$, the Nevanlinna first fundamental theorem:

$$T\left(r, \frac{1}{f-a}\right) = T(r, f) + O(1)$$

for any complex number a, the Nevanlinna second fundamental theorem:

$$T(r, f) \leq \sum_{j=1}^{3} \bar{N}\left(r, \frac{1}{f - a_j}\right) + S(r, f)$$

for any three distinct values $a_j \in \mathbf{P}$, and the logarithmic derivative lemma:

$$m\left(r, \frac{f_{z_j}}{f}\right) = S(r, f),$$

where $S(r, f)$ denotes a quantity satisfying that $S(r, f) = o\{T(r, f)\}$ as $r \to \infty$ outside a possible exceptional set of r of finite Lebesgue measure, if f is nonconstant. Let us also recall that the order $\rho(f)$ of a meromorphic function f is defined as

$$\rho(f) = \limsup_{r \to \infty} \frac{\log^+ T(r, f)}{\log r}.$$

A meromorphic function f in $\mathbf{C}^n$ is of finite λ-type if $T(r, f) \leq A\lambda(Br)$ for some constants $A, B > 0$ and large r, where $\lambda : \mathbf{R}^+ \to \mathbf{R}^+$ is a nondecreasing growth function. Finite λ-type generalizes the traditional concept of finite type for which $\lambda = r^\rho (\rho > 0)$. If $\lambda = \log r$, then an entire (meromorphic) function f is of finite λ-type if and only if f is a polynomial (rational function). If λ is a constant, then an entire (meromorphic) function f is of finite λ-type if and only if f is a constant. We refer to [12] for an extensive study of finite λ-type functions.

3 A Comparison on Characteristics of a Meromorphic Function and Its Partial Derivatives

We give in this section a comparison result on Nevanlinna characteristics of a meromorphic function and its partial derivatives, which will be needed in proving Theorem 4 on entire and meromorphic solutions of (1) and (2). Theorem 3 and Corollary 1 below are more than what we need for the proof of Theorem 4. We include them as stated due to their own independent interests.

In one variable, it is well known that $T(r, f)$ can be bounded by $T(r, f')$ ([2], [28], p. 93): If f is meromorphic in $\mathbf{C}$ with $f(0) \neq \infty$, then for $r > 0$ and $\tau > 1$,

$$T(r, f) \leq C\left\{T\left(\tau r, f'\right) + \log^+ r + \log^+ \left|f(0)\right| + 1\right\}, \tag{4}$$

where $C > 0$ is a constant depending only on τ. It is easy to see that this result is false for meromorphic functions f in $\mathbf{C}^n$ when $n > 1$ if f' is replaced by a partial derivative of f. For example, consider the function $f = e^{e^{z_1}} + e^{z_2}$ in $\mathbf{C}^2$. It is clear that its characteristic $T(r, f)$ cannot be bounded by the characteristic $T(r, f_{z_2})$. In fact, f has infinite order, while its partial derivative $f_{z_2} = e^{z_2}$ has order 1.

It turns out that we have the following comparison theorem in $\mathbf{C}^n$, which coincides with (4) when $n = 1$:

Theorem 3 *Let* f *be a meromorphic function in* $\mathbf{C}^n$ *and analytic at* 0. *Then for* $r > 0$ *and* $\tau > 1$, *we have that*

$$T(r, f) \leq C\left\{\sum_{j=1}^{n} T(\tau r, f_{z_j}) + \log^+ r + \log^+ \left|f(0)\right| + 1 + \log n\right\},$$

where $C > 0$ *is a constant depending only on* τ.

Proof Let $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_n)$ be a unit vector in $\mathbf{C}^n$. We denote $f|_\zeta(z) = f(\zeta z) :$ $\mathbf{C} \to \mathbf{C}$, the lifting of f to the plane via the map $z \to \zeta z$. Note that

$$\frac{d}{dz} f\bigg|_\zeta (z) = \zeta_1 f_{z_1}|_\zeta(z) + \cdots + \zeta_n f_{z_n}|_\zeta(z).$$

Thus, using the arithmetic properties of the characteristic function, we deduce that

$$T\left(r, \frac{d}{dz} f\bigg|_\zeta\right) \leq \sum_{j=1}^{n} T(r, f_{z_j}|_\zeta) + \log n.$$

Applying (4) to the one-variable function $f|_\zeta$, there is a constant C depending only on τ (it is important that the constant C in (4) depends only on τ so that it is independent of the vector ζ) such that

$$\begin{aligned} T(r, f|_\zeta) &\leq C\left\{T\left(\tau r, \frac{d}{dz} f\bigg|_\zeta\right) + \log^+ r + \log^+ \left|f(0)\right| + 1\right\} \\ &\leq C\left\{\sum_{j=1}^{n} T(\tau r, f_{z_j}|_\zeta) + \log^+ r + \log^+ \left|f(0)\right| + 1 + \log n\right\}. \end{aligned}$$

We next use the following result (see [12], p. 335): If f is meromorphic in $\mathbf{C}^n$ and analytic at 0 with $f(0) \neq 0$, then $T(r, f) = \int_S T(r, f|_\zeta)\, d\eta_\zeta$ for $r \geq 0$, where S is the unit sphere, and the integral is normalized so that $\int_S d\eta_\zeta = 1$. Assume for the moment that

$$f(0) f_{z_1}(0) f_{z_2}(0) \cdots f_{z_n}(0) \neq 0.$$

Then we can apply this result to $f, f_{z_1}, f_{z_2}, \ldots, f_{z_n}$ to deduce that

$$\begin{aligned} T(r, f) &= \int_S T(r, f|_\zeta)\, d\eta_\zeta \\ &\leq C\left\{\sum_{j=1}^{n} \int_S T(\tau r, f_{z_j}|\zeta)\, d\eta_\zeta + \log^+ r + \log^+ \left|f(0)\right| + 1 + \log n\right\} \\ &= C\left\{\sum_{j=1}^{n} T(\tau r, f_{z_j}) + \log^+ r + \log^+ \left|f(0)\right| + 1 + \log n\right\}. \qquad (5) \end{aligned}$$

Thus, the theorem holds in this case. Now, if

$$f(0)f_{z_1}(0)f_{z_2}(0)\cdots f_{z_n}(0)=0,$$

then we can choose constants a_j with $|a_j|<1$ so that

$$F:=f+a_1z_1+a_2z_2+\cdots+a_nz_n+a_0,$$

which is f plus a linear function, satisfies the condition

$$F(0)F_{z_1}(0)F_{z_2}(0)\cdots F_{z_n}(0)\neq 0.$$

Then we can apply the already proved result (5) for F to obtain that

$$T(r,F)\le C\left\{\sum_{j=1}^{n}T(\tau r,F_{z_j})+\log^+ r+\log^+\left|F(0)\right|+1+\log n\right\}. \tag{6}$$

On the other hand, in view of the fact that $|a_j|<1$ and by the arithmetic properties of the characteristic function, we have that

$$\left|F(0)\right|=\left|f(0)+a_0\right|\le\left|f(0)\right|+1, \tag{7}$$

$$T(r,F_{z_j})=T(r,f_{z_j}+a_j)\le T(r,f_{z_j})+\log 2, \tag{8}$$

and when $|z|\le r$,

$$\begin{aligned}&\log^+|a_1z_1+a_2z_2+\cdots+a_nz_n+a_0|\\&\quad\le\log^+(nr+1)\le\log^+ r+\log n+\log 2,\end{aligned}$$

which implies that

$$T(r,a_1z_1+a_2z_2+\cdots+a_nz_n+a_0)\le\log^+ r+\log n+\log 2$$

and thus that

$$\begin{aligned}T(r,f)&=T\big(r,F-(a_1z_1+a_2z_2+\cdots+a_nz_n+a_0)\big)\\&\le T(r,F)+T(r,a_1z_1+a_2z_2+\cdots+a_nz_n+a_0)+\log 2\\&\le T(r,F)+\log^+ r+\log n+2\log 2. \qquad (9)\end{aligned}$$

Now, by (9), (6), (8), and (7), we deduce that

$$T(r,f)\le C_1\left\{\sum_{j=1}^{n}T(\tau r,f_{z_j})+\log^+ r+\log^+\left|f(0)\right|+1+\log n\right\},$$

where C_1 is a constant depending only on τ. This completes the proof. □

In one variable, it is well known that f and f' have the same order. This is false in $\mathbf{C}^n$ when $n > 1$ for f and one of its partial derivatives, as we see from the example preceding Theorem 3. However, Theorem 1 enables us to give the relation between the order of f and the orders of its partial derivatives as follows.

Corollary 1 *If f is meromorphic in $\mathbf{C}^n$, then $\rho(f) = \max_{1\le j\le n}\{\rho(f_{z_j})\}$.*

Proof We may assume that f is nonconstant. Denote $\sigma := \max_{1\le j\le n}\{\rho(f_{z_j})\}$. First, suppose that $\sigma < +\infty$. Then for any $\varepsilon > 0$, $T(r, f_{z_j}) \le r^{\sigma+\varepsilon}$ for large r and each $1 \le j \le n$. Without loss of generality, we may assume that f is analytic at 0. (Otherwise, we may consider the function $F(z) := f(z_0 + z)$ for some z_0 at which f is analytic. The characteristic of F and that of f differ by a constant multiple (see, e.g. [12], p. 335]), and thus f and F have the same order.) By Theorem 1 we obtain that

$$\log T(r, f) < (\sigma + \varepsilon)\log r + O(1)$$

for large r. Thus, $\rho(f) \le \sigma$. To show that $\sigma \le \rho(f)$, we use the logarithmic derivative lemma: $m(r, \frac{f_{z_j}}{f}) = o\{T(r, f)\}$ outside a possible exceptional set of r of finite Lebesgue measure, from which it follows that

$$\begin{aligned} T(r, f_{z_j}) &= N(r, f_{z_j}) + m(r, f_{z_j}) \\ &\le N(r, f_{z_j}) + m\left(r, \frac{f_{z_j}}{f}\right) + m(r, f) \le CT(r, f) \end{aligned} \tag{10}$$

for large r outside a possible exceptional set, denoted by E, of r of finite Lebesgue measure, denoted by l, where $C > 0$ is a constant. If $r \in E$, then there exists a $r_1 \notin E$ with $r \le r_1$ and $r_1 - r \le l + 1$ since the measure of E is l. Then by (10), which holds at r_1, we have that

$$T(r, f_{z_j}) \le T(r_1, f_{z_j}) \le CT(r_1, f) \le CT(r + l + 1, f)$$

for large r. Combining this with (10), we see that for large r without any exceptional set, we always have that $T(r, f_{z_j}) \le CT(r + l + 1, f)$. Hence,

$$\begin{aligned} \rho(f_{z_j}) &= \limsup_{r\to\infty} \frac{\log^+ T(r, f_{z_j})}{\log r} \\ &\le \limsup_{r\to\infty} \frac{\log^+ T(r + l + 1, f) + \log^+ C}{\log(r + l + 1)} \frac{\log(r + l + 1)}{\log r} = \rho(f). \end{aligned}$$

Therefore, $\sigma \le \rho(f)$. We thus have proved that $\sigma = \rho(f)$ when $\sigma < +\infty$.

If $\sigma = +\infty$, then it is easy to deduce from (10) that $\rho(f) = +\infty$. Thus, we also have that $\sigma = \rho(f)$. This proves the theorem. □

4 Solutions of (1) an (2)

Lemma 1 *Let a_1, a_2, and a_3 be nonzero meromorphic functions in $\mathbf{C}^n$, and m_1, m_2 positive integers satisfying that $\frac{1}{m_1} + \frac{1}{m_2} < 1$. If f_1 and f_2 are meromorphic solutions of the equation $a_1 f_1^{m_1} + a_2 f_2^{m_2} = a_3$ in $\mathbf{C}^n$, then for $j = 1, 2$,*

$$T(r, f_j) \leq C_j \bar{N}(r, f_j) + O\left\{T\left(r, \frac{a_1}{a_3}\right) + T\left(r, \frac{a_2}{a_3}\right)\right\} + S(r, f_j), \tag{11}$$

where $C_j = \frac{1}{m_j(1-\frac{1}{m_1}-\frac{1}{m_2})}$.

Proof Since $\frac{a_1}{a_3} f_1^{m_1} + \frac{a_2}{a_3} f_2^{m_2} = 1$, we have by Nevanlinna's first and second fundamental theorems that

$$\begin{aligned}
&T\left(r, f_1^{m_1}\right) \\
&\quad\leq T\left(r, \frac{a_1}{a_3} f_1^{m_1}\right) + T\left(r, \frac{a_1}{a_3}\right) + O(1) \\
&\quad\leq \bar{N}\left(r, \frac{a_1}{a_3} f_1^{m_1}\right) + \bar{N}\left(r, \frac{1}{\frac{a_1}{a_3} f_1^{m_1}}\right) + \bar{N}\left(r, \frac{1}{\frac{a_1}{a_3} f_1^{m_1} - 1}\right) + S\left(r, \frac{a_1}{a_3} f_1^{m_1}\right) \\
&\qquad + T\left(r, \frac{a_1}{a_3}\right) + O(1) \\
&\quad= \bar{N}\left(r, \frac{a_1}{a_3} f_1^{m_1}\right) + \bar{N}\left(r, \frac{1}{\frac{a_1}{a_3} f_1^{m_1}}\right) + \bar{N}\left(r, \frac{1}{\frac{a_2}{a_3} f_2^{m_2}}\right) + S\left(r, \frac{a_1}{a_3} f_1^{m_1}\right) \\
&\qquad + T\left(r, \frac{a_1}{a_3}\right) + O(1) \\
&\quad\leq \bar{N}(r, f_1) + \frac{1}{m_1} N\left(r, \frac{1}{f_1^{m_1}}\right) + \frac{1}{m_2} N\left(r, \frac{1}{f_2^{m_2}}\right) + O\left\{T\left(r, \frac{a_1}{a_3}\right)\right. \\
&\qquad \left. + T\left(r, \frac{a_2}{a_3}\right)\right\} + S(r, f_1) \\
&\quad\leq \bar{N}(r, f_1) + \frac{1}{m_1} T\left(r, f_1^{m_1}\right) + \frac{1}{m_2} T\left(r, f_2^{m_2}\right) + O\left\{T\left(r, \frac{a_1}{a_3}\right) + T\left(r, \frac{a_2}{a_3}\right)\right\} \\
&\qquad + S(r, f_1) \\
&\quad\leq \bar{N}(r, f_1) + \frac{1}{m_1} T\left(r, f_1^{m_1}\right) + \frac{1}{m_2} T\left(r, f_1^{m_1}\right) + O\left\{T\left(r, \frac{a_1}{a_3}\right) + T\left(r, \frac{a_2}{a_3}\right)\right\} \\
&\qquad + S(r, f_1),
\end{aligned}$$

which implies that

$$T(r, f_1) \leq \frac{1}{m_1\left(1 - \frac{1}{m_1} - \frac{1}{m_2}\right)}\bar{N}(r, f_1) + O\left\{T\left(r, \frac{a_1}{a_3}\right) + T\left(r, \frac{a_2}{a_3}\right)\right\} + S(r, f_1).$$

The same conclusion holds for $T(r, f_2)$. This proves the lemma. □

Remark 1 The main feature of Lemma 1 is the explicit constant C_j. It allows us to treat both entire and meromorphic solutions in a precise way and implies/unifies various known results.

(a) If f_j is entire, the term $\bar{N}(r, f_j)$ disappears in (11) so that the growth of f_j is controlled by that of the coefficients whenever $\frac{1}{m_1} + \frac{1}{m_2} < 1$. Thus, Lemma 1 gives the results of [19, 24] mentioned above. The condition $\frac{1}{m_1} + \frac{1}{m_2} < 1$ is precise in that the lemma fails when $\frac{1}{m_1} + \frac{1}{m_2} = 1$, i.e., $m_1 = m_2 = 2$, since there are transcendental entire solutions $f = \sin h,\ g = \cos h$ to the equation $f^2 + g^2 = 1$, where h is any nonconstant entire function. Even so, such an estimate is still possible when $m_1 = m_2 = 2$ under certain conditions (see [15]).

(b) When $m_1 \geq 3, m_2 \geq 3$ with $(m_1, m_2) \neq (3, 3)$, it is easy to see that $C_j = \frac{1}{m_j(1 - \frac{1}{m_1} - \frac{1}{m_2})} < 1$ and thus that

$$T(r, f_j) \leq O\left\{T\left(r, \frac{a_1}{a_3}\right) + T\left(r, \frac{a_2}{a_3}\right)\right\} + S(r, f_j).$$

Hence, the growth of f_j is controlled by that of the coefficients of the equation; for instance, if the coefficients $a_j (1 \leq j \leq 3)$ are constant, then it is immediate to see that $T(r, f_j) = O(1)$, i.e., the solutions f_1, f_2 are constant. Thus, it gives the results of [5, 27] mentioned above. The constant C_j in Lemma 1 is precise in the sense that the conclusion of the lemma fails when $(m_1, m_2) = (3, 3)$, since there are transcendental meromorphic solutions f, g (given by Weierstrass elliptic functions) to the equation $f^3 + g^3 = 1$ (see [1]).

(c) The constant C_j also allows us to treat the case where one of m_1, m_2 is equal to 2. It is clear that $C_1 < 1$ when $m_1 > 4,\ m_2 = 2$ or $C_2 < 1$ when $m_2 > 4,\ m_1 = 2$, which, in particular, implies that meromorphic solutions to $f_1^{m_1} + f_2^{m_2} = 1$ must reduce to constant for these m_1, m_2. In the cases $m_1 = 2, 3, 4$ and $m_2 = 2$ (or $m_2 = 2, 3, 4$ and $m_1 = 2$), nonconstant meromorphic solutions to $f_1^{m_1} + f_2^{m_2} = 1$ may exist.

When $m_1 = m_2 = 2$, the entire functions given in (a) or the meromorphic functions f, g given in (3) provide such an example. When $m_1 = 3, m_2 = 2$, one may choose the periods of a Weierstrass elliptic function $\wp$ so that $f = -4^{\frac{1}{3}}\wp$ and $g = i\wp'$ satisfy the equation $f^3 + g^2 = 1$ (see, e.g., [10]). When $m_1 = 4, m_2 = 2$, one may also choose the periods of a Weierstrass elliptic function $\wp$ such that $f = 2\frac{\wp}{\wp'}$ and $g = \wp'$ satisfy $f^4 + g^2 = 1$ (see, e.g., [9], p. 974).

We next prove the following result using Lemma 1, Theorem 3, and Corollary 1.

Theorem 4 *Let a_1, a_2 and a_3 be nonzero meromorphic functions in $\mathbf{C}^2$, and p, q two positive integers satisfying that $p \geq 2,\ q \geq 2$ with $(p, q) \neq (2, 2)$. Let u be an entire solution of the nonlinear partial differential equation $a_1 u_{z_1}^p + a_2 u_{z_2}^q = a_3$ in $\mathbf{C}^2$. Then u must be of finite λ_1-type (resp. finite order $\leq \rho$ or linear) if a_1, a_2, a_3 are of finite λ-type (resp. finite order $\leq \rho$ or constant), where $\lambda_1 = \lambda + \log r$.*

If $p \geq 3,\ q \geq 3$ with $(p, q) \neq (3, 3)$ or $p = 2,\ q > 4$ or $q = 2,\ p > 4$, then the above conclusion also holds for meromorphic solutions u.

A similar statement holds for functional equations $a_1 f^p + a_2 g^q = a_3$ with $\lambda_1 = \lambda$ (cf. the proof below). Note, however, that for partial differential equations in Theorem 4, λ_1 cannot be replaced by λ, as seen from the entire/meromorphic solutions u of the eikonal equation $u_{z_1}^2 + u_{z_2}^2 = 1$, where the coefficients a_j are of finite λ-type with λ being constant, while u is linear, which is of finite $\log r$-type. The term $\log r$ comes exactly from the same term on the right-hand side of the inequality in Theorem 3.

Proof Since $f = u_{z_1}$ and $g = u_{z_2}$ are entire solutions of the functional equation $a_1 f^p + a_2 g^q = a_3$, by Lemma 1 (cf. Remark 1) we have that

$$T(r, f) + T(r, g) = O\left\{\sum_{j=1}^{3} T(r, a_j)\right\} + O(1) \tag{12}$$

outside a possible exceptional set, denoted by E, of r of finite Lebesgue measure, denoted by l. If the functions a_j $(1 \leq j \leq 3)$ are of finite λ-type, then there are positive constants A and B such that $\sum_{j=1}^{3} T(r, a_j) \leq A\lambda(Br)$ for all large r. Thus by (12), for large r outside the exceptional set E, we have that

$$T(r, f) + T(r, g) \leq C\lambda(Br) \tag{13}$$

for some $C > 0$. If $r \in E$, then there exists an $r_1 \notin E$ with $r \leq r_1$ and $r_1 - r \leq l + 1$. Thus, by (13), we have that

$$\begin{aligned} T(r, f) + T(r, g) &\leq T(r_1, f) + T(r_1, g) \\ &\leq C\lambda(Br_1) \leq C\lambda\big(B(r + l + 1)\big) \leq C\lambda(2Br) \end{aligned}$$

for large r. This, together with (13), yields that

$$T(r, f) + T(r, g) \leq C\lambda(2Br) \tag{14}$$

for all large r without exceptional sets of r. This shows that $f = u_{z_1}$ and $g = u_{z_2}$ are of finite λ-type. Without loss of generality, we may assume that u is analytic at 0 (otherwise, we may consider $g(z) := u(\zeta_0 + z)$ for a ζ_0 at which u is analytic; cf. the proof of Corollary 1). Then by Theorem 3 we deduce that u is of finite λ_1-type, where $\lambda_1 = \lambda + \log r$.

If the functions a_j $(1 \le j \le 3)$ are of finite order $\le \rho$, then for any $\varepsilon > 0$, $\sum_{j=1}^{3} T(r, a_j) \le r^{\sigma+\varepsilon}$ for large r. By (12),

$$T(r, f) + T(r, g) \le Cr^{\sigma+\varepsilon} \tag{15}$$

for large r outside the exceptional set E. If $r \in E$, then for an $r_1 \notin E$ with $r \le r_1$ and $r_1 - r \le l + 1$ as above, we have that

$$T(r, f) + T(r, g) \le T(r_1, f) + T(r_1, g) \le C(r + l + 1)^{\sigma+\varepsilon} \le C2^{\sigma+\varepsilon} r^{\sigma+\varepsilon}$$

for large r. This inequality is also true for $r \notin E$ by (15). Thus, $f = u_{z_1}$ and $g = u_{z_2}$ are of finite order $\le \rho$, too. Then by Corollary 1, u is of finite order $\le \rho$.

If the functions a_j $(1 \le j \le 3)$ are constant, we then have that $T(r, f) + T(r, g) = O(1)$ by (12), which implies that $f = u_{z_1}$ and $g = u_{z_2}$ are constant, and thus u is linear.

The conclusions of the theorem for meromorphic solutions u of the partial differential equation under the given condition on p, q may be shown in exactly the same way as above in view of the fact that in these cases, $C_j < 1$ in Lemma 1 (cf. Remark 1). □

5 Some Variations of the Eikonal Equation

As is well known, there is no foolproof ways to solve general nonlinear partial differential equations. To conclude the paper, we discuss some variations of the eikonal equation $u_{z_1}^2 + u_{z_2}^2 = 1$, which present and illustrate different treatments.

First, take the equations $u_{z_1}^2 + u_{z_2}^2 = u^2$ and $u_{z_1}^2 + u_{z_2}^2 = u^3$ as examples. The first one can immediately reduce to the functional equation satisfying the condition in Theorem 1: $f^2 + g^2 = 1$, where $f = \frac{u_{z_1}}{u}$ and $g = \frac{u_{z_2}}{u}$ satisfying that $f_{z_2} = g_{z_1}$. Thus, f, g are constant by Theorem 1, which implies, by integration, that $u = ce^{c_1 z_1 + c_2 z_2}$, where c_1, c_2, c are constants satisfying that $c_1^2 + c^2 = 1$ (see also [17] and [7]). However, the second partial differential equation cannot be treated in this way. Nevertheless, if u is a meromorphic solution of the equation $u_{z_1}^2 + u_{z_2}^2 = u^n$, it is easy to see that

$$\begin{aligned} nT(r, u) = T\big(r, u^n\big) = T\big(r, u_{z_1}^2 + u_{z_2}^2\big) &= m\big(r, u_{z_1}^2 + u_{z_2}^2\big) + N\big(r, u_{z_1}^2 + u_{z_2}^2\big) \\ &\le m\left(r, \frac{u_{z_1}^2 + u_{z_2}^2}{u^2}\right) + m\big(r, u^2\big) + 4N(r, u) \\ &\le 2m(r, u) + 4N(r, u) + S(r, u) \end{aligned}$$

by the logarithmic derivative lemma. When u is entire, the term $N(r, u)$ disappears. Thus, the above inequality clearly implies the following:

Proposition 1 *The partial differential equation $u_{z_1}^2 + u_{z_2}^2 = u^n$ has no nonzero entire solutions when $n > 2$ and has no nonzero meromorphic solutions when $n > 4$.*

The condition on n is precise since the result fails when $n = 1, 2$ for entire solutions and also fails when $n = 1, 2, 3, 4$ for meromorphic solutions. When $n = 1$, $u = \frac{1}{4}(z_1^2 + z_2^2)$ is an entire solution. When $n = 2$, an entire solution was given above. When $n = 3$, $u = \frac{4}{z_1^2+z_2^2}$ is a meromorphic solution. When $n = 4$, $u = \frac{\sqrt{2}}{z_1+z_2}$ is a meromorphic solution.

We next consider the nonlinear partial differential equation of tubular surfaces (see, e.g., [3], p. 27 and p. 95)

$$u^2\big(u_{z_1}^2 + u_{z_2}^2 + 1\big) = 1, \tag{16}$$

or $u_{z_1}^2 + u_{z_2}^2 = \frac{1}{u^2} - 1$. The two-parameter family of the spheres $(x-a)^2 + (y-b)^2 + u^2 = 1$ is a complete integral of the equation $u^2(u_x^2 + u_y^2 + 1) = 1$. The envelopes of one-parameter family of spheres of radius 1 whose center moves along a curve $y = w(x)$ in the x, y plane are tubular surfaces whose axis is $y = w(x)$. The equation clearly has solutions $u = 1, -1$. We will show that they are the only entire (meromorphic) solutions. This can be proved in different ways. The method used below applies to the following more general partial differential equations

$$P(u, u_{z_1}, \ldots, u_{z_n}, u_{z_1z_1}, u_{z_1z_2}, \ldots)Q(u, u_{z_1}, \ldots, u_{z_n}, u_{z_1z_1}, u_{z_1z_2}, \ldots) = 1 \tag{17}$$

in $\mathbf{C}^n$, where P and Q are any nonconstant polynomials in u and its partial derivatives of arbitrary orders (P and Q can have slowly growing coefficients) with

$$P(0, u_{z_1}, \ldots, u_{z_n}, u_{z_1z_1}, u_{z_1z_2}, \ldots) = 0.$$

A special example is $P = u^m$, $m \geq 1$. Thus, (16) is a very special case of (17).

Suppose that u is an entire or meromorphic solution of (17). We use the following result in [18] (Lemma 2): $\int_E \log^+ \frac{P}{f}\eta = S(r, f)$ for any nonconstant meromorphic function f in $\mathbf{C}^n$ and any nonconstant polynomial P in f and its partial derivatives, where E is the set of ζ on the sphere $\mathbf{S}_r$ in $\mathbf{C}^n$ with radius r and centered at the origin such that $|f(\zeta)| \leq 1$ and η is the volume form on the sphere normalized so that the total measure of the sphere is 1. Then by (17) and the definition of the proximity function $m(r, f)$ (cf. [18]),

$$\begin{aligned} m\left(r, \frac{1}{u^2}\right) &\leq \int_E \log^+ \frac{1}{u^2}\eta = \int_E \log^+\left(\frac{P}{u}\frac{Q}{u}\right)\eta \\ &\leq \int_E \log^+ \frac{P}{u}\eta + \int_E \log^+ \frac{Q}{u}\eta = S(r, u), \end{aligned}$$

where the set E is defined as above with $|u(\zeta)| \leq 1$ on E. But, by (17) and the assumption on P we see that u does not vanish. Thus, $N(r, \frac{1}{u^2}) = 0$. We then obtained that

$$2T(r, u) = T\left(r, \frac{1}{u^2}\right) + O(1) = m\left(r, \frac{1}{u^2}\right) + N\left(r, \frac{1}{u^2}\right) = S(r, u),$$

which implies that u is a constant. Substituting u into (17), we obtain the following:

Proposition 2 *If u is a meromorphic solution of* (17) *in* $\mathbf{C}^n$, *then u must be a constant c satisfying that*

$$P(c, 0, \ldots, 0, \ldots)Q(c, 0, \ldots, 0, \ldots) = 1.$$

In particular, the meromorphic solutions of (16) *in* $\mathbf{C}^2$ *are exactly $u = 1$ and $u = -1$.*

References

1. Baker, I.N.: On a class of meromorphic functions,. Proc. Am. Math. Soc. **17**, 819–822 (1966)
2. Chuang, C.: Sur la comparaison de la croissance d'une fonction méromorphe et de celle de sa dérivée. Bull. Sci. Math. **75**, 171–190 (1951)
3. Courant, R., Hilbert, D.: Methods of Mathematical Physics. Partial Differential Equations, vol. II. Interscience, New York (1962)
4. Gundersen, G., Hayman, W.: The strength of Cartan's version of Nevanlinna theory. Bull. Lond. Math. Soc. **36**, 433–454 (2004)
5. Gross, F.: On the equation $f^n + g^n = 1$. Bull. Am. Math. Soc. **72**, 86–88 (1966) (erratum, **72**, 576 (1966))
6. Granville, A., Tucker, T.J.: It's as easy as abc. Not. Am. Math. Soc. **49**, 1224–1231 (2002)
7. Han, Q.: On complex analytic solutions of the partial differential equation $u_{z_1}^m + u_{z_2}^m = u^m$. Houst. J. Math. **35**, 277–289 (2009)
8. Hemmati, J.: Entire solutions of first-order nonlinear partial differential equations. Proc. Am. Math. Soc. **125**, 1483–1485 (1977)
9. Huber, A.: A novel class of solutions for a nonlinear third order wave equation generated by the Weierstrass transformation. Chaos Solitons Fractals **28**, 972–978 (2006)
10. Hurwitz, A., Courant, R.: Funktionentheorie, vol. 4. Springer, Berlin (1964)
11. Khavinson, D.: A note on entire solutions of the eikonal equation. Am. Math. Mon. **102**, 159–161 (1995)
12. Kujala, R.: Functions of finite λ-type in several complex variables. Trans. Am. Math. Soc. **161**, 327–358 (1971)
13. Jategaonkar, A.V.: Elementary proof of a theorem of P. Motel on entire functions. J. Lond. Math. Soc. **40**, 166–170 (1965)
14. Lang, S.: Old and new conjectured diophantine inequalities. Bull. Am. Math. Soc. **23**, 37–75 (1990)
15. Li, B.Q.: On certain functional and partial differential equations. Forum Math. **17**, 77–86 (2005)
16. Li, B.Q.: Entire solutions of certain partial differential equations and factorization of partial derivatives. Trans. Am. Math. Soc. **357**, 3169–3177 (2005)
17. Li, B.Q.: On meromorphic solutions of $f^2 + g^2 = 1$. Math. Z. **258**, 763–771 (2008)
18. Li, B.Q.: Uniqueness of entire functions sharing four small functions. Am. J. Math. **119**, 841–858 (1997)
19. Montel, P.: Leçons sur les familles normales de functions analytiques et leurs applications. Gauthier-Villars, Paris (1927)
20. Saleeby, E.G.: Entire and meromorphic solutions of Fermat type partial differential equations. Analysis **19**, 369–376 (1999)
21. Shabat, B.V.: Introduction to Complex Analysis, part II, Functions of Several Variables. Translation Mathematical Monographs, vol. 110. American Mathematical Society, Providence (1992)
22. Shabat, B.V.: Distribution of Values of Holomorphic Mappings. Translation Mathematical Monographs, vol. 61. American Mathematical Society, Providence (1985)

23. Stoll, W.: Introduction to the Value Distribution Theory of Meromorphic Functions. Springer, New York (1982)
24. Toda, N.: On the functional equation $\sum_{i=0}^{p} a_i f_i^{n_i} = 1$. Tohoku Math. J. **23**, 289–299 (1971)
25. Taylor, R., Wiles, A.: Ring-theoretic properties of certain Hecke algebra. Ann. Math. **141**, 553–572 (1995)
26. Wiles, A.: Modular elliptic curves and Fermat's last theorem. Ann. Math. **141**, 443–551 (1995)
27. Yang, C.C.: A generalization of a theorem of P. Montel on entire functions. Proc. Am. Math. Soc. **26**, 332–334 (1970)
28. Yang, L.: Value Distribution Theory. Springer, Berlin (1993)

An Analogue of the Galois Correspondence for Foliations

Bernard Malgrange

Abstract For transverse parallelisms without first integral, I give a result similar to the Galois correspondence in the differential Galois theory of Kolchin.

1 Introduction

In this note, I discuss a few simple remarks on the "nonlinear differential Galois theory," as developed in [3] and [4]. A different theory, essentially equivalent, is due to Umemura [8, 9].

For simplicity, I limit myself to a special, but important case, the "transverse parallelisms"; see definition in Sect. 3. This case permits also a comparison with the theory of "strongly normal extensions" of Kolchin [2]. In a preliminary section, I review some facts on (algebraic) parallelisms. I have tried to make this note as independent as possible of [3] and [4], and as an introduction to these papers.

2 Parallelisms

2.1 Let X be a smooth and separated algebraic variety over $\mathbb{C}$. Unless explicitly stated, I suppose that X is irreducible.

Let $n = \dim X$, and let L be a Lie algebra over $\mathbb{C}$ of dimension n. Denote by θ_X (resp. Ω_X) the sheaf of vector fields (resp. differential 1-forms) on X. By definition, an L-parallelism on X is a linear map $u : L \to \Gamma(X, \theta_X)$ verifying

(i) u has constant rank n on X;
(ii) u commutes with the brackets (the second one is, of course, the Lie bracket on vector fields).

To the memory of my friend Leon Ehrenpreis.

B. Malgrange (✉)
Institut Fourier, BP 74, 38402 Saint Martin d'Hères, France
e-mail: bernard.malgrange@ujf-grenoble.fr

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_14, © Springer-Verlag Italia 2012

Let $M = L^*$ be the dual of L, with differential $d\colon M \to \Lambda^2 M$ equal to minus the transpose of the bracket of L. It is equivalent to give a map $M \xrightarrow{u^*} \Gamma(X, \Omega_X)$ of constant rank n and commuting with the differentials. It is also equivalent to give a form $\Omega \in \Gamma(X, \Omega_X) \otimes_{\mathbb{C}} L$ with the obvious condition of rank and the integrability condition $d\Omega + [\Omega, \Omega] = 0$.

The standard analytic description of this situation is the following: let X^{an} be the analytic variety defined by X, and $\widetilde{X}^{an}$ its universal covering with base-point $a \in X(\mathbb{C})$. Let also G be a connected Lie group over $\mathbb{C}$, whose Lie algebra Lie G is isomorphic to L. Then, the foliation of $G \times X$ given by the components of $dgg^{-1} + \Omega$ gives a "developing map" $\widetilde{X}^{an} \to G$ (defined up to right multiplication by G) and a monodromy map $\pi_1(\widetilde{X}^{an}, a) \to G$, which are the basis of the study of the given parallelism.

This situation has been much studied (and also its generalization with G replaced by a homogeneous space G/H. For simplicity, I will not consider this case here). But, as far as I know, the corresponding algebraic case seems to have been little studied. There are several differences.

(a) L is not necessarily the Lie algebra of an algebraic group. In the traditional terminology, the "third fundamental theorem" is not true in algebraic context. However, one has a weaker version.

Theorem 1 *For any Lie algebra L over $\mathbb{C}$, there exists an algebraic L-parallelism.*

This theorem is due to Deligne; see the proof in [4], Appendix B. A simple example is the following (loc.cit.): let L be the three-dimensional Lie algebra defined by $[\xi, \eta] = -\lambda\eta$, $[\xi, \zeta] = -\mu\zeta$, $[\eta, \zeta] = 0$, with $\lambda\mu \neq 0$, $\lambda/\mu \notin \mathbb{Q}$. One verifies easily that L is not the Lie algebra of an algebraic group. However, one has an L-parallelism on $\mathbb{C}^3$, with $\xi = \frac{\partial}{\partial x} + \lambda y \frac{\partial}{\partial y} + \mu z \frac{\partial}{\partial z}$, $\eta = \frac{\partial}{\partial y}$, $\zeta = \frac{\partial}{\partial z}$.

(b) Suppose that L is the Lie algebra of an algebraic group G that we can suppose to be connected.

Then, on $G \times X$, we can consider again the foliation defined by $dgg^{-1} + \Omega$ or, equivalently, the connection form $g^{-1}dg + adg^{-1}\Omega$.

It would be equivalent to consider any flat G-connection Π on $G \times X$ (G is supposed to act on the right) and a section $X \to G \times X$ transversal to the foliation defined by Π. In general, this situation is not easy to describe.

[For instance, for $G = Gl(n)$, we would need a generalization of Riemann–Hilbert correspondence to irregular singularities at infinity, plus conditions for the existence of a transversal section.]

2.2 From now on, I will work birationnally on X, i.e., I will replace freely X by a (Zariski) open dense set; a "parallelism" on X will mean a parallelism in the preceding sense on an open dense set U; and I will identify two parallelisms on U and V if they coincide on $U \cap V$.

Now, if (L, Ω) is such a parallelism on X, I define the *pseudogroup* Z associated to (L, Ω) as the pseudogroup on X fixing Ω.

For a precise definition of algebraic pseudogroups, see [4]. But, here it is sufficient to say that the solutions (algebraic, analytic, or formal) of Z are the leaves (algebraic, etc.) of the foliation F_Z on $X \times X$ defined by the coefficients of $p_2^*\Omega - p_1^*\Omega$, (p_1, p_2) the projections $X \times X \to X$.

For simplicity, I abbreviate the notations, denoting by x, ω, etc. the objects on the first component of $X \times X$ and $\bar{x}, \bar{\omega}$, etc. the same object on the second component. So the foliation F_Z is defined by $\overline{\Omega} - \Omega$; equivalently, if $\xi_1, \dots, \xi_n$ is a basis of $u(L)$, this foliation is defined by $\xi_1 + \bar{\xi}_1, \dots, \xi_n + \bar{\xi}_n$.

Now, denote by $\mathbb{C}(X)$ the field of rational functions on X, and the same for $X \times X$. Note that $\mathbb{C}(X)$, equipped with $\xi_1, \dots, \xi_n$, is a "field with operators" in the sense of Bialynicki-Birula [1], but not a "differential field" in the sense of Kolchin [2], since the ξ_i's do not commute in general. Therefore, we cannot use here the theory of "strongly normal extensions" of Kolchin, but we can use its generalization by [1]. This gives the following theorem.

Theorem 2 *Let (X, L, Ω) be a variety equipped with a parallelism. With the preceding notations, let $\mathbb{C}(X \times X)_c \subset \mathbb{C}(X \times X)$ be the field of "first integrals" or "constants" of F_Z (i.e., the $f \in \mathbb{C}(X \times X)$ verifying $(\bar{\xi}_i + \xi_i)(f) = 0$ $(1 \leq i \leq n)$. Then, the following conditions are equivalent*:

(i) *(X, L, Ω) is birationally equivalent to $(X, \operatorname{Lie} G, g^{-1}dg)$ for some algebraic group G.*
(ii) *$\mathbb{C}(X \times X)$ is generated by $\mathbb{C}(X) \otimes 1$ (= the functions of the first variable) and $\mathbb{C}(X \times X)_c$.*

Of course, (i) $\Rightarrow$ (ii) is trivial. The converses uses mainly the birational characterization of algebraic groups due to A. Weil. See [1].

Remark 1 The conditions of Theorem 2 imply that the foliation F_Z is "algebraically integrable", i.e., F_Z is determined by its first integrals.

The converse is false: if $X \rightsquigarrow G$ is a finite rational morphism, the pull-back of the standard parallelism on G is a parallelism on X, and the corresponding foliation is algebraically integrable. I do not know if the converse is true.

More generally, in all the examples that I know, $\operatorname{spec} \mathbb{C}(X \times X)_c$ admits a natural structure of birational algebraic group, but I have no general proof of this fact.

Here is a simple example. I will have to use the following theorem of Rosenlicht [6].

Theorem 3 *Let X be an algebraic variety, and let ξ be a vector field $\neq 0$ on X. Consider, on $X \times \mathbb{C}$, the vector field $\xi + f\frac{\partial}{\partial t}$, $f \in \mathbb{C}(t)$, $f \neq 0$. Then this vector field has no first integral depending on t (i.e., it has the same first integrals as ξ), except perhaps in the following cases*:

(i) $\frac{1}{f} = g'$, $g \in \mathbb{C}(t)$;
(ii) $\frac{1}{f} = c\frac{g'}{g}$, $c \in \mathbb{C}$, $g \in \mathbb{C}(t)$.

Takes now $X = \mathbb{C}$ and put $\xi = f\frac{\partial}{\partial x}$, $f \in \mathbb{C}(x)$, $f \neq 0$. On $X \times X$, the foliation F_Z is given by $f(x)\frac{\partial}{\partial x} + f(\bar{x})\frac{\partial}{\partial \bar{x}}$. If we are not in the exceptional cases, this vector field has no first integral except the constants. Iterating, we find also that on $\mathbb{C}^n$, the vector field $\Sigma f(x_i)\frac{\partial}{\partial x_i}$ has no first integral except the constants. On the other hand, if $\frac{1}{f} = g'$, then $g(\bar{x}) - g(x)$ is a generator of $\mathbb{C}(X \times X)_c$. It has obviously a structure of (birational) additive group. The case $\frac{1}{f} = c\frac{g'}{g}$ is similar, with the multiplicative group.

2.3 The pseudogroup Z has a Lie algebra, more precisely a D-Lie algebra, i.e., a system of linear PDEs on θ_X, whose space of solutions is stable by Lie bracket. Explicitly, this system is given by $L_\xi \Omega = 0$, L the Lie derivative.

This D-Lie algebra should not be confused with $u(L)$. It is a D-Lie algebra, whose (analytic) solutions are not necessarily rational. The relation between both is the following: suppose, by restricting X, that u has constant rank. Then, take $a \in X(\mathbb{C})$. There is an obvious isomorphism of the tangent space T_a with L. On the other hand, one verifies that the projection on T_a gives an isomorphism with T_a of the space of formal solutions at a of Lie Z. Now, the two Lie algebra structures on T_a that we have obtained are not equal, but opposed.

To prove this, it is sufficient to consider the case $X = G$ with $a = e$ (the general situation reduces to this one by considering a germ at a of developing map, in the analytic context). Now the first isomorphism (resp. the second) is the isomorphism of T_e with the right-invariant (resp. the left-invariant) vector fields.

Here is a simple example, already considered in [4]. Take $X = \mathbb{C}^2$ with the parallelism given by $\{\frac{\partial}{\partial y}, \frac{\partial}{\partial x} - y\frac{\partial}{\partial y}\}$ or, equivalently, by the dual basis $\{dy + y\,dx, dx\}$. The D-Lie algebra is given by $L_\xi dx = L_\xi(dy + y\,dx) = 0$. A base of solutions is $\frac{\partial}{\partial x}$, $e^{-x}\frac{\partial}{\partial y}$; this is not algebraic.

On the other hand, there is no pseudogroup Z on $\mathbb{C}^2$ such that Lie Z has exactly $\{\frac{\partial}{\partial y}, \frac{\partial}{\partial x} - y\frac{\partial}{\partial y}\}$ as a basis of solutions. Since these vector fields are linear affine, one can prove the following: find such a pseudogroup would be equivalent to find an algebraic subgroup of Aff(2) with Lie algebra $\{\frac{\partial}{\partial y}, \frac{\partial}{\partial x} - y\frac{\partial}{\partial y}\}$. But this is impossible: such a Lie algebra must contain the semi-simple and the nilpotent parts of $\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}$, i.e., $y\frac{\partial}{\partial y}$ and $\frac{\partial}{\partial x}$.

2.4 To end this section, a few simple remarks. They are essentially well known to the experts; see, e.g., [7].

For X, smooth algebraic variety over $\mathbb{C}$, denote by Aut X the groupoid of germs of analytic isomorphisms $(X^{an}, a) \xrightarrow{\sim} (X^{an}, b)$ with $a, b \in X(\mathbb{C})$. Roughly speaking, an algebraic pseudogroup on X is a subgroupoid of Aut X defined by algebraic partial differential equations. The precise definition is in terms of jets: for $k \geq 0$, let $J_k^* X$ be the variety of k-jets of invertible maps $X \to X$. A pseudogroup will be defined by a projective system of closed subvarieties $Z_k \subset J_k^* X$. For precise conditions, I refer to [4].

Now $Z = \{Z_k\}$ will be called *transitive* if $Z_0 = X \times X$. One proves that this condition is equivalent to the following: after restricting X if necessary, for any pair $a, b \in X(\mathbb{C})$, there exists a formal (or an analytic) solution of Z with source a and target b.

$Z = \{Z_k\}$ will be called "*of finite type*" (some people say "rigid") if the projections $Z_{k+1} \to Z_k$ are bijective for $k \gg 0$.

For instance, if Z is the pseudogroup associated to a parallelism on X, one verifies easily that it is transitive of finite type; the maps $Z_{k+1} \to Z_k$ are even bijective for $k \geq 0$. Incidentally, this implies that Z has no transitive strict subgroupoid (adding equations to Z would imply to add equations of order 0).

Now, it turns out that all pseudogroups transitive of finite type on X can be represented in some suitable frame bundle of X (or "prolongation of X") by a parallelism. This will be the case, e.g., for the pseudogroup of automorphisms of an affine structure, a projective structure, a Riemannian structure with constant curvature, etc. This means that parallelisms are an important special case of Lie pseudogroups.

Conversely, given a parallelism L, say on Y, it would be interesting to see if there exists some X with a dominant map $Y \to X$ such that the pseudogroup associated to L "descend" to a pseudogroup on X. For that purpose, may be, considerations similar to the proof of Theorem 2 could be useful. I will not look at this question here.

3 Transverse Parallelisms

3.1 This notion is defined in terms of differential forms: let X be a smooth variety of dimension n as before, and let M be a Lie coalgebra of dimension $m \leq n$. By definition, a transverse parallelism of type M on X is a map $M \xrightarrow{v} \Gamma(X, \Omega_X)$, verifying

(i) $v(M)$ has constant rank m on X;
(ii) v commutes with the differentials.

Of course, if $m = n$, this is just a parallelism as considered in Sect. 2. If $m < n$, it defines a foliation F of codimension m, with a "transverse parallelism" (roughly speaking, a parallelism on the transversals to F, invariant by the flow, or holonomy, of F).

We will be interested in the transverse parallelisms which have no rational first integral except the constants (I recall that, we work birationally on X). We call these parallelisms "transitive," since their differential Galois pseudogroup (see definition below) is transitive. Note that, if $m = n$, the parallelism is never transitive.

Let L be the Lie algebra dual to M, and let L' be a Lie subalgebra of L. This is equivalent to give $M' = L'^{\perp}$ verifying $dM' \subset M' \wedge M$ (I call this a "coideal" of M). Given a parallelism (M, v) on X, L' (or M') defines a foliation G to X, bigger than F (i.e., the leaves of G contain those of F). Now we have the following simple, but important observation.

Proposition 1 *Suppose that the parallelism* (M, v) *transitive, and denote by* F *the corresponding foliation. Then, all the foliations* G *of* X *bigger than* F *are obtained in this way.*

Let G be such a foliation; let N be the conormal bundle of the foliation [with basis $v(M)$], and let $N' \subset N$ the subbundle defined by G. Then N' is an F-transversal structure, i.e., N' is stable by the vector fields tangents to F. On the other hand, (M, v) gives a canonical trivialization $N \simeq X \times M$.

I claim that N' is compatible with this trivialization; in other words, there exists a subspace M' of M such that, in the preceding trivialization, one has $N' = X \times M'$.

We can suppose that condition (i) is satisfied on the whole X. Choose $a \in X(\mathbb{C})$, and choose a decomposition $N(a) = N_1(a) \oplus N_2(a)$ such that $N'(a)$ is the graph of a map $N_1(a) \to N_2(a)$. The preceding trivialization extends this decomposition to a decomposition $N = N_1 \oplus N_2$, and a decomposition $M = M_1 \oplus M_2$, with $N_i = X \times M_i$. For a general $b \in X(\mathbb{C})$, $N'(b)$ is the graph of a map $N_1(b) \to N_2(b)$. Finally, we obtain a rational function $X \rightsquigarrow \mathrm{Hom}(M_1, M_2)$, stable by the vector fields of F. Therefore this map is constant.

Now the fact that $M' = N'(a)$ is a coideal of M follows from the Frobenius condition on G.

3.2 We will give now some extensions and comments of the preceding result. Here is a first extension in a "tannakian" spirit. Let again (M, v) be a transitive transverse parallelism; let F be the corresponding foliation, and N the conormal bundle. Call "construction on N" a direct sum P of a finite family $N^{\otimes p} \otimes (N^*)^{\otimes q}$ and denote by Q the corresponding construction on M. One has a trivialization $P \simeq X \times Q$. Now, if Q' is a subvector space of Q, $X \times Q'$ is a subvector bundle of P, stable by the vector fields tangent to F; and all subvector bundles of P stable by these vector fields are obtained in this way. The proof is similar to Proposition 1.

3.3 To have a more general and systematic extension of Proposition 1, it is convenient to use the terminology of [3] and [4]. Given a foliation F on a variety X, the "Galois pseudogroup of F" is by definition the smallest pseudogroup on X, whose D-Lie algebra contains in its solutions all the vector fields tangent to F. For the proof that such a "smallest pseudogroup, etc." exists, I refer to loc.cit. Now, if (M, v) is a transverse transitive parallelism on X, with foliation F, one has the following result (loc.cit.).

Theorem 4 *The Galois pseudogroup of* F *is the pseudogroup fixing* $v(M)$.

Let $\pi_1, \ldots, \pi_m$ be a basis of M, and put $\omega_i = v(\pi_i)$. If ξ is a vector field (algebraic, analytic, or formal) tangent to F, one has $\langle \xi, \omega_i \rangle = 0$. Denoting by L_ξ the

Lie derivative, one has $L_\xi \omega_i = d\langle \xi, \omega_i \rangle + i_\xi \, d\omega_i = 0$ (the first term of the second member vanishes by definition; the second because of the expression of $d\omega_i$).

Consider now the pseudogroup Z fixing the ω_i's, and denote by Z' the Galois pseudogroup of F. The preceding argument shows that $Z' \subset Z$. Now, if we had $Z' \neq Z$, a transverse modification of the arguments of Sect. 2.4 would prove that Z' contains nontrivial equations of order 0. But this is equivalent to the existence of nontrivial first integrals. This proves the theorem.

This result reproves immediately Proposition 1 and its extension Sect. 3.2. Let, for instance, N' be a subbundle of the conormal bundle N, stable by the vector fields of the foliation. Then, the pseudogroup leaving N' stable must contain Z. If $Z_1 \subset J_1^* X$ is the equation of order one of Z, this means that N' is stable by Z_1. But Z_1 is the subbundle of $J_1^* X$ leaving F invariant and is equal transversally to the family of isomorphisms $N(a) \xrightarrow{\sim} N(b)$, $a, b \in X(\mathbb{C})$ deduced from (M, v). This gives the required statement.

More generally, Theorem 4 gives similarly a description of all the transverse structures of F, and not only of the structures of order one, as in Proposition 1 and Sect. 3.2. I will omit the details.

We can also ask the following question: given a foliation F on X, does there exist a "normal form," i.e., a prolongation to some frame bundle over X, which is a transitive transversal parallelism? The answer is a transverse analogue to the result of Sect. 2.4. The foliation F should have no nontrivial first integral, and its Galois pseudogroup must be "transversally finite," i.e., its restriction on transversals must be of finite type; see the details in [4].

If F has no first integrals but the Galois pseudogroup is of infinite type, the theory of Lie pseudogroups gives a more or less similar normal form, but with infinite-dimensional Lie algebras; see [7]. The intransitive case is more sophisticated; I refer for it to [4].

3.4 It is interesting to compare the preceding results, especially Sect. 3.2, with the standard theory of "strongly normal extensions" of Kolchin: see [2], and also [1].

I will consider only a special case of this theory. Let G be an algebraic connected group over $\mathbb{C}$, and let $p\colon X \to S$ be a G-principal bundle (G operates on right). I suppose X, S, p smooth, and S connected.

Let Ω be a G-connection on p. I recall that this means a form on X with values on $\mathcal{G} = \operatorname{Lie} G$ verifying:

(i) The restriction of Ω to the (closed) fibers of p is equal to the image of the left Maurer–Cartan form $g^{-1}dg$ (I recall that this is defined without ambiguity).
(ii) If R_g is the right multiplication by $g \in G(\mathbb{C})$, one has $R_g^* \Omega = adg^{-1}\Omega$. I suppose also the connection *flat*, i.e., verifying $d\Omega + [\Omega, \Omega] = 0$.

Then, if M is the dual of $\mathcal{G}$, Ω defines a transverse M-parallelism on X. All the preceding considerations can therefore be applied here.

Let us translate this situation in Kolchin's language. Let k (resp. K) be the field $\mathbb{C}(S)$ (resp. $\mathbb{C}(X)$) of rational functions on S (resp. X). If we want, we can choose a commutative basis $\delta_1, \ldots, \delta_p$ ($p = \dim S$) of the derivations $\operatorname{Der} k/\mathbb{C}$ (for instance, we take a dominant projection $S \to \mathbb{C}^p$, and we lift the partial derivatives $\frac{\partial}{\partial x_i}$ of $\mathbb{C}^p$). Then, $(k;\ \delta_1, \dots, \delta_p)$ is a differential field in the sense of Kolchin, and Ω determines an extension $(\bar{\delta}_1, \dots, \bar{\delta}_p)$ of $(\delta_1, \dots, \delta_p)$ to X, which makes K a differential extension of k.

Suppose now that the parallelism defined by Ω is *transitive*. It is equivalent to say that K_c, the field of constants of $(K, \bar{\delta}_1, \dots, \bar{\delta}_p)$, is equal to $k_c = \mathbb{C}$. Then, we are in the situation of strongly normal extensions. In this situation we have the Galois correspondence (loc. cit.).

Theorem 5 *There is a bijection between the subgroups $G' \subset G$ and the differential fields K' with $k \subset K' \subset K$. Given G', K' is the subfield of K fixed by G', and, given K', G' is the subgroup of G such that $G'(\mathbb{C})$ fixes K'. Moreover, the connected subgroups of G correspond to the K' algebraically closed in K.*

This result has many analogies with Proposition 1. However, it is both less general and more precise.

(i) In the situation of Sect. 3.4, given a Lie subalgebra $\mathcal{G}'$ of $\mathcal{G}$, we have a foliation F' of X bigger that the foliation F defined by Ω. If $\mathcal{G}'$ comes from a subgroup $G' \subset G$, which we can suppose to be connected, then, may be after restricting S, there exists Y with dominant maps $X \xrightarrow{p'} Y \longrightarrow S$ factorizing p, and p' is a G' principal bundle. If we put $K' = \mathbb{C}(Y)$, the foliation F' on X is the inverse image of the foliation of Y defined by the differential structure of K'. But, if $\mathcal{G}'$ is not the Lie algebra of a subgroup of G, no such "descent" exists.

(ii) Take, more generally, the situation of a transitive M-parallelism on X, and denote by F the corresponding foliation. Then, any Lie subalgebra of M^* (or any coideal of M) gives a foliation F' on X bigger than F. This is somewhat similar to Theorem 5, since, from a transverse point of view, (X, F') can be considered as a "quotient" of (X, F). Then we have here an analogue of the Galois correspondence.

But, in general, there is no corresponding "descent" for several reasons.

(a) A priori, we have nothing analogous to S. We can get one by choosing a dominant projection $f\colon X \to S$ generically transverse to F (in particular, $\dim S$ is equal to the dimension of the leaves of F).

But the situation will depend strongly of the chosen projection. Take, for instance, the transverse parallelism defined by $dy - \frac{dx}{x}$ on $\mathbb{C}^2$. If we take the projection $(x, y) \mapsto x$, we find the situation of a strongly normal extension, with $G = G_a$. If we project by $(x, y) \mapsto y$, we find a strongly normal extension with $G = G_m$. If we take the projection $(x, y) \mapsto t = y + x$, we have no group structure in the fibers; for instance, in the fiber $t = 1$, we have the parallelism given by $\frac{dx}{x} + dx$, or $\frac{x}{1+x}\frac{d}{dx}$, which is in the nonexceptional case of Theorem 3.

(b) Suppose that we are in the situation of (a), with an S chosen. Given a Lie subalgebra L' of $L = M^*$, it gives a foliation F' on X bigger than F (F' is also defined by $M' = L'^{\perp} \subset M$). But, in general there is no factorization $X \to Y \to S$

analogous to (i). A necessary condition for the existence of such a factorization is that F' is "algebraically integrable relatively to S" [this is defined as algebraic integrability, but with Ω_X replaced by $\Omega_{X/S}$; equivalently, the foliation defined by $v(M')$ and $\Omega_S \circ p$, p the projection $X \to S$, is algebraically integrable].

It would be interesting to study more systematically this situation, and in particular, to have examples of the phenomena which can occur.

3.5 I come back to the situation of Sect. 3.4, i.e., a transitive transverse parallelism given by a connection the G-principal bundle $p : X \to S$; as in Sect. 3.4, I denote by Ω the connection form.

The group G is usually called the *Galois group*. Another form of G is the "intrinsic group" $\widetilde{G}$, introduced by Katz (see a discussion of this question in [5]). This is the group over S of automorphisms of p commuting with the action of G. $\widetilde{G}$ is naturally equipped with a connection deduced from Ω, which makes it an algebraic differential group.

Consider now, on the same situations, the objects of the nonlinear theory: the Lie algebra $\mathcal{G}$, the form Ω on X with value in $\mathcal{G}$, and the Galois pseudogroup $Z = \{Z_k\}$, $Z_k \in J_k^* X$.

The relation of these objects with the preceding ones is as follows.

(a) $\mathcal{G}$ is the Lie algebra of G (by definition).
(b) Given $s \in S$, the restriction of Z to the fiber $X(s)$ is the family of automorphisms commuting with the action of G [follows from the form of the restriction $\Omega / X(s)$; as usual, I identify a pseudogroup to its solutions]. This is isomorphic noncanonically to G; the isomorphism depends on the choice of a point $x \in X(s)(\mathbb{C})$.
(c) The pseudogroup associated to the differential group $\widetilde{G}$ (see [4], Appendix A) is simply the subpseudogroup Z' of Z of transformations $x \mapsto \bar{x}$ of Z commuting with p and fixing $s = p(x)$. The only point to verify is that the connection is the right one. This is a statement on jets of order one; therefore, I can work in the formal completion of a fiber of p, or also in the analytic category. A fortiori, I can suppose the fibration trivial, i.e., $X = S \times G$. Denote by ω the restriction of Ω to $S \times \{e\}$; then, we have $\Omega = g^{-1}dg + g^{-1}\omega g$. The pseudogroup Z' is given by the transformations $(s, g) \mapsto (s, \bar{g})$ fixing Ω. Writing $\bar{g} = \gamma g$, we find $\gamma^{-1}d\gamma + \gamma^{-1}\omega\gamma = \omega$, or $d\gamma = \gamma\omega - \omega\gamma$. This is the equation of the connection of $\widetilde{G}$.

Now, in the general situation of a transitive transverse parallelism, or more generally, of a general foliation, the things are, roughly speaking, as follows: to the Galois group of the classical theory of Kolchin corresponds the Lie algebra of the parallelism (or, in general, the "virtual group" in the sense of [4]), and to the intrinsic group corresponds the Galois pseudogroup.

The distinction between two versions of the "Galois group" seems to have been considered for the first time by Vessiot [10], with his "groupe de rationalité" and his "groupe spécifique.'. I thank J.-P. Ramis and G. Casale for this reference.

References

1. Bialynicki-Birula, A.: On the Galois theory of fields with operators. Am. J. Math. **84**, 89–109 (1962)
2. Kolchin, E.R.: Differential Algebra and Algebraic Groups. Academic Press, San Diego (1973)
3. Malgrange, B.: Le groupoïden de Galois d'un feuilletage. Monogr. Enseign. Math. **38**, 465–501 (2001)
4. Malgrange, B.: Pseudogroupes de Lie et théorie de Galois différentielle. http://www.ihes.fr/prepublicationsm/10/11
5. Pillay, A.: Algebraic D-groups and differential Galois theory. Pac. J. Math. **216**, 343–360 (2004)
6. Rosenlicht, M.: The non-minimality of the differential closure. Pac. J. Math. **52**, 529–537 (1974)
7. Singer, I.M., Sternberg, S.: The infinite groups of Lie and Cartan, Part I: The transitive groups. J. Anal. Math. **15**, 1–114 (1965)
8. Umemura, H.: Differential Galois theory of infinite dimension. Nagoya Math. J. **114**, 59–134 (1996)
9. Umemura, H.: Sur l'équivalence des théories de Galois différentielles générales. C. R. Acad. Sci. Paris, série 1, **346**, 1155–1158 (2008)
10. Vessiot, E.: Sur la réductibilité aux dérivées partielles non linéaires du premier ordre, à une fonction inconnue. Ann. Sci. Éc. Norm. Super. 3ème série **32**, 137–160 (1915)

A Quantitative Version of Carathéodory's Theorem for Convex Sets

Reinhold Meise and Alan Taylor

Abstract Carathéodory's theorem for compact convex sets $K \subset \mathbb{R}^m$ shows that every point x of K lies in the convex hull of $m+1$ extreme points of K; that is, in the m-simplex with vertices at $m+1$ extreme points. However, it need not be the case that if x is a positive distance away from the boundary of K, then x is a positive distance away from the boundary of one of these simplices. Here, we show that if K has only finitely many extreme points, then there are a finite set $F \subset \partial K$ and a constant $c > 0$ such that if $x \in K$ is of distance $\delta > 0$ from the boundary of K, then x belongs to one of the m-simplices with vertices from F and is of distance at least $c\delta$ from its boundary.

1 Introduction

Let K be a polytope in $\mathbb{R}^m$, i.e., a compact convex set in $\mathbb{R}^m$ with finitely many extreme points, say $x_1, \ldots, x_p$. Let us also suppose that K has a nonempty interior so that $p \geq m+1$. Carathéodory's theorem implies that each point x in K can be written as a convex combination of at most $m+1$ of these extreme points. If one is allowed to use convex combinations of all p extreme points, then a quantitative version of this fact is also true. Namely, each $x \in K$ that is of distance greater than $\delta > 0$ from the boundary of K can be written as a convex combination of extreme points

$$x = \sum_{i=1}^{p} \lambda_i x_i, \qquad \sum_{i=1}^{p} \lambda_i = 1, \quad \lambda_i \geq 0,\ 1 \leq i \leq p,$$

Dedicated to the memory of Leon Ehrenpreis.

R. Meise (✉)
Mathematisches Institut, Heinrich-Heine-Universität, Universitätstraße 1, 40225 Düsseldorf, Germany
e-mail: meise@math.uni-duesseldorf.de

A. Taylor
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: taylor@umich.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_15, © Springer-Verlag Italia 2012

in which all of the coefficients are at least proportional to δ. To see this, one can assume that $0 \in K$ and that the sum of the extreme points x_i is equal to 0. Then if $0 \neq x = \sum \lambda_i x_i \in K$ is of distance δ from the boundary of K, then $x^* = (1 + p\varepsilon)x$ also belongs to K, where $\varepsilon = \frac{\delta}{p|x|} \geq \frac{\delta}{p\,\mathrm{diam}(K)}$. Therefore,

$$x = \frac{x^*}{(1 + p\varepsilon)} = \frac{x^* + 0}{(1 + p\varepsilon)} = \sum_{i=1}^{p} \frac{(\lambda_i + \varepsilon)}{(1 + p\varepsilon)} x_i,$$

and the coefficients in this expansion all satisfy

$$\frac{\lambda_i + \varepsilon}{1 + p\varepsilon} \geq \frac{\varepsilon}{1 + p\varepsilon} = \frac{\delta}{p(|x| + \delta)} \geq \frac{\delta}{2p\,\mathrm{diam}(K)}, \quad 1 \leq i \leq p.$$

The question therefore arises if it is possible to do this using only the "correct number," i.e., $m + 1$, of these points.

A simple example shows that this is not the case. Namely, if K is a rectangle in the plane, then the center of the rectangle is of large distance from the boundary, but it does not belong to the interior of any of the triangles formed by choosing three of the vertices. The obstruction is that the interior of K is not the union of the interiors of the simplices determined by $3 = m + 1$ of the vertices.

There is a similar example for K in three dimensions. Namely, let K denote a pyramid with rectangular base and one additional vertex, say at height 1 above the base. Consider the line segment joining the center of the base to this vertex. Then points on this line at a small distance δ above the base cannot be in the interior of any of the 3-simplices formed by taking four of the five vertices as corners. Thus, it may not be just the "interior" obstruction that must be removed but also the "interiors of the faces" obstruction.

However, we can prove that some finite subset of the boundary of K will work for the quantitative Carathéodory theorem.

Theorem 1.1 (Quantitative Carathéodory theorem) *Let K be a polytope in $\mathbb{R}^m$. Then there are a finite set $F \subset \partial K$ and a constant $c > 0$ such that whenever $x \in K$ is of distance $\delta > 0$ from ∂K, there is an m-simplex S with its $m + 1$ vertices chosen from the set F such that $x \in S$ and the distance from x to the boundary of S is at least $c\delta$.*

Since Carathéodory's theorem has a simple and elegant proof, one might expect that this quantitative version does as well. We have been unable to find such an argument. Rather, the proof given here goes by "brute force." First, we handle points near the boundary of K by showing that if a point $x \in K$ is "near to" a k-dimensional face but "far from" $(k - 1)$-dimensional faces, then it lies in the simplex with $k + 1$ points chosen from a finite subset of the boundary of the nearby face and $m - k$ points chosen from another finite subset of the boundary of K chosen by looking at the projection of K parallel to the nearby face. Second, we show that this allows us to find a finite set $F \subset \partial K$ that works for all the points sufficiently near to the boundary of K. Last, we use a compactness argument to show that the remaining points in the interior can be handled by making the finite set F contain enough points to avoid the "interior point" obstruction mentioned earlier.

The paper is concluded in Sect. 4 by giving the corresponding version of the theorem for cones. In fact, this version is what led us to the question studied here because we suspect that it has application in our ongoing work to characterize the analytic varieties in $\mathbb{C}^n$ that satisfy the local Phragmén–Lindelöf condition at the origin. Such a characterization for analytic surfaces in $\mathbb{R}^3$ was given in [1], and we are still studying higher-dimensional cases.

For general reference and terminology regarding convex sets, we refer to [2].

The authors also wish to thank the referee for several helpful comments.

2 Points Near the Boundary of *K*

It is convenient to have a name for finite sets with the property of the theorem.

Definition 2.1 Let K be a compact convex subset of $\mathbb{R}^m$. A finite subset F of ∂K is said to be *robust* for K if there exists a positive constant $c > 0$ such that given any ball $B(x, \delta) \subset K$, there is an m-simplex S with vertices in F such that $B(x, c\delta) \subset S$.

Remark 1 If a vector $x = \sum_{i=0}^{m} \lambda_i f_i$ is a convex combination of vectors $f_0, \ldots, f_m$ that are the vertices of the simplex, then each coefficient λ_i is of the same magnitude as the distance of the vector x from the "opposite" facet; i.e., the distance to the hyperplane spanned by the other m vertices. For example, if N is the unit vector orthogonal to the hyperplane spanned by the vectors $f_1 - f_0, \ldots, f_{m-1} - f_0$, then the distance of x from this facet is

$$N \cdot (x - f_0) = N \cdot \left(\sum_{i=0}^{m} \lambda_i (f_i - f_0) \right) = \lambda_m \big(N \cdot (f_m - f_0) \big),$$

so λ_m is proportional to this distance. Therefore, the condition $B(x, c\delta) \subset S$ is equivalent to giving a lower bound proportional to δ for the coefficients of x as a convex combination of the vertices of the simplex.

Remark 2 A robust set F must contain all the extreme points of K. So, only polytopes can have finite robust subsets.

Theorem 1.1 can then be restated as: *Every polytope in* $\mathbb{R}^m$ *has a robust subset*.

We are going to find a robust set F for K by constructing it in pieces. In this section, we explain how to find sets F that work for points near the boundary of K. In the next section, we will show how to cover points that are far from the boundary of K.

The main step in the proof is the following lemma, which shows that appropriate sets F exist for balls that lie very near k-dimensional faces of K but far from $(k-1)$-dimensional faces of K. Because the proof seems quite technical, let us sketch an outline here. It proceeds by induction on the dimension. The main idea is to write points x that are "near" a k-dimensional face φ_k of K in the form $x = x' + x''$, where x' lies in the face φ_k, and x'' is the orthogonal complement of x to x'. The point x'' is small and lies in the polytope $K'' = K''(\varphi_k)$ in $\mathbb{R}^{m-k}$ obtained by projecting K onto the orthogonal complement of the flat containing φ_k. Apply the induction hypothesis to x'' in K'' and prove that, since x'' is small, one of the vertices of the $(m-k)$-simplex containing x'' is 0. Moreover, since x'' is small, the sum of the remaining $m-k$ coefficients in the convex combination for x'' is some small $\varepsilon > 0$. Then we write $x' = (1-\varepsilon)y'$, where y' is close to x', so we can apply the induction hypothesis to y' with respect to φ_k to represent it as a convex combination of $k+1$ points. This will give the desired decomposition of x as a convex combination of $m+1$ points.

Lemma 2.2 *Suppose that Theorem 1.1 holds for all polytopes in all dimensions smaller than m. Let φ_k be a face of K of dimension k, where $0 \le k < m$. Then for each $C \ge 1$, there exist $\delta_0 > 0$, $c > 0$, $C_1 > 0$, and a finite set $F = F(\varphi_k) \subset \partial K$ which contains the extreme points of K such that for each $0 < \delta \le \delta_0$ and each $x \in K$, the following three conditions* (*when $k = 0$, disregard* (ii))

(i) $\operatorname{dist}(x, \varphi_k) \le C \operatorname{dist}(x, \partial K)$,
(ii) $\operatorname{dist}(x, \partial\varphi_k) \ge C_1 \operatorname{dist}(x, \partial K)$, *and*
(iii) $B(x, \delta) \subset K$

imply the existence of $x_0, \dots, x_m \in F$ such that the simplex $S(x_0, \dots, x_m)$ generated by these vectors contains $B(x, c\delta)$.

Proof It suffices to prove the statement for $\delta = \operatorname{dist}(x, \partial K)$, since $x \in K$ and $B(x, \delta) \subset K$ if and only if $\operatorname{dist}(x, \partial K) \ge \delta$. We can assume that coordinates are chosen in such a way that the affine subspace of dimension k containing φ_k is $\mathbb{R}^k \times \{0\} = \{(y_1, \dots, y_k, 0, \dots, 0)\}$. Denote by $\pi : \mathbb{R}^m \to \mathbb{R}^{m-k}$ the projection mapping onto the $(m-k)$-dimensional subspace orthogonal to $\mathbb{R}^k \times \{0\}$,

$$\pi(y_1, \dots, y_m) = (y_{k+1}, \dots, y_m),$$

so that $\pi(K)$ is a convex polytope in $\mathbb{R}^{m-k}$. Let us also use the notation $y = (y', y'')$ for the first k and last $m-k$ components of the vector $y \in \mathbb{R}^m$.

First, we claim that if $x = (x', x'') \in K$ with $(x', 0) \notin \varphi_k$, then $\operatorname{dist}(x, \varphi_k) = \operatorname{dist}(x, \partial\varphi_k)$. That is, if the orthogonal projection of x onto the affine hull of φ_k does not lie in φ_k, then the nearest to x point of φ_k lies necessarily in the boundary of φ_k. To prove this, choose the unique point $(x_1', 0) \in \varphi_k$ such that $\operatorname{dist}(x, \varphi_k) = |x - (x_1', 0)|$. Then $(x_1', 0) \in \varphi_k$ and $(x', 0) \notin \varphi_k$, so there exists a point on the line segment joining these two points that lies in $\partial\varphi_k$. The points on this line segment are of the form

$$\big(z', 0\big) = \big(x', 0\big) + \lambda\big(x_1' - x', 0\big), \quad 0 \le \lambda \le 1,$$

and distance from x to each of these points is given by

$$\operatorname{dist}\big(x, (z', 0)\big)^2 = \lambda^2 \big|x_1' - x'\big|^2 + \big|x''\big|^2 \le \big|x_1' - x'\big|^2 + \big|x''\big|^2 = \operatorname{dist}(x, \varphi_k)^2,$$

which shows that $\operatorname{dist}(x, (z', 0)) \le \operatorname{dist}(x, (x_1', 0))$ and hence $\operatorname{dist}(x, \varphi_k) \le \operatorname{dist}(x, \partial\varphi_k)$. Of course, the other inequality is clear.

Assume that $C_1 > C$ and fix $x = (x', x'') \in K$ satisfying conditions (i)–(iii) of the lemma. Then

$$\operatorname{dist}(x, \partial\varphi_k) \ge \frac{C_1}{C} \operatorname{dist}(x, \varphi_k) > \operatorname{dist}(x, \varphi_k).$$

By what we proved in the previous paragraph, this implies that $(x', 0) \in \varphi_k$. In particular, the hypotheses of the lemma, together with the triangle inequality, then imply that

$$\delta = \operatorname{dist}(x, \partial K) \le \operatorname{dist}(x, \varphi_k) = \big|x''\big| \le C \operatorname{dist}(x, \partial K) = C\delta, \tag{1}$$

$$\operatorname{dist}(x, \partial\varphi_k) \ge \operatorname{dist}\big(\big(x', 0\big), \partial\varphi_k\big) \ge \operatorname{dist}(x, \partial\varphi_k) - \big|x''\big| \ge (C_1 - C)\delta = C_1'\delta, \tag{2}$$

where $C_1 > C$ is a constant that will be specified later.

For $y \in \partial K$, let $\mathscr{N}_y$ denote the cone of feasible directions of K at y, i.e., the set of all vectors N that are inward pointing normals to a hyperplane that supports K at the point y. Analytically, this is the condition $(x - y) \cdot N \ge 0$ for all $x \in K$. If y is a point in the relative interior of φ_k and $N \in \mathscr{N}_y$, then N must have the form

$$N = \big(0, N''\big).$$

Further, the cones $\mathscr{N}_y$ are identical as long as y ranges over the interior of φ_k and can be identified with the cone

$$\mathscr{N} = \mathscr{N}(\varphi_k) := \big\{N'' : \big(0, N''\big) \in \mathscr{N}_y\big\} \subset \mathbb{R}^{m-k}.$$

The convex set $\pi(K) \subset \mathbb{R}^{m-k}$ has the property that $0 \in \partial(\pi(K))$, 0 is an extreme point of $\pi(K)$, and the cone of all vectors that are inward-pointing normals to hyperplanes that support $\pi(K)$ at 0 is equal to $\mathscr{N}$. In fact, the extreme rays of the cone $\mathscr{N}$ have the direction of the vectors $\pi(\mathbf{e} - \sigma)$, where σ is an extreme point of φ_k, and $\mathbf{e}$ is an extreme point of K such that $[\sigma, \mathbf{e}]$ is an edge of K.

Let us first suppose that $k \ge 1$. Then both φ_k and $\pi(K)$ are convex polytopes that lie in Euclidean spaces of dimension $< m$. Therefore, there are robust subsets $G_1 \subset \partial\varphi_k$, $G_2 \subset \partial\pi(K)$ and associated positive constants $c_1 > 0$, $c_2 > 0$ such that the conditions of Definition 2.1 are satisfied. Let $F_1 = G_1$, and let $F_2 \subset \partial K$ be any set with the same number of points as G_2 and $\pi(F_2) = G_2$. We claim that the set $F = F_1 \cup F_2$ is a set with the properties of the lemma.

To prove this, recall that $\delta = \operatorname{dist}(x, \partial K)$ and x is a point satisfying the hypotheses of the lemma. Since $B(x, \delta) \subset K$, we see that

$$\pi\big(B(x, \delta)\big) = B\big(\pi(x), \delta\big) \subset \pi(K).$$

We can therefore find $m-k+1$ points $g_0, \dots, g_{m-k} \in G_2$ such that $B(\pi(x), c_2\delta) \subset S(g_0, \dots, g_{m-k})$, the $m-k$ simplex with vertices at $g_0, \dots, g_{m-k}$. The vectors $g_i - g_0$, $1 \le i \le m-k$, must be linearly independent since the simplex contains an open subset of $\mathbb{R}^{m-k}$. Also, by (1), $|x''| \le C\delta$ and $(x', 0) \in \varphi_k$ as pointed out earlier in the proof. Since G_2 is a finite set, if δ is small enough, the only way that we can have $\pi(x) = x'' \in S(g_0, \dots, g_{m-k})$ is for one of the g_i, say g_0, to be the extreme point 0 of $\pi(K)$. Therefore, if δ is small enough and $|y'' - x''| \le c_2\delta$, there are positive numbers $\mu_1, \dots, \mu_{m-k}$ such that $\sum \mu_i \le 1$ and

$$y'' = \sum_{i=1}^{m-k} \mu_i g_i.$$

In fact, the sum of the μ_i must be small. Because, the cone $\mathscr{N}$ contains at least $m-k$ linearly independent unit vectors $N''_1, \dots, N''_{m-k}$, taking the dot product of each of these vectors with the last equation gives a nonsingular system of $m-k$ linear equations in $m-k$ unknowns,

$$A\mu = B,$$

where the $(m-k) \times (m-k)$ matrix $A = [g_i \cdot N_j]$ is nonsingular, and the vector $B = [y'' \cdot N_j]$ has entries of magnitude at most

$$\left|y''\right| \le \left|x''\right| + c_2\delta \le (C + c_2)\delta,$$

where the last inequality is because $|x''| \le C\delta$ by (1). Consequently, there is a constant $C_2 = \|A^{-1}\|(C + c_2) > 0$ such that $\mu_i \le C_2\delta$.

Next make a small adjustment to the positive constants μ_i to write them in the form

$$\mu_i = \frac{\nu_i}{1 + \sum_{j=1}^{m-k} \nu_j}, \quad 1 \le i \le m-k,$$

which then gives us the representation

$$\Lambda y'' = \sum_{i=1}^{m-k} \nu_i g_i, \quad \text{where } \Lambda = 1 + \sum_{i=1}^{m-k} \nu_i.$$

Because the μ_i are small, there are always small ν_i that satisfy this condition, namely

$$\nu_i = \frac{\mu_i}{1 - \sum_{j=1}^{m-k} \mu_j}.$$

It is also clear that there is a constant $C_3 > 0$ such that $\nu_i \le C_3\delta$ when δ is small. In fact, we can take $C_3 = 2C_2$, provided that $\sum_j \mu_j \le 1/2$, which is the case if we make δ small enough.

Suppose now that a point $y \in B(x, c_2\delta)$. Let $f_1, \dots, f_{m-k}$ be points in ∂K with $\pi(f_i) = g_i$. Write $f_i = (h_i, g_i)$. With the coefficients ν_i chosen as above, we have that

$$y = \frac{1}{\Lambda}\left(\sum_{i=1}^{m-k} \nu_i f_i\right) + \frac{1}{\Lambda}\left(\Lambda y' - \sum_{i=1}^{m-k} \nu_i h_i, 0\right). \tag{3}$$

Because $|y - x| \le c_2\delta$, we then also have

$$\begin{aligned}
&\left|\left(\Lambda y' - \sum_{i=1}^{m-k} \nu_i h_i\right) - x'\right| \\
&= \left|(\Lambda - 1)y' + \left(y' - x'\right) - \sum_{i=1}^{m-k} \nu_i h_i\right| \\
&\le \left(\sum_i \nu_i\right)\left|y'\right| + \left|y' - x'\right| + \sum_i \nu_i |h_i| \le (m-k)C_3\delta\left|y'\right| + c_2\delta + C_3\delta \sum_i |h_i| \\
&\le C_4\delta
\end{aligned}$$

for a constant C_4 that depends on C_3 and the size of the largest vectors in K. Define $C_1 = C + (C_4/c_1)$. The last inequality shows that if $z' = \Lambda y' - \sum_{i=1}^{m-k} \nu_i h_i$, then $z' \in B(x', C_4\delta)$. Also, the condition that $\mathrm{dist}(x, \partial\varphi_k) \ge C_1\delta$ implies that

$$\mathrm{dist}\left(x', \partial\varphi_k\right) \ge \mathrm{dist}(x, \partial\varphi_k) - \left|x - \left(x', 0\right)\right| \ge C_1\delta - \left|x''\right| \ge (C_1 - C)\delta \ge \frac{C_4}{c_1}\delta.$$

Therefore, inside the face φ_k, we have $B(x', C_4\delta/c_1) \subset \varphi_k$, and, since the finite set G_1 is robust for the set φ_k with associated constant c_1, there are $k+1$ vectors $w_i = (w_i', 0)$, $0 \le i \le k$, from this set such that $B(x', C_4\delta) \subset S(w_0', \dots, w_k')$. In particular, we have that there exist $\lambda_0, \dots, \lambda_k$ such that $z' = \sum_{i=0}^{m-k} \lambda_i w_i'$ and $\sum_i \lambda_i = 1$. Consequently, from representation (3) we have shown that for all $y \in B(x, c_2\delta)$,

$$y = \frac{1}{\Lambda}\sum_{i=1}^{m-k} \nu_i f_i + \frac{1}{\Lambda}\left(z', 0\right) = \frac{1}{\Lambda}\left(\sum_{i=1}^{m-k} \nu_i f_i + \sum_{i=0}^{k} \lambda_i w_i\right).$$

In this last expression, the sum of the coefficients of all the vectors is equal to 1, since $\Lambda = 1 + \sum_i \nu_i$ and $\sum_i \lambda_i = 1$. Therefore, we have proved that every vector $y \in B(x, c_2\delta)$ lies inside the m-simplex with vertices at $w_0, \dots, w_k, f_1, \dots, f_{m-k}$, which completes the proof of the lemma for the case where $1 \le k \le m-1$.

It remains to treat the case $k = 0$; that is, the edge φ_k is a vertex of the polytope which we take to be equal to 0. In this case, the cone $\mathscr{N}$ is a full cone in $\mathbb{R}^m$ with vertex at the origin. It is no loss of generality to assume that the unit vector $(0, \dots, 0, 1)$ lies in the interior of $\mathscr{N}$ so that every $y = (y_1, \dots, y_m) \in \mathscr{N}^* \setminus \{0\}$, where $\mathscr{N}^*$ is the cone dual to $\mathscr{N}$, satisfies $y_m > 0$. If d is the minimum value of the last coordinate of an extreme point of K other than 0, then $d > 0$, and we have that

$$\mathscr{N}^* \cap \{y_m \le d\} = K \cap \{y_m \le d\}.$$

That is, the “corner of K” that lies near the extreme point 0 looks like the dual cone $\mathscr{N}^*$. If a ball $B(x,\delta) \subset K$ has center in $x_m \le d/2$ and if $\delta \le d/2$, then the central projection mapping from the origin to the hyperplane, $y_m = d$, i.e.,

$$\pi(y) = \frac{dy}{y_m},$$

carries the ball onto a subset of $K_1 = K \cap \{y_m = d\}$, a polytope that lies in a copy of $\mathbb{R}^{m-1}$. Note that the boundary of K_1 (relative to the hyperplane $y_m = d$) is a subset of the boundary of K. So, if we pick a robust subset of K_1, say G, then it will also be a subset of ∂K. Also, there will exist a constant $c > 0$ and m points of G such that $B(\pi(x), cd\delta/x_m) \subset S$ for some $(m-1)$-simplex with vertices in G. If we then look at the m-simplex with the origin added as a vertex, then it will contain the ball $B(x, c\delta)$ which is what we had to show. This completes the proof. □

Corollary 2.3 *Suppose that Theorem 1.1 holds for compact sets in $\mathbb{R}^k$ with $k < m$. Let K be a polytope in $\mathbb{R}^m$. Then there exist constants $c > 0$, $\delta_0 > 0$ and a finite set $F \subset \partial K$ such that for every $x \in K$ and $0 < \delta \le \delta_0$ such that $B(x,\delta) \subset K$ and* $\operatorname{dist}(x, \partial K) \le \delta_0$, *there exist $m+1$ points of F such that $B(x, c\delta)$ is a subset of the m-simplex with vertices at these points.*

Proof We apply Lemma 2.2 m times. Let K^k denote the k-skeleton of K, i.e., the union of all k-dimensional faces of K. If $x \in K$ and $\operatorname{dist}(x, \partial K) = \delta$, then there is an $(m-1)$-dimensional face $\varphi_{m-1} \in K^{m-1}$ such that $\operatorname{dist}(x, \partial K) = \operatorname{dist}(x, \varphi_{m-1})$. Then condition (i) of Lemma 2.2 is satisfied with $C = 1$. Therefore, from the case $k = m-1$ of that lemma, there exist constants $\delta_1 > 0$, $c_1 > 0$, $C_1 > 0$ and a finite subset $F_1 \subset \partial K$ such that $B(x, c_1\delta) \subset S(f_0, \dots, f_m)$ whenever $\delta \le \delta_1$, $d(x, \partial K) = \delta$ and

$$\operatorname{dist}\bigl(x, K^{m-2}\bigr) \ge C_1 \operatorname{dist}(x, \partial K) = C_1 \operatorname{dist}\bigl(x, K^{m-1}\bigr).$$

If x is near the boundary of K but fails to satisfy this latter condition, then

$$\delta = \operatorname{dist}(x, \partial K) \le \operatorname{dist}\bigl(x, K^{m-2}\bigr) \le C_1 \operatorname{dist}(x, \partial K).$$

Then apply the previous lemma with $k = m-2$ to points $x \in K$. It implies that there exist a finite set $F_2 \subset \partial K$ and constants $\delta_2 > 0$, $c_2 > 0$, $C_2 > 0$ such that whenever $x \in K$ satisfies the last condition and in addition

$$\operatorname{dist}\bigl(x, K^{m-3}\bigr) \ge C_2 \operatorname{dist}(x, \partial K),$$

then there exist $m+1$ points $f_0, \dots, f_m$ from the set F_2 such that $B(x, c_2, \delta)$ is a subset of the m-simplex with these vertices. That is, the condition of lemma is satisfied for all points near the $m-2$ skeleton of K but not too near the $m-3$ skeleton of K. The remaining points near the boundary of K therefore satisfy

$$\operatorname{dist}(x, \partial K) \le \operatorname{dist}\bigl(x, K^{m-3}\bigr) \le C_2 \operatorname{dist}(x, \partial K).$$

Continuing in this way, we find finite subsets $F_1, \dots, F_m$ of ∂K and positive numbers $\delta_1, \dots, \delta_m$, $c_1, \dots, c_m$, and $C_1, \dots, C_m$ such that every $x \in \partial K$ with $\operatorname{dist}(x, \partial K) = \delta \le \min\{\delta_1, \dots, \delta_m\}$ satisfies one of the pairs of inequalities

$$\begin{aligned} &\delta = \operatorname{dist}(x, \partial K) \le \operatorname{dist}\bigl(x, K^{m-i}\bigr) \le C_{i-1} \operatorname{dist}(x, \partial K), \\ &\operatorname{dist}\bigl(x, K^{m-(i+1)}\bigr) \ge C_i \operatorname{dist}(x, \partial K) \end{aligned}$$

for $i = 1, \dots, m$, except when $i = m$, the last inequality is vacuous. Thus, $F = F_1 \cup \dots \cup F_m$, and $c = \min\{c_1, \dots, c_m\}$ is a finite set with the required properties. □

3 Points Far from the Boundary of *K*

We first show that the only obstruction to handling the points of K that are far from the boundary of K is the "interior of the faces" obstruction.

Lemma 3.1 *Let K be a polytope in $\mathbb{R}^m$, and F a finite subset of ∂K such that*

$$\operatorname{int}(K) = \bigcup \bigl\{\operatorname{int}(S) : S \textit{ is an } m\textit{-simplex with vertices from } F\bigr\}.$$

Then for each positive number $\eta > 0$, there exists a number $c > 0$ such that whenever $x \in K$ satisfies $B(x, \delta) \subset K$ and $\operatorname{dist}(x, \partial K) \ge \eta$, there exist $m+1$ points of F such that the simplex S with vertices at these points satisfies $B(x, c\delta) \subset S$.

Proof This is a compactness argument. The set K_η of all the points of K of distance $\ge \eta$ from the boundary of K is a compact subset of $\operatorname{int}(K)$. By hypothesis, the interior of K is the union of the interiors of all the m-simplices with vertices in F. Since the interior of each such simplex S is the union of the open sets S_n of points in S of distance at least $1/n$ from the boundary of S, the compact set K_η is also covered by finitely many of these sets S_n. Therefore, there exists an integer N such that each $x \in K_\eta$ lies in a point of one of the simplices and of distance at least $4/N$ from its boundary. Then the ball about x of radius $1/N$ lies within S. If we choose $c = 1/(N \operatorname{diam}(K))$, then whenever $B(x, \delta) \subset K$, we have $\delta \le \operatorname{diam}(K)$, so that $c\delta \le 1/N$, and thus c has the required property when $x \in K_\eta$. □

We also have to prove that there are sets F satisfying the hypotheses. That one cannot take F to be the set of extreme points of K was pointed out in the introduction.

Lemma 3.2 *Suppose that K is a polytope in $\mathbb{R}^m$. Then there is a finite set $F \subset \partial K$ such that*

$$\mathrm{int}(K) = \bigcup \big\{\mathrm{int}(S) : S \textit{ is an } m\textit{-simplex with vertices from } F\big\}.$$

Proof We prove the lemma by induction on m. For $m = 1$, it is obvious. So let us assume that the result has been proved in all dimensions smaller than m. Then each facet ψ (i.e., $(m-1)$-dimensional face of F) contains a finite subset F_ψ of its boundary such that every point in the interior of the face lies in the interior of an $(m-1)$-simplex with vertices from F_ψ. Consequently, if we can find a subset F^1 of ∂K such that given any $x \in \mathrm{int}(K)$, there exists a point $f \in F^1$ and a facet ψ of K such that the ray from f through x intersects the boundary of K at a point $x^* \in \mathrm{int}(\psi)$, then taking the vertices of the $(m-1)$-simplex from the set F_ψ in the facet ψ that contains x^* together with the point f gives us an m-simplex that contains x in its interior. Thus, the union of F^1 with all the sets F_ψ for all facets ψ of K gives a set F with the properties of the lemma.

To find such a set F^1, associate to each $(m-2)$-dimensional face φ of K the flat $\mathscr{L}(\varphi)$; that is, the unique $(m-2)$-dimensional affine subspace of $\mathbb{R}^m$ that contains φ. Then for $f \in \partial K \setminus \mathscr{L}(\varphi)$, let $H(f, \varphi)$ be the unique hyperplane that contains both $\mathscr{L}(\varphi)$ and f. For later reference, note that an equation for this hyperplane can be written in the following way. Choose any $m - 1 = (m-2) + 1$ points from φ_{m-2}, say $y_1, \ldots, y_{m-1}$, such that the vectors $y_j - y_1$, $j = 2, \ldots, m-1$, are linearly independent. The equation is

$$l(x) := l(x, f, \varphi) = \det \begin{bmatrix} x_1 & \ldots & x_m & 1 \\ f_1 & \ldots & f_m & 1 \\ & y_1 & & 1 \\ \vdots & \vdots & \vdots & 1 \\ & y_{m-1} & & 1 \end{bmatrix} = 0. \tag{4}$$

In particular, this shows that $x \in H(f, \varphi)$ if and only if $f \in H(x, \varphi)$.

Let $\mathscr{L} = \bigcup_\varphi \mathscr{L}(\varphi)$ denote the union of the finitely many $(m-2)$-dimensional subspaces $\mathscr{L}(\varphi)$. $\mathscr{L}$ cannot cover any $(m-1)$-dimensional face of K because its dimension is too small. Further define

$$V_f := \bigcup_\varphi H(f, \varphi), \quad f \in \partial K \setminus \mathscr{L},$$

to be the union of these hyperplanes over the $(m-2)$-dimensional faces φ of K. Then V_f is an algebraic variety in $\mathbb{R}^m \subset \mathbb{C}^m$—it is the zero set of the polynomial $P_f(x)$ we get by multiplying together the affine functions (of x) given in (4) that define the hyperplanes. We claim that

$$\bigcap_{f \in \partial K \setminus \mathscr{L}} V_f = \mathscr{L}. \tag{5}$$

If x is a fixed point that does not belong to $\mathscr{L}$ and if φ is an $(m-2)$-dimensional edge, then the set of f that lie on the hyperplane $H(x,\varphi)$ given by (4) cannot contain the entire boundary of K. In fact, unless $x \notin \mathscr{L}$ lies on one of the finitely many hyperplanes that contain an $(m-1)$-dimensional face of K, the intersection of $H(x,\varphi)$ and ∂K is a nowhere dense subset of ∂K since this intersection is at most an $(m-2)$-dimensional subset of ∂K. The union $V_x := \bigcup_\varphi H(x,\varphi)$ over the finitely many $(m-2)$-dimensional faces φ of K then also meets ∂K in a nowhere dense set. Therefore, there exists a point $f \in \partial K \setminus V_x$, so $f \notin V_x$ or, equivalently, $x \notin V_f$. In fact, if x is in the interior of K, this is the case except for f in a nowhere dense subset of ∂K.

Consequently, we have the equality of algebraic varieties (5), and so there must exist finitely many of the values of f, say $f_1, \ldots, f_q$, so that

$$\bigcap_{i=1}^{q} V_{f_i} = \mathscr{L}. \tag{6}$$

We can take for F^1 this finite set of f_i. To prove that it has the required property, choose $x \in \mathrm{int}(K)$. Since $x \notin \bigcap V_{f_i}$, there exists $f = f_i$ such that $x \notin V_{f_i}$. This implies that the ray from f_i through x must intersect ∂K in the interior of an $(m-1)$-dimensional face. Otherwise, because the boundary of the facets is covered by the $(m-2)$-dimensional faces φ, we would have $x \in H(f_i,\varphi)$ for some φ, contrary to the fact that $x \notin V_{f_i}$. Therefore, the ray meets ∂K in an interior point of some facet, and, as noted in the first paragraph of the proof, this completes the argument. □

4 Proof of the Theorem and Concluding Comments

Proof of Theorem 1.1 We are proving the theorem by induction on the dimension m. For $m = 1$, it is obvious, so assume that $K \subset \mathbb{R}^m$ and that the theorem has been proved for smaller values of m. Let $F_1 \subset \partial K$ be a finite set with the property from Corollary 2.3. Let $F_2 \subset \partial K$ be a finite set with the property of Lemma 3.2. Let $F = F_1 \cup F_2$. Then F is a robust subset of K, with the points x so that the balls $B(x,\delta)$ with x near the boundary of K being subsets of simplices with vertices from F_1 and those x away from the boundary by Lemma 3.1. This completes the proof. □

It is also useful to have the version of the theorem that applies to cones in $\mathbb{R}^m$.

Theorem 4.1 *Let $\mathscr{C}$ be a proper convex cone in $\mathbb{R}^m$ that is spanned by finitely many extreme rays. Then there are constants $c > 0$, $C_1 > 0$ and a finite set $\mathscr{N}$ of unit vectors so that the rays in the direction of $N \in \mathscr{N}$ lie in the boundary of $\mathscr{C}$ such that: if $x \in \mathscr{C}$ is at a positive distance from the boundary of $\mathscr{C}$, there exist m of the* *vectors* $N_1, \ldots, N_m \in \mathcal{N}$ *such that*

$$x = a_1 N_1 + \cdots + a_m N_m$$

and

$$a_i \geq c\,\mathrm{dist}(x, \partial\mathscr{C}), \quad 1 \leq i \leq m.$$

Proof It is no loss of generality to assume that the unit vector $(0, \ldots, 0, 1)$ lies in the interior of $\mathscr{C}$. Then set $K = \mathscr{C} \cap \{x_m = 1\}$, a compact convex polytope in a copy of $\mathbb{R}^{m-1}$. Therefore, there is robust subset F of ∂K and a constant $c > 0$ such that whenever $B(x, \delta) \subset K$, we can choose finitely many vectors $f_1, \ldots, f_m$ from F such that $x = \lambda_1 f_1 + \cdots + \lambda_m f_m$ and $\lambda_i \geq c\delta$ for $1 \leq i \leq m$. Now there is a constant $c_2 > 0$ such that if $x \in \mathscr{C}$ is of distance C from the boundary of $\mathscr{C}$, then $x/x_m \in K$ is of distance at least $c_2 C/x_m$ from the boundary of K. Consequently, we can write

$$x/x_m = \lambda_1 f_1 + \cdots + \lambda_m f_m, \quad \lambda_i \geq cc_2 C/x_m,$$

which clearly implies the representation of the theorem with

$$a_i = x_m \lambda_i \geq cc_2\,\mathrm{dist}(x, \partial\mathscr{C}).$$

This completes the proof. □

Remark It is also true that the version of the theorem for cones implies the one for convex polytopes.

Concluding Remarks It seems clear that the arguments we have made here are very far from optimal. A robust set F necessarily contains all the extreme points of K, but it is not clear how many extra points one has to add. Our argument goes by induction, so one can show that the (large) number of points added can be chosen to depend only on the dimension m and the number p of extreme points. It is also not clear if there is interplay between the size of the constant c and the number of additional boundary points chosen to be in F. We list some of these questions that seem natural to us.

A. Are quadrilaterals the only convex polygons in the plane for which the quantitative Carathéodory theorem fails with the set F chosen to be the set of extreme points? For planar convex polygons, if the set of extreme points is augmented by one additional boundary point, is it always a robust set?
B. Is the only obstruction to the quantitative Carathéodory theorem the interior point obstruction? Namely, if the interior of K is equal to the union of the interiors of all the simplices with vertices chosen from the extreme points of K, is the set of extreme points then robust?
C. Is there some simple explicit bound for the number of points needed in the set F? Does it depend on more than the dimension m?
D. Can one say more about the size of the constant c? Is it related to the number of points in the set F? For example, could one always choose $c = 1 - \varepsilon$ by making F contain a large number $N(\varepsilon)$ of points?

Acknowledgement We wish to thank Alexander Barvinok for several helpful conversations about this work.

References

1. Braun, R., Meise, R., Taylor, B.A.: The geometry of analytic varieties satisfying the local Phragmén–Lindelöf condition and a geometric characterization of the partial differential operators that are surjective on $\mathscr{A}^{\omega}(\mathbb{R}^4)$. Trans. Am. Math. Soc. **356**, 1315–1383 (2004)
2. Hiriart-Urruty, J.-B., Lemarérchal, C.: Fundamentals of Convex Analysis. Springer, Berlin (2001)

Geometric Path Integrals. A Language for Multiscale Biology and Systems Robustness

Domenico Napoletani, Emanuel Petricoin, and Daniele C. Struppa

Abstract In this paper we suggest that, under suitable conditions, supervised learning can provide the basis to formulate at the microscopic level quantitative questions on the phenotype structure of multicellular organisms. The problem of explaining the robustness of the phenotype structure is rephrased as a real geometrical problem on a fixed domain. We further suggest a generalization of path integrals that reduces the problem of deciding whether a given molecular network can generate specific phenotypes to a numerical property of a robustness function with complex output, for which we give heuristic justification. Finally, we use our formalism to interpret a pointedly quantitative developmental biology problem on the allowed number of pairs of legs in centipedes.

1 Introduction

Leon Ehrenpreis was a singular mathematician. Not only he had a gift and a vision for a deep understanding of mathematics, but he had a passion for the construction of overarching approaches that would allow a general comprehension of vast areas of mathematics. This passion is embodied in his two masterpieces, Fourier Analysis in Several Complex Variables [7] and The Universality of the Radon Transform

This paper is dedicated to the memory of Leon Ehrenpreis 1930–2010.

D. Napoletani (✉) · E. Petricoin
Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA 20110, USA
e-mail: dnapolet@gmail.com

E. Petricoin
e-mail: epetrico@gmu.edu

D. Napoletani · D.C. Struppa
Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA

D.C. Struppa
e-mail: struppa@chapman.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_16, © Springer-Verlag Italia 2012

[8], but is also apparent in the many papers he published, for example, on wide generalizations of the Edge-of-the-Wedge theorem.

Reading Ehrenpreis' works, we are reminded of a beautiful phrase that Kawai, Kashiwara, and Kimura insert in [16], just after the proof of a fundamental theorem on the propagation of solutions of microdifferential equations:

> It is like a view from a mountain peak.

This aspiration to a global view of mathematics that would offer insights, even in advance of a fully realized technical description of that view, is part of what made Ehrenpreis' work uniquely captivating and uniquely fertile. Many mathematicians, of at least three generations, have worked to understand, formalize, explain, refute, demonstrate statements in Ehrenpreis' work. It is for this reason, that Ehrenpreis' student, C.A. Berenstein, once noted in his review [4] of The Universality of the Radon Transform [8] that it is

> a book that is worth studying, although mining may be a more appropriate word, as the reader may find the clues to the keys he is searching for to open up subjects that are seemingly unrelated to this book. Thus, one finds at the end that the title is justified.

It is in this spirit, but with a deep sense of humility and with a full awareness of our limitations, that we would like to present, in this paper, a proposal, a strategy for a way to mathematically understand and describe one of the fundamental problems (the fundamental problem?) of modern biology: how can we understand macroscopic biological traits from our knowledge of molecular level information.

The proposal we make is inspired by the instrument of path integrals, which is probably the most enduring legacy of Richard Feynman [9], and for which we suggest here a tentative generalization to provide a plausible tool for the description of the macroscopic properties of a biological system.

Our specific point of departure lies in an important, and somewhat surprising, fact. In [19], Chap. 9, it is reported that, out of all known species of centipedes, there are about 1000 species with 15 pairs of legs, none with 17 or 19 pairs, several with 21 and 23 pairs, and a few distributed over a very large range from 27 to 191. No centipedes have an even number of pairs of legs, and some species have a stable interspecies number of pairs of legs, while some others display a variability of the number of legs among individuals.

Can we explain in any quantitative way this striking pattern of gaps with respect to the dynamics at fine molecular scale? And how do we express the remarkable robustness of the resulting phenotypes? To be more specific, we state the following:

Problem 1 (Centipedes Segmentation (CS) Problem) Show that it is impossible to have centipedes with an even number of pairs of legs, or with 17 or 19 pairs of legs.

Clearly if we can suitably quantify this problem, we will be able to generalize the question to the full gap structure that we have described in the previous paragraph. Moreover, we take the CS problem simply as emblematic of the variety of developmental biology problems in which we observe strong constraints on the phenotype

([19], p. 86), without having a conceptual frame through which to approach these problems.

In general, molecular biology approaches to developmental biology problems such as the CS problem require the construction of an appropriate map that can relate microscopic variables with macroscopic outputs. But, since a mechanistic way to relate these variables is often absent, and it is indeed problematic even to quantify phenotype properties [12], it is challenging to express these questions in a proper mathematical setting. Even though we are aware of the dangers of suggesting a general frame that is not fully developed, we still believe (in view of the difficulty of the problems at hand) that it is worthwhile to attempt to set up a language and a set of techniques through which these problems might be approached.

Ideally, we would need to study entire sets of models at once, since we would expect the micro/macro maps to be stable under wide variations of the parameters in the model of the microscopic dynamics. At the same time, such tremendous model variability should be understandable in a compact, possibly analytical setting to have any hope of providing computationally feasible answers.

In this paper we suggest that supervised learning and the resulting classification functions [11] can provide the basis to formulate at the microscopic level questions on the phenotype, such as the centipedes segmentation problem, provided that the classification function satisfies some suitable growth conditions. The problem of phenotype robustness can then be rephrased as a problem of real geometry on a fixed domain. General methods to solve such problems are still in their infancy [5], and we propose a generalization of path integrals that allows us to reduce the problem of class belonging and of phenotype robustness to specific questions on functions with complex output. Finally, we will show how to reduce the CS problem to a problem on the global properties of these functions. Other problems about the restricted variability of phenotypes in developmental biology could be formulated in similar ways.

2 A Geometrical Robustness Condition

Let $X = (X_1, \dots, X_N)$ be the activation level of a set of proteins (genes, metabolites, or combinations of) at some time t_0 and assume that we have access to the derivatives $\dot{X} = (\dot{X}_1, \dots, \dot{X}_N)$ of those levels at the same time t_0 (in practice this means that we measure the proteins at two very close time points); and suppose that the biological samples from which $(X, \dot{X})$ is measured can be classified in a set of M classes $C_1, \dots, C_M$.

We assume that the protein measurements are taken at the embryo stage of development of the individuals in the CS problem. Clearly these measurements will be some sort of average of the activity levels of several cells [6], even though single cell measurements can be envisioned [13]. We further assume that the underlying network of interactions is stable (i.e., no essential parameter variation that changes significantly the dynamics) over a short time range. Without excluding in principle other possible state variables, we focus our attention on protein networks as these are believed to be evolutionary more stable [22].

Call $B(X, \dot{X})$ an instance of the biological class associated to $(X, \dot{X})$ (in the case of the CS problem, this will be the number of pairs of legs of an adult individual centipede). We assume that we have access to a training set of instances for each class, so that we can build a special type of classifier that has the following structure:

Definition 1 (Interval Classifier) A function $F(X, \dot{X})$ is an interval classifier for the classification problem with classes S_m, $m = 1, \ldots, M$, if it satisfies $m - 1 < F(X, \dot{X}) < m$ when $B(X, \dot{X}) \in C_m$, $m = 1, \ldots, M$.

Though not strictly necessary at this point, it is useful for our subsequent analysis of Propositions 1 and 2 to require that F is bounded at infinity and analytic, and therefore we assume for simplicity that the classifier function F is a neural network with exponential sigmoidal activation functions ([14], Chap. 10).

Remark 1 At a fundamental level, it is not necessarily possible to identify a subset of molecular variables that is indeed predictive for the phenotype characteristic that we are interested in. This problem is not exclusive to our setting, but it is a major difficulty in all approaches that try to bridge molecular biology with the study of phenotype characteristics. The very existence of an accurate classifier F depends on the identification of such variables. Note that our setting requires to estimate not only state variables, but their derivatives as well. In classical rational mechanics, state variables and their first derivative are sufficient to characterize a system for all future times (in a variational, Lagrangian setting [1]). While first derivatives are also sufficient for our formal analysis of phenotype classification problems, it would be interesting to understand how many derivatives are truly necessary to have effective classifiers for these types of problems.

Note that the way we define the multiclass classifier F is not the standard one, in which an M-class problem is usually approached by having a vector of M output classification functions ([14], p. 331). For reasons that will be clear when we reinterpret analytically the CS problem in Problem 2 of the last section, we not only use a single function F for the multiclass problem, but we also require that all instances belonging to a certain class must be within a given interval. This superimposes a stronger metrical structure on the classification problem.

Remark 2 In the context of neural networks, the request of interval classifier for training instances within each class transforms the unconstrained optimization problem usually associated to finding the classifier function ([14], p. 335) into a constrained optimization problem.

Take now a slack variable Y and consider the function $F(X, \dot{X}) - Y$, with $(X, \dot{X}) \in D \times \dot{D}$, where D is the set of biologically meaningful conditions for X, and $\dot{D}$ is the set of biologically meaningful conditions for $\dot{X}$. Then if $B(X, \dot{X}) \in C_1$, there exists Y with $0 < Y < 1$ such that $F(X, \dot{X}) - Y = 0$, so that the condition $B(X, \dot{X}) \in C_1$ can be rewritten as $F(X, \dot{X}) - Y = 0$, $(X, \dot{X}) \in D \times \dot{D}$, $0 < Y < 1$. Similarly, $B(X, \dot{X}) \in C_m$ becomes $F(X, \dot{X}) - Y = 0$, $(X, \dot{X}) \in D \times \dot{D}$, $m - 1 < Y < m$.

It is reasonable to suppose that X is in fact a state variable of an ordinary differential equation (ODE) network $\dot{x} = f(x, a_0)$, $x = (x_1, \ldots, x_N)$, $f = (f_1, \ldots, f_N)$, f a vector of polynomials in x, modeling ODEs with polynomials or power functions has proven itself to be very flexible for systems of molecular reactions [21]. We further ask that $f(x, a)$ is an analytic function in a. As the condition of analytic structure of the classifier F itself, the analyticity of $f(x, a)$ in a will be important in the justification of Proposition 2.

A network of biological significance will usually depend on a large number of parameters that will depend on the environment where the variable of the network actually act and live [10]; this is the reason we allow a dependence from the parameter vector a in the ODE network. Write the dependence of $\dot{x}$ from $f(x, a)$ explicitly in $F(X, \dot{X}) - Y = 0$, i.e., $F(X, f(X, a)) - Y = 0$. The condition of class belonging can be written as:

Definition 2 (Network Classification) A network $\dot{x} = f(x, a_0)$ generates phenotypes belonging to class C_m if

$$\exists X \in D, \quad m - 1 < Y < m : F\big(X, f(X, a_0)\big) - Y = 0. \tag{1}$$

Equation (1) is a condition for the network $\dot{x} = f(x, a_0)$ to give rise to states that belong to one of the classes we are considering. Note that the domain D of X constrains the domain $\dot{D}$ through the relation $\dot{D} = f(D)$.

The macroscopic phenotypic states of an organism are believed to be robust under wide ranging changes of parameters [10]. Therefore, for a realistic network, (1) should be satisfied for all parameters in a region A, where A is some sizable neighborhood around a nominal value $a = a_0$ of the parameter a. In other words:

Definition 3 (Class Robustness) A phenotype class C_m is robust if the zeros of the function $F(X, f(X, a)) - Y = 0$ are persistent in a region A of parameters, i.e.,

$$\forall a \in A, \exists X \in D, \quad m - 1 < Y < m : F\big(X, f(X, a)\big) - Y = 0. \tag{2}$$

Remark 3 We assume that, for each a, $\dot{x} = f(x, a)$ is capable of generating $(x, \dot{x})$ belonging to a single class C_m, to avoid, in the CS problem, the paradoxical situation in which the predicted number of pairs of legs can change in a given centipede with time. We assume instead that the embryo is committed to its specific segmentation within a large time frame where we could measure our state variables.

3 Stable Zeros and Path Integrals

In the previous section we described condition (2) that must be satisfied if $\dot{x} = f(x, a)$ is to generate robustly states that belong to a class C_m. The problem with this condition is that it requires identification of zeros of a (nonalgebraic) function over a real domain, and moreover it requires us to establish that these zeros are stable under a wide variation of parameters. This is problematic as it is difficult to establish the existence of solutions of real equations on domains, even in the algebraic case [3].

In this section we show that a generalized path integral [9] can be built in such a way that a specific condition on this integral corresponds to the verification or falsification of (2) over a domain. Path integrals have the remarkable property of giving information, in a single analytical object, about global, collective properties of physical systems, a point of view especially stressed in condensed matter field theory literature [2], and it is this ability that we will try to mirror in the setting of network analysis. We start by building a path integral that is related to (1). Essentially, we will build a domain G and a function L such that if (1) is satisfied, then there is at least a path connecting two points in G. This path (and a small tubular neighborhood thereof, with squeezed endpoints) will dominate the path integral that we are building, and it will allow us to make qualitative conclusions on the value of the integral when (1) is verified. We mirror then this analysis for (2). We first go through the technical building of the path integral, before we explain its heuristic interpretation.

The condition $x = (x_1, \ldots, x_N) \in D$ in (1) can be explicitly written as a condition on each variable, i.e., $d_{nb} < x_n < d_{nt}$, where d_{nb} and d_{nt} are lower and upper bounds on the biologically meaningful values that variable x_n can assume; in principle these values can be measured over repeated in vitro experiments. We can always change variables $x_n \to \tilde{x}_n$ so that $-1 < \tilde{x}_n < 1$. This is accomplished by setting

$$\tilde{x}_n = \frac{2}{d_{nt} - d_{nb}}(x_n - d_{nb}) - 1.$$

Similarly, we can force $-1 < \tilde{y} < 1$ by setting $\tilde{y} = 2(y - (m-1)) - 1$. These are invertible linear transforms, so we can write

$$\exists \tilde{x}, \tilde{y} : \tilde{F}\big(\tilde{x}, \tilde{f}(\tilde{x}, a)\big) - \left(\frac{\tilde{y} - 1}{2} + (m-1)\right) = 0,$$
$$-1 < \tilde{x}_n < 1, \; -1 < \tilde{y} < 1, \; a = a_0 \tag{3}$$

for some transformed functions $\tilde{F}$, $\tilde{f} = (\tilde{f}_1, \ldots, \tilde{f}_N)$ which are obtained from F and f by replacing x with $\tilde{x}$ and y with $\tilde{y}$. We further simplify notation by defining the analytical function

$$H(\tilde{x}, \tilde{y}, a, m) = \tilde{F}\big(\tilde{x}, \tilde{f}(\tilde{x}, a)\big) - \left(\frac{\tilde{y} - 1}{2} + (m-1)\right). \tag{4}$$

So we can rewrite condition (3) as

$$\exists \tilde{x}, \tilde{y} : H(\tilde{x}, \tilde{y}, a, m)^2 = 0, \quad \text{on } -1 < \tilde{x}_n < 1,\ -1 < \tilde{y} < 1,\ a = a_0. \tag{5}$$

We square the function H for purposes that will be clear in the following (see (7) and its justification). The domain restriction on $\tilde{x}_n$ and $\tilde{y}$ can also be written as $(\tilde{x}, \tilde{y}) \in \mathscr{S}_1 \times \cdots \times \mathscr{S}_1$, where $\mathscr{S}_1$ is the unit interval $[-1, 1]$, and we take the cartesian product $N + 1$ times. We now introduce a spherical extension of this domain in such a way that on every section of the extended sphere we can formulate a condition similar to (5).

To build the spherical extension, first suppose we work with a single variable $\tilde{x}_n$, keeping all other variables constant. We embed each point $\tilde{x}_n$ in the disk $\mathscr{S}_1$ in the space R^2 with the map $\tilde{x}_n \to (\tilde{x}_n, 0)$. We want then a basic way to map points $(\tilde{x}_n, 0)$ on $(\mathscr{S}_1, 0)$ to points $(\tilde{x}_{nz}, z)$ in the slices $(*, z)$ for z in $-1 < z < 1$, and moreover we want the whole set $(\mathscr{S}_1, 0)$ to be mapped to the points $(-1, 0)$, $(1, 0)$ in the limit of $z \to \pm 1$.

One way to achieve this embedding is through maps $\tilde{x}_{nz} = \tilde{x}_n \sqrt{1 - z^2}$. Conversely, any point $(\tilde{x}_{nz}, z)$, in $\mathscr{D}_n = \{-1 < z < 1, -\sqrt{1 - z^2} < \tilde{x}_{nz} < \sqrt{1 - z^2}\}$ can be mapped to a point in $\mathscr{S}_1$ by setting $\tilde{x}_n = \tilde{x}_{nz} \frac{1}{\sqrt{1-z^2}}$. If we do a similar mapping for all $\tilde{x}_n$, and for $\tilde{y}$ as well, the function $H(\tilde{x}, \tilde{y}, a, m)^2$ can be extended to the following function of variables $(\tilde{x}_z, \tilde{y}_z, z, a, m)$:

$$L(\tilde{x}_z, \tilde{y}_z, z, a, m) = H\left(\tilde{x}_z \frac{1}{\sqrt{1 - z^2}}, \tilde{y}_z \frac{1}{\sqrt{1 - z^2}}, a, m\right)^2, \tag{6}$$

where $\tilde{x}_z = (\tilde{x}_{1z}, \ldots, \tilde{x}_{Nz})$, and $\tilde{Y}_z \in \mathscr{D}_y$ with $\mathscr{D}_y = \{-1 < z < 1, -\sqrt{1 - z^2} < \tilde{y}_z < \sqrt{1 - z^2}\}$. The same value of z is used to define all components of $\tilde{x}_z$ and $\tilde{y}_z$, so we can define the domain of all points in the spherical extension as

$$\mathscr{D} = \big\{(\tilde{x}_z, \tilde{y}_z, z), -1 < z < 1, -\sqrt{1 - z^2)} < \tilde{x}_{nz} < \sqrt{1 - z^2},$$
$$-\sqrt{1 - z^2} < \tilde{y}_z < \sqrt{1 - z^2}\big\}.$$

We are now ready to introduce the generalization of the path integral that we aimed for:

Definition 4 (Geometric Path Integral) Let $\gamma(t)$, $0 \le t \le 1$, be a path in $\mathscr{D}$ with $\gamma(0) = (0, 0, -1)$ and $\gamma(1) = (0, 0, 1)$. Let Γ be the set of all such paths, and let $D\gamma$ be a suitable measure on Γ. We define the geometric path integral associated to condition (5), and dependent on a parameter $h > 0$, as

$$P(a, m, h) = \int_{\gamma \in \Gamma} e^{\frac{i}{h} \int_0^1 L(\gamma(t), a, m)\, dt}\, D\gamma. \tag{7}$$

The components of $|\tilde{x}_z/\sqrt{1-z^2}|$ and $|\tilde{y}_z/\sqrt{1-z^2}|$ are all bounded by 1 in $\mathscr{D}$; moreover, with respect to $\tilde{x}_z/\sqrt{1-z^2}$, L is essentially the square of (a linear transform of) a neural network classifier F with exponential sigmoidal activation functions ([14], p. 225), while, with respect to $\tilde{y}_z/\sqrt{1-z^2}$, L is the square of a linear function. This implies that L is bounded and continuous on all paths in Γ, except possibly at the endpoints of the paths themselves, where it may be only bounded, and the integral of L on each path is well defined.

Remark 4 The choice of the appropriate measure $D\gamma$ that ensures the convergence of path integrals is very delicate, and it will require further investigation in the context of geometric path integrals for appropriate classes of integrands L. Note, however, that the geometrical path integral is defined with respect to paths in a compact set, and this is a scenario where the standard path integrals are amenable to rigorous convergence results [15].

In order to understand the motivation for the integral in (7), we consider the case in which L (and therefore H) has a set of zeros in $\mathscr{D} \bigcap \{(\tilde{x}_z, \tilde{y}_z, z) : z = 0\}$; specifically we can assume that $L(\tilde{x}_0, \tilde{y}_0, 0, a, m) = 0$. We can then build a full path γ in $\mathscr{D}$ such that $L(\gamma(t), a, m) = 0$ for every t in $[0, 1]$, just by taking suitable mappings of $(\tilde{x}_0, \tilde{y}_0)$ in $\mathscr{D}$ for all values of $-1 < z < 1$.

Now, following standard heuristic arguments for semi-classical approximations of path integrals ([2, 9], Chap. 3, [20]), we expect the following result to hold. We denote by $\Re f$ and $\Im f$ the real and imaginary part of f, respectively.

Proposition 1 (Geometric Path Integral Real Dominance Conditions) *If the network $\dot{x} = f(x, a)$ can generate states belonging to class C_m, then $\Re(P(a, m, h)) > 0$, $\Re(P(a, m, h)) \gg \Im(P(a, m, h))$ for all positive values h sufficiently close to zero.*

Heuristic justification: Since $L = H^2$ in (6) is a quadratic function, and H is linear in y, all the first derivatives of L vanish only when L itself is zero. Moreover, if $\dot{x} = f(x, a)$ can generate states belonging to class C_m, from Definition 2 we know that $F(X, f(X, a)) - Y = 0$ has a solution in the domain that establishes class belonging, and therefore $H = 0$ has an appropriate solution as well (see (4) and (5)). This implies that there is a path $\gamma_0 \in \Gamma$ such that L is identically zero on γ_0 and that the functional $S(\gamma, a, m) = \int_0^1 L(\gamma(t), a, m)\, dt$ has first-order functional derivatives equal to zero as well, at $\gamma = \gamma_0$. The path γ_0 is therefore an extremal path for $S(\gamma, a, m)$. In the limit as $h \to 0$, the extremal paths, and quadratic fluctuations around them, will dominate the geometric path integral, since all other nonextremal contributions to $P(a, m, h)$ will mostly cancel each other out because of the much faster phase interference of the corresponding exponential integrals in $P(a, m, h)$. *For near-extremal paths in a neighborhood of extremal paths, we have $e^{\frac{\imath}{h}\int_0^1 L(\gamma(t),a,m)\,dt} \approx e^{\frac{\imath}{h}\int_0^1 0\,dt} \approx 1$, and they will provide a large, real positive contribution to $P(a, m, h)$, so that $P(a, m, h) \approx p + \imath q$ with p positive and $p \gg q$, if h is sufficiently small, up to a multiplicative phase factor that does not depend on* *the choice of* $S(\gamma, a, m)$, *but only on quadratic local fluctuations of near-extremal paths around the paths for which* $S(\gamma, a, m) = 0$ [15], *and that can be factored out.*

Remark 5 In standard path integrals, there may be a different change of phase for each of the individual contributions of extremal paths to the overall integral ([20], Chap. 17). Essentially, this is due to the fact that extremal paths may not be globally minima of the functionals that replace $S(\gamma, a, m)$ in standard path integrals. No such problem arises for the extremal paths used in our heuristic justification, since they all achieve the very minimum (zero) value allowed for the functional $S(\gamma, a, m)$ itself.

Remark 6 For geometric path integrals, the extremal paths are not isolated, when they exist. This may require techniques from functional field integrals (i.e., higher-dimensional path integrals, see [2], Chap. 4) for the detailed construction of semi-classical types of approximations in the limit $h \to 0$.

The computation of $P(a, m, h)$ is a global approach to identify zeros of functions in a specific real domain. Indeed there is a dependence on the original domain of biologically meaningful conditions that is hidden in the definition of the function L. However, we really want to know whether these zeros are persistent in a full measure subset $\tilde{A}$ of a domain A of parameters. Because of this additional requirement, we need one more step before we can fully express condition (2) with the geometric path integral formalism. This is achieved by taking an ordinary integral of a function of $P(a, m, h)$ with respect to the parameter vector a in the domain A where we want to enforce robustness as in (2).

Definition 5 (Robustness Function) The robustness function $R(m, h)$ associated to class C_m is, for $h > 0$,

$$R(m, h) = \int_A P(a, m, h) e^{-\frac{1}{h}(\Im P(a,m,h))^2} \, da. \tag{8}$$

This definition of robustness may formally remind the reader of the one proposed by Kitano in [17]. The two proposals, however, are substantially different, since Kitano considers a space of perturbations and defines a measure of robustness through integration on that space.

What is crucial for our interpretation of the CS problem is the fact that $R(m, h)$ inherits the real dominance conditions from $P(a, m, h)$, namely:

Proposition 2 (Robustness Function Real Dominance Conditions) *If a phenotype class* C_m *is robust for a region of parameters* $\tilde{A} \subseteq A$, *then the robustness function* $R(m, h)$ *satisfies the real dominance conditions, i.e.,* $\Re(R(m, h)) \gg \Im(R(m, h))$, *and* $\Re(R(m, h)) > 0$ *for all positive values h sufficiently close to zero.*

Heuristic justification: We make the assumption that the imaginary part of $P(a, m, h)$ goes to zero fast enough as $h \to 0$ if $P(a, m, h)$ satisfies the real dominant conditions, more particularly we assume that $|\Im P(a, m, h)| \approx h^{1/2+\varepsilon}$ with

$\varepsilon > 0$ for h small. Now, from Definition 3, if a phenotype class C_m is robust, then there are persistent zeros of the function $F(X, f(X,a)) - Y = 0$ in the appropriate domain, and $P(a,m,h)$ satisfies $\Re(P(a,m,h)) \gg \Im(P(a,m,h))$ and $\Re(P(a,m,h)) > 0$ for all a in some region $\tilde{A} \subseteq A$. Therefore for a in such region $\tilde{A}$, $P(a,m,h)$ will give large, real positive contributions to $R(m,h)$ for h that goes to zero, since the exponential in (8) will converge to 1. Suppose instead that we are in a region $\bar{A} \subseteq A$ where the functional $S(\gamma,a,m) = \int_0^1 L(\gamma(t),a,m)\,dt$ has no extremal paths for all $a \in \bar{A}$. Note that, reverting to a coordinate representation for L, for every small Δa, $L(\tilde{x}_z, \tilde{y}_z, z, a, m)$ and $L(\tilde{x}_z, \tilde{y}_z, z, a + \Delta a, m)$ will be equal on at most a finite number of points in $\mathscr{D}$, since we asked that $f(x,a)$ was analytical in a, and F is also assumed analytical in its arguments. The differences, small, but located almost everywhere in $\mathscr{D}$, between $L(\tilde{x}_z, \tilde{y}_z, z, a, m)$ and $L(\tilde{x}_z, \tilde{y}_z, z, a + \Delta a, m)$ will be enhanced in the limit as $h \to 0$, leading to large differences in the phases of $P(a,m,h)$ and $P(a+\Delta a,m,h)$. Therefore, nearby geometric path integrals in $\bar{A} \subseteq A$ will have uncorrelated phases for h that goes to 0. In particular, for each h, the set of points in $\bar{A}$ for which $\Im P(a,m,h) = 0$ is of measure zero, and therefore this set can be removed when computing the integral in (8). For all remaining $a \in \bar{A}$, the exponential in (8) will suppress to 0 the contribution of the corresponding $P(a,m,h)$ to $R(m,h)$ in the limit as $h \to 0$. We can conclude that the contributions to $R(m,h)$ from path integrals in $\bar{A}$ will be subject to strong phase interference, and also that their individual contributions to $R(m,h)$ will have norm that converges to 0. Putting together this result with the real dominant contributions from regions $\tilde{A}$ of A for which $S(\gamma,a,m)$ has extremal paths, we conclude that $R(m,h)$ will satisfy the real dominant conditions for all h sufficiently close to 0.

Remark 7 While the real dominant conditions of Propositions 1 and 2 are only necessary conditions to the existence of zeros and persistent zeros for H, respectively, these conditions are likely to be sufficient for a generic H. In the absence of extremal paths for $S(\gamma,a,m)$, it is unlikely that real dominant conditions would hold for all h sufficiently close to zero, as in the limit the phase of $P(a,m,h)$ becomes increasingly uncorrelated as a function of both h and a. Also, if $e^{-\frac{1}{h}(\Im P(a,m,h))^2}$ in (8) is substituted by $e^{-\frac{1}{h}(\frac{\Im P(a,m,h)}{\Re P(a,m,h)})^2}$, the condition $|\Im(P(a,m,h))| \approx h^{1/2+\varepsilon}$ in the justification of Proposition 2 could be substituted by the weaker $|\Im(P(a,m,h))/\Re(P(a,m,h))| \approx h^{1/2+\varepsilon}$, at the price of a slightly more complicated argument.

Remark 8 The real dominance conditions for geometric path integrals seem to portend a method to establish the existence of solutions of equations (in particular real equations) in bounded domains that does not depend on constraints on the signs of first derivatives. More specifically, suppose that we want to know whether $g(x) = 0$ has zeros in a domain D. Then we can substitute the function H in (5) with $H(x,y) = (g(x) - y)^2$ (we have no dependence from a and m in this setting, and no change of variables). The boundaries for x can be inferred directly from D, and we take y in the domain $D_y(\varepsilon) = \{y : -\varepsilon < y < \varepsilon\}$ for $\varepsilon > 0$. We can use this function

$H(x, y)$ in the definition of the geometric path integral, so that the partial derivatives of the corresponding function L in (7) are all zeros only when $g(x) - y = 0$. Therefore, if $g(x)$ has zeros in the domain D, then $g(x) - y = 0$ at least for some $(x, y) \in D \times D_y(\varepsilon)$, and a real dominance condition on the geometric path integral will hold on $D \times D_y(\varepsilon)$ for all ε sufficiently small. We will explore this important application of our technique in a subsequent paper.

4 Centipedes Segmentation Problem Reinterpreted

We come back now to the problem that motivated this work. How to interpret a pattern of allowed changes in phenotype on the basis of the structure of the underlying molecular network? In the context of the centipedes' segmentation problem, we assume that the network $\dot{x} = f(x, a)$ is essentially the same for all species of centipedes, with only the set of parameters a changing from one species to the other. This assumption is not unreasonable if we think that the same species of centipedes can display different individuals with different number of segments, showing that there is, in the same network, the potential for variable segmentations. Moreover, we would expect the process of segmentation to be evolutionary stable ([19], p. 53).

Classify now the networks in such a way that the classification function mirrors the quantitative phenotype structure. In the setting of the CS problem, if $\dot{x} = f(x, a)$ gives rise to a phenotype with 15 pairs of legs, we assume that $15 - 1 < F(X, \dot{X}) < 15$ for all states $(X, \dot{X})$ arising from that network. Effectively, we treat F as a nonlinear regression model, which predicts the number of pairs of legs from state variables. Except for the fact that we do not simply want to know what is the output of F under a specific input $(X, \dot{X})$, we must assure that some suitable $(X, \dot{X})$ can be generated stably from the network $\dot{x} = f(x, a)$. We use the robustness function R to formulate a quantitative version of the CS problem as follows:

Problem 2 (CS Problem Reinterpreted) Let $\dot{x} = f(x, a)$ be an analytic network, polynomial in x, with $a \in A$ and $x \in D$ that describes the molecular dynamics of relevant signaling compounds in centipedes' embryos. Show that there is no interval classifier F, with growth conditions compatible with the geometric path integral definition in (7), and trained on a set of actual data for known centipedes segmentation classes $C_{m_1}, \ldots, C_{m_k}$, such that, for all h sufficiently close to zero, $R(m, h)$ satisfies the real dominant conditions for m even, $m = 17, 19$.

We assume that the integral defining R is taken over a very large domain A, so that we can suppose that different segmentation phenotypes correspond to different regions of parameters within A.

According to the network path integral formalism, Problem 2 is equivalent to stating that it is not possible to find sizable volumes of parameters in A such that $\dot{x} = f(x, a)$ always gives rise to state variables $(X, \dot{X})$ such that $F(X, \dot{X})$ is even or equal to 17, 19, when $F(X, \dot{X})$ is trained to properly predict the allowed, known number of pairs of legs of centipedes. The quantitative interpretation of the CS problem is not dependent on a specific classifier, and it rather enforces some properties on any classifier that we may derive from experimental data.

We defined interval classifiers in Sect. 2 exactly to be able to achieve this compact interpretation of the CS problem, and the function $R(m, h)$ is linked to the corresponding class C_m just by the single parameter m that, in principle, could be treated now as a continuous variable.

Remark 9 The geometric path integral formulation of the CS problem allows us to comment on some essential differences between mathematization in biology and in physics. The theoretical tools used to solve problems can be similar in the two fields, as we suggested with the development of the geometric path integral formalism. But in the biological setting we lack the ability to unify our understanding of multiple problems: the functionals in the geometric path integral interpretation of the CS problem are derived from the classifiers found with supervised learning, and therefore they are not amenable to interpretation. It is as if every question that we may ask about phenotypes requires its own theory and associated geometric path integral, not reducible, even in principle, to simpler geometric path integrals.

5 Challenges Ahead

In this section we highlight some of the major problems that need to be addressed regarding geometric path integrals and their applications.[1]

First, it is to be seen how known analytical techniques to evaluate and approximate path integrals [18] apply to the highly nonstandard geometrical path integrals derived from biological classification problems. At the very least, we would expect numerical approximation of these integrals to be possible and hopefully less computationally intensive than an actual resolution of the associated geometrical problem in (2), especially for very large domains.

Moreover, path and functional field integrals are powerful qualitative tools to describe the global state of large systems [2], and similar methods for geometric path integrals may allow us to rule out real dominant conditions for entire families of functions H. In particular, to approach the geometric path integral interpretation of the CS problem, we would need methods that can constrain effectively the sign of the real part of $R(m, h)$ for large spaces of classifiers trained on a set of experimental data. It would also be important to develop the theory of geometric path integrals to allow for a precise estimate of the size of the parameters for which (2) is satisfied. This parameter size can vary for different classes, and therefore a careful estimate could be used to compare the relative robustness of different classes.

We point out that a network may be constrained by several classifier functions if different phenotype characteristics are dependent on it. The function L in the

[1] Refer to Remarks 5 an 6 for outstanding issues related to the definition of geometric path integrals.

geometric path integral (7) can be extended to these cases by taking a sum of squares of the classifier functions where each of them requires the introduction of a new slack variable, and all heuristic arguments that lead to the real dominance conditions can be repeated in this generalized case as well.

Finally, we note that if the classifier function F is fixed, it is possible to ask questions on the topology of the networks $\dot{x} = f(x, a)$ that are compatible with the real dominance conditions for each specific class. In other words, the analytical structure of suitable geometric path integrals may encode and shed light on the structure of the topologies of molecular networks that are compatible with some given phenotypic outcomes.

Acknowledgement We would like to thank Mirco Mannucci, Roman Buniy, and the referee for very useful comments.

References

1. Arnold, V.I.: Mathematical Methods of Classical Mechanics, 2nd edn. Springer, New York (2010)
2. Altland, A., Simons, B.D.: Condensed Matter Field Theory. Cambridge University Press, Cambridge (2010)
3. Basu, S., Pollack, R., Roy, M.F.: Algorithms in Real Algebraic Geometry. Springer, New York (2003)
4. Berenstein, C.A.: Review of "The universality of the Radon transform" by L. Ehrenpreis, MR2019604 (2007d:58047)
5. Bochnak, J., Coste, M., Roy, M.F.: Real Algebraic Geometry. Springer, New York (1998)
6. Calvert, V., Tang, Y., Boveia, V., Wulfkuhle, J., Schutz-Geschwender, A., Olive, M., Liotta, L., Petricoin, E.: Development of multiplexed protein profiling and detection using near infrared detection of reverse-phase protein microarrays. Clin. Proteomics **1**, 81–90 (2004)
7. Ehrenpreis, L.: Fourier Analysis in Several Complex Variables. Wiley-Interscience, New York (1970)
8. Ehrenpreis, L.: The Universality of the Radon Transform. Oxford University Press, Oxford (2003)
9. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)
10. Gunawardena, J.: Models in systems biology: The parameter problem and the meanings of robustness. In: Lodhi, H., Muggleton, S. (eds.) Elements of Computational Systems Biology. Wiley, New York (2009)
11. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning, 2nd edn. Springer, New York (2008)
12. Houle, D., Govindaraju, D.R., Omholt, S.: Phenomics: The next challenge. Nat. Rev. Genet. **11**, 855–866 (2010)
13. Irish, J.M., Kotecha, N., Nolan, G.P.: Mapping normal and cancer cell signalling networks: towards single-cell proteomics. Nat. Rev., Cancer **6**, 146–155 (2006)
14. Izenman, A.J.: Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer, New York (2008)
15. Kumano-go, N., Fujiwara, D.: Feynman path integrals and semiclassical approximation. RIMS Kokyuroku Bessatsu **B5**, 241–263 (2008)
16. Kashiwara, M., Kawai, T., Kimura, T.: Foundations of Algebraic Analysis. Princeton Mathematical Series, vol. 37. Princeton University Press, Princeton (1986)

17. Kitano, H.: Towards a theory of biological robustness. Mol. Syst. Biol. **3**, 137 (2007)
18. Kleinert, H.: Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets, 4th edn. World Scientific, Singapore (2004)
19. Minelli, A.: The Development of Animal Form. Cambridge University Press, Cambridge (2003)
20. Schulman, L.S.: Techniques and Applications of Path Integration. Dover, New York (2005)
21. Voit, E.O.: Computational Analysis of Biochemical Systems. Cambridge University Press, Cambridge (2000)
22. Wray, G.A.: Evolutionary dissociations between homologous genes and homologous structures. In: Bock, G.R., Cardew, G. (eds.) Homology, pp. 189–206. Wiley, Chichester (1999)

Bounded Cohomology for Solutions of Systems of Differential Equations: Applications to Extension Problems

Irene Sabadini and Daniele C. Struppa

Abstract In this paper we expand on some ideas originally put forward by Ehrenpreis in his monograph (Fourier Analysis in Several Complex Variables, Wiley Interscience, New York, 1970), and we show how to extend approximate solutions to the Cauchy–Fueter system in n variables.

1 Introduction

One of the early successes of Ehrenpreis' approach to the study of systems of partial differential equations was his new and surprising proof of the famous Hartogs' theorem on the removability of compact singularities for holomorphic functions of $n \geq 2$ complex variables [7]. Ehrenpreis' proof is elegant and contains many of the ideas which will be relevant in this paper, and for this reason, we offer it here in a concise version. The reader may consult [15, 18] for the history of the approaches to the Hartogs' theorem.

Theorem 1 *Let K be a compact convex subset of $\mathbb{C}^n$, and let $f : \mathbb{C}^n \setminus K \to \mathbb{C}$ be a holomorphic function. Then there exists a unique entire function $\tilde{f} : \mathbb{C}^n \to \mathbb{C}$ such that $\tilde{f} = f$ on $\mathbb{C}^n \setminus K$.*

Proof To begin with, we consider a hyperfunction g which extends f to all of $\mathbb{C}^n$. Let $D_i = \partial/\partial \bar{z}_i$, $i = 1, \ldots, n$, and let $g_i := D_i g$. Clearly, the hyperfunctions g_i are supported in K, and

$$D_i g_j = D_j g_i \tag{1}$$

This paper is dedicated to the memory of the late Professor Leon Ehrenpreis.

I. Sabadini (✉)
Dipartimento di Matematica, Politecnico di Milano, Via E. Bonardi 9, 20133 Milano, Italy
e-mail: irene.sabadini@polimi.it

D.C. Struppa
Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA
e-mail: struppa@chapman.edu

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_17, © Springer-Verlag Italia 2012

for all pairs i, j. Since $D_i g_j$ also have compact support, we can take the Fourier transform of these hyperfunctions, and from (1) we obtain $z_i G_j = z_j G_i$, where we have indicated with G_i the Fourier transform of g_i. Therefore the function $H := G_j/z_j = G_i/z_i$ is holomorphic outside the origin of $\mathbb{C}^n$. By the so-called second Riemann removability singularity theorem (which stipulates that an analytic singularity of codimension bigger than or equal to two can be removed) [12], we immediately obtain that H is entire. The Paley–Wiener theorem for the Fourier transform of hyperfunctions supported by a compact convex set K implies that for every positive ε, there exists A_ε such that $|G_j(z)| \leq A_\varepsilon \exp(H_K(z) + \varepsilon|z|)$, where $z = (z_1, \dots, z_n)$, and H_K is the supporting function of the compact convex set K. Therefore, the Ehrenpreis–Malgrange lemma shows that H itself satisfies the same inequality and therefore is the Fourier transform of a hyperfunction h supported in K as well. The proof is concluded by defining $\tilde{f} = g - h$. □

Remark 1 The proof we have given invokes the use of hyperfunctions because this allows the extension of f to g without losing anything along the boundary of K. This is not necessary, however, and in fact Ehrenpreis worked directly in the C^∞ setting.

Remark 2 It is important to note that the result in [7] is actually stated for solutions to any system

$$\begin{cases} P_1(D)f = 0, \\ \dots \\ P_r(D)f = 0, \end{cases} \tag{2}$$

where the polynomials $P_1, \dots, P_r$, symbols of the operators $P_1(D), \dots, P_r(D)$, have no common factors. The proof we have indicated above carries over with no changes. One should also point out that Ehrenpreis mentions in [7, 8], and Palamodov proves in [14] that the same result can also be obtained for a large class of rectangular systems satisfying an algebraic condition, which in the case of system (2), with one unknown function, reduces to the request that the polynomials have no common factors.

Remark 3 As Ehrenpreis himself has pointed out, this result can be rephrased in terms of sheaf cohomology. Specifically, if $\mathcal{O}$ denotes the sheaf of germs of holomorphic functions in n complex variables, then the proof we have described before shows that compact singularities can be removed if one can prove the vanishing of the first cohomology group with compact support $H^1_c(U, \mathcal{O})$ for any open convex set U in $\mathbb{C}^n$. In the case in which a system different from the Cauchy–Riemann system is considered, the sheaf $\mathcal{O}$ needs to be replaced by the sheaf of solutions of the system itself.

Let us discuss explicitly this interpretation. Let f be a solution in $\mathbb{R}^n \setminus K$ of a system of linear differential operators with constant coefficients. We denote by $\mathbf{P}(D)$ the column vector $(P_1(D), \dots, P_r(D))^t$. As in the proof that we outlined before, f can be extended to a hyperfunction g over $\mathbb{R}^n$, and one can then consider the family $g_i := P_i(D)g$. Assume now that we know that $H_c^1(\mathbb{R}^n, \mathscr{S}) = 0$ where $\mathscr{S}$ is the sheaf of hyperfunction solutions to the system $\mathbf{P}(D)f = 0$. By Theorem 4.6 in [13] the vanishing of this cohomology group is equivalent to the request that the kernel of the matrix operator $\mathbf{Q}(D)$ associated to the first syzygies of $\mathbf{P}(D)$ satisfies $\mathrm{Ker}(\mathbf{Q}(D)) = \mathrm{Im}(\mathbf{P}(D))$ on the space $\mathscr{B}^r$ of r-vectors whose components are compactly supported hyperfunctions in $\mathbb{R}^n$. Since $\mathbf{g} = (g_1, \ldots, g_r)^t$ belongs to the kernel of $\mathbf{Q}(D)$, and by the vanishing of the compact cohomology, we obtain that there is an element h compactly supported in $\mathbb{R}^n$ such that $P_i(D)h = g_i$. It is then immediate to verify that the function $\tilde{f} := g - h$ is the required extension. In actuality, the proof of the theorem shows directly the vanishing of the quotient $\frac{\mathrm{Ker}(\mathbf{Q}(D))}{\mathrm{Im}(\mathbf{P}(D))}$, and we can see how the vanishing of the cohomology would imply the vanishing of the quotient. Specifically, the collection $\{P_j(D)g_i\}$ is a 1-cocycle, and therefore the vanishing of the cohomology implies the existence of a family of compactly supported hyperfunctions $\{k_i\}$ such that $P_j(D)g_i = k_i - k_j$. By taking the Fourier transform, and using again capital letters G_j, K_i to denote the Fourier transforms of g_j, k_i, we obtain $P_j G_i = K_i - K_j$, and one can then formally define $H := \frac{K_i - K_j}{P_i P_j}$ which is a well-defined entire function because $\frac{K_i - K_j}{P_i P_j} = \frac{P_j G_i}{P_i P_j} = \frac{G_i}{P_i}$, which is independent of the index i. If we define the function h to be the anti-Fourier transform of H and set $\tilde{f} = g - h$, we obtain the statement.

Remark 4 Note that Ehrenpreis in his proof does not explicitly construct the family $\{k_i\}$ but uses the cocycle $\{P_j(D)g_i\}$ to construct directly the function h.

Remark 5 Since the Cauchy–Riemann system is elliptic, the sheaf of hyperfunctions solutions to the system is exactly the sheaf of holomorphic functions.

Remark 6 The vanishing of $H_c^1(U, \mathscr{O})$ for all open convex sets U in $\mathbb{C}^n$ is of course a standard fact in the theory of several complex variables.

In [8] Ehrenpreis proved a refinement of these results which applies to what he calls "approximate solutions" to the system of differential equations, see [8], Chap. XI. He does so for systems of the form (2) in which $\mathbf{P}(D)$ is a column $(P_1(D), \ldots, P_r(D))^t$ of linear partial differential operators with constant coefficients.

Theorem 2 *Let $\Omega_0 = \Omega_2 \setminus \Omega_1$ be a subset of $\mathbb{R}^n$ such that Ω_1 and Ω_2 are open, convex, and bounded sets in $\mathbb{R}^n$, with $\overline{\Omega}_1 \subset \Omega_2$. Let $\partial\Omega_i$ be the boundary of Ω_i $(i = 1, 2)$, and let $\mathscr{U}(\partial\Omega_1)$ be a neighborhood of $\partial\Omega_1$ in Ω_2. Let $\mathscr{E}$ denote the sheaf of infinitely differentiable functions, and let B be a set of functions in $\mathscr{E}(\Omega_0 \cup \mathscr{U}(\partial\Omega_1))$ such that*

$$\big\{\mathbf{P}(D)f \mid f \in B\big\} \text{ is bounded in } \mathscr{E}^r\big(\Omega_0 \cup \mathscr{U}(\partial\Omega_1)\big).$$

If the polynomials P_i have no common factors then, for any $f \in B$, there exists $\tilde{f} \in \mathscr{E}(\Omega_2)$ such that:

(a) $\{f - \tilde{f} : f \in B\}$ *is a bounded set in* $\mathscr{E}(\Omega_0 \cup \mathscr{U}'(\partial\Omega_1))$ *for some neighborhood* $\mathscr{U}'$ *of* $\partial\Omega_1$,
(b) $\{\mathbf{P}(D)\tilde{f}\}$ *is bounded in* $\mathscr{E}^r(\Omega_2)$.

The first goal of this paper is to offer a cohomological interpretation of Ehrenpreis' Theorem 2 in terms of what we call *bounded cohomology* to which we dedicate the next section. The second goal is to expand on the ideas of [19] and show how Ehrenpreis' theorem can be formulated for an important concrete case in which a matrix system is considered, specifically for the Cauchy–Fueter system in $4n$ variables. The case $n = 2$ was treated in [19].

2 Bounded Cohomology

Our main tool to give a cohomological interpretation of Theorem 2 will be a notion of cohomology that we call *bounded cohomology*. It is constructed by taking as elements the standard cochains with infinitely differentiable coefficients and as equivalence relation the relation we introduce in the next definition.

Definition 1 Let $\Omega \subseteq \mathbb{R}^n$, and let $\mathscr{V} = \{\Omega_i\}_{i\in I}$ be an open covering of Ω. We will say that a family of sections $\{f_i\}_{i\in I}$, $f_i \in \mathscr{E}(\Omega_i)$, is *equal up to bounded terms* to $\{g_i\}_{i\in I}$, $g_i \in \mathscr{E}(\Omega_i)$, and we will write

$$\{f_i\}_{i\in I} \doteq \{g_i\}_{i\in I}$$

if $f_i - g_i$ is bounded in the topology of $\mathscr{E}(\Omega_i)$, $\forall i \in I$.

It is immediate to prove the following result:

Proposition 1 *The relation of equality up to bounded terms defines an equivalence relation on* $\mathscr{E}(\Omega)$.

Let us consider the set of 0-cochains $C^0(\mathscr{V}, \mathscr{E}) = \prod_i \mathscr{E}(\Omega_i)$ of $\mathscr{E}$ and, more in general, the set of p-cochains $C^p(\mathscr{V}, \mathscr{E}) = \prod_{i_0\neq i_1\neq\cdots\neq i_p} \mathscr{E}(\Omega_{i_0} \cap \cdots \cap \Omega_{i_p})$. A p-cochain in $C^p(\mathscr{V}, \mathscr{E})$ is denoted, as customary, as $f = \{f_{i_0\ldots i_p}\}$ where $f_{i_0\ldots i_p} \in \mathscr{E}(\Omega_{i_0} \cap \cdots \cap \Omega_{i_p})$. Let us define the set $C^p_b(\mathscr{V}, \mathscr{E})$ of p-cochains with coefficients in $\mathscr{E}$ modulo the equality $\doteq$:

$$C^p_b(\mathscr{V}, \mathscr{E}) = C^p(\mathscr{V}, \mathscr{E})/\doteq,$$

where we use the subscript "b" to emphasize that we consider p-cochains up to bounded terms. We define the coboundary operator as usual by

$$\delta: \ C_b^p(\mathscr{V}, \mathscr{E}) \to C_b^{p+1}(\mathscr{V}, \mathscr{E}),$$
$$\delta(f)_{i_0 \dots i_{p+1}} \doteq \sum_{j=0}^{p+1} (-1)^j f_{i_0 \dots \widehat{i_j} \dots i_{p+1} | \Omega_{i_0} \cap \dots \Omega_{i_p}},$$

where $\widehat{i_j}$ means that the index i_j is omitted.

Definition 2 A p-cochain $f \in C_b^p(\mathscr{V}, \mathscr{E})$ such that $\delta(f) \doteq 0$ is called a p-cocycle, and the set of p-cocycles is denoted by $Z_b^p(\mathscr{V}, \mathscr{E})$. A p-cochain f such that $f \doteq \delta(g)$ is called a p-coboundary, and the set of p-coboundaries is denoted by $B_b^p(\mathscr{V}, \mathscr{E})$.

Definition 3 We call the pth bounded cohomology group and denote by $H_b^p(\mathscr{V}, \mathscr{E})$ the quotient

$$H_b^p(\mathscr{V}, \mathscr{E}) = \frac{Z_b^p(\mathscr{V}, \mathscr{E})}{B_b^p(\mathscr{V}, \mathscr{E})}.$$

Remark 7 In particular, we have

$$Z_b^1(\mathscr{V}, \mathscr{E}) = \big\{ \{f_{ij}\} \in C_p^1(\mathscr{V}, \mathscr{E}) \mid \{f_{ij} + f_{ji}\} \doteq \{0\} \big\},$$
$$B_b^1(\mathscr{V}, \mathscr{E}) = \big\{ \{f_{ij}\} \in C_p^1(\mathscr{V}, \mathscr{E}) \mid \{f_{ij}\} \doteq \{f_i - f_j\} \big\},$$

so that $H_b^1(\mathscr{V}, \mathscr{E}) = Z_b^1(\mathscr{V}, \mathscr{E}) / B_b^1(\mathscr{V}, \mathscr{E})$. The vanishing of $H_b^1(\mathscr{V}, \mathscr{E})$ means that for any choice of $\{f_{ij} + f_{ji}\}$ which are bounded in $\mathscr{E}(\Omega)$, there exist f_i, f_j such that $f_{ij} = (f_i - f_j) + b_{ij}$, where b_{ij} is bounded in $\mathscr{E}(\Omega)$.

Remark 8 The previous definitions can be repeated also in the case of p-cochains with compact support. In that case we obtain the bounded cohomology with compact support that we denote by $H_{c,b}^1(\mathscr{V}, \mathscr{E})$. By taking a suitable covering of Ω we get the bounded cohomology with compact support $H_{c,b}^1(\Omega, \mathscr{E})$.

We now show how this language can be useful to rephrase Theorem 2.

Theorem 3 *Let $\Omega_0 = \Omega_2 \setminus \Omega_1$ be a subset of $\mathbb{R}^n$ such that Ω_1 and Ω_2 are open, convex, and bounded sets in $\mathbb{R}^n$, with $\overline{\Omega}_1 \subset \Omega_2$. Let $\partial\Omega_i$ be the boundary of Ω_i $(i = 1, 2)$, and let $\mathscr{U}(\partial\Omega_1)$ be a neighborhood of $\partial\Omega_1$ in Ω_2. Let $\mathscr{S}$ be the sheaf of infinitely differentiable functions that are solutions to the system $\mathbf{P}(D) f = 0$. Let B be a set of functions in $\mathscr{E}(\Omega_0 \cup \mathscr{U}(\partial\Omega_1))$ such that*

$$\big\{\mathbf{P}(D) f \mid f \in B\big\} \text{ is bounded in } \mathscr{E}^r\big(\Omega_0 \cup \mathscr{U}(\partial\Omega_1)\big).$$

If $H_{c,b}^1(\Omega_2, \mathscr{S}) = 0$, then, for any $f \in B$, there exists $\tilde{f} \in \mathscr{E}(\Omega_2)$ such that:

(a) *$\{f - \tilde{f} : f \in B\}$ is a bounded set in $\mathscr{E}(\Omega_0 \cup \mathscr{U}'(\partial\Omega_1))$ for some neighborhood $\mathscr{U}'$ of $\partial\Omega_1$,*
(b) *$\{\mathbf{P}(D)\tilde{f}\}$ is bounded in $\mathscr{E}^r(\Omega_2)$.*

Proof For every $f \in B$, we can construct its extension g to Ω_2 and define $g_j = P_j(D)g$ which can be assumed to be with compact support. By hypothesis we have that the family $\{P_i(D)g_j - P_j(D)g_i\}$ is bounded and compactly supported. The same is therefore true for the family of its Fourier transforms $\{P_i G_j - P_j G_i\}$. Now assume that $H^1_{c,b}(\Omega_2, \mathscr{S}) = 0$. This implies the existence of a family $\{K_i\}$ such that $P_j G_i \dot{=} K_i - K_j$. By defining as before h as the anti-Fourier transform of $\frac{K_i - K_j}{P_i P_j}$, we obtain that the family of extensions $\{\tilde{f} = g - h \mid f \in B\}$ is such that $\{f - \tilde{f} \mid f \in B\}$ is bounded as requested and $\mathbf{P}\tilde{f}$ is bounded as well. □

3 The Cauchy–Fueter System

In this section we will prove an analogue of Theorem 2 for the solutions of the so-called Cauchy–Fueter system in several variables. Let us begin by recalling some basic definitions. We will denote by $\mathbb{H}$ the skew field of quaternions $q = x_0 + ix_1 + jx_2 + kx_3$, where x_ℓ is real for $\ell = 0, \dots, 3$, and i, j, k are such that $i^2 = j^2 = k^2 = -1$ and $ij = k,\ jk = i, ki = j, ij = k,\ jk = i,\ ki = j$. From a vector-space point of view, $\mathbb{H}$ can be seen as $\mathbb{R}^4$, but the space $\mathbb{H}$ of quaternions admits a noncommutative multiplicative structure that differentiates it from $\mathbb{R}^4$.

There are several notions that extend the concept of holomorphicity to functions defined on open sets in $\mathbb{H}$. In this paper we deal with the one introduced by R. Fueter back in the 1930s, see [9, 20], and which has been extensively studied in the literature.

Definition 4 Let Ω be an open subset of $\mathbb{H}$, and let $f : \Omega \to \mathbb{H}$ be a real differentiable function. We say that f is *regular* if it satisfies the so-called *Cauchy–Fueter equation*

$$\frac{\partial f}{\partial \bar{q}} := \frac{\partial f}{\partial x_0} + i\frac{\partial f}{\partial x_1} + j\frac{\partial f}{\partial x_2} + k\frac{\partial f}{\partial x_3} = 0. \tag{3}$$

Equation (3) can be rewritten it by taking advantage of the explicit expression of the $\frac{\partial}{\partial \bar{q}}$ operator and of the representation of f in its real components $f = f_0 + if_1 + jf_2 + kf_3$. From this point of view, (3) becomes the following 4×4 system of differential equations in the unknown variables f_0, f_1, f_2, f_3:

$$\begin{aligned}
&\frac{\partial f_0}{\partial x_0} - \frac{\partial f_1}{\partial x_1} - \frac{\partial f_2}{\partial x_2} - \frac{\partial f_3}{\partial x_3} = 0,\\
&\frac{\partial f_0}{\partial x_1} + \frac{\partial f_1}{\partial x_0} - \frac{\partial f_2}{\partial x_3} + \frac{\partial f_3}{\partial x_2} = 0,\\
&\frac{\partial f_0}{\partial x_2} + \frac{\partial f_1}{\partial x_3} + \frac{\partial f_2}{\partial x_0} - \frac{\partial f_3}{\partial x_1} = 0,\\
&\frac{\partial f_0}{\partial x_3} - \frac{\partial f_1}{\partial x_2} + \frac{\partial f_2}{\partial x_1} + \frac{\partial f_3}{\partial x_0} = 0.
\end{aligned}$$

In [6] the authors showed how the notion of regularity can be naturally defined for functions in several quaternionic variables and demonstrated that such functions enjoy important properties that make their theory quite similar to the theory of holomorphic functions in several complex variables. To begin with, we give the following definition.

Definition 5 Let Ω be an open subset of $\mathbb{H}^n$, and let $f : \Omega \to \mathbb{H}$ be a differentiable function. Let $(q_1, \ldots, q_n)$ denote the variable in $\mathbb{H}^n$. We say that the function f is regular if it satisfies the Cauchy–Fueter system of differential equations

$$\frac{\partial f}{\partial \bar{q}_1} = \frac{\partial f}{\partial \bar{q}_2} = \cdots = \frac{\partial f}{\partial \bar{q}_n} = 0,$$

with obvious meaning of the symbols. The right quaternionic vector space of functions that are regular on the open set Ω is denoted by $\mathscr{R}(\Omega)$.

Remark 9 The assignment $\Omega \to \mathscr{R}(\Omega)$ is a sheaf of right vector spaces on $\mathbb{H}$.

Similarly to what is done in the case of one variable, the Cauchy–Fueter system can be rewritten as a system of $4n$ differential equations in four unknown functions $f_0, \ldots, f_3$. Thus, we now have what is usually referred to as an overdetermined system of differential equations. As we did in a series of articles, see [1–4], which culminated in [6], we used extensively the ideas of Ehrenpreis and Palamodov to develop the theory of its solutions.

In particular, in [16], we described the compatibility conditions on $(g_1, \ldots, g_n)$ for the nonhomogeneous Cauchy–Fueter system

$$\frac{\partial f}{\partial \bar{q}_1} = g_1, \ \ldots, \ \frac{\partial f}{\partial \bar{q}_n} = g_n \tag{4}$$

to be solvable. These conditions correspond, from the algebraic point of view, to the first syzygies of the module associated to the system and are described in the following proposition.

Proposition 2 *System* (4) *admits a solution* f *if and only if the datum* $\mathbf{g} = (g_1, \ldots, g_n)^t$ *satisfies*:

1. *For each of the* $2\binom{n}{2}$ *ordered pairs of indices* r, s, $1 \le r, s \le n$,

$$\partial_{\bar{q}_r} \partial_{q_s} g_s - \partial_{\bar{q}_s} \partial_{q_s} g_r = 0.$$

2. *For each of the* $\binom{n}{3}$ *triples of indices* h, r, s, $1 \le h, r, s \le n$,

$$\partial_{q_h} \partial_{\bar{q}_r} g_s + \partial_{q_r} \partial_{\bar{q}_h} g_s - \partial_{\bar{q}_s} \partial_{q_r} g_h - \partial_{\bar{q}_s} \partial_{q_h} g_r = 0$$

and

$$\partial_{q_r} \partial_{\bar{q}_s} g_h + \partial_{q_s} \partial_{\bar{q}_r} g_h - \partial_{\bar{q}_h} \partial_{q_r} g_s - \partial_{\bar{q}_h} \partial_{q_s} g_r = 0.$$

3. *For each of the* $\binom{n}{3}$ *triples of indices* h, r, s, $1 \le h, r, s \le n$,

$$\big(D'_{q_r}\partial_{\bar q_s} - D'_{q_s}\partial_{\bar q_r}\big)g_h + \big(D'_{q_s}\partial_{\bar q_h} - D'_{q_h}\partial_{\bar q_s}\big)g_r + \big(D'_{q_h}\partial_{\bar q_r} - D'_{q_r}\partial_{\bar q_h}\big)g_s = 0$$

and

$$\big(D''_{q_r}\partial_{\bar q_s} - D''_{q_s}\partial_{\bar q_r}\big)g_h + \big(D''_{q_s}\partial_{\bar q_h} - D''_{q_h}\partial_{\bar q_s}\big)g_r + \big(D''_{q_h}\partial_{\bar q_r} - D''_{q_r}\partial_{\bar q_h}\big)g_s = 0,$$

where $\partial_{\bar q_i} = \frac{\partial}{\partial q_{i0}} + i\frac{\partial}{\partial q_{i1}} + j\frac{\partial}{\partial q_{i2}} + k\frac{\partial}{\partial q_{i3}}$, $\partial_{q_i} = \frac{\partial}{\partial q_{i0}} - i\frac{\partial}{\partial q_{i1}} - j\frac{\partial}{\partial q_{i2}} - k\frac{\partial}{\partial q_{i3}}$, $D'_{q_i} = -i\frac{\partial}{\partial q_{i2}} + \frac{\partial}{\partial q_{i3}}$, *and* $D''_{q_i} = \frac{\partial}{\partial q_{i3}} + j\frac{\partial}{\partial q_{i1}}$.

In the sequel we will denote by $\mathbf{Q}(D)$ the matrix of differential operators such that $\mathbf{Q}(D)\mathbf{g} = 0$ corresponds to the matrix form of the relations in Proposition 2.

We now give the notion of approximate solutions to the Cauchy–Fueter system (see also [19]):

Definition 6 Let Ω be a bounded open set in $\mathbb{R}^{4n}$ (or equivalently a bounded open set in $\mathbb{H}^n$), and let B be a set of quaternionic valued infinitely differentiable functions in Ω. We say that B is a set of approximate solutions to the Cauchy–Fueter system on Ω if the set

$$\left\{\frac{\partial f}{\partial \bar q_1}, \ldots, \frac{\partial f}{\partial \bar q_n} : f \in B\right\}$$

is a bounded set in the topology of the space $\mathscr{E}^n(\Omega)$ of ordered n-tuples of infinitely differentiable functions on Ω.

Note that regular functions are approximate solutions as well, since the set above would be reduced to the zero set. In [19], Struppa has proved the following theorem.

Theorem 4 *Let* $\Omega_0 = \Omega_2 \setminus \Omega_1$ *be a subset of* $\mathbb{H}^2$ *such that* Ω_1 *and* Ω_2 *are open, convex, and bounded sets in* $\mathbb{H}^2$, *with* $\overline{\Omega}_1 \subset \Omega_2$. *Let* $\partial\Omega_i$ *be the boundary of* Ω_i $(i = 1, 2)$, *and let* $\mathscr{U}(\partial\Omega_1)$ *be a neighborhood of* $\partial\Omega_1$ *in* Ω_2. *Let B be a set of functions in* $\mathscr{E}(\Omega_0 \cup \mathscr{U}(\partial\Omega_1))$ *such that*

$$\left\{\frac{\partial f}{\partial \bar q_1}, \frac{\partial f}{\partial \bar q_2} : f \in B\right\} \textit{ is bounded in } \mathscr{E}^2\big(\Omega_0 \cup \mathscr{U}(\partial\Omega_1)\big).$$

Then, for any $f \in B$, *there exists* $\tilde f \in \mathscr{E}(\Omega_2)$ *such that*:

- $\{f - \tilde f : f \in B\}$ *is a bounded set in* $\mathscr{E}(\Omega_0 \cup \mathscr{U}'(\partial\Omega_1))$ *for some neighborhood* $\mathscr{U}'$ *of* $\partial\Omega_1$,
- $\{\frac{\partial \tilde f}{\partial \bar q_1}, \frac{\partial \tilde f}{\partial \bar q_2}\}$ *is bounded in* $\mathscr{E}^2(\Omega_2)$.

We will now show how to extend this result to the case of several quaternionic variables.

Theorem 5 *Let $\Omega_0 = \Omega_2 \setminus \Omega_1$ be a subset of $\mathbb{H}^n$ such that Ω_1 and Ω_2 are open, convex, and bounded sets in $\mathbb{H}^n$, with $\overline{\Omega}_1 \subset \Omega_2$. Let $\partial\Omega_i$ be the boundary of Ω_i $(i = 1, 2)$, and let $\mathscr{U}(\partial\Omega_1)$ be a neighborhood of $\partial\Omega_1$ in Ω_2. Let B be a set of functions in $\mathscr{E}(\Omega_0 \cup \mathscr{U}(\partial\Omega_1))$ such that*

$$\left\{\frac{\partial f}{\partial \bar{q}_1}, \ldots, \frac{\partial f}{\partial \bar{q}_n} : f \in B\right\} \text{ is bounded in } \mathscr{E}^n\big(\Omega_0 \cup \mathscr{U}(\partial\Omega_1)\big).$$

Then, for any $f \in B$, there exists $\tilde{f} \in \mathscr{E}(\Omega_2)$ such that:

(a) $\{f - \tilde{f} : f \in B\}$ *is a bounded set in $\mathscr{E}(\Omega_0 \cup \mathscr{U}'(\partial\Omega_1))$ for some neighborhood $\mathscr{U}'$ of $\partial\Omega_1$,*
(b) $\{\frac{\partial \tilde{f}}{\partial \bar{q}_1}, \ldots, \frac{\partial \tilde{f}}{\partial \bar{q}_n}\}$ *is bounded in $\mathscr{E}^n(\Omega_2)$.*

Proof We begin by extending the function f to a function g, infinitely differentiable on the entire space Ω_2, and such that g is equal to f on the set $\Omega_0 \cup \mathscr{U}'$ for some small neighborhood $\mathscr{U}' \subset \mathscr{U}$ of $\partial\Omega_1$, and we set $\Omega_3 = \Omega_1 \setminus \mathscr{U}'$. Let us define

$$g_r = \frac{\partial g}{\partial \bar{q}_r}, \quad r = 1, \ldots, n,$$

and consider the set

$$\{g_1, \ldots, g_n : f \in B\},$$

which is bounded in $\mathscr{E}(\Omega_0 \cup \mathscr{U}')$. We now multiply the functions g_r by a C^∞ function χ which is identically one on Ω_3 and which is compactly supported in Ω_2; we obtain new functions (for the sake of simplicity, we will still denote them by g_r), which are now compactly supported in Ω_2. The set $\{g_1, \ldots, g_n : f \in B\}$ is therefore a bounded set of compactly supported functions. We now consider the first syzygies of the system of n Cauchy–Fueter operators described in Proposition 2, and we apply them to $g_1, \ldots, g_n$. We obtain the set

$$\begin{aligned}
\mathscr{F}_B = \big\{&\partial_{\bar{q}_r}\partial_{q_s} g_s - \partial_{\bar{q}_s}\partial_{q_s} g_r,\ \partial_{q_h}\partial_{\bar{q}_r} g_s + \partial_{q_r}\partial_{\bar{q}_h} g_s - \partial_{\bar{q}_s}\partial_{q_r} g_h - \partial_{\bar{q}_s}\partial_{q_h} g_r,\\
&\partial_{q_r}\partial_{\bar{q}_s} g_h + \partial_{q_s}\partial_{\bar{q}_r} g_h - \partial_{\bar{q}_h}\partial_{q_r} g_s - \partial_{\bar{q}_h}\partial_{q_s} g_r,\\
&\big(D'_{q_r}\partial_{\bar{q}_s} - D'_{q_s}\partial_{\bar{q}_r}\big) g_h + \big(D'_{q_s}\partial_{\bar{q}_h} - D'_{q_h}\partial_{\bar{q}_s}\big) g_r + \big(D'_{q_h}\partial_{\bar{q}_r} - D'_{q_r}\partial_{\bar{q}_h}\big) g_s,\\
&\big(D''_{q_r}\partial_{\bar{q}_s} - D''_{q_s}\partial_{\bar{q}_r}\big) g_h + \big(D''_{q_s}\partial_{\bar{q}_h} - D''_{q_h}\partial_{\bar{q}_s}\big) g_r\\
&+ \big(D''_{q_h}\partial_{\bar{q}_r} - D''_{q_r}\partial_{\bar{q}_h}\big) g_s : f \in B\big\},
\end{aligned}$$

which is bounded in $\mathscr{D}(\Omega_4)$, where Ω_4 is a bounded convex set such that $\bar{\Omega}_3 \subset \bar{\Omega}_4 \subset \Omega_1$. We can now take, formally, the Fourier transform of the elements of $\mathscr{F}_B$. Since we know that the Fourier transform is a topological isomorphism, we denote by G_r the Fourier transform of g_r for $r = 1, \ldots, n$, we obtain (with obvious symbolism) that the new set

$$\mathscr{S}_B := \big\{q_r\bar{q}_sG_s - q_s\bar{q}_sG_r,\ \ \bar{q}_hq_rG_s + \bar{q}_rq_hG_s - q_s\bar{q}_rG_h - q_s\bar{q}_hG_r,$$
$$\bar{q}_rq_sG_h + \bar{q}_sq_rG_h - q_h\bar{q}_rG_s - q_h\bar{q}_sG_r,$$
$$\big(z'_rq_s - z'_sq_r\big)G_h + \big(z'_sq_h - z'_hq_s\big)G_r + \big(z'_hq_r - z'_rq_h\big)G_s,$$
$$\big(z''_rq_s - z''_sq_r\big)G_h + \big(z''_sq_h - z''_hq_s\big)G_r + \big(z''_hq_r - z''_rq_h\big)G_s,$$
$$r, s = 1,\dots,n,\ r \neq s \neq h\ :\ f \in B\big\}$$

is bounded in $\hat{\mathscr{D}}(\Omega_4)$. That this is indeed the case can be confirmed by writing explicitly the operators (in terms of their real coordinates), then taking the Fourier transform of the various components, and finally regrouping in quaternionic form.

Denote by $\mathfrak{B}$ the multiplicity variety associated to the equation $q_1 = 0$ (we refer to [8] and [17] for a detailed definition of such a variety). Let us now look at the set $\tilde{\mathscr{S}_B} = \mathscr{S}_B \setminus \{q_r\bar{q}_1G_1 - q_1\bar{q}_1G_r, r = 1,\dots,n\}$. The elements in $\tilde{\mathscr{S}_B}$ do not vanish identically on the variety defined by $q_1 = 0$. Let us formally set $q_1 = 0$ in all the elements in $\tilde{\mathscr{S}_B}$, and let us denote this new set by $(\tilde{\mathscr{S}_B})_{|q_1=0}$. From this fact we immediately deduce that

$$(\tilde{\mathscr{S}_B})_{|q_1=0} \text{ is bounded in } \hat{\mathscr{D}}(\Omega_4)(\mathfrak{B}).$$

Let us consider in $(\tilde{\mathscr{S}_B})_{|q_1=0}$ the element $q_2\bar{q}_2G_1$. If we denote by $\mathfrak{V}$ the subvariety of $\mathfrak{B}$ where $|\bar{q}_2q_2| \geq 1$, then we obtain that $\{G_1 : f \in B\}$ is bounded in the topology of $\hat{\mathscr{D}}(\Omega_4)(\mathfrak{V})$, and so, since the set where $|\bar{q}_2q_2| \geq 1$ is a sufficient set according to [8], we deduce that the set $\{G_1 : f \in B\}$ is actually bounded in all of $\hat{\mathscr{D}}(\Omega_4)(\mathfrak{B})$. Note that a similar argument can be made if we look at other components of the set $\mathscr{S}$, which are bounded on the multiplicity variety associated to $q_r = 0$, $r = 2,\dots,n$. By the Fundamental Principle of Ehrenpreis [8], but see also [5, 11], (actually by the Extension Theorem that implies the Fundamental Principle, [10, 17]) we know that every function G holomorphic on the multiplicity variety $\mathfrak{B}$ can be extended to an entire function $\tilde{G}$ satisfying on $\mathbb{C}^n$ the same bounds that G satisfies on $\mathfrak{B}$. Thus we can say that there exists a function $\tilde{G}_1$ such that the set $\{\tilde{G}_1 : f \in B\}$ is bounded and, for some entire function $H_1 \in \hat{\mathscr{D}}(\Omega_4)$,

$$G_1 - \tilde{G}_1 = q_1 \cdot H_1.$$

Now let $h_1 \in \mathscr{D}(\Omega_4)$ be the anti-Fourier transform of H_1, and let $\tilde{g}_1 \in \mathscr{D}(\Omega_4)$ be the anti-Fourier transform of $\tilde{G}_1$. Then

$$\frac{\partial h_1}{\partial \bar{q}_2} - g_2 = \left(\frac{\partial h_2}{\partial \bar{q}_2} - g_2\right) + \left(\frac{\partial h_1}{\partial \bar{q}_2} - \frac{\partial h_2}{\partial \bar{q}_2}\right) = -\tilde{g}_2 + \left(\frac{\partial h_1}{\partial \bar{q}_2} - \frac{\partial h_2}{\partial \bar{q}_2}\right).$$

Now notice that

$$\Delta_1\frac{\partial h_1}{\partial \bar{q}_2} - \Delta_1\frac{\partial h_2}{\partial \bar{q}_2} = \frac{\partial \Delta_1 h_1}{\partial \bar{q}_2} - \Delta_1\frac{\partial h_2}{\partial \bar{q}_2} = \frac{\partial}{\partial \bar{q}_2}\frac{\partial}{\partial q_1}\frac{\partial h_1}{\partial \bar{q}_1} - \Delta_1\frac{\partial h_2}{\partial \bar{q}_2}.$$

This last expression is, up to a bounded function, nothing but

$$\frac{\partial}{\partial \bar{q}_2}\frac{\partial g_1}{\partial q_1} - \Delta_1 g_2,$$

which is therefore bounded.

Since differential operators (and in particular the Laplacian) are topological isomorphisms (see Theorem 6.5 in [8]), we deduce that

$$\frac{\partial h_1}{\partial \bar{q}_2} - \frac{\partial h_2}{\partial \bar{q}_2}$$

is also a bounded set. But since the set $\{\tilde{g}_r \ : \ f \in B\}$ is a bounded set, so is the set

$$\left\{\frac{\partial h_1}{\partial \bar{q}_2} - g_2 : f \in B\right\}.$$

If we now define $\tilde{f} = g - h_1$, we conclude that the function $\tilde{f}$ is the one requested in the statement of the theorem. □

We conclude the paper with the cohomological version of this same theorem.

Theorem 6 *Let $\Omega_0 = \Omega_2 \setminus \Omega_1$ be a subset of $\mathbb{H}^n$ such that Ω_1 and Ω_2 are open, convex, and bounded sets in $\mathbb{H}^n$, with $\overline{\Omega}_1 \subset \Omega_2$. Let $\partial\Omega_i$ be the boundary of Ω_i $(i = 1, 2)$, and let $\mathscr{U}(\partial\Omega_1)$ be a neighborhood of $\partial\Omega_1$ in Ω_2. Let B be a set of functions in $\mathscr{E}(\Omega_0 \cup \mathscr{U}(\partial\Omega_1))$ such that*

$$\left\{\frac{\partial f}{\partial \bar{q}_1}, \ldots, \frac{\partial f}{\partial \bar{q}_n} : f \in B\right\} \text{ is bounded in } \mathscr{E}^n\big(\Omega_0 \cup \mathscr{U}(\partial\Omega_1)\big).$$

If $H^1_{c,b}(\Omega_2, \mathscr{R}) = 0$, then for any $f \in B$, there exists $\tilde{f} \in \mathscr{E}(\Omega_2)$ such that:

(a) *$\{f - \tilde{f} : f \in B\}$ is a bounded set in $\mathscr{E}(\Omega_0 \cup \mathscr{U}'(\partial\Omega_1))$ for some neighborhood $\mathscr{U}'$ of $\partial\Omega_1$,*
(b) *$\{\frac{\partial \tilde{f}}{\partial \bar{q}_1}, \ldots, \frac{\partial \tilde{f}}{\partial \bar{q}_n}\}$ is bounded in $\mathscr{E}^n(\Omega_2)$.*

Proof For every $f \in B$, we can construct its extension g to Ω_2 and define $\tilde{\mathbf{g}} = (\frac{\partial g}{\partial \bar{q}_1}, \ldots, \frac{\partial g}{\partial \bar{q}_n})$. It is immediate to see that $\tilde{\mathbf{g}}$ belongs to the kernel of the operator $\mathbf{Q}(D)$ associated to the first syzygies and that we can assume that $\tilde{\mathbf{g}}$ is with compact support. By hypothesis, $H^1_{c,b}(\Omega_2, \mathscr{R}) = 0$, and so $\tilde{\mathbf{g}} = (\frac{\partial h}{\partial \bar{q}_1}, \ldots, \frac{\partial h}{\partial \bar{q}_n})$ for a suitable $h \in \mathscr{E}(\Omega_2)$. The family of extensions $\{\tilde{f} = g - h \mid f \in B\}$ now satisfies the theorem. □

Remark 10 In [6], the authors have discussed the explicit form of the syzygies for a wide variety of systems of differential equations. In addition to the study of the Cauchy–Fueter system, for example, the authors have identified the syzygies for the

Dirac systems, for the Moisil–Theodorescu system, and for many of their variations. It is quite simple to verify that the theorems proved in this paper for the Cauchy–Fueter system extend indeed to all the systems treated in [6]. It is possible, and worth investigating, that in fact the study of bounded cohomology may allow the proof of a general theorem on the extension of sets of approximate solutions for a large class of rectangular systems of differential equations.

References

1. Adams, W.W., Berenstein, C.A., Loustaunau, P., Sabadini, I., Struppa, D.C.: Regular functions of several quaternionic variables and the Cauchy–Fueter complex. J. Geom. Anal. **9**, 1–15 (1999)
2. Adams, W.W., Loustaunau, P.: Analysis of the module determining the properties of regular functions of several quaternionic variables. Pac. J. Math. **196**, 1–15 (2000)
3. Adams, W.W., Loustaunau, P., Palamodov, V.P., Struppa, D.C.: Hartogs' phenomenon for polyregular functions and projective dimension of related modules. Ann. Inst. Fourier **47**, 623–640 (1997)
4. Berenstein, C.A., Sabadini, I., Struppa, D.C.: Boundary values of regular functions of quaternionic variables. In: Deville, R., Esterle, J., Petkov, V., Sebbar, A., Yger, A. (eds.) Complex Analysis, Harmonic Analysis and Applications, pp. 220–232. Addison Wesley/Longman, Reading (1996)
5. Björk, J.E.: Rings of Differential Operators. North Holland, Amsterdam (1979)
6. Colombo, F., Sabadini, I., Sommen, F., Struppa, D.C.: Analysis of Dirac Systems and Computational Algebra. Progress in Mathematical Physics, vol. 39. Birkhäuser, Boston (2004)
7. Ehrenpreis, L.: A new proof and an extension of Hartogs' theorem. Bull. Am. Math. Soc. **67**, 507–509 (1961)
8. Ehrenpreis, L.: Fourier Analysis in Several Complex Variables. Wiley Interscience, New York (1970)
9. Fueter, R.: Die Funktionentheorie der Differentialgleichungen $\Delta u = 0$ und $\Delta\Delta u = 0$ mit vier reellen Variablen. Comment. Math. Helv. **7**, 307–330 (1934)
10. Hansen, S.: Localizable analytically uniform spaces and the fundamental principle. Trans. Am. Math. Soc. **264**, 235–250 (1981)
11. Hörmander, L.: An Introduction to Complex Analysis in Several Variables. 3rd edn. North-Holland Mathematical Library, vol. 7. North-Holland, Amsterdam (1990)
12. Kaup, L., Kaup, B., Barthel, G.: Holomorphic Functions of Several Variables: An Introduction to the Fundamental Theory. De Gruyter Studies in Mathematics, vol. 3. De Gruyter, Berlin (1983)
13. Komatsu, H.: Relative Cohomology of Sheaves of Solutions of Differential Equations. Springer Lecture Notes in Mathematics, vol. 287, pp. 192–261. Springer, Berlin (1973)
14. Palamodov, V.P.: Linear Differential Operators with Constant Coefficients. Springer, New York (1970)
15. Range, M.: Extension phenomena in multidimensional complex analysis: correction of the historical record. Math. Intell. **24**, 4–12 (2002)
16. Sabadini, I., Struppa, D.C., Sommen, F., Van Lancker, P.: Complexes of Dirac operators in Clifford algebras. Math. Z. **239**, 293–320 (2002)
17. Struppa, D.C.: The Fundamental Principle for Systems of Convolution Equations. Mem. Amer. Math. Soc., vol. 273 (1981)

18. Struppa, D.C.: The first eighty years of the Hartogs' theorem. In: Sem. di Geometria, Univ. di Bologna, pp. 127–209 (1988)
19. Struppa, D.C.: Extension of sets of approximate solutions to the Cauchy–Fueter system in two variables. Complex Var. Elliptic Equ. (2012, to appear). doi:10.1080/17476933.2011.584130
20. Sudbery, A.: Quaternionic analysis. Math. Proc. Phil. Soc. **85**, 199–225 (1979)

On Two Lacunary Series and Modular Curves

Ahmed Sebbar

Abstract We study, from different points of view, the two series $\chi_+(z) = \sum_{n\geq 0} z^{2^n}$ and $\chi_-(z) = \sum_{n\geq 0}(-1)^n z^{2^n}$. We show that the first series is related to the Jacobi theta function and the second is related to the Dedekind eta function and to the modular curve $X_0(14)$. We also present another approach to a celebrated identity of Hardy.

1 Introduction

As far as I know, Leon Ehrenpreis has never visited Bordeaux, but his name was very familiar to us in the 1980s, thanks to our teacher Roger Gay. Other names were also familiar: André Martineau, Bernard Malgrange, and Lars Hörmander. It happened that some theorems carry one or simultaneously two or three of these names. The main topics we studied were: Analytic functionals, division of distributions by a polynomial, existence of a fundamental solution for an arbitrary differential operator with constant coefficients and especially the fundamental principle. The book *Fourier Analysis in Several Complex Variables* of Leon Ehrenpreis [9], together with *Introduction to Complex Analysis in Several Variables* of Lars Hörmander [14], was our main concern. Several of us thought that *Fourier Analysis* contained many problems, good for very high-level *doctorat d'état*, the French doctoral dissertation at that time.

But the total mathematics of Leon Ehrenpreis is much broader and has an impact in many areas. In the present paper I will discuss some ideas on lacunary series and automorphic functions or forms, subjects that interested Leon too.

We study, in some details, the algebraic, analytic, and arithmetic properties of the two lacunary series

$$\chi_+(z) = \sum_{n\geq 0} z^{2^n}, \qquad \chi_-(z) = \sum_{n\geq 0}(-1)^n z^{2^n}, \quad |z| < 1, \tag{1}$$

A. Sebbar (✉)
Institut de Mathématiques de Bordeaux, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence cedex, France
e-mail: ahmed.sebbar@math.u-bordeaux1.fr

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_18, © Springer-Verlag Italia 2012

already considered under a different angle in [18]. I am pleased to observe that the present point of view is close, in some points, to Ehrenpreis work [10]. The series (1) verify the functional equations

$$\chi_+(z) = z + \chi_+\big(z^2\big), \qquad \chi_-(z) = z - \chi_-\big(z^2\big), \tag{2}$$

and the main objective is to (try to) study their singularities on the unit circle. To do so, we first try to find a relationship between $\chi_\pm$ and automorphic functions, in particular theta functions, while they seem unrelated. The main results are: The first power series is related to the fourth powers of the Jacobi theta functions, and the second is related to an eta product and to the modular curve $X_0(14)$. This is the content of main Theorems 6.5 and 6.6 of this work. Theorem 3.1 gives a new presentation of a celebrated identity of Hardy.

It is surprising in several respects that the functions χ_+ and χ_- are related. One is tempted to think that there must be a hidden relationship between $\chi_+(z)$ and $\chi_-(z)$, despite their differences. Among the things which separate χ_+ and theta functions, there is the behavior at 1. When x is real and tends to 1, $x < 1$,

$$\sum_{n=0}^{\infty} x^{n^2} \sim \frac{\sqrt{\pi}}{2\sqrt{1-x}} \sim \frac{\sqrt{\pi}}{2\sqrt{-\log x}}, \qquad \sum_{n=0}^{\infty} x^{2^n} \sim -\frac{\log(1-x)}{\log 2}.$$

Our main reference for modular forms and theta functions is Zagier [27] and Ono [20], and for elliptic curves is Silvermann [24]. For the sequel, we set $z = e^{2i\pi\tau}$.

Definition 1.1 The theta function, associated to the Dirichlet character ψ is the series given by

$$\theta_\psi(\tau) = \theta_\psi(z) = \sum_n \psi(n) e^{2i\pi n^2\tau} = \sum_n \psi(n) z^{n^2}$$

if ψ is even, and if ψ is odd, by

$$\theta_\psi(\tau) = \theta_\psi(z) = \sum_n \psi(n) n e^{2i\pi n^2\tau} = \sum_n \psi(n) n z^{n^2}.$$

The summations are over positive integers, unless ψ is a trivial character, in which case the summation is over all integers.

In the case of a trivial character, we write $\theta_\psi = \theta$, and we know that

$$\theta(\tau) = \prod_{n=1}^{\infty}\big(1 - z^{2n}\big)\big(1 + z^{2n-1}\big)^2 = \sum_{-\infty}^{\infty} z^{n^2}.$$

The Dedekind eta function is

$$\eta(\tau) = z^{\frac{1}{24}} \prod_{n=1}^{\infty}\big(1 - z^n\big) = z^{\frac{1}{24}} \sum_{-\infty}^{\infty} (-1)^n z^{\frac{3n^2+n}{2}} = \sum_{n=1}^{\infty} \left(\frac{12}{n}\right) z^{\frac{n^2}{24}}, \tag{3}$$

where $(\frac{12}{\bullet})$ is the Legendre–Jacobi symbol. Hence the eta function is a theta function, and in fact if $\chi_{12} = (\frac{12}{\bullet})$, then

$$\eta = \theta_{\chi_{12}},$$
$$\chi_{12}(m) = \left(\frac{12}{m}\right) = \begin{cases} 1 & \text{for } m \equiv \pm 1 \bmod 12 \\ -1 & \text{for } m \equiv \pm 5 \bmod 12. \end{cases} \tag{4}$$

On the other hand, we have the identity [20]

$$\theta(z) = \theta(\tau) = \eta^5(2\tau)\eta^{-2}(\tau)\eta^{-2}(4\tau),$$

which can be considered as an inversion of (3).

2 Algebraic Differential Equations, Differential Operators of Infinite Order

In this section we compare the differential properties of $\chi_\pm(z)$ and $\theta(z)$. An algebraic differential equation (ADE) is an equation of the form

$$P\big(z, \omega, \omega', \dots, \omega^{(k)}\big) = 0,$$

where P is a polynomial over $\mathbb{C}$ in all its variables [15]; k is called the order of the equation, and its degree is the degree of the polynomial $P(X, X_0, X_1, \dots, X_k)$. A power series which satisfies an ADE is called differentially algebraic. A classical theorem of Maillet, Ostrowski, and Popken asserts that if $\sum f_{n_k} z^{n_k}$, $f_{n_k} \in \mathbb{C}$, $f_{n_k} \neq 0$ is differentially algebraic, then necessarily $\limsup \frac{n_{k+1}}{n_k} < \infty$. Nevertheless:

Proposition 2.1 *The functions $\chi_+(z)$ and $\chi_-(z)$ are not differentially algebraic.*

This result can be explained in two ways; the first is through the functional equation (2) as in [19] or [15]. It should be noted that the lacunarity of a differentially algebraic power series is specified by the following result of Maillet [17].

Theorem 1 (Maillet) *Let $\sum_{m=0}^{\infty} b_m z^m$ be a given series, formal solution of a differential equation of order k and degree μ. We can find a fixed number τ_0, independent of m, such that for large m, if b_m is nonzero, the previous nonzero coefficient has an index greater than or equal to $\frac{(m+\tau_0)}{\mu}$.*

The function $\theta(z)$, instead, is differentially algebraic. It satisfies the Jacobi differential equation

$$\big(y^2 y''' - 15 y y' y'' + 30 y'^3\big)^2 + 32\big(y y'' - 3y'^2\big)^3 = y^{10}\big(y y'' - 3y'^2\big)^2,$$

where "$\,'\,$" stands for $\frac{1}{2i\pi}\frac{d}{d\tau} = z\frac{d}{dz}$, $z = e^{2i\pi\tau}$.

On the other hand, following Sato, Kawai, and Kashiwara [23], consider

$$Q_1 = \frac{d}{d\tau} \prod_{n=1}^{\infty} \left(1 - \frac{1}{2i\pi n^2} \frac{d}{d\tau}\right).$$

Using

$$\sinh \zeta = \zeta \prod_{n=1}^{\infty} \left(1 + \frac{\zeta^2}{\pi^2 n^2}\right),$$

we have that

$$Q_1 = \sqrt{2i\pi \frac{d}{d\tau}} \sinh \sqrt{2i\pi \frac{d}{d\tau}}$$

or

$$Q_1 = \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} \left(2i\pi \frac{d}{d\tau}\right)^{j+1}.$$

Q_1 is a well-defined differential operator of infinite order, with

$$Q_1\left(\frac{d}{d\tau}\right)\theta(\tau) = 0.$$

Similarly, we have the following:

Proposition 2.2 *The infinite product*

$$h(\tau) = \prod_{n=0}^{\infty} \left(1 - \frac{\tau}{2^n}\right)$$

is a unique solution of the functional equation

$$h(\tau) = (1 - \tau) h\left(\frac{\tau}{2}\right), \quad h(0) = 1.$$

It represents an entire function of zero exponential type with the following expansion:

$$h(\tau) = 1 + \sum_{n=1}^{\infty} \frac{2^n}{(1-2^n)(1-2^{n-1})\cdots(1-2)} \tau^n,$$

and hence

$$h\left(\frac{d}{d\tau}\right) = 1 + \sum_{n=1}^{\infty} \frac{2^n}{(1-2^n)(1-2^{n-1})\cdots(1-2)} \frac{d^n}{d\tau^n}$$

is a well-defined operator of infinite order with

$$h\left(\frac{d}{d\tau}\right)\chi_+(z) = 0, \qquad h\left(\frac{d}{d\tau}\right)\chi_-(z) = 0.$$

The functions $\chi_\pm$ and θ share the following analytic property: the associated Dirichlet series are solutions to infinite-order differential operators, and hence the lines of convergence are natural boundaries. It is possible to show that, actually, the Dirichlet series associated to (1) cannot be continued neither analytically nor quasi-analytically across the line of convergence $\Im\tau = 0$. This means that the power series have the unit circle as an analytic and quasi-analytic natural boundary [18].

The functions $\chi(z)$ and $\theta(z)$ share some arithmetical properties. Mahler [16] showed that for every $d \geq 2$, the value of $\sum_{n\geq 0} z^{-d^n}$ at any nonzero algebraic number is transcendental. For theta functions, the situation is richer, and we refer to [5] and [3] for various results and conjectures.

3 New Look at Hardy's Identity Concerning the Series χ_+

We would like to interpret Hardy's identity [11]

$$\begin{aligned}\sum_{n\geq 0} x^{2^n} &= \sum_{n\geq 1} \frac{(-1)^{n-1}(\log 1/x)^n}{n!(2^n-1)} - \frac{1}{\log 2}\log(\log 1/x) + \frac{1}{2} - \frac{\gamma}{\log 2} \\ &\quad - \frac{1}{\log 2}\sum_{k\neq 0}\Gamma\left(\frac{-2ki\pi}{\log 2}\right)(\log 1/x)^{2ki\pi/\log 2}, \quad 0 < x < 1, \end{aligned} \tag{5}$$

from another point of view by giving a description in terms of difference equations. We consider the power series

$$G(z) = G(z,a) = \sum_{n\geq 1} \frac{1}{a^n - 1}\frac{z^n}{n!}, \tag{6}$$

where $a \neq 1$ is a given positive real number. The function G is an entire function solution of the functional equation

$$G(z,a) = -G\left(\frac{z}{a}, \frac{1}{a}\right). \tag{7}$$

We introduce a new variable u by $z = -a^u$ and the function

$$\mathscr{H}(u) = -zG'(z). \tag{8}$$

Then

$$\mathscr{H}(u+1) - \mathscr{H}(u) = j_a(u) = a^u e^{-a^u}. \tag{9}$$

The function j_a belongs to the Schwartz space $\mathscr{S}(\mathbb{R})$ of infinitely differentiable functions f such that $u^n f^{(m)}$ is bounded on $\mathbb{R}$. In particular, each of the two series $\sum_{m\geq 0} j_a(u+m)$ and $\sum_{m\geq 1} j_a(u-m)$ converge uniformly on compact sets of $\mathbb{R}$.

Theorem 3.1 *Equation (9) has three entire solutions $\mathscr{K}(u)=\mathscr{K}(u,a)$, $\mathscr{K}_1(u)=\mathscr{K}_1(u,a)$, and $\mathscr{K}_2(u)=\mathscr{K}_2(u,a)$, with*

1. $\mathscr{K}(u)=-\sum_{n\geq 0}\frac{(-1)^{n+1}}{a^{n+1}-1}\frac{a^{(n+1)u}}{n!}$;
2. $\mathscr{K}_1(u)=\sum_{m\geq 1} j_a(u-m)$;
3. $\mathscr{K}_2(u)=-\sum_{m\geq 0} j_a(u+m)$;
4. $\mathscr{K}(u)=\mathscr{K}_1(u)$ *if* $a>1$; $\mathscr{K}(u)=\mathscr{K}_2(u)$ *if* $a<1$;
5. $\mathscr{K}_1(u)-\mathscr{K}_2(u)=P(u)$ *is a periodic function of period* 1 *given by*

$$P(u)=\sum_{-\infty}^{+\infty}\Gamma\left(1-\frac{2in\pi}{\log a}\right)e^{2n\pi iu};$$

6. *The map which carries $\mathscr{K}_1(u)$ on $\mathscr{K}_2(u)$ is represented by the functional relation (7)* $G(z,a)=-G(\frac{z}{a},\frac{1}{a})$;
7. *The following identity holds for all* $a>1$:

$$\sum_{n\geq 1}\frac{(-1)^n}{a^n-1}\frac{a^{nu}}{n!}=-\sum_{m\geq 1}\left(e^{-a^{u-m}}-1\right).$$

The proof of the theorem is based on the difference equation (9). Since the second term j_a is in the Schwartz space $\mathscr{S}(\mathbb{R})$, we can solve this equation by right iteration, left iteration, or localization. More explicitly, let D be the differentiation operator $\frac{d}{dx}$. By Taylor's formula, (9) takes the form

$$\left(e^D-1\right)\mathscr{H}(u)=j_a(u)$$

or

$$\left(1-e^{-D}\right)\mathscr{H}(u)=j_a(u-1).$$

The formal inverse of the operator $(1-e^{-D})$ is $\sum_{n\geq 0} e^{-nD}$ so that (9) has the solution $\mathscr{K}_1(u)=\sum_{m\geq 1} j_a(u-m)$, which is in fact an entire function, the series being a series of holomorphic functions on $\mathbb{C}$ uniformly convergent on compact sets.

Equation (9) written in the form $(1-e^D)\mathscr{H}(u)=-j_a(u)$ can be solved formally by right iteration $\mathscr{K}_2(u)=-\sum_{m\geq 0} j_a(u+m)$, and we obtain another entire solution.

The third method to solve (9) is the localization method. We define the operator $D^{-1}=\frac{1}{D}$ by

$$D^{-1}g(u)=\int_0^u g(x)\,dx;$$

then (9) takes the form

$$\left(\frac{e^D-1}{D}\right)\mathscr{H}(u)=\int_0^u j_a(x)\,dx. \tag{10}$$

The function $\frac{z}{e^z-1}$ is holomorphic in the disc centered at the origin and of radius 2π and has the power series expansion

$$\frac{z}{e^z-1}=\sum_{n=0}^{\infty}\frac{b_n}{n!}z^n, \quad |z|<2\pi.$$

The coefficients b_n are the Bernoulli numbers

$$b_0=1, \quad b_1=-\frac{1}{2}, \quad b_2=\frac{1}{6}, \quad b_3=0, \quad b_4=\frac{1}{30}, \quad \dots.$$

A formal solution of (10) is

$$\mathscr{K}_3(u)=D^{-1}j_a(u)+\sum_{n\le 1}\frac{b_n}{n!}D^nD^{-1}j_a(u),$$

and an easy calculation shows that

$$\mathscr{K}_3(u)=a^u\sum_{m\ge 0}\frac{(-1)^m a^{um}}{m!(a^{m+1}-1)}.$$

We observe that we have recovered the function $\mathscr{K}_3(u)=\mathscr{K}(u)$ from which we derived the difference equation (9).

The difference $\mathscr{K}_1(u)-\mathscr{K}_2(u)$ is a solution of the homogeneous equation

$$\mathscr{H}(u+1)-\mathscr{H}(u)=0,$$

so it is a periodic function of period 1. The Fourier coefficients can be computed as in the preceding section. The map $a\mapsto\frac{1}{a}$ is a homeomorphism from (0, 1) onto $(1,+\infty)$, so we can suppose in the sequel that $a>1, z=a^u$. Relation (7) gives $-zG'(z,a)=\frac{z}{a}G'(\frac{z}{a},\frac{1}{a})$, which is equivalent to $\mathscr{K}_1(u)=\mathscr{K}_2(1-u,\frac{1}{a})$. Finally, if we integrate from $-\infty$ to u the identities

$$\sum_{m\ge 1}j_a(x-m)=\mathscr{K}_1(u)=\mathscr{K}(u)$$

and use

$$\int_{-\infty}^{u}j_a(x-m)dx=\frac{-1}{\log a}\left(e^{-a^{u-m}}-1\right),$$

we obtain the last statement of the proposition.

Remark 1 We have seen that Hardy's identity can be considered as a connection formula between two solutions to the same difference equation. In principle, other connection formulas could be obtained by considering other solutions. However, according to the asymptotic formula

$$\left|\Gamma(x+iy)\right| = \sqrt{2\pi}e^{-\frac{1}{2}\pi|y|}|y|^{x-\frac{1}{2}}\big(1+r(x,y)\big),$$

where as $|x+iy|$ tends to ∞, $r(x,y)$ tends to 0, uniformly in the strip $|x| \le \alpha$, where α is a constant, the connection

$$\mathscr{K}_1(u) - \mathscr{K}_2(u) = P(u) = \sum_{-\infty}^{+\infty} \Gamma(1-2n\pi i A)e^{2n\pi i u}$$

with $A = \frac{1}{\log a}$ is very small on the real axis; two different solutions of the same difference equation can be very close. This has been remarked in another context by Ramanujan and Hardy (see the discussion in [4]; Entry 17) and more recently by Tricomi [25], who found that for $a = 2$ or $A = \frac{1}{\log 2} = 1.4427\ldots$, one has that, for all $x \in \mathbb{R}$, $|P(x) - A| < 0.0000143\ldots$, whereas Berndt quotes in [4] the approximation

$$\log 2\left(\sum_{n=0}^{\infty} 2^n e^{-2^n x} + \sum_{n=0}^{\infty} \frac{(-x)^n}{(2^{n+1}-1)n!}\right) \approx \frac{1}{x}\left(1 + 0.0000098844\cos\left(\frac{2\pi\log x}{\log 2} + 0.872711\right)\right).$$

Now we would like to investigate how the functional relations (7),(8) can be linked to modular properties of some Eisenstein series. The function (6),

$$G(z) = \sum_{n\ge 1} \frac{1}{a^n - 1}\frac{z^n}{n!},$$

is an entire function of order 1 and of exponential type $\frac{1}{a}$, and its Borel transform is

$$B(z,a) = \sum_{n\ge 1} \frac{1}{a^n-1} z^{-(n+1)}, \quad |z| > \frac{1}{a}.$$

Following a presentation of Bochner, we show that the function

$$P(x) = G(\log x) = G(\log x, a) = \sum_{n\ge 1} \frac{1}{a^n-1}\frac{(\log x)^n}{n!}, \quad x \in \mathbb{C}^*, \qquad (11)$$

verifies a functional equation almost like the Riemann zeta function.

Definition 3.1 A function $F(x)$ is residual if the three following conditions are fulfilled:

1. It is defined and differentiable for $0 < x < \infty$, and for some $\gamma > 0$, we have $F(x) = O(x^{-\gamma})$ as x tends to 0 and $F(x) = O(x^{\gamma})$ as x tends to $+\infty$, so that we may introduce for complex s the Mellin transforms

$$\xi_r(s) = \int_0^1 F(x)x^{s-1}\,dx, \qquad \xi_l(s) = \int_1^{+\infty} F(x)x^{s-1}\,dx;$$

   ξ_r is defined in a right half-plane, and ξ_l in a left-half plane.
2. These two functions can be continued into one another in a domain D which is the exterior of a bounded set S.
3. If we denote the joint value of the continued function by $\xi_0(s)$, $s = \sigma + i\tau$, then we have

$$\lim_{|\tau| \mapsto +\infty} \xi_0(\sigma + i\tau) = 0$$

   uniformly on finite intervals $\sigma_1 \le \sigma \le \sigma_2$.

This definition is motivated by Hecke's theory on modular forms [26] that will be also considered in Sect. 6. That the function $P(x)$ is residual is a consequence of the fact that $G(z)$ is of exponential type $\frac{1}{a}$ and of order 1 and from the following theorem of Bochner [6].

Theorem 3.2 *A function $F(x)$ defined for $0 < x < \infty$ is residual if and only if it can be represented as a series*

$$F(x) = \sum_{n\ge 0} \frac{c_n}{n!}\left(\log\frac{1}{x}\right)^n, \qquad \gamma_0 = \limsup |c_n|^{\frac{1}{n}} < \infty.$$

The function $\xi_0(s)$ is the Borel transform $\xi_0(s) = \sum_{n\ge 0} \frac{c_n}{s^{n+1}}$ of the entire function of exponential type $F(e^{-y}) = \sum_{n\ge 0} \frac{c_n}{n!} y^n$.

Instead of $\xi_0(s)$, we consider

$$\xi(z,a) = \xi(z) = \sum_{n\ge 1} \frac{1}{a^n - 1} z^n, \qquad |a| > 1. \tag{12}$$

This function has some interesting properties. We first prove the following:

Proposition 3.3 *The function $\xi(z,a)$ has a meromorphic continuation to the whole plane with simple poles at a^n, $n \in \mathbb{N}^*$, and with residues $-a^n$.*

In fact, from the definition we see that, for $|z| < 1$,

$$\xi(az,a) - \xi(z,a) = -\frac{z}{z-1},$$

so that

$$\xi(z,a) = z\sum_{n\geq 1}\frac{1}{a^n - z}.$$

The last series converges uniformly on compact sets of $\mathbb{C}\setminus\{a^n, n\in\mathbb{N}^*\}$, and the result follows.

Let $B(t)$ be the function defined for $0 < t < \infty$ by

$$B(t) = \frac{1}{4} - \frac{\pi}{12}t + 2\pi t\frac{\partial\xi}{\partial z}\left(1, e^{2\pi t}\right).$$

Proposition 3.4 *The function B verifies the functional relation*

$$B(t) + B\left(\frac{1}{t}\right) = 0.$$

In particular,

$$\sum_{n\geq 1}\frac{n}{e^{2\pi n} - 1} = \frac{1}{24} - \frac{1}{8\pi} = 0,0018779\ldots.$$

The last equality appears in Ramanujan [22] (p. 326, question 387). The proof of the proposition is an immediate consequence of the modular transformation law of the Eisenstein series

$$G_2(\tau) = -\frac{1}{24} + \sum_{n\geq 1}\frac{n}{e^{-2i\pi n\tau} - 1} = -\frac{1}{24} + \sum_{n\geq 1}\sigma_1(n)e^{2i\pi n\tau}, \quad \sigma_1(n) = \sum_{d|n} d, \tag{13}$$

according to which for all $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we have

$$G_2\left(\frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right) = (\gamma\tau + \delta)^2 G_2(\tau) - \frac{\gamma(\gamma\tau + \delta)}{4i\pi}.$$

We obtain the desired formula by choosing $\alpha = \delta = 0$, $\beta = -1$, $\gamma = 1$, and $\tau = it$, $t > 0$.

Remark 2 The series G_2 in (13) verifies the Chazy equation

$$D^3y + 24yD^2y - 36(Dy)^2 = 0; \quad Dy = \frac{1}{2i\pi}\frac{d}{d\tau},$$

and hence it is differentially algebraic.

Now we show that the basic function (12) is related to the Eisenstein series G_2. We shall suppose that $q \in \mathbb{C}$, $|q| < 1$, and introduce

$$\chi(z,q) = \sum_{n \geq 1} \frac{z^n}{1-q^n} \tag{14}$$

and

$$f(z,q) = \sum_{n \geq 1} \frac{z^n}{n(1-q^n)}. \tag{15}$$

We clearly have $\chi(z,q) = z\frac{\partial}{\partial z} f(z,q)$. For $|z| < 1$, we have

$$\chi(z,q) = \xi\left(\frac{z}{q}, \frac{1}{q}\right)$$

and, for $|z| > q$ and $a = \frac{1}{q}$,

$$B(z,a) = \frac{1}{z}\xi\left(\frac{1}{z}, a\right) = \frac{1}{z}\chi\left(\frac{q}{z}, q\right).$$

Theorem 3.5 *Let $F(z,q)$ be defined for $|z| < 1$ by*

$$F(z,q) = \frac{1}{(1-q)(1-qz)(1-q^2z)\cdots}.$$

Then

$$\chi(z,q) = z\frac{F'(z,q)}{F(z,q)}.$$

The main idea of the proof is in [12], where it is shown that

$$e^{f(z,q)} = F(z,q)$$

with the following power series expansion:

$$F(z,q) = 1 + \sum B_k(q)z^k, \quad B_k(q) = \frac{1}{(1-q)(1-q)(1-q^2)\cdots(1-q^n)}.$$

By taking the inverse Borel transform we obtain

$$G(z,a) = \frac{1}{2i\pi}\int_{\mathscr{C}_0} B(w,a)e^{zw}\,dw,$$

where $\mathscr{C}_0$ is a circle positively oriented, centered at the origin, and with radius strictly greater than 1. An easy manipulation shows that finally, for every $z \in \mathbb{C}$,

$$G(z,a) = \frac{1}{2i\pi}\int_{\mathscr{C}_1} \frac{\frac{\partial}{\partial x}F(x,q)}{F(x,q)}e^{\frac{qz}{x}}\,dx;$$

here $\mathscr{C}_1$ is a circle positively oriented, centered at the origin, and with radius strictly smaller than 1. This is the desired link between the central function of this work $G(z,a)$ and the partition function.

Remark 3.1 The power series (14) and (15) defined initially for $a \in \mathbb{C}$, $|a| \neq 1$, can also be considered for $a = e^{i\pi\alpha}$ with irrational α. Although denominators $1 - e^{i\pi n\alpha}$ never vanish, their modulus $|1 - e^{i\pi n\alpha}|$ can become very small (phenomenon of small divisors). The result is that the power series (14) and (15) may have any radius of convergence ρ, $0 \le \rho \le 1$, according to the arithmetic nature of α [12].

4 Decomposition Theorems, Singularity at $x = 1$

The function $\chi_+(x) = \sum_{n\ge 0} x^{2^n}$, $0 < x < 1$ solves the functional equation (2),

$$f(x) = f\left(x^2\right) + x.$$

The difference of two particular solutions to it is a periodic function of $\log\log\frac{1}{x}$. Another solution of (2) is

$$\sum_{n\ge 1} \frac{(-1)^{n-1}(\log 1/x)^n}{n!(2^n - 1)} - \frac{1}{\log 2}\log(\log 1/x),$$

leading to the identity of Hardy (5),

$$\sum_{n\ge 0} x^{2^n} = \sum_{n\ge 1} \frac{(-1)^{n-1}(\log 1/x)^n}{n!(2^n - 1)} - \frac{1}{\log 2}\log(\log 1/x) + p(x)$$

with

$$p(x) = \frac{1}{2} - \frac{\gamma}{\log 2} - \frac{1}{\log 2}\sum_{k\neq 0} \Gamma\left(\frac{-2ki\pi}{\log 2}\right)(\log 1/x)^{2ki\pi/\log 2}.$$

A much more general result in this direction is the following [1]. If

$$\phi(t) = \sum_{k\ge 1} a_k t^k, \qquad f_\phi(x) = \sum_{n\in\mathbb{Z}} \phi\left(x^{\theta^n}\right), \qquad D_\phi(s) = \sum_{k\ge 1} \frac{a_k}{k^s},$$

then the following formula holds:

$$f_\phi(x) = -\frac{1}{\log\theta} D'_\phi(0) + \frac{1}{\log\theta} \sum_{n\in\mathbb{Z}\setminus\{0\}} \Gamma\left(-\frac{2\pi m}{\log\theta}\right) D_\phi\left(-\frac{2\pi m}{\log\theta}\right) \exp\left(2\pi m \frac{\log\log x^{-1}}{\log\theta}\right).$$

5 Singularities on the Circle

Theorem 5.1 (Wigert's theorem [10]) *Let $g(z)$ be an entire function of class $(1,0)$, that is, an entire function of at most zero type and of order one*:

$$\forall\epsilon > 0, \exists C_\epsilon > 0, \quad |g(z)| \le C_\epsilon e^{\epsilon|z|}.$$

The function defined by the series

$$f(z) = C + \sum_{n\ge1} g(n) z^n, \quad |z| < 1, \tag{16}$$

extends analytically to all the plane (including at ∞) except at $z = 1$. Conversely, if $f(z)$ is a function with the above regularity properties, then there is a function $g(z)$ of class $(1,0)$ such that (16) holds in the unit disc. Moreover if $g(z)$ is a polynomial, $f(z)$ is a rational function of $\frac{1}{1-z}$.

It is possible to interpret Wigert's theorem slightly differently by using a differential operator of infinite order and a Poisson kernel for the unit disc. Because the entire function $g(z)$ is of exponential type zero, it defines a differential operator of infinite order $g(\mathscr{D})$, $\mathscr{D} = \frac{d}{d\zeta}$. We set $g(z) = \sum_{n\ge0} b_n z^n$, $z \in \mathbb{C}$, $z = e^\zeta$, and we obtain by absolute convergence

$$\sum_{n\ge0} g(n) z^n = g(\mathscr{D}) \frac{1}{1-e^\zeta}, \quad \mathscr{D} = \frac{d}{d\zeta} = zD. \tag{17}$$

By the main property of differential operators of infinite order (as local operators) and by the holomorphy of the function $\frac{1}{1-e^\zeta}$ in $\mathbb{C} \setminus 2i\pi\mathbb{Z}$, we see that $\sum_{n\ge0} g(n) z^n$ extends analytically to all $\mathbb{C} \setminus \{1\}$. This is Wigert's theorem (16).

We can think of (17) as a representation theorem that we extend in the following manner. To $g(z) = \sum_{n\ge0} b_n z^n$ we associate a second infinite-order operator

$$P(D) = \frac{1}{2} \sum_{n\ge0} (-i)^n b_n \frac{d^n}{dz^n}.$$

We denote, as usual, by H_r the classical Poisson kernel

$$H_r(\theta - t) = \frac{e^{it} + re^{i\theta}}{e^{it} - re^{i\theta}} = \frac{e^{it} + z}{e^{it} - z}, \quad z = re^{i\theta}, \ 0 < r < 1.$$

Let, for $n \ge 0$, $\phi_n(z) = \mathscr{D}^n \frac{1}{1-z} = \sum_{p\ge0} p^n z^p$, $|z| < 1$. Then

$$f(z) = \sum_{n\ge0} b_n \phi_n(z).$$

The functions $\phi_n(z)$ and the Poisson kernel are related by

$$H_r(\theta) = \frac{1+z}{1-z} = 2\phi_0(z) - 1, \qquad H_r^{(n)}(\theta) = \frac{d^n}{d\theta^n} H_r(\theta) = 2i^n \phi_n\big(re^{i\theta}\big), \quad n \geq 1.$$

In conclusion,

$$f(z) = \frac{b_0}{2} + \frac{1}{2} \sum_{n \geq 0} b_n (-i)^n H_r^{(n)}(\theta).$$

We now make the following simple observation: Let $(m_n)_{n\geq 0}$ a sequence of positive real numbers. For $|z| \neq 1$, we clearly have

$$\frac{1+z^{m_n}}{1-z^{m_n}} = \frac{1+z^{m_0}}{1-z^{m_0}} + \sum_{\nu=1}^{n} \left\{ \frac{1+z^{m_\nu}}{1-z^{m_\nu}} - \frac{1+z^{m_{\nu-1}}}{1-z^{m_{\nu-1}}} \right\}.$$

We deduce, for $m_\nu = 2^\nu,\ \nu \geq 0$, that

$$\frac{1+z}{1-z} + \frac{2z}{z^2-1} + \frac{2z^2}{z^4-1} + \frac{2z^4}{z^8-1} + \cdots = \psi(z),$$

where $\psi(z) = 1$ if $|z| < 1$, $\psi(z) = -1$ if $|z| > 1$, or

$$\sum_{\nu \geq 0} \frac{1}{z^{2^\nu} - z^{-2^\nu}} = \phi(z) = \begin{cases} \frac{z}{z-1} & \text{if } |z| < 1, \\ \frac{1}{z-1} & \text{if } |z| > 1. \end{cases}$$

Introducing the Möbius function μ, we obtain the decomposition of the function χ in Lambert series (see Lemma 6.4)

$$\chi(z) = {\sum_{n \geq 1}}' \mu(n) \frac{z^{2^n}}{1-z^{2^n}}, \tag{18}$$

the summation being only on positive odd integers. This is not a decomposition of Mittag-Leffler type. Moreover, we observe that the series $f(z) = \sum_{n\geq 1}' \frac{z^{2^n}}{1-z^{2^n}}$ verifies, similarly to (2),

$$f(z) = \frac{z}{1-z} + f\big(z^2\big).$$

Remark 5.1 The function $\theta^4(\tau)$ admits also the well-known decomposition in Lambert series ([21], p. 198), similar to (18),

$$\theta^4(\tau) = 1 + 8 \sum_{m=1}^{\infty} \frac{nq^n}{1+(-1)^n q^n}, \qquad q = e^{2i\pi\tau}. \tag{19}$$

But this decomposition is not really of Mittag-Leffler type.

We seek to find a generalized Mittag-Leffler decomposition into partial fractions, each term giving rise to an essential singularity at the root of unity, $z = e^{\frac{2i\pi h}{k}}$. To this end, we work first with the theta function. We recall some properties of singular series, as developed by Hardy (see the complete list of references in [2]). Let $r_s(n)$ denote the number of representations of the positive integer n as a sum of s squares, that is,

$$1 + \sum_{n=1}^{\infty} r_s(n) e^{2i\pi n\tau} = \left(\sum_{-\infty}^{\infty} e^{2in\pi\tau} \right)^s = \theta(\tau)^s, \quad \Im\tau > 0.$$

Hardy proved that for $s = 5, 6, 7, 8,\ r_s(n) = \rho_s(n)$ is given by

$$r_s(n) = \rho_s(n) = \frac{\pi^{\frac{s}{2}}}{\Gamma(\frac{s}{2})} n^{\frac{s}{2}} \mathscr{S}_s(n). \tag{20}$$

$\mathscr{S}_s(n)$ is called singular series and is given more explicitly by

$$\mathscr{S}_s(n) = \sum_{k=1}^{\infty} B_k(n), \quad B_k(n) = k^{-\frac{s}{2}} \sum_{h \bmod 2k} \eta(h,k) e^{i\pi n \frac{h}{k}}, \tag{21}$$

where $\eta(h,k) = 0$ if $(h,k) > 1$, and if $(h,k) = 1$,

$$\eta(h,k) = \frac{1}{2} k^{-\frac{1}{2}} \sum_{j \bmod 2k} e^{i\pi h \frac{j^2}{k}} = \frac{1}{2} \left\{ 1 + (-1)^{hk} \right\} k^{-\frac{1}{2}} \sum_{j=1}^{k} e^{i\pi h \frac{j^2}{k}}. \tag{22}$$

$\eta(h,k)$ is an eighth root of unity if h and k are relatively prime and of opposite parity; otherwise, $\eta(h,k)$ is zero. For reasons that will become clear later, we are concerned with the case $s = 4$, a case qualified as exceptional by Hardy and worked out by Bateman [2], who shows that if

$$\Psi_4(\tau) = 1 + \sum_{n=1}^{\infty} \rho_s(n) e^{2i\pi n\tau},$$

then

$$\Psi_4(\tau) = \frac{1}{2} \sum_{k} \sum_{(h,k)=1, h \equiv k+1 (\mathrm{mod}\ 2)} \frac{(-1)^k}{(h - k\tau)^2}. \tag{23}$$

We can remove the condition $(h,k) = 1$ by using Mertens' theorem that the Cauchy product of a convergent series and an absolutely convergent series is convergent to the product of the sums. Multiplying both sides of (23) by $\frac{\pi^2}{8} = \sum_{n \geq 0} \frac{1}{(2n+1)^2}$, we obtain

$$\frac{\pi^2}{8} \Psi_4(\tau) = \frac{1}{2} \sum_{k} \sum_{h \equiv k+1 (\mathrm{mod}\ 2)} \frac{(-1)^k}{(h - k\tau)^2}, \tag{24}$$

from which, by using modular forms properties of $\Psi_4(\tau)$ and $\theta^4(\tau)$, we obtain that $\Psi_4(\tau) = \theta^4(\tau)$. Hence, we have the following:

Theorem 5.2 *The function $\theta^4(\tau)$ admits the generalized Mittag-Leffler decomposition into partial fractions*

$$\theta^4(\tau) = \frac{4}{\pi^2} \sum_k \sum_{h \equiv k+1 (\mathrm{mod}\ 2)} \frac{(-1)^k}{(h - k\tau)^2}, \tag{25}$$

which is a good substitute for (19).

6 A Brief Summary of the Modularity Theorem

6.1 Elliptic Curves

In this section, we give an idea on the relation connecting some lacunary series (like χ_+, χ_-) and/or elliptic curves and modular forms, taking into account, of course, the existence of a strong relationship between these two last concepts. We wish to develop this theme in a later publication. The reason is that the moduli of elliptic curves are expressible in terms of modular forms of the parameter τ, $\Im\tau > 0$. Moreover, in the 1950s a precise relation between elliptic curves and modular forms was formulated, first by Taniyama, then by Shimura and Weil. We refer to [7] for the details. Let E be an elliptic curve defined over Q, say by an equation $y^2 = 4x^3 - ax - b$ with rational integers a, b. For every prime p not dividing the discriminant $\Delta = a^3 - 27b^2$, we get an elliptic curve over the finite field F_p with this equation. We therefore have its zeta function, the numerator of which has the shape $1 - a_p t + pt^2$, with a_p defined by counting the number of solutions to the congruence $y^2 \equiv 4x^3 - ax - b \bmod p$,

$$1 - a_p + p = \sharp(F_p).$$

Note that $\sharp(F_p)$ is actually one more than the number of solutions to the congruence, since E has one point at infinity in the projective plane. Following Hasse, we consider the infinite product

$$L(E, s) = \prod_p \left(1 - a_p^{p-s} + p^{1-2s}\right)^{-1}.$$

Then Wiles' theorem, conjectured by Taniyama, Shimura, and Weil, is the following [7]:

Theorem 6.1 (Modularity theorem) *Let E be an elliptic curve E over $\mathbb{Q}$ with conductor N. Then there exists $f(\tau)$, a cusp form of weight 2 for $\Gamma_0(N)$, such that $L(f, s) = L(E, s)$.*

In a very simplified way, Wiles' theorem says the following: Let an elliptic curve E be defined by an equation of the form $f(x, y) = 0$, $f(x, y) \in \mathbb{Z}[x, y]$, and for any prime p not dividing its discriminant, let $E(F_p)$ denote the number of solutions to the congruence $f(x, y) = 0 \bmod p$ including the point(s) at infinity, written in the form $E(F_p) = 1 + p - a_p(E)$. Then the integer $a_p(E)$ is the$p$th Fourier coefficient of a cusp form of weight 2, associated to $\Gamma_0(N)$. Here, N is the conductor of E.

6.2 η-Products

We first recall some properties of the Dedekind η-function (3), $\eta(\tau) = e^{\frac{2i\pi\tau}{24}} \prod_{n=1}^{\infty}(1 - e^{2i\pi n\tau})$. It verifies [21]

$$\eta(\tau + 1) = \eta(\tau), \quad \eta\left(-\frac{1}{\tau}\right) = \sqrt{-i\tau}\,\eta(\tau),$$

and for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$,

$$\eta\left(\frac{a\tau + b}{c\tau + d}\right) = \epsilon(a, b, c, d)(c\tau + d)^{\frac{1}{2}}\eta(\tau),$$

where

$$\epsilon(a, b, c, d) = e^{\frac{ib\pi}{12}} \quad \text{for } c = 0 \text{ and } d = 1,$$
$$\epsilon(a, b, c, d) = e^{i\pi(\frac{a+d}{12c} - S(d,c) - \frac{1}{4})} \quad \text{for } c > 0$$

and

$$S(h, k) = \sum_{n=1}^{k-1} \frac{n}{k}\left(\frac{hn}{k} - \left[\frac{hn}{k}\right] - \frac{1}{2}\right).$$

Proposition 6.2 *The η-function is differentially algebraic.*

In fact, its logarithmic derivative is related to the function G_2 in (13) by $\frac{D\eta}{\eta} = -G_2$, $D = \frac{1}{2i\pi}\frac{d}{d\tau}$, so the Chazy equation gives

$$D^4\eta\eta^3 - 28D^3\eta D\eta\eta^2 + 12D^2\eta(D\eta)^2\eta + 33\left(D^2\eta\right)^2\eta^2 - 18(D\eta)^4 = 0,$$

as claimed.

The interest in the η-function lies in the following theorem [20], p. 18:

Theorem 6.3 *Let N be a positive integer, and let $f(z)=\prod_{\delta|N}\eta^{r_d}(dz)$ be an η-product, $r_d\in\mathbb{Z}$. If N is such that*

$$\sum_{d|N} dr_d = 0 \mod 24,$$

$$\sum_{d|N} \frac{Nr_d}{d} = 0 \mod 24,$$

then

$$f\left(\frac{az+b}{cz+d}\right)=\chi(d)(cz+d)^k f(z)$$

for all

$$\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in\Gamma_0(N)=\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in SL_2(\mathbb{Z}) : c\equiv 0 \bmod N\right\},$$

where $k=\frac{1}{2}\sum_{d|N} r_d$. The character χ is defined by the Legendre–Jacobi symbol

$$\chi(d)=\left(\frac{(-1)^k s}{d}\right),\qquad s=\prod_{d|N} d^{r_d}.$$

More interesting for us is the example of the η-product

$$S(\tau)=\eta(\tau)\eta(2\tau)\eta(7\tau)\eta(14\tau). \tag{26}$$

It is one of the exactly 30 cusp forms, with multiplicative coefficients, of the forms

$$\prod_{k=1}^{s}\eta^{t_k}(a_k\tau),\ a_k,t_k\in\mathbb{N},$$

discovered by Dummit, Kisilevskii, and McKay [8]. It is a newform [7, 20] of weight 2 whose Mellin transform agrees with the Hasse–Weil L-function of the (isogeny class of the) elliptic curve of equation

$$y^2+xy+y=x^3-x.$$

This curve is modular, of conductor 14, and is generally denoted by $X_0(14)$. We will soon meet this curve by another method.

6.3 Main Results

We continue to use the previous notation except that z is replaced by q, according to a current usage:

$$\chi_+(q) = \sum_{n\geq 0} q^{2^n}, \qquad \chi_-(q) = \sum_{n\geq 0} (-1)^n q^{2^n}, \quad |q| < 1.$$

We fix an odd positive integer l. Let f be a modular form in $M_2(2l)$, that is, of weight 2 on the group $\Gamma_0(2l)$. We assume that f is an eigenform of the Hecke operator $T(2)$, which is defined by

$$T(2) : f = \sum_{n\geq 0} a_f(n) q^n \to \sum_{n\geq 0} a_f(2n) q^n.$$

It is known that the only eigenvalues of $T(2)$ on newforms in $M_2(2n)$ are ± 1 [24]. Let ϵ be the eigenvalue of f, and assume that it equals ± 1. Then, for all n,

$$a_f(2n) = \epsilon a_f(n).$$

In other words, we have

$$f = a_f(0) + \sum_{n \text{ odd}} a_f(n) \chi_\epsilon\big(q^n\big).$$

Now for an arithmetical function $\alpha(n)$ with $\alpha(1) = 1$, we define $\hat{\alpha}(n)$ by

$$\sum_{n\geq 1} \frac{\hat{\alpha}(n)}{n^s} = \left(\sum_{n\geq 1} \frac{\alpha(n)}{n^s}\right)^{-1}.$$

We have the following lemma, easily proved by taking Mellin transforms.

Lemma 6.4 *Let F, G be two formal power series in q, and let $\alpha(n)$ be an arithmetical function with $\alpha(1) = 1$. Then*

$$G(q) = \sum_{n\geq 1} \alpha_{(n)} F\big(q^n\big)$$

if and only if

$$F(q) = \sum_{n\geq 1} \hat{\alpha}(n) G\big(q^n\big).$$

We know assume that f is normalized, i.e., $a_f(1) = 1$. Then we can apply the lemma to deduce

$$\chi_\epsilon(q) = \sum_{n\geq 1, nodd} \hat{\alpha}_f(n)\big(f(nz) - a_f(o)\big).$$

We look at some examples. The first nonzero space is $M_2(2)$. It is one-dimensional, and $T(2)$ is the identity. Let E be the unique normalized modular form in $M_2(2)$. It has the following formula [13]:

$$E(z)=\frac{1}{24}+\sum_{n\geq 0}\Bigg(\sum_{d|n,nodd}\Bigg)q^n=\frac{1}{24}\big(\theta_0^4+\theta_1^4\big),$$

where, with a classical notation for two other Jacobi theta functions,

$$\theta_0(q)=\sum_{n\equiv 0 \bmod 2} q^{\frac{n^2}{4}}, \qquad \theta_1(q)=\sum_{n\equiv 1 \bmod 2} q^{\frac{n^2}{4}}.$$

In particular, we have

$$\hat{\alpha}_E(n)=\sum_{d|n,nodd}\mu(d)\mu\left(\frac{n}{d}\right)d.$$

Summarizing, we find (compare with [18]):

Theorem 6.5

$$\chi_+(q)=\frac{1}{24}\sum_{n\geq 1}\hat{\alpha}_E(n)\big(\theta_0^4(nz)+\theta_1^4(nz)-1\big).$$

Remark 3 The right Mittag-Leffler decomposition of $\chi_+(q)$ is obtained by inserting (25) into the equality of Theorem 6.5.

We now consider the function χ_-. The first even level where $T(2)$ has eigenvalue -1 is $l=14$. The space $M_2(14)$ is four-dimensional. It contains a unique cusp form S, which is also the unique eigenform of $T(2)$ with eigenvalue -1. We can normalize it so that $a_S(1)=1$. Its first 100 Fourier coefficients are

S	1	2	3	4	5	6	7	8	9	10
0	1	−1	−2	1	0	2	1	−1	1	0
10	0	−2	−4	−1	0	1	6	−1	2	0
20	−2	0	0	2	−5	4	4	1	−6	0
30	−4	−1	0	−6	0	1	2	−2	8	0
40	6	2	8	0	0	0	−12	−2	1	5
50	−12	−4	6	−4	0	−1	−4	6	−6	0
60	8	4	1	1	0	0	−4	6	0	0
70	0	−1	2	−2	10	2	0	−8	8	0
80	−11	−6	−6	−2	0	−8	12	0	−6	0
90	−4	0	8	12	0	2	−10	−1	0	−5

It can be described in closed form using the Dedekind eta function by

$$S = \eta(z)\eta(2z)\eta(7z)\eta(14z)$$

already met in (26). This is a concrete example of the Modularity Theorem 6.1. The L-series of S is the L-function of the elliptic curve $X_0(14)$:

$$E : y^2 + xy + y = x^3 - x$$

by

$$\sum_{n\geq 1} \frac{a_S(n)}{n^s} = \prod_p \frac{1}{1 - [p - \sharp E(\mathbb{F}_p)]p^{-s} + p^{1-2s}},$$

where

$$\sharp E(\mathbb{F}_p) = \left\{(x, y) \in \mathbb{F}_p : y^2 + xy + y = x^3 - x\right\}.$$

The following result is an unpublished result, obtained with N.P. Skoruppa:

Theorem 6.6 *With $q = e^{2i\pi z}$, $\Im z > 0$, and $\chi_-(q) = \sum_{n\geq 0}(-1)^n q^{2^n}$, we have*

$$\chi_-(q) = \sum_{n\geq 1, n odd} \hat{\alpha}_S(n)\eta(nz)\eta(2nz)\eta(7nz)\eta(14nz).$$

Like in Theorem 6.5, we can rewrite the identity in the last theorem by using four theta functions and identity (4). To this end, we introduce the following:

Definition 6.1 For each $a \in \mathbb{N}$, we define the theta series $\theta_{6,a}$ by

$$\theta_{6,a}(q) = \sum_{n\in\mathbb{Z}, n\equiv a \bmod 12} q^{\frac{n^2}{24}}.$$

According to (4), $\eta = \theta_{6,1} - \theta_{6,5}$, and hence, we have the following:

Corollary 6.7

$$\sum_{k\geq 0}(-1)^k q^{2^k}$$

$$= \sum_{n\geq 1,\ n\ odd} \hat{\alpha}_S(n) \tag{27}$$

$$\sum_{a_1,\dots,a_4\equiv 1,5 \bmod 12} \left(\frac{a_1 a_2 a_3 a_4}{3}\right) \theta_{6,a_1}(nz)\theta_{6,a_2}(2nz)\theta_{6,a_3}(7nz)\theta_{6,a_4}(14nz). \tag{28}$$

Here $(\frac{n}{3}) = \epsilon \mod 3$ *with* $-1 \leq \epsilon \leq +1$.

Acknowledgements I want to thank Roger Gay for the discussions we had on Ehrenpreis' mathematics, Michel Mendès France for the discussions we had on Hardy's identity and Nils-Peter Skoruppa for the discussions we had on Hecke's theory.

References

1. Balazard, M., Mendès France, M., Sebbar, A.: Variations on a theme of Hardy's. Ramanujan J. **9**(1–2), 203–213 (2005)
2. Bateman, P.T.: On the representations of a number as the sum of three squares. Trans. Am. Math. Soc. **71**, 70–101 (1951)
3. Berenstein, C.A., Kawai, T., Struppa, D., Takei, Y.: Exponential representation of a holomorphic solution of a system of differential equations associated with the theta-zerovalue. In: Morimoto, M., Kawai, T. (eds.) Structure of Solutions of Differential Equations, pp. 89–102. World Scientific, Singapore (1996)
4. Berndt, B.: Ramanujan's Notebooks, Part IV. Springer, New York (1994)
5. Bertrand, D.: Theta functions and transcendence. International Symposium on Number Theory, Madras (1996). Ramanujan J. **1**(4), 339–350 (1997)
6. Bochner, S.: Connection between functional equations and modular relations and functions of exponential type. J. Indian Math. Soc. **16**, 99–102 (1952)
7. Diamond, F., Shurman, J.: A First Course in Modular Forms. Graduate Texts in Math., vol. 228. Springer, Berlin (2005)
8. Dummit, D., Kisilevsky, H., McKay, J.: Multiplicative Products of η-Functions. Contemp. Math., vol. 45, pp. 89–98. Am. Math. Soc., Providence (1985)
9. Ehrenpreis, L.: Fourier Analysis in Several Complex Variables. Wiley Interscience, New York (1970)
10. Ehrenpreis, L.: Function theory for Rogers–Ramanujan-like partition identities. Contemp. Math. **143**, 259–320 (1993)
11. Hardy, G.H.: On certain oscillating series. Q. J. Math. **38**, 269–288 (1927)
12. Hardy, G.H., Littlewood, J.E.: Notes on the theory of series. XXIV. A curious power series. Proc. Camb. Philos. Soc. **42**, 85–90 (1946)
13. Hirzebruch, F., Berger, T., Jung, R.: Manifolds and Modular Forms. Aspects of Math., vol. 20. Vieweg, Wiesbaden (1992)
14. Hörmander, L.: Introduction to Complex Analysis in Several Variables. North Holland, Amsterdam (1973)
15. Lipshitz, L., Rubel, L.A.: A gap theorem for power series solutions of algebraic differential equations. Am. J. Math. **108**(5), 1193–1213 (1986)
16. Mahler, K.: Arithmetische Eigenschaften der Lösungen einer Klasse von funktional Gleichungen. Math. Ann. **101**, 342–366 (1929)
17. Maillet, E.: Introduction à la théorie des nombres transcendants et des proprietés arithmétiques des fonctions. Gauthier-Villars, Paris (1906)
18. Mendès France, M., Sebbar, A.: Pliages de papiers, fonctions thêta et méthode du cercle. Acta Math. **183**, 101–139 (1999)
19. Moore, E.H.: Concerning transcendentally transcendent functions. Math. Ann. **48**, 49–74 (1897)
20. Ono, K.: The Web of Modularity: Arithmetic of the Coefficients of Modular Forms and q-Series. CBMS Reg. Conf. Ser. Math., vol. 102. Am. Math. Soc., Providence (2004)
21. Rademacher, H.: Topics in Analytic Number Theory. Die Grundlehren der math. Wissenschaften, vol. 169. Springer, Berlin (1973)
22. Ramanujan, S.: In: Hardy, G.H., Seshu Aiyar, P.V., Wilson, B.M. (eds.) Collected Papers. Chelsea, New York (1962)
23. Sato, M., Kawai, T., Kashiwara, M.: Linear differential equations of infinite order and theta functions. Adv. Math. **47**, 300–325 (1983)

24. Silvermann, J.H.: The Arithmetic of Elliptic Curves. Graduate Texts in Mathematics, vol. 106. Springer, Berlin (1986)
25. Tricomi, F.: A new entire function related to a well-known noncontinuable power series. Comm. Pures Appl. Math. **5**, 213–222 (1952)
26. Weil, A.: Remarks on Hecke's lemma and its use. In: Iyanaga, S. (ed.) Algebraic Number Theory, Intern. Symposium, Kyoto 1976, Jap. Soc. for the promotion of Science, pp. 267–274 (1977), Collected Works III pp. 405–412
27. Zagier, D.: Elliptic modular forms and their applications. In: The 1-2-3 of Modular Forms. Springer, Berlin (2008)

PT Symmetry and Weyl Asymptotics

Johannes Sjöstrand

Abstract For a class of PT-symmetric operators with small random perturbations, the eigenvalues obey Weyl asymptotics with probability close to 1. Consequently, when the principal symbol is nonreal, there are many nonreal eigenvalues.

1 Introduction

PT-symmetry has been proposed as an alternative for self-adjointness in quantum physics [1, 2]. Thus for instance, if we consider a Schrödinger operator on $\mathbf{R}^n$,

$$P = -h^2 \Delta + V(x), \tag{1}$$

the usual assumption of self-adjointness (implying that the potential V is real valued) can be replaced by that of PT-symmetry:

$$V \circ \upsilon = \overline{V}, \tag{2}$$

where $\upsilon : \mathbf{R}^n \to \mathbf{R}^n$ is an isometry with $\upsilon^2 = 1 \neq \upsilon$. If we introduce the **p**arity operator $U_\upsilon u(x) = u(\upsilon(x))$ and the **t**ime reversal operator $\Gamma u = \overline{u}$, then this can be written

$$[P, U_\upsilon \Gamma] = 0. \tag{3}$$

Under mild additional technical assumptions, it is easy to see that the spectrum of a PT-symmetric operator is invariant under reflection in the real axis. However, in order to build PT-symmetric quantum physics, it seems important that the spectrum be real, so a natural mathematical question is then to determine when so is the case. Results on reality and nonreality of the spectrum of PT-symmetric operators can be found in [2, 6, 7, 12].

In memory of Leon Ehrenpreis.

J. Sjöstrand (✉)
Institut de Mathématiques de Bourgogne, Université de Bourgogne, UMR 5584 du CNRS, 9, avenue Alain Savary, BP 47870, 21078 Dijon cedex, France
e-mail: johannes.sjostrand@u-bourgogne.fr

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_19, © Springer-Verlag Italia 2012

The purpose of this note is to show that in a probabilistic sense "most" non-self-adjoint PT-symmetric operators that are symmetric in the sense of (7) have their eigenvalues distributed according to the Weyl law, and hence many of their eigenvalues are nonreal. As a matter of fact, this will be a rather easy adaptation of general results on the Weyl asymptotics for non-self-adjoint operators with small random perturbations [4], [9–11], [5, 13, 14], where the last three references are the ones that we shall use directly. For technical reasons, we will state our results for elliptic operators on compact manifolds, but it would be easy to adapt the results of [13] in order to treat Schrödinger operators on $\mathbf{R}^n$.

The addition of small random perturbations has the effect of destroying (uniform) analyticity (if the unperturbated operator has analytic coefficients). A very interesting question is to give criteria for PT symmetric operators with analytic coefficients to have real spectrum.

The plan of the paper is the following: In Sect. 2 we treat the semi-classical case, and in Sect. 3 we treat the case of large eigenvalues.

2 The Semi-classical Case

Let X be a compact smooth manifold of dimension n. Let $\upsilon : X \to X$ be a smooth involution; $\upsilon^2 = \mathrm{id}$, with $\upsilon \neq \mathrm{id}$. Fix a smooth positive density dx on X which is invariant under υ and let us take L^2 norms with respect to dx. Let P be a differential operator on X of order $m \geq 2$ with smooth coefficients, so that in local coordinates,

$$P = \sum_{|\alpha| \leq m} a_\alpha(x; h)(hD_x)^\alpha, \quad a_\alpha(\cdot; h) \in C^\infty. \tag{4}$$

Here $0 < h \ll 1$ is the semi-classical parameter, and we assume that

$$a_\alpha(x; h) - a_\alpha(x; 0) = \mathscr{O}(h) \tag{5}$$

locally uniformly and similarly for all its derivatives. We also assume for simplicity that $a_\alpha(x; h) = a_\alpha(x)$ is independent of h when $|\alpha| = m$. Let

$$p(x, \xi) = \sum_{|\alpha| \leq m} a_\alpha(x; 0)\xi^\alpha, \qquad p_m(x, \xi) = \sum_{|\alpha| = m} a_\alpha(x)\xi^\alpha.$$

We assume that $p_m(x, \xi) \neq 0$ on $T^*X \setminus 0$, so that P is elliptic in the classical sense. We also assume that

$$p_m\big(T^*X\big) \neq \mathbf{C}. \tag{6}$$

Assume that P is symmetric,

$$P = \Gamma P^* \Gamma =: P^{\mathrm{t}}, \tag{7}$$

and that

$$PU = UP^*, \quad \text{where } Uu(x) = U_\upsilon u(x) := u\big(\upsilon(x)\big),\ \Gamma u(x) = \overline{u(x)}. \tag{8}$$

This means that P is PT-symmetric:

$$[U\Gamma, P] = 0. \tag{9}$$

In addition to the PT-symmetry property (9), we have assumed in (7) that P is symmetric.

Example 1 $P = -h^2\Delta + V(x)$ on $\mathbf{T}^n$, where $\Re V$ is even, and $\Im V$ is odd, $V(-x) = \overline{V}(x)$. Then P is symmetric and PT-symmetric with $\upsilon(x) = -x$.

Let $\widetilde{R}$ be an auxiliary h-independent positive elliptic second-order differential operator on X which commutes with U. We also assume that $\widetilde{R}$ is real or, equivalently, that

$$\big[\Gamma, \widetilde{R}\big] = 0. \tag{10}$$

Then $\widetilde{R}$ has an orthonormal basis of real eigenfunctions e_j such that $Ue_j = (-1)^{k(j)}e_j$ where $k(j) = 1$ or $k(j) = -1$. We say that e_j is even in the first case and odd in second case. Put $\epsilon_j = e_j$ when e_j is even and $\epsilon_j = ie_j$ when e_j is odd. Then $\{\epsilon_j\}$ is also an orthonormal basis, and a linear combination $V = \sum \alpha_j\epsilon_j$ is PT symmetric iff the coefficients α_j are real: $U(V) = \overline{V}$.

In order to formulate our result, we shall follow [14], where we treated a situation without any extra symmetry.

Let $\Omega \Subset \mathbf{C}$ be open, simply connected, and not entirely contained in $\Sigma(p) := p(T^*X)$. Let $V_z(t) := \mathrm{vol}\,(\{\rho \in \mathbf{R}^{2n};\ |p(\rho) - z|^2 \le t\})$. For $\kappa \in]0, 1]$ and $z \in \Omega$, we consider the property that

$$V_z(t) = \mathcal{O}\big(t^\kappa\big), \quad 0 \le t \ll 1. \tag{11}$$

Since $r \mapsto p(x, r\xi)$ is a polynomial of degree m in r with nonvanishing leading coefficient, we see that (11) holds with $\kappa = 1/(2m)$.

By $B_{\mathbf{R}^d}(0, r)$ we denote the open ball in $\mathbf{R}^d$ with center 0 and radius r. Let q_ω be a random potential of the form

$$q_\omega(x) = \sum_{0<\mu_k\le L} \alpha_k(\omega)\epsilon_k(x), \quad \alpha(\omega) = \big(\alpha_k(\omega)\big)_{0<\mu_k\le L} \in B_{\mathbf{R}^D}(0, R), \tag{12}$$

where $\mu_k > 0$ are the square roots of the eigenvalues of $h^2\widetilde{R}$, so that $h^2\widetilde{R}\epsilon_k = \mu_k^2\epsilon_k$. We choose $L = L(h)$, $R = R(h)$ in the interval

$$\begin{aligned} &h^{\frac{\kappa-3n}{s-\frac{n}{2}-\epsilon}} \ll L \le Ch^{-M}, \quad M \ge \frac{3n-\kappa}{s-\frac{n}{2}-\epsilon}, \\ &\frac{1}{C}h^{-(\frac{n}{2}+\epsilon)M+\kappa-\frac{3n}{2}} \le R \le Ch^{-\widetilde{M}}, \quad \widetilde{M} \ge \frac{3n}{2} - \kappa + \left(\frac{n}{2}+\epsilon\right)M, \end{aligned} \tag{13}$$

for some $\epsilon \in]0, s - \frac{n}{2}[$, $s > \frac{n}{2}$, so by Weyl's law for the large eigenvalues of elliptic self-adjoint operators, the dimension D in (12) is of the order of magnitude $(L/h)^n$. We introduce the small parameter $\delta = \tau_0 h^{N_1+n}$, $0 < \tau_0 \leq \sqrt{h}$, where

$$N_1 := \widetilde{M} + sM + \frac{n}{2}. \tag{14}$$

The randomly perturbed PT symmetric operator is

$$P_\delta = P + \delta h^{N_1} q_\omega =: P + \delta Q_\omega. \tag{15}$$

Here (cf. [13]) the exponent N_1 has been chosen so that we have uniformly for $h \ll 1$ and q_ω as above:

$$\left\| h^{N_1} q_\omega \right\|_{L^\infty} \leq \mathscr{O}(1) h^{-n/2} \left\| h^{N_1} q_\omega \right\|_{H_h^s} \leq \mathscr{O}(1),$$

where H_h^s is the natural semi-classical Sobolev space discussed in Section 2 of [14] with a norm equivalent to the standard norm in H^s for each fixed $h > 0$.

The random variables $\alpha_j(\omega)$ will have a joint probability distribution

$$P(d\alpha) = C(h) e^{\Phi(\alpha;h)} L(d\alpha), \tag{16}$$

where for some $N_4 > 0$,

$$|\nabla_\alpha \Phi| = \mathscr{O}\big(h^{-N_4}\big), \tag{17}$$

and $L(d\alpha)$ is the Lebesgue measure. ($C(h)$ is the normalizing constant, assuring that the probability of $B_{\mathbf{R}^D}(0, R)$ is equal to 1.)

We also need the parameter

$$\epsilon_0(h) = \left(h^\kappa + h^n \ln \frac{1}{h}\right)\left(\ln \frac{1}{\tau_0} + \left(\ln \frac{1}{h}\right)^2\right) \tag{18}$$

and assume that $\tau_0 = \tau_0(h)$ is not too small, so that $\epsilon_0(h)$ is small. Recall that $\Omega \Subset \mathbf{C}$ is open, simply connected, and not entirely contained in $\Sigma(p)$. The main result of this section is the following:

Theorem 1 *Under the assumptions above, let $\Gamma \Subset \Omega$ have smooth boundary, let $\kappa \in]0, 1]$ be the parameter in (12), (13), (18), and assume that (11) holds uniformly for z in a neighborhood of $\partial\Gamma$. Then there exists a constant $C > 0$ such that for $C^{-1} \geq r > 0$ and $\widetilde{\epsilon} \geq C\epsilon_0(h)$, we have with probability*

$$\geq 1 - \frac{C\epsilon_0(h)}{r h^{n+\max(n(M+1), N_4+\widetilde{M})}} e^{-\frac{\widetilde{\epsilon}}{C\epsilon_0(h)}} \tag{19}$$

that

$$\left|\#\big(\sigma(P_\delta)\cap\Gamma\big)-\frac{1}{(2\pi h)^n}\mathrm{vol}\left(p^{-1}(\Gamma)\right)\right|$$
$$\leq\frac{C}{h^n}\left(\frac{\widetilde{\epsilon}}{r}+C\left(r+\ln\left(\frac{1}{r}\right)\mathrm{vol}\left(p^{-1}\big(\partial\Gamma+D(0,r)\big)\right)\right)\right).\tag{20}$$

Here $\#(\sigma(P_\delta)\cap\Gamma)$ *denotes the number of eigenvalues of* P_δ *in* Γ, *counted with their algebraic multiplicities.*

In the introduction of [13] there is a discussion about the choice of parameters which applies here also: Very roughly, if τ_0 is equivalent to some high power of h, then up to some power of $\ln(1/h)$, ϵ_0 is of the order of magnitude h^κ. Now choose $\widetilde{\epsilon}=h^{\kappa-\kappa_0}$ for some $\kappa_0\in]0,\kappa[$. When $\kappa>1/2$, then the volume in (20) is $\mathscr{O}(r^\beta)$ with $\beta=2\kappa-1>0$, and more generally we may assume that it is $\mathscr{O}(r^\beta)$ for some $\beta>0$. Then we choose r to be a suitable power of h and obtain that the right-hand side in (20) is $\mathscr{O}(h^{\gamma-n})$ for some $\gamma>0$. With these choices of the parameters, we also see that the probability in (19) is very close to 1.

Proof of Theorem 1 We just have to make some small modifications in the proof of the main result in [14] (which in turn is a modification of the proof in [13]) and only mention the points where a difference appears. The proof in the two cited papers (see also the lecture notes [15]) uses three ingredients:

(1) The construction of a special perturbation of the form δq_ω with q_ω as in (12) but with α in the *complex* ball $B_{\mathbf{C}^D}(0,R)$ for which we have nice lower bounds on the small singular values of P_δ in (15), see Proposition 7.3 in [13] and Proposition 5.1 in [14].
(2) A complex variable argument in the α variables using the existence of the special perturbation in step (1), which permits to conclude that we have nice lower bounds on a relative determinant for $P_\delta-z$, with probability close to 1.
(3) Application of a proposition about the number of zeros of holomorphic functions with exponential growth. (See also [16] for an improved version of this proposition, not yet fully exploited.)

In the present situation we want our special perturbation $\delta q_\omega(x)$ to be PT-symmetric, that is, we want the coefficients α in (12) to be real. All the parts of the proofs in step 1 immediately carry over to the case of real α except the following result which is the basic ingredient in the iterative process leading to the propositions mentioned above:

Let $e_1,\dots,e_N$ be an ON family in $L^2(X)$ such that

$$\left\|\sum_1^N\lambda_je_j\right\|_{H_h^s}\leq\mathscr{O}(1)\|\lambda\|_{\mathbf{C}^N},$$

where the constant $\mathscr{O}(1)$ is independent of the family and especially of N. Then there exists

$$q = \sum_{0<\mu_j\le L} \alpha_j \epsilon_j, \quad \alpha_j \in \mathbf{C}, \tag{21}$$

with $\|\alpha\|_{\mathbf{C}^D} \le R$ with the parameters as in (13), such that

$$\|q\|_{H_h^s} \le \mathscr{O}(1) h^{-\frac{n}{2}} N L^{s+\frac{n}{2}+\epsilon}$$

and such that the matrix

$$M_q = \left(\int q(x) e_j(x) e_k(x) \right) dx)_{1\le j,k\le N}$$

and its singular values

$$\|M_q\| = s_1(M_q) \ge \cdots \ge s_N(M_q)$$

satisfy

$$\begin{aligned} &\|M_q\| \le \mathscr{O}(1) N h^{-n}, \\ &s_k(M_q) \ge h^n/\mathscr{O}(1) \quad \text{for } 1 \le k \le N/2. \end{aligned} \tag{22}$$

(See (6.23), (7.20), (7.23) in [13].)

Write $q = q_1 + i q_2$ where $q_1 = \sum (\Re \alpha_j) \epsilon_j$, $q_2 = \sum (\Im \alpha_j) \epsilon_j$, so that q_1 and q_2 are PT-symmetric. The upper bounds on $\|q\|_{H_h^s}$ and on $\|M_q\|$ follow from the bound $\|\alpha\| \le R$ and therefore carry over to q_j. Since $M_q = M_{q_1} + i M_{q_2}$, we can apply the Ky Fan inequalities [8] and get

$$\frac{h^n}{\mathscr{O}(1)} \le s_{2k-1}(M_q) \le s_k(M_{q_1}) + s_k(M_{q_2}), \quad 1 \le k \le \frac{N}{4}.$$

Since the singular values are enumerated in decreasing order, it follows that for j equal to 1 or 2, we have

$$s_k(M_{q_j}) \ge \frac{h^n}{2\mathscr{O}(1)}, \quad 1 \le k \le \frac{N}{4}. \tag{23}$$

This means that step 1 can be carried out, and we get a PT symmetric operator P_δ as in Proposition 5.1 in [14]; the only slight difference is that rather than taking θ in $]0, 1/4[$, we have to confine this parameter to the smaller interval $]0, 1/8[$.

Step 2 now follows from Remark 8.3 in [13], where the main point is the reality of the coefficients α_j, while the assumption of reality of the basis elements is not necessary and was made there only because we had in mind a real perturbation.

Step 3 can be carried out without any modifications. □

3 Weyl Asymptotics for Large Eigenvalues

Let P^0 be an elliptic differential operator on X of order $m \geq 2$ with smooth coefficients and with principal symbol $p_m(x,\xi)$. In local coordinates we get, using standard multi-index notation,

$$P^0 = \sum_{|\alpha|\leq m} a^0_\alpha(x) D^\alpha, \qquad p_m(x,\xi) = \sum_{|\alpha|=m} a^0_\alpha(x)\xi^\alpha. \tag{24}$$

Recall that the ellipticity of P^0 means that $p_m(x,\xi) \neq 0$ for $\xi \neq 0$. We assume that

$$p_m\big(T^*X\big) \neq \mathbf{C}. \tag{25}$$

As before, we assume symmetry,

$$\big(P^0\big)^* = \Gamma P^0 \Gamma, \tag{26}$$

and that

$$P^0 U = U\big(P^0\big)^*, \tag{27}$$

with $U = U_\upsilon$ as in Sect. 2.

Let $\widetilde{R}$ be a reference operator as in and around (10) and define ϵ_j as there. Write

$$\widetilde{R}\epsilon_j = \big(\mu^0_j\big)^2 \epsilon_j, \quad 0 < \mu^0_0 < \mu^0_1 \leq \mu^0_2 \leq \cdots, \tag{28}$$

so that $\mu_k = h\mu^0_k$ where μ_k are given after (12). Our randomly perturbed operator is

$$P^0_\omega = P^0 + q^0_\omega(x), \tag{29}$$

where ω is the random parameter, and

$$q^0_\omega(x) = \sum_0^\infty \alpha^0_j(\omega)\epsilon_j. \tag{30}$$

Here we assume that $\alpha^0_j(\omega)$ are independent real Gaussian random variables of variance σ^2_j and mean value 0:

$$\alpha^0_j \sim \mathscr{N}\big(0,\sigma^2_j\big), \tag{31}$$

where

$$\big(\mu^0_j\big)^{-\rho} e^{-(\mu^0_j)^{\frac{\beta}{M+1}}} \lesssim \sigma_j \lesssim \big(\mu^0_j\big)^{-\rho}, \tag{32}$$

$$M = \frac{3n - \frac{1}{2}}{s - \frac{n}{2} - \epsilon}, \quad 0 \leq \beta < \frac{1}{2}, \quad \rho > n, \tag{33}$$

where s, ρ, ϵ are fixed constants such that

$$\frac{n}{2} < s < \rho - \frac{n}{2}, \qquad 0 < \epsilon < s - \frac{n}{2}.$$

Let $H^s(X)$ be the standard Sobolev space of order s. As we saw in [5] (where the random variables α_j^0 were complex valued), $q_\omega^0 \in H^s(X)$ almost surely since $s < \rho - \frac{n}{2}$. Hence $q_\omega^0 \in L^\infty$ almost surely, implying that P_ω^0 has purely discrete spectrum.

Consider the function $F(w) = \arg p_m(w)$ on S^*X. For given $\theta_0 \in S^1 \simeq \mathbf{R}/(2\pi\mathbf{Z})$, $N_0 \in \dot{\mathbf{N}} := \mathbf{N} \setminus \{0\}$, we introduce the property $P(\theta_0, N_0)$:

$$\sum_1^{N_0} \left|\nabla^k F(w)\right| \neq 0 \quad \text{on } \left\{w \in S^*X;\, F(w) = \theta_0\right\}. \tag{34}$$

Notice that if $P(\theta_0, N_0)$ holds, then $P(\theta, N_0)$ holds for all θ in some neighborhood of θ_0. Also notice that if X is connected and X, p are analytic and the analytic function F is nonconstant, then there exists $N_0 \in \dot{\mathbf{N}}$ such that $P(\theta_0, N_0)$ holds for all θ_0.

We can now state the main result of this section, which is an adaptation of the main result of [5].

Theorem 2 *Assume that $m \geq 2$. Let $0 \leq \theta_1 \leq \theta_2 \leq 2\pi$ and assume that $P(\theta_1, N_0)$ and $P(\theta_2, N_0)$ hold for some $N_0 \in \dot{\mathbf{N}}$. Let $g \in C^\infty([\theta_1, \theta_2];]0, \infty[)$ and put*

$$\Gamma^g_{\theta_1,\theta_2;0,\lambda} = \left\{re^{i\theta};\, \theta_1 \leq \theta \leq \theta_2,\ 0 \leq r \leq \lambda g(\theta)\right\}.$$

Then for every $\delta \in]0, \frac{1}{2} - \beta[$, there exists $C > 0$ such that almost surely there exists $C(\omega) < \infty$ such that for all $\lambda \in [1, \infty[$,

$$\left|\#\left(\sigma\left(P_\omega^0\right) \cap \Gamma^g_{\theta_1,\theta_2;0,\lambda}\right) - \frac{1}{(2\pi)^n} \operatorname{vol} p_m^{-1}\left(\Gamma^g_{\theta_1,\theta_2;0,\lambda}\right)\right| \\ \leq C(\omega) + C\lambda^{\frac{n}{m} - \frac{1}{m}(\frac{1}{2} - \beta - \delta)\frac{1}{N_0+1}}. \tag{35}$$

The proof actually allows us to have almost surely a simultaneous conclusion for a whole family of θ_1, θ_2, g:

Theorem 3 *Assume that $m \geq 2$. Let Θ be a compact subset of $[0, 2\pi]$. Let $N_0 \in \mathbf{N}$ and assume that $P(\theta, N_0)$ holds uniformly for $\theta \in \Theta$. Let $\mathcal{G}$ be a subset of $\{(g, \theta_1, \theta_2);\, \theta_j \in \Theta, \theta_1 \leq \theta_2, g \in C^\infty([\theta_1, \theta_2];]0, \infty[)\}$ with the property that g and $1/g$ are uniformly bounded in $C^\infty([\theta_1, \theta_2];]0, \infty[)$ when (g, θ_1, θ_2) varies in $\mathcal{G}$. Then for every $\delta \in]0, \frac{1}{2} - \beta[$, there exists $C > 0$ such that almost surely there exists $C(\omega) < \infty$ such that for all $\lambda \in [1, \infty[$ and all $(g, \theta_1, \theta_2) \in \mathcal{G}$, we have estimate* (35).

Condition (32) allows us to choose σ_j decaying faster than any negative power of μ_j^0. Then from the discussion below it will follow that $q_\omega(x)$ is almost surely a smooth function. A rough and somewhat intuitive interpretation of Theorem 3 is then that for almost every PT-symmetric elliptic operator of order ≥ 2 with smooth coefficients on a compact manifold which satisfies conditions (25), (26), and (27), the large eigenvalues distribute according to Weyl's law in sectors with limiting directions that satisfy a weak nondegeneracy condition.

Proof of Theorem 2 As already mentioned, the theorem is a variant of Theorem 1.1 in [5]. The difference is just that we now use real random variables in the perturbation q_ω^0 in order to assure the PT-symmetry while in [5] they were complex. The proof in [5] used a reduction to the semi-classical case where the main result of [14] could be applied. The proof of Theorem 2 is an immediate modification of that proof, where we replace the main result in [14] by Theorem 1. The only point where the use of real Gaussian random variables instead of complex ones causes a slight change is the use of (4.10) in [5] that was established in [3], where we have to replace the denominator 2 by 4 in the case of real random variables. That was also proved by Bordeaux Montrieux in [3], Proposition 2.5.4. □

Acknowledgements Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-08-BLAN-0228-01 ainsi que d'une bourse FABER du conseil régional de Bourgogne.

References

1. Bender, C.M., Boettcher, S., Meisinger, P.N.: $\mathscr{PT}$-symmetric quantum mechanics. J. Math. Phys. **40**(5), 2201–2229 (1999)
2. Bender, C.M., Mannheim, P.D.: $\mathscr{PT}$ symmetry and necessary and sufficient conditions for the reality of energy eigenvalues. Phys. Lett. A **374**(15–16), 1616–1620 (2010)
3. Bordeaux Montrieux, W.: Loi de Weyl presque sûre et résolvante pour des opérateurs différentiels non-autoadjoints. Thesis, CMLS, Ecole Polytechnique (2008). http://pastel.archives-ouvertes.fr/pastel-00005367/en
4. Bordeaux Montrieux, W.: Loi de Weyl presque sûre pour un système différentiel en dimension 1. Ann. Henri Poincaré **12**(1), 173–204 (2011)
5. Bordeaux Montrieux, W., Sjöstrand, J.: Almost sure Weyl asymptotics for non-self-adjoint elliptic operators on compact manifolds. Ann. Fac. Sci. Toulouse **19**(3–4), 567–587 (2010)
6. Caliceti, E., Graffi, S., Sjöstrand, J.: Spectra of $\mathscr{PT}$-symmetric operators and perturbation theory. J. Phys. A, Math. Gen. **38**(1), 185–193 (2005)
7. Caliceti, E., Graffi, S., Sjöstrand, J.: $\mathscr{PT}$ symmetric non-selfadjoint operators, diagonalizable and non-diagonalizable, with real discrete spectrum. J. Phys. A, Math. Theor. **40**(33), 10155–10170 (2007)
8. Gohberg, I.C., Krein, M.G.: Introduction to the Theory of Linear Non-selfadjoint Operators. Translations of Mathematical Monographs, vol. 18. AMS, Providence (1969)
9. Hager, M.: Instabilité spectrale semiclassique pour des opérateurs non-autoadjoints. I. Un modèle. Ann. Fac. Sci. Toulouse **15**(2), 243–280 (2006)
10. Hager, M.: Instabilité spectrale semiclassique d'opérateurs non-autoadjoints. II. Ann. Henri Poincaré **7**(6), 1035–1064 (2006)

11. Hager, M., Sjöstrand, J.: Eigenvalue asymptotics for randomly perturbed non-selfadjoint operators. Math. Ann. **342**(1), 177–243 (2008)
12. Shin, K.C.: On the reality of the eigenvalues for a class of $\mathscr{P}\mathscr{T}$-symmetric oscillators. Commun. Math. Phys. **229**(3), 543–564 (2002)
13. Sjöstrand, J.: Eigenvalue distribution for non-self-adjoint operators with small multiplicative random perturbations. Ann. Fac. Sci. Toulouse **18**(4), 739–795 (2009)
14. Sjöstrand, J.: Eigenvalue distribution for non-self-adjoint operators on compact manifolds with small multiplicative random perturbations. Ann. Fac. Sci. Toulouse **19**(2), 277–301 (2010)
15. Sjöstrand, J.: Spectral properties of non-self-adjoint operators. Lecture notes from Evian les Bains, 8–12 juin, 2009, Actes des Journées d'é.d.p. d'Évian 2009, see: http://jedp.cedram.org:80/jedp-bin/feuilleter?id=JEDP_2009___ also: arXiv:1002.4844
16. Sjöstrand, J.: Counting zeros of holomorphic functions of exponential growth. J. Pseudodifferential Oper. Appl. **1**(1), 75–100 (2010). arXiv:0910.0346

Complex Gradient Systems

Giuseppe Tomassini and Sergio Venturini

Abstract Let $\widetilde{M}$ be a complex manifold of complex dimension $n+k$. We say that the functions $u_1, \ldots, u_k$ and the vector fields $\xi_1, \ldots, \xi_k$ on $\widetilde{M}$ form a *complex gradient system* if $\xi_1, \ldots, \xi_k, J\xi_1, \ldots, J\xi_k$ are linearly independent at each point $p \in \widetilde{M}$ and generate an integrable distribution of $T\widetilde{M}$ of dimension $2k$ and $\mathrm{d}u_\alpha(\xi_\beta) = 0$, $\mathrm{d}^c u_\alpha(\xi_\beta) = \delta_{\alpha\beta}$ for $\alpha, \beta = 1, \ldots, k$. We prove a Cauchy theorem for such complex gradient systems with initial data along a CR-submanifold of type (n, k). We also give a complete local characterization for the complex gradient systems which are *holomorphic* and *abelian*, which means that the vector fields $\xi_\alpha^c = \xi_\alpha - iJ\xi_\alpha$, $\alpha = 1, \ldots, k$, are holomorphic and satisfy $[\xi_\alpha^c, \bar{\xi}_\beta^c] = 0$ for each $\alpha, \beta = 1, \ldots, k$.

1 Introduction

Let $\widetilde{M}$ be a complex manifold, and $T\widetilde{M}$ its (real) tangent space endowed by its complex structure J.

In [7] the authors introduced a geometric tool named *one-dimensional calibrated foliation* on the complex manifold $\widetilde{M}$. It consists of a real function $u : \widetilde{M} \to \mathbb{R}$ and a vector field $\xi \in \Gamma(\widetilde{M}, T\widetilde{M})$ which satisfy the conditions

$$[\xi, J\xi] = 0,$$
$$\mathrm{d}u(\xi) = 0,$$
$$\mathrm{d}^c u(\xi) = 1.$$

Here $[\cdot, \cdot]$ is the Poisson Lie bracket between vector fields, and $\mathrm{d}^c u(\xi) = -\mathrm{d}u(J(\xi))$.

G. Tomassini (✉)
Scuola Normale Superiore, Piazza dei Cavalieri, 7, 56126 Pisa, Italy
e-mail: g.tomassini@sns.it

S. Venturini
Dipartimento Di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, 40127 Bologna, Italy
e-mail: venturin@dm.unibo.it

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_20, © Springer-Verlag Italia 2012

Among the results proved in [7] there is a Cauchy-like theorem for one-dimensional calibrated foliation (see Theorem 3.1), which states the following: if $M \subset \widetilde{M}$ is a real hypersurface and ξ_0 is a vector field on M which is transversal to the holomorphic tangent space to M, then, under the assumption that the integral curves $t \mapsto \gamma(t)$ of ξ_0 are real analytic, there exists a (unique) one-dimensional calibrated foliation (ξ, u), defined in a suitable neighbourhood of M in $\widetilde{M}$, such that the vector field ξ extends ξ_0.

The notion of one-dimensional calibrated foliation was motivated by the problem of finding for M an equation satisfying the homogeneous complex Monge–Ampère equation. The method has subsequently been applied in [8, 9] to prove the existence of adapted complex structures on the symplectization of a pseudo-Hermitian manifold. The key point in proving the existence of a calibrated foliation is the construction of a function $\widetilde{G}(z, p) : D \to \widetilde{M}$, where $D \subset M \times \mathbb{C}$ is an open neighbourhood of $M \times \{0\}$, which is holomorphic in z for each $p \in M$, and for real $z = t$, the map $t \mapsto \widetilde{G}(t, p)$ is the integral curve of the vector field ξ_0 such that $\widetilde{G}(0, p) = p$. This idea goes back to Duchamp and Kalka (see [3]).

The purpose of this paper is to provide a natural higher-dimensional generalization of the notion of one-dimensional calibrated foliations.

Let $\widetilde{M}$ be a complex manifold of complex dimension $n + k$. We say that the functions $u_1, \dots, u_k$ and the vector fields $\xi_1, \dots, \xi_k$ on $\widetilde{M}$ form a *complex gradient system* (of dimension k) if

$$\xi_1, \dots, \xi_k, J(\xi_1), \dots, J(\xi_k)$$

are linearly independent at each point $p \in \widetilde{M}$ and generate an integrable distribution of $T\widetilde{M}$ of dimension $2k$ and

$$\mathrm{d}u_\alpha(\xi_\beta) = 0, \qquad \mathrm{d}^c u_\alpha(\xi_\beta) = \delta_{\alpha\beta}$$

for $\alpha, \beta = 1, \dots, k$. Here $\delta_{\alpha\beta}$ is the usual Kronecker symbol.

In a more intrinsic way a complex gradient system is given by a real vector space $\mathscr{V}$ of dimension k, a linear monomorphism $\rho : \mathscr{V} \to \Gamma(\widetilde{M}, T\widetilde{M})$, the *representation* map, and a map $U : \widetilde{M} \to \mathscr{V}$, the *gradient map*, which satisfy

$$\mathrm{d}U\big(\rho(V)\big) = 0,$$
$$\mathrm{d}^c U\big(\rho(V)\big) = V$$

for each $V \in \mathscr{V}$.

The name "complex gradient system" (instead of "calibrated foliation") arises from the fact that there are examples of triples $(\mathscr{V}, \rho, U)$ where $\mathscr{V} = \mathfrak{g}$ is the Lie algebra of a Lie group G which is a compact real form of a reductive complex Lie group $G^{\mathbb{C}}$ and the map U (with the identification of $\mathfrak{g} = \mathfrak{g}^*$ with its dual $\mathfrak{g}^*$ by the Killing form) is the moment map associated to a symplectic action of G. It is customary in the symplectic geometry to call such a moment map as "gradient map" (see e.g. [4]) and hence the name "complex gradient system".

For a general reference on symplectic geometry and moment maps theory, see e.g. [6] and [1]. Basic definitions and notions in CR-geometry can be found in [2].

Let now describe in more details the content of the paper.

In Sect. 2 contains the elementary properties of a complex gradient system. In particular, any complex gradient system satisfies the formal commutativity property

$$\left[\rho^{\mathbb{C}}(V), \rho^{\mathbb{C}}(W)\right] = 0$$

for all $V, W \in \mathscr{V}$, where $\rho^{\mathbb{C}}(V) = \frac{1}{2}(\rho(V) - iJ\rho(V))$ (and similarly $\rho^{\mathbb{C}}(, W) = \frac{1}{2}(\rho(W) - iJ\rho(W)))$ is the complex vector field of type $(1, 0)$ naturally associated to the real vector field $\rho(V)$ (resp. $\rho(W)$). See Theorem 1.

In Sect. 3 we solve a Cauchy problem for a complex gradient system on a complex manifold $\widetilde{M}$ of dimension $n + k$ with initial data on a CR-submanifold of $\widetilde{M}$ of type (n, k) (Theorem 2).

In Sect. 4 we give a couple of examples applying our construction to the case of the complexification of a real Lie group G. In particular, we find explicitly the complex gradient system associated to the standard representation of the Lie algebra $\mathfrak{g}$ of G as left-invariant vector fields on G.

Finally, in the last section we give a complete local description of any *abelian holomorphic* complex gradient system $(\mathscr{V}, \rho, U)$, where abelian means that

$$\left[\rho^{\mathbb{C}}(V), \bar{\rho^{\mathbb{C}}}(W)\right] = 0$$

for each pair of vectors $V, W \in \mathscr{V}$, and holomorphic means that $\rho^{\mathbb{C}}(V)$ is a holomorphic vector field on $\widetilde{M}$ for each $V \in \mathscr{V}$. See Theorem 3 for details.

2 Complex Gradient Systems

Let $\widetilde{M}$ be a complex manifold of complex dimension $n + k$, $T\widetilde{M}$ its (real) tangent space endowed by its complex structure J. Let $\mathscr{V}$ be a real vector space, and let $\rho : \mathscr{V} \to \Gamma(\widetilde{M}, T\widetilde{M})$ be a linear map. We denote by $\mathscr{D}_\rho^{\mathbb{R}} \subset T\widetilde{M}$ the distribution generated by the vector fields of the form $\rho(V)$, $V \in \mathscr{V}$. We also denote by $\mathscr{D}_\rho^{\mathbb{C}} \subset T\widetilde{M}$ the distribution

$$\mathscr{D}_\rho^{\mathbb{C}} = \mathscr{D}_\rho^{\mathbb{R}} + J\left(\mathscr{D}_\rho^{\mathbb{R}}\right)$$

generated by the vector fields of the form $\rho(V)$ and $J(\rho(V))$, $V \in \mathscr{V}$.

Definition 1 A *complex gradient system* of dimension k on $\widetilde{M}$ is a triple

$$(\mathscr{V}, \rho, U)$$

where:

1. $\mathscr{V}$ is a real vector space of dimension k;

2. $\rho : \mathscr{V} \to \Gamma(\widetilde{M}, T\widetilde{M})$ is an $\mathbb{R}$-linear map;
3. $U : \widetilde{M} \to \mathscr{V}$ is a smooth map

which satisfies

(i) for each $V \in \mathscr{V}$, the vector field $\rho(V)$ is smooth, and we have the identities

$$\mathrm{d}U\big(\rho(V)\big) = 0,$$
$$\mathrm{d}^c U\big(\rho(V)\big) = V;$$

(ii) the distribution $\mathscr{D}_\rho^{\mathbb{C}} \subset T\widetilde{M}$ is integrable.

The maps ρ and U are said respectively the *representation* and the *gradient map* of the complex gradient system $(\mathscr{V}, \rho, U)$.

If $\{V_1, \dots, V_k\}$ is a basis of $\mathscr{V}$, we set

$$\xi_1 = \rho(V_1), \dots, \xi_k = \rho(V_k),$$

and for some smooth functions $u_1, \dots, u_k$, we have

$$U = u_1 V_1 + \cdots + u_k V_k.$$

Then $(\mathscr{V}, \rho, U)$ is a complex gradient system if, and only if,

$$\mathrm{d}u_\alpha(\xi_\beta) = 0, \quad \alpha, \beta = 1, \dots, k,$$
$$\mathrm{d}^c u_\alpha(\xi_\beta) = \delta_{\alpha\beta}, \quad \alpha, \beta = 1, \dots, k,$$

and $\{\xi_1, J\xi_1, \dots, \xi_k, J\xi_k\}$ is a basis of an integrable distribution.

Proposition 1 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system on the manifold $\widetilde{M}$. Then*

$$\mathscr{D}_\rho^{\mathbb{R}} = \mathscr{D}_\rho^{\mathbb{C}} \cap \ker \mathrm{d}U,$$
$$\mathscr{D}_\rho^{\mathbb{C}} = \mathscr{D}_\rho^{\mathbb{R}} \oplus J\mathscr{D}_\rho^{\mathbb{R}},$$
$$T\widetilde{M} = J\mathscr{D}_\rho^{\mathbb{R}} \oplus \ker \mathrm{d}U.$$

Proof Let $p \in \widetilde{M}$ and $v \in T_p\widetilde{M}$. Assume that $v \in \mathscr{D}_\rho^{\mathbb{C}}$. Then there are $V, W \in \mathscr{V}$ such that $v = \rho(V)_p + J\rho(W)_p$. It follows that

$$\mathrm{d}U(v) = \mathrm{d}U\big(\rho(V)_p\big) + \mathrm{d}U\big(J\rho(W)_p\big) = -W,$$

whence

$$v \in \ker \mathrm{d}U \iff W = 0 \iff v = \rho(V)_p \in \mathscr{D}_\rho^{\mathbb{R}}.$$

This proves the first assertion of the proposition.

As for the second one, it suffices to prove that $\mathscr{D}_\rho^{\mathbb{R}} \cap J\mathscr{D}_\rho^{\mathbb{R}} = 0$. Let $v \in \mathscr{D}_\rho^{\mathbb{R}} \cap J\mathscr{D}_\rho^{\mathbb{R}}$. Then $v = \rho(V)_p = JW_p$ for some $V, W \in \mathscr{V}$. We then have

$$0 = \mathrm{d}U\big(\rho(V)_p\big) = \mathrm{d}U\big(J\rho(W)_p\big) = -W,$$

and hence $v = J\rho(W)_p = 0$, as required.

Let now $v \in T_p\widetilde{M}$ be arbitrary and set $V = \mathrm{d}U(v)$, $w = \rho(V)_p$. Clearly, $v = (v - Jw) + Jw$. Observe that

$$\mathrm{d}U(v - Jw) = \mathrm{d}U(v) - \mathrm{d}U(Jw) = V - \mathrm{d}U\big(J\rho(V)_p\big) = V - V = 0,$$

i.e. $v - Jw \in \ker \mathrm{d}U$ and $Jw \in J\mathscr{D}_\rho^{\mathbb{R}}$.

If $v \in J\mathscr{D}_\rho^{\mathbb{R}} \cap \ker \mathrm{d}U$, then $v = J\rho(W)_p$ for some $W \in \mathscr{V}$. It follows that

$$0 = \mathrm{d}U(v) = \mathrm{d}U\big(J\rho(W)_p\big) = -W,$$

and hence $v = J\rho(W)_p = J\rho(0)_p = 0$. This proves the last assertion of the proposition. □

Definition 2 Given a complex gradient system $(\mathscr{V}, \rho, U)$, we denote by $\mathscr{H}_\rho$ the distribution $\ker \mathrm{d}U \cap \ker \mathrm{d}^c U$.

Proposition 2 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system on the manifold $\widetilde{M}$. Then*

$$\ker \mathrm{d}U = \mathscr{D}_\rho^{\mathbb{R}} \oplus \mathscr{H}_\rho,$$

$$T\widetilde{M} = \mathscr{D}_\rho^{\mathbb{R}} \oplus J\mathscr{D}_\rho^{\mathbb{R}} \oplus \mathscr{H}_\rho.$$

Proof The second equality easily follows from the first in view of the equality $T\widetilde{M} = J\mathscr{D}_\rho^{\mathbb{R}} \oplus \ker \mathrm{d}U$ proved in the last proposition. So it suffices to prove that $\ker \mathrm{d}U = \mathscr{D}_\rho^{\mathbb{R}} \oplus \mathscr{H}_\rho$.

By definition we have $\mathscr{D}_\rho^{\mathbb{R}} \subset \ker \mathrm{d}U$ and by construction $\mathscr{H}_\rho \subset \ker \mathrm{d}U$, so that $\mathscr{D}_\rho^{\mathbb{R}} + \mathscr{H}_\rho \subset \ker \mathrm{d}U$.

Let $v \in \ker \mathrm{d}U$ be arbitrary. Set $V = \mathrm{d}^c U(v)$ and $w = \rho(V)_p$. Then we have immediately $v = (v - w) + w$ and $w, v - w \in \ker \mathrm{d}U$. Observe that

$$\mathrm{d}^c U(v - w) = \mathrm{d}^c U(v) - \mathrm{d}^c U(w) = V - \mathrm{d}^c U\big(\rho(V)_p\big) = V - V = 0,$$

that is, $v - w \in \ker \mathrm{d}^c U$ and $w \in \mathscr{D}_\rho^{\mathbb{R}}$.

Assume now that $v \in \mathscr{D}_\rho^{\mathbb{R}} \cap \ker \mathrm{d}^c U$. Then $v = \rho(W)_p$ for some $W \in \mathscr{V}$. It follows that

$$0 = \mathrm{d}^c U(v) = \mathrm{d}^c U\big(\rho(W)_p\big) = W,$$

and hence $v = \rho(W)_p = \rho(0)_p = 0$. The proof is now complete. □

Proposition 3 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system on the manifold $\widetilde{M}$, and $V, W \in \mathscr{V}$. Then*

$$\big[\rho(V), \rho(W)\big], \big[\rho(V), J\rho(W)\big], \big[J\rho(V), J\rho(W)\big] \in \Gamma\big(\widetilde{M}, \mathscr{D}_\rho^{\mathbb{R}}\big).$$

Moreover,

$$\begin{aligned} \mathrm{dd}^c U\big(\rho(V), \rho(W)\big) &= -\mathrm{d}^c U\big(\big[\rho(V), \rho(W)\big]\big),\\ \mathrm{dd}^c U\big(J\rho(V), J\rho(W)\big) &= -\mathrm{d}^c U\big(\big[\rho(V), \rho(W)\big]\big),\\ \mathrm{dd}^c U\big(\rho(V), J\rho(W)\big) &= -\mathrm{d}^c U\big(\big[\rho(V), J\rho(W)\big]\big). \end{aligned}$$

Proof Let ξ_1 be either $\rho(V)$ or $J\rho(V)$, and ξ_2 be either $\rho(W)$ or $J\rho(W)$. Then $\xi_1(U)$ and $\xi_2(U)$ are constant functions, and hence,

$$\xi_2\big(\xi_1(U)\big) = \xi_1\big(\xi_2(U)\big) = 0.$$

This easily implies that $[\xi_1, \xi_2] \in \Gamma(\widetilde{M}, \ker \mathrm{d}U)$. By the definition of a complex gradient system, $\mathscr{D}_\rho^{\mathbb{C}}$ is an integrable distribution, so, in view of the last proposition, we have $\mathscr{D}_\rho^{\mathbb{R}} = \mathscr{D}_\rho^{\mathbb{C}} \cap \ker \mathrm{d}U$, and hence $[\xi_1, \xi_2] \in \Gamma(\widetilde{M}, \mathscr{D}_\rho^{\mathbb{R}})$.

Using again the equalities $\xi_2(\xi_1(U)) = \xi_1(\xi_2(U)) = 0$, we also obtain

$$\begin{aligned} \mathrm{dd}^c U(\xi_1, \xi_2) &= \xi_1\big(\xi_2(U)\big) - \xi_2\big(\xi_1(U)\big) - \mathrm{dd}^c U\big([\xi_1, \xi_2]\big)\\ &= -\mathrm{dd}^c U\big([\xi_1, \xi_2]\big). \end{aligned}$$

This completes the proof of the proposition. □

Corollary 1 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system on the manifold $\widetilde{M}$, and $V, W \in \mathscr{V}$. Then*

$$\begin{aligned} \rho\big(\mathrm{dd}^c U\big(\rho(V), \rho(W)\big)\big) &= -\big[\rho(V), \rho(W)\big],\\ \rho\big(\mathrm{dd}^c U\big(J\rho(V), J\rho(W)\big)\big) &= -\big[\rho(V), \rho(W)\big],\\ \rho\big(\mathrm{dd}^c U\big(\rho(V), J\rho(W)\big)\big) &= -\big[\rho(V), J\rho(W)\big]. \end{aligned}$$

Proof Apply ρ to both sides of the last three equalities of the previous proposition and use the identity $\mathrm{d}^c U(\rho(V)) = V$. □

Corollary 2 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system on the manifold $\widetilde{M}$. Then the distribution $\mathscr{D}_\rho^{\mathbb{R}} \subset T\widetilde{M}$ is integrable.*

Corollary 3 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system of dimension k on the complex manifold $\widetilde{M}$ of dimension $n + k$. For every $V \in \mathscr{V}$, the level set $U^{-1}(V)$ of the smooth function U is either empty, or it is a* CR-*submanifold of type (n, k).*

Proposition 4 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system on the manifold $\widetilde{M}$. For every $V, W \in \mathscr{V}$, one has*

$$\begin{aligned}\big[J\rho(V), J\rho(W)\big] &= \big[\rho(V), \rho(W)\big],\\ \big[J\rho(V), \rho(W)\big] &= -\big[\rho(V), J\rho(W)\big].\end{aligned}$$

Proof Since $\widetilde{M}$ is a complex manifold, the complex structure J is integrable, and hence,

$$J\big[\rho(V), \rho(W)\big] - J\big[J\rho(V), J\rho(W)\big] = \big[J\rho(V), \rho(W)\big] + \big[\rho(V), J\rho(W)\big].$$

The right side of such an equality belongs to $\Gamma(\widetilde{M}, J\mathscr{D}_\rho^{\mathbb{R}})$, while the second belongs to $\Gamma(\widetilde{M}, \mathscr{D}_\rho^{\mathbb{R}})$. Since $J\mathscr{D}_\rho^{\mathbb{R}} \cap \mathscr{D}_\rho^{\mathbb{R}} = 0$, it follows that

$$\begin{aligned}J\big[\rho(V), \rho(W)\big] - J\big[J\rho(V), J\rho(W)\big] &= 0,\\ \big[J\rho(V), \rho(W)\big] + \big[\rho(V), J\rho(W)\big] &= 0,\end{aligned}$$

and the assertion follows. □

Let $T_{\mathbb{C}}\widetilde{M} = \mathbb{C} \otimes_{\mathbb{R}} T\widetilde{M}$ denote the complexification of $T\widetilde{M}$, and $T_{\mathbb{C}}^{(1,0)}\widetilde{M}$ the subbundle of the tangent vector of type $(1, 0)$.

Definition 3 Let $(\mathscr{V}, \rho, U)$ be a complex gradient system. The *complexified representation*

$$\rho^{\mathbb{C}} : \mathscr{V} \to \Gamma\big(\widetilde{M}, T_{\mathbb{C}}^{(1,0)}\widetilde{M}\big)$$

is defined for each $V \in \mathscr{V}$ by

$$\rho^{\mathbb{C}}(V) = \frac{1}{2}\big(\rho(V) - iJ\big(\rho(V)\big)\big).$$

With a little abuse of language we say that $\rho^{\mathbb{C}}$ is holomorphic if $\rho^{\mathbb{C}}(V)$ is a holomorphic vector field on $\widetilde{M}$ for each $V \in \mathscr{V}$.

With this notation the last proposition can be restated as follows.

Theorem 1 *Let $(\mathscr{V}, \rho, U)$ be a complex gradient system. Then for each $V, W \in \mathscr{V}$, we have*

$$\big[\rho^{\mathbb{C}}(V), \rho^{\mathbb{C}}(W)\big] = 0.$$

3 A Cauchy Problem

Let $\widetilde{M}$ be a complex manifold of complex dimension $n+k$. Let $M \subset \widetilde{M}$ be a CR-submanifold of $\widetilde{M}$ of type (n, k).

Definition 4 Let $\mathscr{V}$ be a real vector space. A linear map $\rho_0 : \mathscr{V} \to \Gamma(M, TM)$ is said to be CR-*transverse* if for each $V \in \mathscr{V} \setminus \{0\}$ and each $p \in M$, we have $J(\rho_0(V)(p)) \notin T_pM$.

Given a vector field $X \in \Gamma(M, TM)$, we denote by $\mathrm{Exp}_p(X)$ the exponential mapping associated to the vector field X: if $\gamma(t)$ is the integral curve of the vector field X such that $\gamma(0) = p$, then $\mathrm{Exp}_p(X) = \gamma(1)$.

Let $\rho_0 : \mathscr{V} \to \Gamma(M, TM)$ be a linear map of real vector spaces. The *flow* associated to ρ_0 is defined for $p \in M$ and $V \in \mathscr{V}$ by

$$G_{\rho_0}(p, V) = \mathrm{Exp}_p\big(\rho_0(V)\big).$$

G_{ρ_0} is a smooth map which is well defined in an open neighbourhood of $M \times \{0\}$ in $M \times \mathscr{V}$.

Let $\mathscr{V}^{\mathbb{C}}$ denote the complexification $\mathbb{C} \otimes_{\mathbb{R}} \mathscr{V}$ of the real vector space $\mathscr{V}$. We then say that the flow G_{ρ_0} is *uniformly (real) analytic* if there exist an open neighbourhood $D \subset M \times \mathscr{V}^{\mathbb{C}}$ of $M \times \{0\}$ and a smooth function

$$\widetilde{G}_{\rho_0} : D \to M,$$

which coincides with G_{ρ_0}, on $M \times \mathscr{V}$, and for each $p \in M$, the map defined on the open set

$$D_p = \big\{V \in \mathscr{V}^{\mathbb{C}} \mid (p, V) \in D\big\}$$

by

$$V \mapsto \widetilde{G}_{\rho_0}(p, V)$$

is holomorphic.

The map $\widetilde{G}_{\rho_0}$ will be called the *complexification* of the flow G_{ρ_0}.

Let F_{ρ_0} denote the restriction of $\widetilde{G}_{\rho_0}$ to $\widetilde{D} = D \cap M \times i\mathscr{V}$. As it is immediately seen, shrinking the domain D if necessary, the map F_{ρ_0} is a diffeomorphism between $\widetilde{D}$ and $F_{\rho_0}(\widetilde{D})$ if, and only if, the map ρ_0 is CR-transverse.

In this case we denote by $U_{\rho_0} : F_{\rho_0}(\widetilde{D}) \to \mathscr{V}$ the unique map satisfying

$$U_{\rho_0}\big(F_{\rho_0}(p, iV)\big) = V$$

for each $p \in M$ and each $V \in \mathscr{V}$.

Observe that the map U_{ρ_0} vanishes exactly on M, so we will refer to it as to the *equation* of M associated to ρ_0.

We say that a complex gradient system $(\mathscr{V}, \rho, U)$ *extends* ρ_0 if it is defined in a open neighbourhood $N \subset \widetilde{M}$ of M and for every $p \in M$ and $V \in \mathscr{V}$,

$$\rho(V)(p) = \rho_0(V)(p).$$

If $(\mathscr{V}, \rho_1, U)$, $(\mathscr{V}, \rho_2, U)$ are two extensions of ρ_0 such that, for every $V \in \mathscr{V}$, the sections $\rho_1(V)$, $\rho_2(V)$ coincide on N, then we write $\rho_{1|N} = \rho_{2|N}$.

Theorem 2 *Let $\widetilde{M}$ be a complex manifold of complex dimension $n+k$, $M \subset \widetilde{M}$ a* CR*-submanifold of $\widetilde{M}$ of type (n,k). Let $\mathscr{V}$ be a real vector space, and $\rho_0 : \mathscr{V} \to \Gamma(M, TM)$ a* CR*-transverse linear map such that the distribution $\mathscr{D}^{\mathbb{R}}_{\rho_0}$ is integrable. Assume that the associated flow G_{ρ_0} is uniformly real analytic and let U_{ρ_0} be the associated equation.*

Then there exists an open neighbourhood $N \subset \widetilde{M}$ of M and an $\mathbb{R}$-linear map

$$\rho : \mathscr{V} \to \Gamma(N, T\widetilde{M})$$

such that $(\mathscr{V}, \rho, U_{\rho_0})$ is a complex gradient system which extends ρ_0.

The map ρ is unique in a neighbourhood of M, that is, if

$$\rho_1 : \mathscr{V} \to \Gamma(N_1, T\widetilde{M}), \qquad \rho_2 : \mathscr{V} \to \Gamma(N_2, T\widetilde{M})$$

are $\mathbb{R}$-linear maps such that $(\mathscr{V}, \rho_1, U_{\rho_0})$ and $(\mathscr{V}, \rho_2, U_{\rho_0})$ are complex gradient systems which extend ρ_0, then $\rho_{1|N} = \rho_{2|N}$ for a suitable open neighbourhood $N \subset N_1 \cap N_2$ of M.

Proof It is not restrictive to assume that the map $\widetilde{G}_{\rho_0}$ is a diffeomorphism between $\widetilde{D} = D \cap M \times i\mathscr{V}$ and $F_{\rho_0}(\widetilde{D})$.

Fix a basis $\{V_1, \ldots, V_k\}$ of $\mathscr{V}$ and set

$$\xi^0_\alpha = \rho(V_\alpha), \quad \alpha = 1, \ldots, k.$$

Let J denote the complex structure on $T\widetilde{D}$ induced by the pullback of the complex structure on $TF_{\rho_0}(\widetilde{D}) \subset T\widetilde{M}$ under the map F_{ρ_0}.

It is not restrictive to identify the neighbourhood $F_{\rho_0}(\widetilde{D})$ of M in $\widetilde{M}$ with the domain $\widetilde{D} \subset M \times i\mathscr{V}$. We also identify $M \times i\mathscr{V}$ with $M \times \mathbb{R}^k$ by

$$M \times \mathbb{R}^k \ni (p, u_1, \ldots, u_k) \mapsto (p, iu_1V_1 + \cdots + iu_kV_k) \in M \times i\mathscr{V}.$$

Let

$$U = (u_1, \ldots, u_k) : M \times \mathbb{R}^k \to \mathbb{R}^k \simeq \mathscr{V}$$

be the projection on the second factor.

We will prove the existence of the required complex gradient system showing that there exist vector fields

$$\xi_\alpha, \quad \alpha = 1, \ldots, k,$$

defined in a suitable neighbourhood N of $M \times \{0\}$ in $\widetilde{D}$ such that

$$\mathrm{d}u_\alpha(\xi_\beta) = 0, \quad \alpha, \beta = 1, \ldots, k,$$

and

$$\mathrm{d}^c u_\alpha(\xi_\beta) = \delta_{\alpha\beta}, \quad \alpha, \beta = 1, \ldots, k.$$

Let $\tilde{\xi}^0_\alpha$, $\alpha = 1, \ldots, k$, denote the vector fields on $\widetilde{D}$ which coincide with ξ^0_α on $M \times \{0\}$ and are invariant under the action $\mathscr{V} \times (M \times i\mathscr{V})$ given by

$$\big(W, (p, V)\big) \mapsto (p, W + V).$$

Let also $\mathscr{D}$ denote the distribution on $T(M \times i\mathscr{V}) \approx T(M \times \mathbb{R}^k)$ generated by the vector fields

$$\tilde{\xi}^0_1, \ldots, \tilde{\xi}^0_k, \frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_k}.$$

$\mathscr{D}$ is completely integrable, and the maximal integral submanifolds of $\mathscr{D}$ are of the form $S \times \mathbb{R}^k$, where S is a maximal integral submanifold of the distribution $\mathscr{D}^{\mathbb{R}}_\rho$.

By construction, the intersection of each maximal integral submanifold of $\mathscr{D}$ with the domain $\widetilde{D}$ is a complex submanifold of $\widetilde{D}$ of complex dimension k. Moreover, for each $p \in M$ and each $\alpha = 1, \ldots, k$, we have

$$J\big(\tilde{\xi}^0_\alpha\big)(p) = \frac{\partial}{\partial u_\alpha}(p).$$

Let $P = (p_{\alpha\beta})$, $Q = (q_{\alpha\beta})$ be the square matrices of order k with entry smooth function on $\widetilde{D}$ defined by

$$p_{\alpha\beta} = J\big(\tilde{\xi}^0_\beta\big)(u_\alpha),$$
$$q_{\alpha\beta} = J\left(\frac{\partial}{\partial u_\beta}\right)(u_\alpha).$$

Observe that for each $p \in M$, the matrices $P((p, 0))$ and $Q((p, 0))$ are respectively the identity matrix and the zero matrix of order k.

Let N be the open neighbourhood of $M \times \{0\}$ in $\widetilde{D}$ defined by

$$N = \big\{(p, u_1, \ldots, u_k) \in \widetilde{D} \mid \det P\big((p, u_1, \ldots, u_k)\big) \neq 0\big\}.$$

Denote by $A = (a_{\alpha\beta})$ the matrix $P^{-1}Q$, and set

$$\xi_\alpha = -J\left(\frac{\partial}{\partial u_\alpha}\right) + \sum_{\beta=1}^{k} a_{\beta\alpha} J\big(\tilde{\xi}^0_\beta\big), \quad \alpha = 1, \ldots, k.$$

Then

$$J(\xi_\alpha) = \frac{\partial}{\partial u_\alpha} - \sum_{\beta=1}^{k} a_{\beta\alpha}\tilde{\xi}^0_\beta, \quad \alpha = 1, \ldots, k,$$

and, in view of the J-invariance of the distribution $\mathscr{D}$, it follows that

$$\xi_1, \ldots, \xi_k, \; J(\xi_1), \ldots, J(\xi_k)$$

generate the distribution $\mathscr{D}$ on N. It is easy to check that

$$\mathrm{d}u_\alpha(\xi_\beta) = 0, \quad \alpha, \beta = 1, \ldots, k,$$

and

$$\mathrm{d}^c u_\alpha(\xi_\beta) = \delta_{\alpha\beta}, \quad \alpha, \beta = 1, \ldots, k,$$

as required.

In order to prove the uniqueness of the map ρ, assume that the complex gradient systems $(\mathscr{V}, \rho_1, U_{\rho_0})$, $(\mathscr{V}, \rho_2, U_{\rho_0})$ extend ρ_0 and set $\gamma = \rho_1 - \rho_2$.

We are going to prove that, after shrinking N if necessary, $\gamma_{|N} = 0$ showing before that the complex distributions $\mathscr{D}^{\mathbb{C}}_{\rho_1}$ and $\mathscr{D}^{\mathbb{C}}_{\rho_2}$ associated respectively to ρ_1 and ρ_2 coincide near to $M \times \{0\}$.

By hypothesis, the distribution $\mathscr{D}^{\mathbb{R}}_{\rho_0}$ is integrable, and its maximal integral submanifold are real submanifolds of M of (real) dimension k. For every $p \in M$, consider the maximal integral submanifolds S_1, S_2 through p of $\mathscr{D}^{\mathbb{C}}_{\rho_1}$ and $\mathscr{D}^{\mathbb{C}}_{\rho_2}$, respectively. Since ρ_1 and ρ_2 both extend ρ_0, it follows that

$$S^{\mathbb{R}} = S_1 \cap M = S_2 \cap M$$

is the maximal integral (real) submanifold of (real) dimension k of the distribution $\mathscr{D}^{\mathbb{R}}_{\rho_0}$ through p. In view of the hypothesis of CR-transversality, $S^{\mathbb{R}}$ is a totally real submanifold of S_1 and S_2. It follows that $S_1 = S_2$.

We have so proved that the maximal integral submanifolds of the distributions $\mathscr{D}^{\mathbb{C}}_{\rho_1}$ and $\mathscr{D}^{\mathbb{C}}_{\rho_2}$ which meet the submanifold M are the same, and consequently, after shrinking N if necessarily, it follows that the distributions $\mathscr{D}^{\mathbb{C}}_{\rho_1}$ and $\mathscr{D}^{\mathbb{C}}_{\rho_2}$ coincide on N.

Let now $V \in \mathscr{V}$ be an arbitrary vector. Then, $\gamma(V) \in \Gamma(N, \mathscr{H}_{\rho_1})$, and the above argument shows that $\gamma(V) \in \Gamma(N, \mathscr{D}^{\mathbb{C}}_{\rho_1})$. Since $T\widetilde{M} = \mathscr{D}^{\mathbb{R}}_{\rho_1} \oplus J\mathscr{D}^{\mathbb{R}}_{\rho_1} \oplus \mathscr{H}_{\rho_1}$ it follows that $\gamma(V)_{|N} = 0$, and this ends the proof $V \in \mathscr{V}$ being arbitrary. □

When $k = \dim \mathscr{V} = 1$, the result above is contained in [7, Theorem 3.1], where a stronger uniqueness result was obtained. Namely, if (ξ_1, u_1) and (ξ_2, u_2) are two one-dimensional calibrated foliations such that ξ_1 and ξ_2 both extend ξ_0 along the hypersurface M, then $\xi_1 = \xi_2$ and $u_1 = u_2$ in a neighbourhood of M.

Such a uniqueness result does not hold for a general complex gradient system. Indeed, consider $\widetilde{M} = \mathbb{C}$, $M = \mathbb{R}$, and

$$\xi_0 = \frac{\partial}{\partial x}.$$

Then our construction yields the gradient map

$$U(z) = U(x + iy) = -\Re(z) = -y$$

and the vector field

$$\xi = \frac{\partial}{\partial x},$$

and also the pair (ξ_1, U_1), where

$$U_1(z) = U_1(x + iy) = e^{-y} - 1$$

and

$$\xi_1 = e^y \frac{\partial}{\partial x},$$

is a complex gradient system which extends ξ_0.

In the case $k = 1$ the condition $[\xi, J\xi] = 0$, which is in the definition of one-dimensional calibrated foliation given in [7], ensures the uniqueness for the Cauchy problem (see [7, Theorem 3.1]). It is not clear which is the right condition (if any) to add in order to guarantee the uniqueness also in this noncommutative setting.

See also the examples given in the next section.

4 Lie Groups

Let $G^{\mathbb{C}}$ be a complex Lie group of (complex) dimension k which is the complexification of a real Lie group G; G is a totally real submanifold of $G^{\mathbb{C}}$. Let $\mathfrak{g}^{\mathbb{C}}$ and $\mathfrak{g}$ be the Lie algebras of $G^{\mathbb{C}}$ and G, respectively.

We identify $\mathfrak{g}$ (resp. $\mathfrak{g}^{\mathbb{C}}$) with the tangent space to G ($G^{\mathbb{C}}$) at the origin, and for each $V \in \mathfrak{g}$ (resp. $V \in \mathfrak{g}^{\mathbb{C}}$), we denote by L_V the corresponding left-invariant vector field on G (resp. $G^{\mathbb{C}}$). The complexification of the flow associated to the map $\mathfrak{g}^{\mathbb{C}} \ni V \mapsto L_V$ is the map

$$G \times \mathfrak{g}^{\mathbb{C}} \ni (g, V) \mapsto g \exp(V) \in G^{\mathbb{C}},$$

exp being the standard exponential map $\exp : \mathfrak{g}^{\mathbb{C}} \to G^{\mathbb{C}}$. Let us denote by $(\mathfrak{g}, \rho, U)$ the complex gradient system which extends $V \mapsto L_V$. Then we have the identity

$$U\big(g \exp(-iV)\big) = V.$$

If $G^{\mathbb{C}}$ is a complex reductive Lie group and G is a compact real form for $G^{\mathbb{C}}$, then we have the Cartan decomposition of $G^{\mathbb{C}}$

$$G \times \mathfrak{g} \to G^{\mathbb{C}}$$
$$(g, V) \mapsto g \exp(iV).$$

In this case the algebra $\mathfrak{g}$ admits a definite metric B, invariant under the adjoint representation Ad_G of G, inducing an isomorphism between the Lie algebra $\mathfrak{g}$ and its dual $\mathfrak{g}^*$. With this identification, the gradient map U is (up to the sign) the moment map associated to a symplectic action of G on $G^{\mathbb{C}}$. See e.g. [5] for details.

This example explains our terminology "*complex gradient system*".

We would like to point out that in general, as shown by the examples below, the representation ρ of the complex gradient system extending the left representation $V \mapsto L_V$ is not the restriction of the left representation of $\mathfrak{g}^{\mathbb{C}}$.

Let $G^{\mathbb{C}}$ be the matrix Lie group of the matrices of the form

$$\begin{pmatrix} 1 & z_1 & z_3 \\ 0 & 1 & z_2 \\ 0 & 0 & 1 \end{pmatrix}$$

with $z_\alpha = x_\alpha + i y_\alpha \in \mathbb{C}$, $\alpha = 1, 2, 3$, and let G be the corresponding group with $z_i \in \mathbb{R}$.

Then $\mathfrak{g}$ is given by the matrices of the form

$$\begin{pmatrix} 0 & u_1 & u_3 \\ 0 & 0 & u_2 \\ 0 & 0 & 0 \end{pmatrix}$$

with $u_\alpha \in \mathbb{R}$, $\alpha = 1, \dots, 3$.

Put

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then E_1, E_2, E_3 is a basis of $\mathfrak{g}$. Denoting $L_\alpha = L_{E_\alpha}$, we then have

$$L_1 = \frac{\partial}{\partial x_1},$$
$$L_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3} + y_1 \frac{\partial}{\partial y_3},$$
$$L_3 = \frac{\partial}{\partial x_3},$$
$$J L_1 = \frac{\partial}{\partial y_1},$$

$$JL_2 = \frac{\partial}{\partial y_2} + x_1 \frac{\partial}{\partial y_3} - y_1 \frac{\partial}{\partial x_3},$$
$$JL_3 = \frac{\partial}{\partial y_3}$$

Some computation yields for the gradient map U the expression

$$U(z_1, z_2, z_3) = -y_1 E_1 - y_2 E_2 - (y_3 + x_1 y_2) E_3,$$

and the representation ρ is given by

$$\rho(E_1) = \widetilde{E}_1 = \frac{\partial}{\partial x_1} + y_2 \frac{\partial}{\partial y_3} = L_1 + y_2 J(L_3),$$
$$\rho(E_2) = \widetilde{E}_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3} = L_2 - y_1 J(L_3),$$
$$\rho(E_3) = \widetilde{E}_3 = \frac{\partial}{\partial x_3} = L_3.$$

Observe that

$$[\widetilde{E}_1, \widetilde{E}_2] = \widetilde{E}_3, \quad [\widetilde{E}_1, \widetilde{E}_3] = [\widetilde{E}_2, \widetilde{E}_3] = 0,$$
$$[J\widetilde{E}_1, J\widetilde{E}_2] = \widetilde{E}_3, \quad [J\widetilde{E}_1, J\widetilde{E}_3] = [J\widetilde{E}_2, J\widetilde{E}_3] = 0,$$
$$[\widetilde{E}_i, J\widetilde{E}_j] = 0, \quad i, j = 1, 2, 3.$$

It follows that the representation $\rho : \mathfrak{g} \to \Gamma(G^{\mathbb{C}}, TG^{\mathbb{C}})$ is a Lie algebra isomorphism and the gradient map U is a harmonic function.

Let now $G^{\mathbb{C}}$ be the matrix Lie group of the matrices of the form

$$\begin{pmatrix} z_1 & z_2 \\ 0 & 1 \end{pmatrix}$$

with $z_1, z_2 \in \mathbb{C}$, $z_1 \neq 0$, and let G be the corresponding group with $z_1, z_2 \in \mathbb{R}$.

The Lie algebra $\mathfrak{g}$ of G is given by the matrices of the form

$$\begin{pmatrix} u_1 & u_2 \\ 0 & 0 \end{pmatrix}$$

with $u_1, u_2 \in \mathbb{R}$. The matrices

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

form a basis of the Lie algebra $\mathfrak{g}$ which satisfies the relation

$$[E_1, E_2] = E_2.$$

The corresponding left-invariant vector fields on $G^{\mathbb{C}}$ are given by

$$L_1 = x_1 \frac{\partial}{\partial x_1} + y_1 \frac{\partial}{\partial y_1},$$

$$L_2 = x_1 \frac{\partial}{\partial x_2} + y_1 \frac{\partial}{\partial y_1},$$

$$JL_1 = -y_1 \frac{\partial}{\partial x_1} + x_1 \frac{\partial}{\partial y_1},$$

$$JL_2 = -y_1 \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial y_2}.$$

After some computations we obtain that the gradient map is given by

$$U(z_1, z_2) = -\theta_1 E_1 - \frac{y_2 \theta_1}{y1} E_2,$$

where

$$\theta_1 = \arctan \frac{y_1}{x_1},$$

and the representation ρ satisfies

$$\rho(E_1) = \widetilde{E}_1 = x_1 \frac{\partial}{\partial x_1} + y_1 \frac{\partial}{\partial y_1} + y_2 \left(\frac{x_1}{y_1} - \frac{1}{\theta_1} \right) \frac{\partial}{\partial x_2} + y_2 \frac{\partial}{\partial y_2},$$

$$\rho(E_2) = \widetilde{E}_2 = \frac{y_1}{\theta_1} \frac{\partial}{\partial x_2}.$$

Observe that

$$[\widetilde{E}_1, \widetilde{E}_2] = [J\widetilde{E}_1, J\widetilde{E}_2] = \widetilde{E}_2$$

and

$$[\widetilde{E}_1, J\widetilde{E}_1] = \frac{2y_2}{y_1} \left(\frac{x_1}{y_1} - \frac{1}{\theta_1} \right) \widetilde{E}_2,$$

$$[\widetilde{E}_1, J\widetilde{E}_2] = -\left(\frac{x_1}{y_1} - \frac{1}{\theta_1} \right) \widetilde{E}_2,$$

$$[\widetilde{E}_2, J\widetilde{E}_2] = 0,$$

namely the representation $\rho : \mathfrak{g} \to \Gamma(G^{\mathbb{C}}, TG^{\mathbb{C}})$ is a Lie algebra isomorphism, but the gradient map U is not a harmonic function, and the vector fields $\widetilde{E}_1$, $\widetilde{E}_2$, $J\widetilde{E}_1$ and $J\widetilde{E}_2$ are not a basis of a Lie sub-algebra of $\Gamma(G^{\mathbb{C}}, TG^{\mathbb{C}})$.

5 The Holomorphic Abelian Case

Let $(\mathscr{V}, \rho, U)$ be a complex gradient system.

With a little abuse of language, we say that such a complex gradient system is *holomorphic* if $\rho^{\mathbb{C}}(V)$ is a holomorphic vector field on $\widetilde{M}$ for each $V \in \mathscr{V}$.

We also say that it is *abelian* if

$$\left[\rho^{\mathbb{C}}(V), \rho^{\bar{\mathbb{C}}}(W)\right] = 0$$

for each pair of vectors $V, W \in \mathscr{V}$. Such a condition is equivalent to say that for each pair of vectors $V, W \in \mathscr{V}$, one has

$$\left[\rho(V), \rho(W)\right] = \left[\rho(V), J\rho(W)\right] = \left[J\rho(V), J\rho(W)\right] = 0.$$

Consider now a domain $\Omega \subset \mathbb{C}^n$ and let $F : \Omega \to \mathbb{R}^k$ be a smooth function. We associate to F a complex gradient system as follows.

Set $\widetilde{M} = \Omega \times \mathbb{C}^k$, $\mathscr{V} = \mathbb{R}^k$ and define

$$U(z, w) = F(x, y) - u,$$

where $z = x + iy$ and $w = t + iu$ with $x, y \in \mathbb{R}^n$ and $t, u \in \mathbb{R}^k$. Finally consider the linear map $\rho : \mathbb{R}^k \to \Gamma(\widetilde{M}, T\widetilde{M})$ characterized by the conditions

$$\rho(e_\alpha) = \frac{\partial}{\partial t_\alpha}, \quad \alpha = 1, \dots, k,$$

where $e_1, \dots, e_k$ is the canonical basis of $\mathbb{R}^k$.

It is easy to show that this complex gradient system is holomorphic and abelian, and the aim of the next theorem is to prove that it is the local model of any holomorphic abelian complex gradient system. Namely the following is true.

Theorem 3 *Let $\widetilde{M}$ be a complex manifold of complex dimension $n + k$. Let $(\mathbb{R}^k, \rho, U)$ be a holomorphic abelian complex gradient system on $\widetilde{M}$. Then for each point p, there exist a complex coordinate system*

$$z = (z_1, \dots, z_n), \quad w = (w_1, \dots, w_k),$$

$z_\mu = x_\mu + iy_\mu$, $\mu = 1, \dots, n$, $w_\alpha = t_\alpha + iu_\alpha$, $\alpha = 1, \dots, k$, $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, $u = (u_1, \dots, u_k)$, *and a smooth* (*vector*) *function F depending only on x and y such that*

$$\rho(e_\alpha) = \frac{\partial}{\partial t_\alpha}, \quad \alpha = 1, \dots, k,$$
$$U(z, w) = F(x, y) - u.$$

Proof Let

$$g_t^1, \dots, g_t^k, \ h_t^1, \dots, h_t^k$$

be the (local) one parameter group of transformation of $\widetilde{M}$ generated by the vector fields

$$\xi_1, \ldots, \xi_k, J(\xi_1), \ \ldots, J(\xi_k).$$

By hypotheses the Lie brackets between all pairs of vector fields among $\xi_1, \ldots, \xi_k$ and $J\xi_1, \ldots, J\xi_k$ are zero, and hence the transformations $g_t^1, \ldots, g_t^k$ and $h_t^1, \ldots, h_t^k$ commute each other.

Let $p \in \widetilde{M}$ be fixed, and let $z_1, \ldots, z_{n+k}$ be a complex coordinates system around p, where $z_\mu = x_\mu + i y_\mu$, $\mu = 1, \ldots, n+k$.

After reordering the coordinates we may suppose that

$$\frac{\partial}{\partial x_1}, \frac{\partial}{\partial y_1}, \ \ldots, \ \frac{\partial}{\partial x_n}, \frac{\partial}{\partial y_n}, \ \xi_1, \ J\xi_1, \ \ldots, \ \xi_k, \ J\xi_k$$

generate the tangent space to $\widetilde{M}$ at each point in a suitable neighbourhood of p.

For $\alpha = 1, \ldots, k$, set $w_\alpha = t_\alpha + i u_\alpha$ and define

$$G^\alpha_{w_\alpha} = g^\alpha_{t_\alpha} \circ h^\alpha_{u_\alpha}.$$

Then the map

$$(z_1, \ldots, z_n, w_1, \ldots, w_k) \mapsto G^1_{w_1} \circ \cdots \circ G^k_{w_k}(z_1, \ldots, z_n, 0, \ldots, 0)$$

is a diffeomorphism φ between an open set of $\mathbb{C}^{n+k}$ and a suitable neighbourhood U of p in $\widetilde{M}$, that is,

$$x_1, \ y_1, \ \ldots, \ x_n, \ y_n, \ t_1, \ u_1, \ \ldots, \ t_k, \ u_k$$

is a real coordinate system on U.

Since the maps $G^1_{w_1} \ldots G^k_{w_k}$ commute each other, it follows that with respect to such a coordinate system we have

$$\xi_\alpha = \rho(e_\alpha) = \frac{\partial}{\partial t_\alpha}$$

for $\alpha = 1, \ldots, k$.

We now prove

$$z_1, \ldots, z_n, \ w_1, \ldots, w_k$$

are complex coordinates on U, showing that the diffeomorphism φ is in fact a biholomorphism.

Since, by hypotheses, $(\mathbb{R}^k, \rho, U)$ is a holomorphic abelian complex gradient system, it follows that $G^1_{w_1} \ldots G^k_{w_k}$ are holomorphic local diffeomorphisms and for fixed $w_1, \ldots, w_k$, the map

$$\varphi(z_1, \ldots, z_n, w_1, \ldots, w_k)$$

is holomorphic with respect to the variables $z_1, \ldots, z_n$. Moreover, for $\alpha = 1, \ldots, k$, the maps g_{t_α} and h_{u_α} commute, and hence the map $w_\alpha \mapsto G^\alpha_{w_\alpha}(\cdot)$ is holomorphic with respect to w_α. On the other hand, since the maps $G^1_{w_1} \ldots G^k_{w_k}$ commute each other, the map φ is holomorphic with respect to the variable w_α, $\alpha = 1, \ldots, k$, when the variables $z_1, \ldots, z_n$ and $w_1, \ldots, w_{\alpha-1}, w_{\alpha+1}, \ldots, w_k$ are fixed.

Thus the map φ is separately holomorphic in each variable and hence is holomorphic.

Finally, let $U = (U_1, \ldots, U_k) : \widetilde{M} \to \mathbb{R}^k$ be the gradient map and consider the map $F = (F_1, \ldots, F_k) : U \to \mathbb{R}^k$ defined by

$$F_\alpha(z, w) = F_\alpha(x, y, t, u) = U_\alpha(z, w) + u_\alpha, \quad \alpha = 1, \ldots, k.$$

We end the proof showing that the map F does not depend on the variables t and u. Indeed, for $\alpha, \beta = 1, \ldots, k$, we have

$$\frac{\partial F_\alpha}{\partial t_\beta} = \xi_\beta(U_\alpha) = 0$$

and

$$\frac{\partial F_\alpha}{\partial u_\beta} = J(\xi_\beta)(U_\alpha) + \delta_{\alpha\beta} = -\delta_{\alpha\beta} + \delta_{\alpha\beta} = 0.$$

□

References

1. da Silva, A.C.: Lectures on Symplectic Geometry. Lecture Notes in Mathematics, vol. 1764. Springer, Berlin (2001)
2. Dragomir, S., Tomassini, G.: Differentiable Geometry and Analysis on CR Manifolds. Birkhauser, Basel (2006)
3. Duchamp, T., Kalka, M.: Singular Monge–Ampère foliations. Math. Ann. **325**, 187–209 (2003)
4. Heinzner, P., Schuetzdeller, P.: Convexity properties of gradient maps. Adv. Math. **225**(3), 1119–1133 (2010)
5. Heinzner, P., Schwarz, G.: The Cartan decomposition of the moment map. Math. Ann. **337**, 197–232 (2007)
6. McDuff, D., Salomon, D.: Introduction to Symplectic Topology. Oxford Mathematical Monographs. Oxford University Press, London (1995)
7. Tomassini, G., Venturini, S.: Contact geometry of one dimensional holomorphic foliations. Indiana J. Math. (2009, to appear). arXiv:0907.5082v1
8. Tomassini, G., Venturini, S.: Adapted complex tubes on the symplectization of pseudo-Hermitian manifolds. (2010). arXiv:1002.4558
9. Tomassini, G., Venturini, S.: Adapted complex tubes on the symplectization of pseudo-Hermitian manifolds. Arkive Math. **96**, 77–83 (2011)

Coleff–Herrera Currents Revisited

Alekos Vidras and Alain Yger

Abstract In the present paper, we describe the recent approach to residue currents by Andersson, Björk, and Samuelsson (Andersson in Ann. Fac. Sci. Toulouse Math. Sér. 18(4):651–661, 2009; Björk in The Legacy of Niels Henrik Abel, pp. 605–651, Springer, Berlin, 2004; Björk and Samuelsson in J. Reine Angew. Math. 649:33–54, 2010), focusing primarily on the methods inspired by analytic continuation (that were initiated in a quite primitive form in Berenstein et al. in Residue Currents and Bézout Identities. Progress in Mathematics, vol. 114, Birkhäuser, Basel, 1993). Coleff–Herrera currents (with or without poles) play indeed a crucial role in Lelong–Poincaré-type factorization formulas for integration currents on reduced closed analytic sets. As revealed by local structure theorems (which can also be understood as global when working on a complete algebraic manifold due to the GAGA principle), such objects are of algebraic nature (antiholomorphic coordinates playing basically the role of "inert" constants). Thinking about division or duality problems instead of intersection ones (especially in the "improper" setting, which is certainly the most interesting), it happens then to be necessary to revisit from this point of view the multiplicative inductive procedure initiated by Coleff and Herrera (Lecture Notes in Mathematics, vol. 633, Springer, Berlin, 1978), this being the main objective of this presentation. In homage to the pioneering work of Leon Ehrenpreis, to whom we are both deeply indebted, and as a tribute to him, we also suggest a currential approach to the so-called Nœtherian operators that remain the key stone in various formulations of Leon's Fundamental Principle.

1 From Poincaré–Leray to Coleff–Herrera Construction

Let $\mathcal{X}$ be a complex n-dimensional analytic manifold. Consider $M \leq n$ closed hypersurfaces $S_1, \ldots, S_M$ in $\mathcal{X}$ that intersect as a nonempty complete intersec-

A. Vidras (✉)
Department of Mathematics and Statistics, University of Cyprus, Nicosia 1678, Cyprus
e-mail: msvidras@ucy.ac.cy

A. Yger
Institut de Mathématiques, Université de Bordeaux, 33405, Talence, France
e-mail: Alain.Yger@math.u-bordeaux1.fr

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_21, © Springer-Verlag Italia 2012

tion, that is, the closed analytic subset $V = \bigcap_{j=1}^{M} S_j \subset \mathcal{X}$ is purely $(n-M)$-dimensional (all its irreducible components have complex dimension $n-M$). When $S_1, \dots, S_M$ are assumed to be smooth and moreover to intersect transversally, a well-known construction by Leray [22] (see also [1]) leads to the construction (from the cohomological point of view) of the iterated Poincaré residue morphism from $H^p(\mathcal{X} \setminus S_1 \cup \cdots \cup S_M, \mathbb{C})$ into $H^{p-M}(V, \mathbb{C})$ (paired with its dual iterated coboundary morphism) when $p \geq M$. Following a currential (instead of cohomological) point of view, the construction proposed by Coleff and Herrera [14] allows us to drop the assumption about smoothness of the S_j's and the fact they intersect transversally, keeping just (for the moment) the complete intersection hypothesis. We propose here to make explicit in this introduction the bridge between such currential construction and J. Leray's approach. In order to do that, we recalls a concept, which is of interest by itself for algebraic reasons, of *multilogarithmic meromorphic form* [7, 25].

Definition 1 Let $\mathcal{X}$ and $S_1, \dots, S_M, V$ be as above. A meromorphic $(p,0)$-form ω on $\mathcal{X}$ $(M \leq p \leq n)$, with polar set contained in $\bigcup_{j=1}^{M} S_j$, is called *multilogarithmic with respect to* $S_1, \dots, S_M$ if and only if, for any $x \in V$, one can find an open neighborhood U_x of x and M holomorphic functions $s_{1,x}, \dots, s_{M,x}$ in U_x such that:

- for any $j = 1, \dots, M$, the hypersurface $S_j \subset \mathcal{X}$ is defined in U_x as $\{s_{j,x} = 0\}$;
- $ds_{1,x} \wedge \cdots \wedge ds_{M,x}$ is not vanishing identically on any irreducible component of V in U_x, that is, the complete intersection $V \cap U_x$ is defined by the $s_{j,x}$, $j = 1, \dots, M$, as a *reduced complete intersection*;
- for any $j = 1, \dots, M$, the differential forms $s_{j,x}\omega$ and $s_{j,x}\, d\omega$ (or, equivalently, $s_{j,x}\omega$ and $ds_{j,x} \wedge \omega$) can be expressed in U_x as $\sum_{l=1}^{M} \omega_l$, where ω_l is a meromorphic form with polar set contained in $\bigcup_{l' \neq l} S_{l'} \cap U_x$.

Consider $\mathcal{X}$, the S_j's and ω as in Definition 1. Let V_{sing} be the set of singular points of V, and let $U = \mathcal{X} \setminus V_{\text{sing}}$. The closed hypersurfaces $\Sigma_j = S_j \cap U \subset U$, $j = 1, \dots, M$ (considered as closed hypersurfaces in U), are smooth and intersect transversally in some open neighborhood $\widetilde{U} \subset U$ of $W = \bigcap_{j=1}^{M} \Sigma_j$. Under these conditions, one can define on the complex submanifold $W \subset \widetilde{U}$ the Leray–Poincaré residue $\text{Res}_{\Sigma_1, \dots, \Sigma_M}[\omega]$ of the meromorphic form ω (considered as multilogarithmic in $\widetilde{U}$ with respect to $\Sigma_1, \dots, \Sigma_M$). Let us recall here this construction. For any $x \in W \subset V$, one can find an open neighborhood U_x in U such that $ds_{1,x} \wedge \cdots \wedge ds_{M,x}$ does not vanish identically on any irreducible component of $V \cap U_x$. If $y \in W \cap U_x$ and $ds_{1,x}(y) \wedge \cdots \wedge ds_{M,x}(y) \neq 0$, then $\{s_{j,x} = 0\}$ is necessarily a reduced equation for the complex submanifold Σ_j about y. In a neighborhood $U_{x,y} \subset U_x$ of such $y \in W \cap U_x$, $ds_{1,x} \wedge \cdots \wedge ds_{M,x}$ does not vanish, and thus we can write a local division formula (iterating with respect to $j = 1, \dots, M$ the division procedure for differential forms, as introduced by G. de Rham and extensively used in [22]):

$$\omega = \left(\bigwedge_{j=1}^{M} \frac{ds_{j,x}}{s_{j,x}} \right) \wedge r_{x,y}[\omega] + \sigma_{x,y}[\omega],$$

where the $(p-M,0)$-form $r_{x,y}[\omega]$, also denoted by $\mathrm{Res}_{\Sigma_1,\dots,\Sigma_M,x}[\omega]$, and the $(p,0)$-form $\sigma_{x,y}[\omega]$ are both meromorphic, of the form $\sum_l \varphi_{l,x,y}$, $\varphi_{l,x,y}$ being a meromorphic form in $U_{x,y}$ with polar set contained in $\bigcup_{l'\neq l} S_{l'}$. The restriction of every $r_{x,y}[\omega]$ to W is a $\bar\partial$-closed, holomorphic $(p-M)$-differential form on the closed submanifold $W\cap U_{x,y}$. All such forms $\mathrm{Res}_{\Sigma_1,\dots,\Sigma_M,x}[\omega]$, $x\in W$, fit together to form a holomorphic, $\bar\partial$-closed form on the closed manifold W, which is precisely the *Poincaré–Leray residue of* ω *on* W and is denoted as $\mathrm{Res}_{\Sigma_1,\dots,\Sigma_M}[\omega]$. Such a holomorphic $(p-M)$-differential form on the manifold $W\subset U$ defines a (p,M)-current on U:

$$\mathrm{Res}_{\Sigma_1,\dots,\Sigma_M}[\omega]\colon \varphi\in\mathcal{D}^{n-p,n-M}(U,\mathbb{C})\mapsto \int_W \mathrm{Res}_{\Sigma_1,\dots,\Sigma_M}[\omega]\wedge\varphi. \qquad (1)$$

The main issue now is to extend (in some standard way) the (p,M)-current (1) to a (p,M)-current T over the whole manifold $\mathcal{X}$ so that $\mathrm{supp}\,T\subset W$ and $\bar\partial T=0$. There are different ways of doing this, but, for reasons of algebraic nature that will be made explicit later on, the one we adopt here is based on the analytic continuation of meromorphic current-valued maps. The use of this approach in different settings is the main theme of the present paper. It is based on an algorithmic construction of $\bar\partial$-closed currents sharing a common holonomy property.

To be more specific, we consider a finite collection $f_1, f_2,\dots,f_m$ of holomorphic functions in an open set $\Omega\subset\mathbb{C}^n$, where $m\le n$, and a collection of natural numbers $q_1,q_2,\dots,q_m\in\mathbb{N}$. We define now the current

$$T^f_{\underline{q},1}=\left[\bar\partial\left(\frac{|f_1|^{2\lambda}}{f_1^{q_1}}\right)\right]_{\lambda_1=0}=\bar\partial\left[(\mathbf{1}-\mathbf{1}_{[f_1=0]})\frac{1}{f_1^{q_1}}\right],$$

where $[f_1=0]$ denotes the principal Weil divisor $\mathrm{div}\,(f_1)$. For a holomorphic function h in Ω, there exists, by the result of Sabbah [26] (completed later on by Gyoja [19]), about any point z in Ω, a local formal Bernstein–Sato equation, originally introduced in [6],

$$\mathcal{Q}_z(\lambda_1,\lambda_2,\zeta,\partial/\partial\zeta)\big[h^{\lambda_2+1}f_1^{\lambda_1}\big]=\prod_\iota(\alpha_{0,\iota}+\alpha_{1,\iota}\lambda_1+\alpha_{2,\iota}\lambda_2)h^{\lambda_2}f_1^{\lambda_1}, \qquad (2)$$

where $\alpha_{0,\iota}\in\mathbb{N}^*$, $(\alpha_{1,\iota},\alpha_{2,\iota})\in\mathbb{N}^2\setminus\{(0,0)\}$. This result extends to the context of two functions a deep result due to Kashiwara [20]. Exploiting this local formal equation (2) in the sense of distributions in a neighborhood U_z of z, one has, by lifting the antiholomorphic polar parts, that

$$\begin{aligned} b_z(\lambda_1,\lambda_2)\left(\frac{|h|^{2\lambda_2}}{h}|f_1|^{2\lambda_1}\right) &= \text{(formally) } b_z(\lambda_1,\lambda_2)\,\overline{h}^{\lambda_2}\overline{f}_1^{\lambda_1}\,f_1^{\lambda_1}h^{\lambda_2-1} \\ &= \overline{\mathcal{Q}}_z(\lambda_1,\lambda_2,\overline{\zeta},\partial/\partial\overline{\zeta})\left[|h|^{2\lambda_2}\frac{\overline{h}}{h}|f_1|^{2\lambda_1}\right], \end{aligned} \qquad (3)$$

whenever $\mathrm{Re}\,\lambda_1\gg 1$ and $\mathrm{Re}\,\lambda_2\gg 1$. Using the fact that any distribution coefficient τ of the current $T^f_{\underline{q},1}$ can be achieved through analytic continuation as $\tau=[\tau_{\lambda_1}]_{\lambda_1=0}$

(where τ_{λ_1} is a distribution coefficient of $\overline{\partial}(|f_1|^{2\lambda_1}/f_1^{q_1})$), we deduce from (3) the identity

$$b_z(0,\lambda_2)\left(\frac{|h|^{2\lambda_2}}{h}\otimes\tau\right)=\overline{\mathcal{Q}}_z(0,\lambda_2,\overline{\zeta},\partial/\partial\overline{\zeta})\left[\left(|h|^{2\lambda_2}\frac{\overline{h}}{h}\right)\otimes\tau\right]$$

(in the sense of distributions about z) for $\operatorname{Re}\lambda_2\gg 1$. Iterating the above identity M times, we get

$$\begin{aligned}b_z(0,\lambda_2)\cdots b_z(0,\lambda_2+M-1)&\left(\frac{|h|^{2\lambda_2}}{h}\otimes\tau\right)\\&=\overline{\mathfrak{Q}}_{z,M}(\lambda_2,\overline{\zeta},\partial/\partial\overline{\zeta})\left[\left(|h|^{2\lambda_2}\frac{\overline{h}^M}{h}\right)\otimes\tau\right]\end{aligned}\tag{4}$$

for some differential operator $\mathfrak{Q}_{z,M}$. Provided that M is sufficiently large, we deduce from (4) that the map

$$\lambda_2\mapsto\frac{|h|^{2\lambda_2}}{h}\,T^f_{\underline{q},1}$$

can be continued as a holomorphic map to some half-plane $\{\operatorname{Re}\lambda_2>-\eta\}$. Furthermore, if u is an invertible holomorphic function in Ω, then any differentiation of $|u|^{2\lambda_2}$ generates λ_2 as a factor. Thus the value of the analytic continuation of

$$\lambda_2\mapsto\frac{|uh|^{2\lambda_2}}{h}\,T^f_{\underline{q},1}=\frac{|h|^{2\lambda_2}}{h}\,|u|^{2\lambda_2}\,T^f_{\underline{q},1}$$

at $\lambda_2=0$ is independent of u. This is a remarkable *holonomy property* allowing us to use the above process iteratively. In particular, the definition of

$$T^f_{\underline{q},2}=\left[\overline{\partial}\left(\frac{|f_2|^{2\lambda_2}}{f_2^{q_2}}\,T^f_{\underline{q},1}\right)\right]_{\lambda_2=0}=\left[\overline{\partial}\left(\frac{|f_2|^{2\lambda_2}}{f_2^{q_2}}\right)\wedge T^f_{\underline{q},1}\right]_{\lambda_2=0}$$

is then justified. In a similar manner, by using a slightly more general form of (2), given by

$$\begin{aligned}\mathcal{Q}_z(\lambda_1,\ldots,\lambda_m,\zeta,\partial/\partial\zeta)&\left[h^{\lambda_m+1}\prod_{j=1}^{m-1}f_j^{\lambda_j}\right]\\&=\left(\prod_\iota\left(\alpha_{\iota 0}+\sum_{j=1}^m\alpha_{\iota j}\lambda_j\right)\right)h^{\lambda_m}\prod_{j=1}^{m-1}f_j^{\lambda_j},\end{aligned}$$

we can construct a current $T^f_{\underline{q},3}$ (for $m=3$) by multiplying the current $T^f_{\underline{q},2}$ with a suitable meromorphic function. We continue this iteration of the analytic continuation process until the current $T^f_{\underline{q},m}$ is constructed. What is important in this approach is that it is algorithmic and essentially algebraic, because of the use of Bernstein–Sato relations. No log resolution of singularities is explicitly involved in the picture. Furthermore, this procedure mimics the Leray iterated residue construction. An interesting application of the above approach is the following:

Proposition 1 *Let $\mathcal{X}$, the S_j's, V, and ω be as before. Let $U = \mathcal{X} \setminus V_{\mathrm{sing}}$, and $\Sigma_j = S_j \cap U$ for $j = 1, \ldots, M$. The closed hypersurfaces $\Sigma_1, \ldots, \Sigma_M$ (in U) are smooth and intersect transversally in some open neighborhood (in U) of $W = \bigcap_{j=1}^{M} \Sigma_j$, which allows us to define the Poincaré–Leray residue $\mathrm{Res}_{\Sigma_1,\ldots,\Sigma_M}[\omega]$ as a $(p-M, 0)$-holomorphic form on the closed submanifold W of U. The associated (p, M)-current $\mathrm{Res}_{\Sigma_1,\ldots,\Sigma_M}[\omega]$ in U, acting as (1), is the restriction of a $\overline{\partial}$-closed (p, M)-current T over $\mathcal{X}$, with $\mathrm{Supp}\, T \subset V$.*

Proof Let $x \in V$, and let U_x be the neighborhood attached to the multilogarithmicity of ω as described in Definition 1. Since ω is a meromorphic form with polar set in $\bigcup_{j=1}^{M} S_j$, one can express ω in U_x as

$$\omega = \frac{\psi_x}{s_{1,x}^{q_{1,x}} \cdots s_{M,x}^{q_{M,x}}},$$

where ψ_x is holomorphic in U_x. Consider the ${}'\mathcal{D}^{(p,M)}(U_x, \mathbb{C})$-valued map defined on $\{\mathrm{Re}\,\lambda_1 \gg 1, \ldots, \mathrm{Re}\,\lambda_M \gg 1\}$ as

$$\begin{aligned}(\lambda_1, \ldots, \lambda_M) \quad \longmapsto \quad R^{s_x,\lambda_1,\ldots,\lambda_M}[\omega] &= \frac{(-1)^{M(M-1)/2}}{(2i\pi)^M} \left(\bigwedge_{j=1}^{M} \overline{\partial} |s_{j,x}|^{2\lambda_j} \right) \wedge \omega \\ &= \frac{1}{(2i\pi)^M} \left(\bigwedge_{j=M}^{1} \overline{\partial} \left(\frac{|s_{j,x}|^{2\lambda_j}}{s_{j,x}^{q_{j,x}}} \right) \right) \wedge \psi_x .\end{aligned}$$

The reverse order of indices expresses here the absorption of the factor $\frac{(-1)^{M(M-1)/2}}{(2i\pi)^M}$. It is known from [27] that the current-valued map

$$(\lambda_1, \ldots, \lambda_M) \longmapsto R^{s_x,\lambda_1,\ldots,\lambda_M}[\omega]$$

can be continued analytically as a function of M complex variables $(\lambda_1, \ldots, \lambda_M)$ to $\{\mathrm{Re}\,\lambda_1 > -\eta, \ldots, \mathrm{Re}\,\lambda_M > -\eta\}$ for some $\eta > 0$. The proof of such a result relies deeply on the use of a log resolution $\widetilde{\mathcal{X}} \xrightarrow{\pi} \mathcal{X}$ such that $\pi^{-1}[\bigcup_j S_j]$ is a hypersurface with normal crossings. The approach we developed above for construction of $\overline{\partial}$-closed (p, M)-current in U_x through the iterated analytic continuation process

$$R^{s_x}[\omega] = \Big[\Big[\cdots\Big[\Big[R_x^{s_x,\lambda_1,\ldots,\lambda_M}[\omega]\Big]_{\lambda_1=0}\Big]_{\lambda_2=0}\cdots\Big]_{\lambda_{M-1}=0}\Big]_{\lambda_M=0}$$

is applied at this point, taking successively λ_1 up to $\{\mathrm{Re}\,\lambda_1 > -\eta_1\}$, then λ_2 up to $\{\mathrm{Re}\,\lambda_2 > -\eta_2\}$, and so on. Note (again) that the argument does not seem (apparently) to require the use of an appropriate log resolution to resolve singularities (namely here that of the hypersurface defined as the zero set of $hf_1 \cdots f_{m-1}$), but this is indeed hidden behind the fact that there exist local Bernstein–Sato equations. The current $R^{s_x}[\omega]$ is also denoted as

$$R^{s_x}[\omega] = \left(\bigwedge_{j=1}^{M} \overline{\partial}\left(\frac{1}{s_{j,x}^{q_{j,x}}}\right)\right) \wedge \psi_x.$$

To show that all $R^{s_x}[\omega]$, for the different U_x, globalize into a $\overline{\partial}$-closed, (p, M)-current over $\mathcal{X}$, we use the holonomy of the currents under consideration. That is, for any holomorphic functions u, h in U_x, with u nonvanishing, the current-valued function

$$\lambda \in \{\mathrm{Re}\,\lambda \gg 1\} \longrightarrow \frac{|uh|^\lambda}{h} R^{s_x}[\omega]$$

can be continued analytically into a half-plane $\{\mathrm{Re}\,\lambda > -\eta\}$, whose value at $\lambda = 0$ is independent of u. The global $\overline{\partial}$-closed (p, M)-current thus obtained is denoted as $R_{[S_1]_{\mathrm{red}},\ldots,[S_M]_{\mathrm{red}}}[\omega]$. This reflects the fact that it depends only on the meromorphic form ω and on the reduced cycles corresponding to the closed hypersurfaces $S_1, \ldots, S_M$ (with respect to this ordering). In a neighborhood $U_{x,y}$ of some $y \in W \cap U_x$, as introduced before, the ${}'\mathcal{D}^{(p,M)}(U_{x,y}, \mathbb{C})$-current-valued map

$$\begin{aligned}
&(\lambda_1, \ldots, \lambda_M) \in \{\mathrm{Re}\,\lambda_j > 1\,;\ j = 1, \ldots, M\} \\
&\quad \longmapsto \frac{1}{(2i\pi)^M} \left(\bigwedge_{j=M}^{1} \overline{\partial}|s_{j,x}|^{2\lambda_j} \right) \wedge \left(\bigwedge_{j=1}^{M} \frac{ds_{j,x}}{s_{j,x}} \right) \wedge r_{x,y}[\omega]
\end{aligned}$$

can be continued as a holomorphic map to $\{\mathrm{Re}\,\lambda_j > -1;\ j = 1, \ldots, M\}$, with value at $\lambda_1 = \cdots = \lambda_M = 0$ the (p, M)-current

$$\varphi \in \mathcal{D}^{(n-p,n-M)}(U_{x,y}, \mathbb{C}) \mapsto \int_{W \cap U_x} \mathrm{Res}_{\Sigma_1,\ldots,\Sigma_M}[\omega] \wedge \varphi.$$

Note that in such a neighborhood $U_{x,y}$, we have, for $\mathrm{Re}\,\lambda_j \gg 1$, $j = 1, \ldots, M$,

$$\begin{aligned}
\left(\bigwedge_{j=M}^{1} \overline{\partial}\left(\frac{|s_{j,x}|^{2\lambda_j}}{s_{j,x}^{q_{j,x}}}\right)\right) \wedge \psi_x &= \left(\bigwedge_{j=M}^{1} \overline{\partial}|s_{j,x}|^{2\lambda_j} \right) \wedge \left(\bigwedge_{j=M}^{1} \frac{ds_{j,x}}{s_{j,x}} \right) \wedge r_{x,y}[\omega] \\
&\quad + \left(\bigwedge_{j=M}^{1} \overline{\partial}|s_{j,x}|^{2\lambda_j} \right) \wedge \sigma_{x,y}[\omega].
\end{aligned}$$

Thus, we obtain, for any $\varphi \in \mathcal{D}^{n-p,n-m}(U_{x,y}, \mathbb{C})$,

$$\left\langle R_{[S_1]_{\mathrm{red}},\ldots,[S_M]_{\mathrm{red}}}[\omega], \varphi \right\rangle = \int_{W \cap U_x} \mathrm{Res}_{\Sigma_1,\ldots,\Sigma_M}[\omega] \wedge \varphi.$$

This comes from the fact that the $'\mathcal{D}^{(p,M)}(U_{x,y},\mathbb{C})$-current-valued map

$$(\lambda_1,\ldots,\lambda_M)\longmapsto\left(\bigwedge_{j=M}^{1}\overline{\partial}|s_{j,x}|^{2\lambda_j}\right)\wedge\sigma_{x,y}[\omega]$$

is holomorphic in $\{\mathrm{Re}\,\lambda_j>-1\,;\ j=1,\ldots,M\}$ and takes the value 0 at $\lambda_1=\cdots=\lambda_M=0$. Finally, using the covering of V by the U_x, $x\in V$, we conclude that the (p,M)-current defined in $U=\mathcal{X}\setminus V_{\mathrm{sing}}$ as

$$\varphi\in\mathcal{D}^{(n-p,n-M)}(U,\mathbb{C})\mapsto\int_W \mathrm{Res}_{\Sigma_1,\ldots,\Sigma_M}[\omega]\wedge\varphi$$

can be continued as the $\overline{\partial}$-closed current $T=R_{[S_1]_{\mathrm{red}},\ldots,[S_M]_{\mathrm{red}}}[\omega]$ over the whole manifold $\mathcal{X}$. Note that the support of T satisfies $\mathrm{supp}\,T\subset V$. □

2 Regular Holonomy of Integration Currents

Let $\mathcal{X}$ be an n-dimensional complex manifold, and $V\subset\mathcal{X}$ be a closed, purely dimensional, reduced, analytic subset of codimension M. El Mir's extension theorem in [17] implies that the integration current $[V]$ is defined as the unique, positive, d-closed, (M,M)- current over $\mathcal{X}$ such that, for any test function $\varphi\in\mathcal{D}^{(n-M,n-M)}(\mathcal{X}\setminus V_{\mathrm{sing}},\mathbb{C})$,

$$\big\langle[V],\,\varphi\big\rangle=\int_V\varphi=\int_{V_{\mathrm{reg}}}\varphi.$$

It is important to point out here that the closed analytic set V is considered as being embedded in the ambient manifold $\mathcal{X}$. This will be revealed to us to be important for two reasons: first, with respect to connections between intersection and divisions problems in $\mathcal{X}$ (that one intends to study jointly), closed analytic subsets in $\mathcal{X}$ need to be understood (and studied) in terms of their defining equations. Second, the Coleff–Herrera sheafs of currents $\mathrm{CH}_{\mathcal{X},V}$ and $\mathrm{CH}_{\mathcal{X},V}(\cdot\,;\,\star S)$ that we will introduce in the two following sections are indeed sheaves of currents in $\mathcal{X}$, with support on V, which depend in a crucial way on the embedding $\iota:V\to\mathcal{X}$. Therefore, instead of working on the complex analytic space $(V,(\mathcal{O}_{\mathcal{X}})_{|V})$, using, for example, a log resolution $\widetilde{V}\xrightarrow{\pi}V$ for some closed hypersurface H_{sing} on V, satisfying $V_{\mathrm{sing}}\subset H_{\mathrm{sing}}$, we will work in the ambient manifold $\mathcal{X}$ and keep as far as possible to methods based on the use of Bernstein–Sato-type functional equations [19, 20, 26]. We will use extensively in this section the methods introduced to prove Proposition 1. These methods allow the possibility to define (in a robust way) the exterior multiplication of the integration current $[V]$ with a semi-meromorphic form ω, whose polar set intersects V along a closed analytic subset W satisfying $\dim W<\dim V$. Recall here that $'\mathcal{D}^{(p,q)}(\mathcal{X},\mathbb{C})$ denotes the space of (p,q)-currents on $\mathcal{X}$, acting on the space $\mathcal{D}^{(n-p,n-q)}(\mathcal{X},\mathbb{C})$ of smooth $(n-p,n-q)$-test forms on $\mathcal{X}$.

Proposition 2 (A holonomy property) *Let $\mathcal{X}$ and $V \subset \mathcal{X}$ be as above. Let $h, u \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$. The ${}'\mathcal{D}^{(M,M)}(\mathcal{X})$-valued map*

$$(\lambda, \mu) \in \{\operatorname{Re}\lambda \gg 1, \operatorname{Re}\mu \gg 1\} \longmapsto |u|^{2\mu} \frac{|h|^{2\lambda}}{h}[V]$$

can be continued analytically as a holomorphic map to the product of half-planes $\{\operatorname{Re}\lambda > -\eta, \operatorname{Re}\mu > -\eta\}$ for some $\eta > 0$. Moreover, if $\overline{V \setminus \{u = 0\}} = V$, then the value of this analytic continuation at $\lambda = \mu = 0$ remains unchanged if one replaces $|u|$ by 1. When $\overline{V \setminus \{h = 0\}} = V$, the construction of the principal value current

$$\frac{1}{h}[V] := \left[\frac{|h|^{2\lambda}}{h}[V]\right]_{\lambda=0}, \tag{5}$$

is "robust" in the following sense:

$$\frac{1}{h}[V] = \left[|u|^{2\mu} \frac{|h|^{2\lambda}}{h}[V]\right]_{\lambda=\mu=0} = \left[|u|^{2\mu} \frac{1}{h}[V]\right]_{\mu=0} \tag{6}$$

for any holomorphic function $u \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ such that $\overline{V \setminus \{u = 0\}} = V$.

Proof The second assertion in the statement of the proposition is a consequence of the first. If $\overline{V \setminus \{u = 0\}} = V$, i.e., $|u|$ does not vanish identically on any component of V (hence $[|u|^{2\mu}]_{\mu=0} \equiv 1$ almost everywhere on such component), one has

$$\left[|u|^{2\mu} \frac{|h|^{2\lambda}}{h}[V]\right]_{\mu=0} = \frac{|h|^{2\lambda}}{h}[V]$$

for $\operatorname{Re}\lambda \gg 1$. Assume the first assertion, namely that the current-valued function (5) is holomorphic in two variables in a product of half-spaces $\{\operatorname{Re}\lambda > -\eta, \operatorname{Re}\mu > -\eta\}$ for some $\eta > 0$. Then, following the analytic continuation in λ up to $\lambda = 0$, we get:

$$\left[\left[|u|^{2\mu} \frac{|h|^{2\lambda}}{h}[V]\right]_{\mu=0}\right]_{\lambda=0} = \left[|u|^{2\mu} \frac{|h|^{2\lambda}}{h}[V]\right]_{\lambda=\mu=0} = \left[\frac{|h|^{2\lambda}}{h}[V]\right]_{\lambda=0}.$$

This proves the second assertion (under the assumption that the first one holds).

In order now to prove the first assertion above, let us reduce the situation to the local one, that is, where $\mathcal{X}$ is a neighborhood Ω of the origin in $\mathbb{C}^n$. One can assume that V (defined in Ω as the common zero set of holomorphic functions $v_1, \ldots, v_k$ in $H(\Omega)$) is the union of a finite number of irreducible components of the complete intersection $\widetilde{V} = \{f_1 = \cdots = f_M = 0\}$, with $df_1 \wedge \cdots \wedge df_M \not\equiv 0$ on each such component ([18], p. 72). Let v be a linear combination of $v_1, \ldots, v_k$ which does not vanish identically on any of the irreducible components of the complete intersection $\widetilde{V}$ that are not irreducible components of V. We introduce from now on the notation $\widetilde{V}^{\mathcal{X}\setminus V}$ to denote the union of the irreducible components of $\widetilde{V}$ which are not entirely contained in V. Let u_{sing} be a holomorphic function in $\mathcal{X}$ such that $\widetilde{V}_{\rm sing} \subset \{u_{\rm sing} = 0\}$ and $u_{\rm sing} \not\equiv 0$ on any irreducible component of $\widetilde{V}$. Let us introduce the differential $(M,0)$-form

$$\omega = \frac{df_1 \wedge \cdots \wedge df_M}{f_1 \cdots f_M}$$

and the $\overline{\partial}$-closed (M,M)-current

$$\frac{T^f_{\underline{1},M}}{(2i\pi)^M} \wedge df_1 \wedge \cdots \wedge df_M = \mathrm{Res}_{[f_1=0]_{\rm red},\dots,[f_M=0]_{\rm red}}[\omega],$$

where the current $T^f_{\underline{1},M}$ is defined by the iterated process

$$T^f_{\underline{1},M} = \left[\overline{\partial}\left(\frac{|f_M|^{2\lambda_M}}{f_M}\right) \wedge \left[\cdots \wedge \left[\overline{\partial}\left(\frac{|f_1|^{2\lambda_1}}{f_1}\right)\right]_{\lambda_1=0} \cdots\right]_{\lambda_{M-1}=0}\right]_{\lambda_M=0}$$

considered in the proof of Proposition 1, where also the notation $\mathrm{Res}_{[\cdot]}[\omega]$ was introduced. Using the Bernstein–Sato equation (2) (here for $M+4$ functions), still in its conjugate form, one can prove that the current-valued function

$$(\lambda,\mu,\nu,\varpi) \mapsto |u_{\rm sing}|^{2\varpi}|v|^{2\nu}|u|^{2\mu}\frac{|h|^{2\lambda}}{h}\mathrm{Res}_{[f_1=0]_{\rm red},\dots,[f_M=0]_{\rm red}}[\omega]$$

can be continued from

$$\left\{(\lambda,\mu,\nu,\varpi);\, \mathrm{Re}\,\lambda \gg 1, \mathrm{Re}\,\mu \gg 1, \mathrm{Re}\,\nu \gg 1, \mathrm{Re}\,\varpi \gg 1\right\}$$

to a product of half-planes

$$\left\{(\lambda,\mu,\nu,\varpi);\, \mathrm{Re}\,\lambda > -\eta, \mathrm{Re}\,\mu > -\eta, \mathrm{Re}\,\nu > -\eta, \mathrm{Re}\,\varpi > -\eta\right\}$$

for some $\eta > 0$. Moreover, when $\mathrm{Re}\,\lambda \gg 1$, $\mathrm{Re}\,\mu \gg 1$, and $\mathrm{Re}\,\varpi \gg 1$, the value at $\nu = 0$ of

$$\nu \longmapsto |u_{\rm sing}|^{2\varpi}\left(1 - |v|^{2\nu}\right)|u|^{2\mu}\frac{|h|^{2\lambda}}{h}\mathrm{Res}_{[f_1=0]_{\rm red},\dots,[f_M=0]_{\rm red}}[\omega]$$

is equal to the current

$$|u_{\rm sing}|^{2\varpi}|u|^{2\mu}\frac{|h|^{2\lambda}}{h}[V].$$

Keeping $\mathrm{Re}\,\lambda \gg 1$ and $\mathrm{Re}\,\mu \gg 1$ and taking the analytic continuation in ϖ up to $\varpi = 0$, we get precisely the current

$$|u|^{2\mu}\frac{|h|^{2\lambda}}{h}[V]. \qquad \square$$

3 "Holomorphic" Coleff–Herrera Sheaves of Currents

Given an n-dimensional analytic manifold $\mathcal{X}$, together with a closed, purely dimensional reduced analytic subset V (of codimension M), the ("holomorphic") Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot,E)$ of E-valued $(0,M)$-currents, where $E \to \mathcal{X}$ denotes a holomorphic bundle of finite rank over $\mathcal{X}$, plays a major role in division or duality problems. The local description of its sections, together with the subsequent properties, suggests how one can profit from the $2n$ local parameters $\zeta_1,\ldots,\zeta_n,\overline{\zeta}_1,\ldots,\overline{\zeta}_n$ instead of just the n "holomorphic" ones $\zeta_1,\ldots,\zeta_n$. Thinking heuristically, the antiholomorphic local coordinates $\overline{\zeta}_1,\ldots,\overline{\zeta}_n$ remain unaffected by the holomorphic differentiations involved in the action of such currents. For example, if $\Delta_1,\ldots,\Delta_M$ are Cartier divisors on $\mathcal{X}$ and $s_1,\ldots,s_M$ denote the corresponding holomorphic sections of the Δ_j's such that the hypersurfaces $s_j^{-1}(0)$ intersect properly (that is, define a nonempty complete intersection on $\mathcal{X}$), then the usual Coleff–Herrera residue $\bigwedge_{j=1}^M \overline{\partial}(1/s_j)$ stands as a global section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot,E)$, where $E = \bigwedge_1^M \mathcal{O}_{\mathcal{X}}(-\Delta_j)$.

The concept and its importance were pointed out by Björk [11, 12]. The original construction of global sections for such sheaves is due to Coleff and Herrera [14]. In this section, we will recall the definition of the sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot,E)$ (following the approach of Björk, Andersson, and Samuelsson [2, 11–13]), together with the local structure of its sections (which justifies their operational properties). Since our objective all along this presentation is to stick to the methods based on analytic continuation (which seems to be a natural way to introduce the objects algebraically, for example, by using the Bernstein–Sato functional equations as (2)), the approach we adopt here follows that developed by Andersson [2].

Definition 2 (The Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot,E)$) Let $\mathcal{X}$, V, E be as above. The ("holomorphic") *Coleff–Herrera sheaf* $\mathrm{CH}_{\mathcal{X},V}(\cdot,E)$ is the sheaf of sections of $(0,M)$ E-valued currents T on $\mathcal{X}$, with support on V, which satisfy the three following conditions:

1. For any holomorphic function u in a neighborhood of V satisfying

$$\overline{V \setminus \{u=0\}} = V,$$

   the current-valued function

$$\lambda \in \{\mathrm{Re}\,\lambda \gg 1\} \longmapsto |u|^{2\lambda} T$$

   can be analytically continued as an holomorphic map to $\{\mathrm{Re}\,\lambda > -\eta\}$ for some $\eta > 0$, and

$$\big[|u|^{2\lambda} T\big]_{\lambda=0} = T$$

   (that is, T satisfies the Standard Extension Property (SEP) with respect to its support V).

2. One has, in the sense of currents,

$$(\mathcal{I}_V)_{\mathrm{conj}} T \equiv 0,$$

   where $(\mathcal{I}_V)_{\mathrm{conj}}$ denotes the complex conjugate of the ideal sheaf of sections of $\mathcal{O}_{\mathcal{X}}$ that vanish on V.
3. The current T is $\overline{\partial}$-closed.

Global sections of this sheaf, that is, elements in $\mathrm{CH}_{\mathcal{X},V}(\mathcal{X}, E)$, are called E-valued *Coleff–Herrera currents* (with respect to V) on $\mathcal{X}$.

Action on integration currents by adjoints of "simple" holomorphic differential operators with values in the dual bundle E^* provides us with an example of Coleff–Herrera sheaf of currents. To be more specific:

Example 1 Let D be a Cartier divisor in $\mathcal{X}$, and U be an open subset of $\mathcal{X}$. A holomorphic differential operator with analytic coefficients $Q_U : C^\infty_{n,n-M}(U, E^*) \to C^\infty_{n-M,n-M}(U, \mathcal{O}_{\mathcal{X}}(D))$ is said to be $(n, n-M)$-*simple* in U if its splits as

$$Q_U[\varphi] = q_U[\varphi] \wedge \omega_U,$$

where q_U denotes a holomorphic differential operator from $C^\infty_{n,n-M}(U, E^*)$ to $C^\infty_{0,n-M}(U, E^*)$, and ω_U is an element of $\Omega^{n-M}_{\mathcal{X}}(U, E \otimes \mathcal{O}_{\mathcal{X}}(D))$, that is, a global section over U of the sheaf of $E \otimes \mathcal{O}_{\mathcal{X}}(D)$-valued $(n-M)$-holomorphic forms. Let us denote as $\mathfrak{D}^{n,n-M}_{\mathcal{X}}(\cdot, E^*, D)$ the sheaf whose sections over $U \subset \mathcal{X}$ are $(n, n-M)$-simple holomorphic differential operators with analytic coefficients from $C^\infty_{n,n-M}(U, E^*)$ into $C^\infty_{n-M,n-M}(U, \mathcal{O}_{\mathcal{X}}(D))$. If $Q_U \in \mathfrak{D}^{n,n-M}_{\mathcal{X}}(U, E^*, D)$, then let Q^*_U be the adjoint operator which transforms elements from ${}'\mathcal{D}^{(M,M)}(U, \mathcal{O}_{\mathcal{X}}(-D))$ into elements in ${}'\mathcal{D}^{(0,M)}(U, E)$ as follows:

$$\langle Q^*_U[T], \varphi\rangle = \langle T, Q_U[\varphi]\rangle, \quad \forall \varphi \in \mathcal{D}^{(n,n-M)}(U).$$

If h_U denotes an holomorphic section of D in U such that $\overline{(V \cap U) \setminus \{h_U^{-1}(0)\}} = V \cap U$ and $Q_U \in \mathfrak{D}^{n,n-M}_{\mathcal{X}}(U, E^*)$, then the current

$$T_U = Q^*_U \left[\frac{[V \cap U]}{h_U}\right]$$

(where $[V \cap U]/h_U$ is defined as in (6), see Proposition 2) fulfills conditions 1 and 2 in Definition 2. This follows from the fact that the current-valued function

$$\mu \in \{\mathrm{Re}\, \mu \gg 1\} \longmapsto |u|^{2\mu} Q^* \left[\frac{1}{h}[V]_{\mathrm{red}}\right]$$

is analytically continued to $\mathrm{Re}\, \mu > -\eta$ and that its value at $\mu = 0$ does not depend on u as soon as $\overline{V \setminus \{u = 0\}} = V$. This shows that $Q^*[\frac{1}{h}[V]_{\mathrm{red}}]$ satisfies both the holonomy property and the standard extension property with respect to V, exactly as $\frac{1}{h}[V]_{\text{red}}$ does. If it is additionally $\overline{\partial}$-closed (which unfortunately cannot be read directly on the operator with meromorphic coefficients Q_U/h_U), then T_U fulfills also condition 3 in Definition 2 and therefore is a global section of the Coleff–Herrera sheaf $\text{CH}_{\mathcal{X},V}(\cdot, E)$ over U.

Let U be an open subset of $\mathcal{X}$. The local structure result established in [2, 11–13] can be stated as follows: when $T \in {}'\mathcal{D}^{(0,M)}(U, E)$ is a $\overline{\partial}$-closed current, $T \in \text{CH}_{\mathcal{X},V}(U, E)$ if and only if, for any $x \in U$, there exists a neighborhood $U_x \subset U$ of x in U, a section $Q_x \in \mathfrak{D}^{n,n-M}(U_x, E^*, \mathbb{C})$, and a holomorphic function h_x in U_x such that $\overline{V \cap U_x} \setminus \{h_x = 0\} = V \cap U_x$ and

$$T_{|U_x} = Q_x^*\left[\frac{[V \cap U_x]}{h_x}\right].$$

The local structure result, besides the fact that it provides a useful local representation of sections of the Coleff–Herrera sheaf $\text{CH}_{\mathcal{X},V}(\mathcal{X}, E)$, also emphasizes that only holomorphic differential operators are involved in the action of such currents (which explains indeed why they do play a role of algebraic nature despite their analytic structure).

It is important also to point out that, when $\mathcal{X} = \mathbb{P}^n(\mathbb{C})$, such a local structure result reflects (thanks to the GAGA principle) into a global structure result in this algebraic setting. The matrix of differential operators $Q_{\mathcal{X},IJ,K}$ involved in the definition of $Q_{\mathcal{X}}$, when expressed in local coordinates $(\zeta_1, \ldots, \zeta_n)$ in some affine chart, as

$$Q_{\mathcal{X}}\left[\Big(\sum_{\substack{J\subset\{1,\ldots,n\}\\ |J|=n-M}} \varphi_I\, d\overline{\zeta}_I\Big) \wedge d\zeta_1 \wedge \cdots \wedge d\zeta_n\right]$$

$$= \sum_{\substack{I,J\subset\{1,\ldots,n\}\\ |I|=|J|=n-M}} \Big(\sum_{\substack{K\subset\{1,\ldots,n\}\\ |K|=n-M}} Q_{\mathcal{X},IJ,K}\Big(\zeta, \frac{\partial}{\partial\zeta}\Big)[\varphi_K]\Big)\, d\overline{\zeta}_J \wedge d\zeta_I,$$

becomes a matrix of differential operators with polynomial coefficients, while the polar factor $h_{\mathcal{X}}$ corresponds here to a polynomial section of the bundle $\mathcal{O}_{\mathcal{X}}(k)$ for some $k \in \mathbb{N}$. Such differential operators with polynomial coefficients are of course reminiscent of the *Nœtherian operators* involved in the formulation of the *Ehrenpreis–Palamodov fundamental principle* [5, 10, 16, 24]. For example, when $P_1, \ldots, P_M$ are M homogeneous polynomials in $[z_0 : \cdots : z_n]$ defining a complete intersection V in $\mathbb{P}^n(\mathbb{C})$, a global section of the Coleff–Herrera sheaf $\text{CH}_{\mathcal{X},V}(\cdot, \bigwedge_1^M \mathcal{O}_{\mathcal{X}}(-\deg P_j))$ can be used to test the membership to the ideal $(P_1, \ldots, P_M)$. Note also that local structure results of this type originally go back to the work of Dolbeault [15].

4 "Meromorphic" Coleff–Herrera Sheaves of Currents

Intersection and division problems (in the case of proper intersection) are intimately connected through the Lelong–Poincaré equation: namely, if $\Delta_1, \dots, \Delta_M$ are M Cartier divisors on a complex manifold $\mathcal{X}$, together with respective metrics $|\ |_j$ and holomorphic sections s_j such that the $s_j^{-1}(0)$ intersect as a nonempty complete intersection $s^{-1}(0)$, then the integration current $[\mathrm{div}\,(s_1) \bullet \cdots \bullet \mathrm{div}\,(s_M)]$ (the operation between cycles being here the intersection product in the proper intersection context) factorizes as

$$\big[\mathrm{div}(s_1) \bullet \cdots \bullet \mathrm{div}(s_M)\big] = \left(\bigwedge_{j=1}^{M} \overline{\partial}(1/s_j)\right) \wedge \mathfrak{d}_1 s_1 \wedge \cdots \wedge \mathfrak{d}_M s_M,$$

where $\bigwedge_{j=1}^{M} \overline{\partial}(1/s_j) \in \mathrm{CH}_{\mathcal{X},s^{-1}(0)}(\mathcal{X}, \bigwedge_1^M \mathcal{O}_{\mathcal{X}}(-\Delta_j))$ is a Coleff–Herrera current independent of the choice of the metrics $|\ |_j$, and $\mathfrak{d}_j$ stands here for the Chern connection on $(\mathcal{O}_{\mathcal{X}}(\Delta_j), |\ |_j)$ (one could in fact replace $\mathfrak{d}_j$ by the de Rham operator d since the choice of the metrics is here irrelevant). Unfortunately, when V denotes an $(n-M)$-purely dimensional, reduced, closed analytic set in $\mathcal{X}$, the integration current $[V]$ cannot usually be factorized (locally about a point $x \in V$) as the product of a section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot, \mathbb{C})$ with a local section of the sheaf $\Omega_{\mathcal{X}}^{n-M}$ of $(n-M)$-abelian forms. A sufficient condition for this to be true is that $\mathcal{O}_{\mathcal{X},x}/\mathcal{I}_{V,x}$ is Cohen–Macaulay (see [3]). In general (see the proof of Proposition 2), in some convenient neighborhood U_x of x, there exists a factorization $[V \cap U_x] = T_{U_x} \wedge \omega_{U_x}$, where $\omega_{U_x} \in \Omega_{\mathcal{X}}^{n-M}(U_x)$, and T_{U_x} is a section in U_x of the meromorphic Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot\,; \star S_x, \mathbb{C})$ defined below (S_x being here a closed hypersurface in U_x such that $\overline{(V \cap U_x) \setminus S_x} = V \cap U_x$). This motivates the enlargement of the concept of Coleff–Herrera sheaf in order to tolerate holomorphic singularities (as we proceed when we enlarge the sheaf $\mathcal{O}_{\mathcal{X}}$ of holomorphic functions in $\mathcal{X}$ by introducing the sheaf $\mathcal{M}_{\mathcal{X}}$ of meromorphic functions on $\mathcal{X}$).

Let $\mathcal{X}$, V, E be as in the previous section. We now add in our list of data a closed hypersurface S in some neighborhood of V (in $\mathcal{X}$) such that $\overline{V \setminus S} = V$. The hypersurface S will play the role of a prescribed polar set for the sections of the sheaves we are about to define.

Definition 3 (The Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot\,; \star S, E)$) Let $\mathcal{X}$, V, E be as in Definition 2, and S be as above. The ("meromorphic") Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot\,; \star S, E)$ is the sheaf of sections of $(0, M)$ E-valued currents on $\mathcal{X}$ ($M = \mathrm{codim}_{\mathcal{X}} V$), with support on V, which satisfy, besides conditions 1 and 2 in Definition 2, the additional condition

$$\mathrm{Supp}\,(\overline{\partial} T) \subset V \cap S. \tag{7}$$

In order to exhibit sections of meromorphic Coleff–Herrera sheaves (see Example 2 below), the following lemma reveals to be essential. The method we use here to prove it illustrates both the power and the flexibility of the analytic continuation method. An alternative approach (based on the regularization of currents and the use of cut-off functions) was proposed in [13].

Lemma 1 *Let V be a purely $(n-M)$-dimensional closed analytic subset in an n-dimensional complex manifold $\mathcal{X}$. Let $u, h, s \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ satisfy*

$$\overline{V \setminus \{h=0\}} = \overline{V \setminus \{s=0\}} = \overline{V \setminus \{u=0\}} = V.$$

Let $Q \in \mathfrak{D}^{n,n-M}_{\mathcal{X}}(\mathcal{X}, \mathbb{C})$ (see Example 1). The $(0, M)$-current-valued map

$$(\mu, \nu) \in \{\mathrm{Re}\,\mu \gg 1, \mathrm{Re}\,\nu \gg 1\} \longmapsto |u|^{2\nu} \frac{|s|^{2\mu}}{s} Q^*\left[\frac{1}{h}[V]\right]$$

extends as an holomorphic map to $\{\mathrm{Re}\,\mu > -\eta, \mathrm{Re}\,\nu > -\eta\}$ for some $\eta > 0$, whose value $\mathcal{T}$ at $\mu = \nu = 0$ is independent of u. The "robust" definition of $\mathcal{T}$ makes it natural to denote it as

$$\mathcal{T} = \frac{1}{s} Q^*\left[\frac{1}{h}[V]\right].$$

The current $\mathcal{T}$ fulfills conditions 1 and 2 in Definition 2.

Proof Let $u \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ and μ, ν be such that $\mathrm{Re}\,\nu \gg 1$ and $\mathrm{Re}\,\mu \gg 1$. Then

$$\begin{aligned}\Big\langle |u|^{2\nu} \frac{|s|^{2\mu}}{s} Q^*\Big[\frac{1}{h}[V]\Big], \varphi\Big\rangle &= \Big\langle \frac{1}{h}[V], Q\Big(\zeta, \frac{\partial}{\partial\zeta}\Big)\Big[|u|^{2\nu}\frac{|s|^{2\mu}}{s}\varphi\Big]\Big\rangle \\ &= \Big\langle \frac{|u|^{2\nu}|s|^{2\mu}}{h}[V], Q\Big[\frac{\varphi}{s}\Big]\Big\rangle + \mu\Big\langle \frac{|u|^{2\nu}|s|^{2\mu}}{h}[V], Q_{u,s}\Big(\mu, \nu, \zeta, \frac{\partial}{\partial\zeta}\Big)[\varphi]\Big\rangle \\ &\quad + \nu\Big\langle \frac{|u|^{2\nu}|s|^{2\mu}}{h}[V], \widetilde{Q}_{u,s}\Big(\mu, \nu, \zeta, \frac{\partial}{\partial\zeta}\Big)[\varphi]\Big\rangle,\end{aligned}$$

where $Q_{u,s}(\mu, \nu, \zeta, \partial/\partial\zeta)$ and $\widetilde{Q}_{u,s}(\mu, \nu, \zeta, \partial/\partial\zeta)$ are the meromorphic differential operators (polynomial in μ, ν) from $C^\infty_{n,n-M}(\mathcal{X})$ into $C^\infty_{n-M,n-M}(\mathcal{X})$, with polar set contained in $\{us = 0\}$. We can rewrite (for some convenient $K \in \mathbb{N}$, namely the order of the differential operator Q)

$$\begin{aligned}&\Big\langle \frac{|u|^{2\nu}|s|^{2\mu}}{h}[V], Q_{u,s}\Big(\mu, \nu, \zeta, \frac{\partial}{\partial\zeta}\Big)[\varphi]\Big\rangle \\ &\quad = \Big\langle \frac{|u|^{2\nu}|s|^{2\mu}}{hu^K s^{K+1}}[V], A_{u,s}\Big(\mu, \nu, \zeta, \frac{\partial}{\partial\zeta}\Big)[\varphi]\Big\rangle \\ &\quad = \Big[\Big\langle \frac{|h|^{2\lambda}}{h}\frac{|u|^{2\nu}|s|^{2\mu}}{u^K s^{K+1}}[V], A_{u,s}\Big(\mu, \nu, \zeta, \frac{\partial}{\partial\zeta}\Big)[\varphi]\Big\rangle\Big]_{\lambda=0},\end{aligned}$$

where $A_{u,s}$ denotes a holomorphic differential operator (polynomial in μ, ν) from $C^\infty_{n,n-M}(\mathcal{X})$ into $C^\infty_{n-M,n-M}(\mathcal{X})$. The same reasoning holds when one replaces $Q_{u,s}$ by $\widetilde{Q}_{u,s}$ with some holomorphic differential operator $\widetilde{A}_{u,s}$ instead of $A_{u,s}$. Note also that

$$Q\left[\frac{\varphi}{s}\right] = \frac{1}{s^M} A\left(\zeta, \frac{\partial}{\partial \zeta}\right)[\varphi],$$

where A is a holomorphic differential operator from the space $C^\infty_{n-M,n-M}$ into $C^\infty_{n-M,n-M}$. The second assertion follows from the fact that, when $\operatorname{Re}\mu \gg 1$, the $(0, M)$-current

$$\frac{|sh|^{2\nu}}{s}\, Q^*\left[\frac{1}{h}[V]\right]$$

is annihilated locally by $(\mathcal{I}_V)_{\mathrm{conj}}$ (since Q is a holomorphic differential operator), which remains indeed true for the current

$$\frac{1}{s} Q^*\left[\frac{1}{h}[V]\right] = \left[\frac{|sh|^{2\nu}}{s} Q^*\left[\frac{1}{h}[V]\right]\right]_{\nu=0}.$$

This current fulfills conditions 1 and 2 in Definition 2. □

Example 2 Lemma 1 allows us to revisit Example 1, introducing possible poles. Let $\mathcal{X}$, V, E, S be as in Definition 3. Let additionally D, Δ be two Cartier divisors on $\mathcal{X}$. Let $U \subset \mathcal{X}$ and h_U, s_U be respectively holomorphic sections of D and Δ in U such that $s_U^{-1}(0) \subset S$ and $\overline{(V \cap U) \setminus h_U^{-1}(0)} = V \cap U$. Let $Q \in \mathfrak{D}^{n,n-M}_{\mathcal{X}}(U, E^*, D)$. Then the $\mathcal{O}_{\mathcal{X}}(-\Delta) \otimes E$-valued current in U

$$\mathcal{T} = \frac{1}{s_U} Q_U^*\left[\frac{[V \cap U]}{h_U}\right]$$

belongs to $\mathrm{CH}_{\mathcal{X},V}(U\ ; \star S, \mathcal{O}_{\mathcal{X}}(-\Delta) \otimes E)$ as soon as Q_U and h_U are such that the current $Q_U^*[[V \cap U]/h_U]$ is $\overline{\partial}$-closed.

Example 2 above provides, in fact, what appears locally to be the description of sections of Coleff–Herrera sheaves, since one has the following proposition (see [13]):

Proposition 3 *Let $\mathcal{X}$ be an n-dimensional complex manifold, V be an $(n - M)$-purely dimensional closed analytic subset, and S be a closed hypersurface in $\mathcal{X}$ such that $\overline{V \setminus S} = V$. Any element $\mathcal{T}$ in $\mathrm{CH}_{\mathcal{X},V}(\mathcal{X}; \star S, \mathbb{C})$ can be locally realized in an open neighborhood U_x of $x \in V$ as $\mathfrak{T} = T_x/s_x$, where T_x is a current in $\mathrm{CH}_{\mathcal{X},V}(U_x, \mathbb{C})$, $s_x \in \mathcal{O}_{\mathcal{X}}(U_x)$ satisfying $s_x^{-1}(0) \cap U_x = S \cap U_x$. This means also that one has*

$$\mathcal{T} = \frac{1}{s_x} Q_x^*\left[\frac{1}{h_x}[V]\right] \tag{8}$$

with $Q_x \in \mathfrak{D}_{\mathcal{X}}^{n,n-M}(U_x, \mathbb{C}, \mathbb{C})$, $h_x \in \mathcal{O}_{\mathcal{X}}(U_x)$, *satisfying* $\overline{(V \cap U_x) \setminus h_x^{-1}(0)} = V \cap U_x$, *the current* $Q_x^*[[V \cap U_x]/h_x]$ *being* $\overline{\partial}$*-closed in* U_x. *Conversely, any* $(0, M)$*-current* $\mathcal{T}$ *over* $\mathcal{X}$ *with support contained in* V *that can be locally expressed about each point* $x \in V$ *(in the ambient manifold* $\mathcal{X}$*) as* (8) *and is* $\overline{\partial}$*-closed outside* S, *belongs to* $\mathrm{CH}_{\mathcal{X},V}(\mathcal{X}; \star S, \mathbb{C})$.

Proof The second assertion follows from Lemma 1 since conditions 1, 2 in Definition 2 and (7) in Definition 3 can be checked locally. If $\mathcal{T} \in \mathrm{CH}_{\mathcal{X},V}(\mathcal{X}; \star S, \mathbb{C})$ and $x \in V$, $\{\sigma_x = 0\}$ being a reduced equation for S in an open neighborhood U_x of x in $\mathcal{X}$, one has $\overline{\partial}(s_x \mathcal{T}) \equiv 0$ in U_x if $s_x = \sigma_x^\gamma$ as soon as $\gamma \in \mathbb{N}$ exceeds strictly the order of $\mathcal{T}$ in $\overline{U_x}$. Therefore, $s_x \mathcal{T}_{|U_x} \in \mathrm{CH}_{\mathcal{X},V}(U_x, \mathbb{C})$ (conditions 1, 2 in Definition 2 remain fulfilled, condition (7) in Definition 3 is now realized). One can check immediately that

$$\frac{1}{s_x} \times (s_x \mathcal{T}_{|U_x}) = \mathcal{T}_{|U_x}$$

(the product on the left-hand side being understood as in Lemma 1), which proves that $\mathcal{T}$ can be represented as (8) in U_x. □

One can adapt the proof of Lemma 1 and Proposition 3 in order to get the following result.

Proposition 4 *Let* $\mathcal{X}$, V, E *be as in Definition* 3. *Let* $T \in \mathrm{CH}_{\mathcal{X},V}(\mathcal{X}, E)$, Δ *be a Cartier divisor on* $\mathcal{X}$, *equipped with a hermitian metric* $|\ |$, *and* s *be a holomorphic section of* Δ. *The* $(0, M)$*-current-valued map*

$$\mu \in \{\mathrm{Re}\,\mu \gg 1\} \longmapsto \frac{|s|^{2\mu}}{s} T$$

extends as a holomorphic map to $\{\mathrm{Re}\,\mu > -\eta\}$ *for some* $\eta > 0$. *Moreover, one has that*

$$\left[\frac{|s|^{2\mu}}{s} T\right]_{\mu=0} \in \mathrm{CH}_{\mathcal{X},V^{\mathcal{X}\setminus s^{-1}(0)}}\big(\mathcal{X}; \star s^{-1}(0), \mathcal{O}_{\mathcal{X}}(-\Delta) \otimes E\big),$$

the current being independent of the choice of the metric on Δ. *Recall that* $V^{\mathcal{X}\setminus s^{-1}(0)}$ *denotes the union of irreducible components of* V *that do not lie entirely in the closed hypersurface* $s^{-1}(0)$.

Proof Since it is sufficient to prove this proposition locally, one can assume that $T = Q^*[[V]/h]$, where $Q \in \mathfrak{D}_{\mathcal{X}}(\mathcal{X}, \mathbb{C})$, and $h \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ is not identically zero on any irreducible component V_ι of V which does not lie entirely in $s^{-1}(0)$. For $\mathrm{Re}\,\mu \gg 1$, one has, since $1/h\, 1_{\mathcal{X}\setminus s^{-1}(0)} \cdot [V] = 1/h[V^{\mathcal{X}\setminus s^{-1}(0)}]$, that

$$\frac{|s|^{2\mu}}{s} T = \frac{|s|^{2\mu}}{s} Q^*\left[\frac{1}{h}[V]\right] = \frac{|s|^{2\mu}}{s} Q^*\left[\frac{1}{h}\big[V^{\mathcal{X}\setminus s^{-1}(0)}\big]\right].$$

We now notice that s does not vanish identically on any irreducible component of $V^{\mathcal{X}\setminus s^{-1}(0)}$, which means $\overline{V^{\mathcal{X}\setminus s^{-1}(0)} \setminus s^{-1}(0)} = V^{\mathcal{X}\setminus s^{-1}(0)}$. Proposition 4 follows immediately from Lemma 1, combined with the second assertion in Proposition 3 (replacing V by $V^{\mathcal{X}\setminus s^{-1}(0)}$). □

Meromorphic E-valued Coleff–Herrera currents (with respect to V and prescribed polar set on S such that $\overline{V \setminus S} = V$) induce, via the $\overline{\partial}$ operator, elements in $\mathrm{CH}_{\mathcal{X},V\cap S}(\cdot, E)$. We present here an alternative proof (based on the analytic continuation) of a key result from [13].

Theorem 1 *The $\overline{\partial}$-operator maps* $\mathrm{CH}_{\mathcal{X},V}(\cdot\,;\star S, E)$ *into* $\mathrm{CH}_{\mathcal{X},V\cap S}(\cdot, E)$.

Remark 1 Note that the morphism above is surjective (at the level of germs at $x \in V$) as soon as $\mathcal{O}_{\mathcal{X},x}/\mathcal{I}_{V,x}$ is Cohen–Macaulay [12].

Proof Since one can reduce the problem to the local situation where E is trivialized, we may assume from now on that E is the trivial bundle $\mathcal{X} \times \mathbb{C}$. Let $\mathcal{T} \in \mathrm{CH}_{\mathcal{X},V}(\mathcal{X}; \star S, \mathbb{C})$. The statement in Theorem 1 amounts to check conditions 1, 2, 3 in Definition 2 locally for the current $\overline{\partial}\mathcal{T}$ (with respect to $V \cap S$). Then we can assume (see Proposition 3) that $\mathcal{X} = U$, where $U = U_x$ is an open neighborhood of a point x in V, $\mathcal{T} = 1/s\, Q^*[[V]/h]$, with $h, s \in \mathcal{O}_{\mathcal{X}}(U)$ satisfying

$$\overline{(V \cap U) \setminus \{h = 0\}} = \overline{(V \cap U) \setminus \{s = 0\}} = V \cap U,$$

and $Q \in \mathfrak{D}^{n,n-M}_{\mathcal{X}}(U, \mathbb{C}, \mathbb{C}) = \mathfrak{D}^{n,n-M}_{\mathcal{X}}(U)$. It is clear that $\overline{\partial}\mathcal{T}$ satisfies condition 3 since $\overline{\partial}^2 = 0$. Since $\mathcal{T} = [|s|^{2\mu}/s\, Q^*[[V]/h]]_{\mu=0}$ (see Lemma 1) and $Q^*[[V]/h]$ is closed (as an element in $\mathrm{CH}_{\mathcal{X},V}(U, \mathbb{C})$), we have

$$\overline{\partial}\mathcal{T} = \left[\overline{\partial}\frac{|s|^{2\mu}}{s} \wedge Q^*\left[\frac{1}{h}[V]\right]\right]_{\mu=0} = \left[\mu|s|^{2\mu}\frac{1}{s}\frac{d\overline{s}}{\overline{s}} \wedge Q^*\left[\frac{1}{h}[V]\right]\right]_{\mu=0}.$$

In order to prove that conditions 1 and 2 in Definition 2 hold for $\overline{\partial}\mathcal{T}$, it is enough to show that, when $u \in \mathcal{O}_{\mathcal{X}}(U)$ does not vanish identically on any irreducible component of $V \cap S \cap U$ (that is, $\{s = u = 0\} \cap V$ is defined as a complete intersection in $V \cap U$), the $(0, M+1)$-current-valued function

$$(\mu, \nu) \in \{\mathrm{Re}\,\mu \gg 1, \mathrm{Re}\,\nu \gg 1\} \longmapsto \mu|u|^{2\nu}|s|^{2\mu}\frac{1}{s}\frac{d\overline{s}}{\overline{s}} \wedge Q^*\left[\frac{1}{h}[V]\right] \tag{9}$$

extends as a holomorphic map to $\{\mathrm{Re}\,\mu > -\eta, \mathrm{Re}\,\nu > -\eta\}$ for some $\eta > 0$, whose value at $\mu = \nu = 0$ is independent of u and is annihilated (as a current) by $(\mathcal{I}_{V\cap S\cap U})_{\mathrm{conj}}$. Shrinking $U = U_x$ about x, if necessary, we can assume that there exists a holomorphic differential operator $\mathcal{Q} \in \mathfrak{D}^{n,n-M-1}_{\mathcal{X}}(U, \mathbb{C}, \mathbb{C})$ such that, for

$\operatorname{Re}\mu \gg 1$, $\operatorname{Re}\nu \gg 1$, and for any $\varphi \in \mathcal{D}^{(n,n-M-1)}(U,\mathbb{C})$, the following identity holds:

$$\mathcal{Q}\left[|u|^{2\nu}|s|^{2\mu}\frac{1}{s}\frac{d\overline{s}}{\overline{s}}\wedge\varphi\right] = \frac{d\overline{s}}{\overline{s}}\wedge\mathcal{Q}\left[|u|^{2\nu}|s|^{2\mu}\frac{1}{s}\varphi\right] \quad \text{on } V_{\rm reg}.$$

This comes from consideration of the facts that multiplication with antiholomorphic functions commutes with the action of holomorphic differential operators and that $\mathcal{Q}$ splits as $\mathcal{Q}[\varphi] = q[\varphi]\wedge\omega$, where q preserves the maximal degree of differential forms (on $V_{\rm reg}$) in $d\overline{\zeta}$, and $\omega \in \Omega^{n-M}_{\mathcal{X}}(U,\mathbb{C})$. Let K be the order of $\mathcal{Q}$. There exist holomorphic differential operators $\mathcal{A}_s$, $\mathcal{A}_{s,u}$, $\widetilde{\mathcal{A}}_{s,u}$ in $\mathfrak{D}^{n,n-M-1}_{\mathcal{X}}(U,\mathbb{C},\mathbb{C})$ (the two last ones depending also polynomially on μ and ν) such that, for any $\operatorname{Re}\mu \gg 1$, $\operatorname{Re}\nu \gg 1$, and for any $\varphi \in \mathcal{D}^{(n,n-M-1)}(U,\mathbb{C})$,

$$\begin{aligned}\mathcal{Q}\left[|u|^{2\nu}|s|^{2\mu}\frac{1}{s}\varphi\right] &= \frac{|u|^{2\nu}|s|^{2\mu}}{s^{K+1}}\mathcal{A}_s\left(\zeta,\frac{\partial}{\partial\zeta}\right)[\varphi]\\ &+ \frac{|u|^{2\nu}\,|s|^{2\mu}}{u^K s^{K+1}}\left(\mu\mathcal{A}_{s,u}\left(\mu,\nu,\zeta,\frac{\partial}{\partial\zeta}\right)+\nu\widetilde{\mathcal{A}}_{s,u}\left(\mu,\nu,\zeta,\frac{\partial}{\partial\zeta}\right)\right)[\varphi].\end{aligned}$$

Consider the $(0, M+1)$-valued maps

$$\begin{aligned}(\mu,\nu)\in\{\operatorname{Re}\mu\gg 1,\operatorname{Re}\nu\gg 1\} &\longmapsto \left[\mu\frac{|u|^{2\nu}\,|s|^{2\mu}\,|h|^{2\lambda}}{s^{K+1}h}\,\mathcal{B}_s\right]_{\lambda=0}\\ (\mu,\nu)\in\{\operatorname{Re}\mu\gg 1,\ \operatorname{Re}\nu\gg 1\} &\longmapsto \left[\mu\frac{|u|^{2\nu}\,|s|^{2\mu}\,|h|^{2\lambda}}{u^K s^{K+1}h}\,\mathcal{B}_{s,u}(\mu,\nu)\right]_{\lambda=0},\end{aligned}\tag{10}$$

where

$$\begin{aligned}\langle\mathcal{B}_s,\varphi\rangle &= \left\langle[V]_{\rm red},\frac{d\overline{s}}{\overline{s}}\wedge\mathcal{A}_s[\varphi]\right\rangle,\\ \big\langle\mathcal{B}_{s,u}(\mu,\nu),\varphi\big\rangle &\\ &= \left\langle[V]_{\rm red},\frac{d\overline{s}}{\overline{s}}\wedge\left(\mu\mathcal{A}_{s,u}\left(\mu,\nu,\zeta,\frac{\partial}{\partial\zeta}\right)+\nu\widetilde{\mathcal{A}}_{s,u}\left(\mu,\nu,\zeta,\frac{\partial}{\partial\zeta}\right)\right)[\varphi]\right\rangle\end{aligned}$$

for all $\varphi \in \mathcal{D}^{(n,n-M-1)}(U,\mathbb{C})$. We claim that both current-valued maps (10) extend as holomorphic maps to $\{\operatorname{Re}\mu > -\eta,\ \operatorname{Re}\nu > -\eta\}$ for some $\eta > 0$. Moreover, the value at $\mu = \nu = 0$ of the first of these maps is annihilated (as a current) by $(\mathcal{I}_{V\cap S\cap U})_{\rm conj}$, while the value at $\mu = \nu = 0$ of the second one equals 0.

Let us assume this claim for the moment and conclude the proof of the theorem. For $\operatorname{Re}\lambda \gg 1$, $\operatorname{Re}\mu \gg 1$, and $\operatorname{Re}\nu \gg 1$, we have

$$\mu|u|^{2\nu}|s|^{2\mu}\frac{1}{s}\frac{d\overline{s}}{\overline{s}}\wedge Q^*\left[\frac{|h|^{2\lambda}}{h}[V]\right]$$
$$=\mu\frac{|u|^{2\nu}|s|^{2\mu}|h|^{2\lambda}}{s^{K+1}h}\mathcal{B}_s+\mu\frac{|u|^{2\nu}|s|^{2\mu}|h|^{2\lambda}}{u^K s^{K+1}h}\mathcal{B}_{s,u}(\mu,\nu). \tag{11}$$

Thus, the current-valued map (9), which can be rewritten because of (11) (for $\operatorname{Re}\lambda\gg 1, \operatorname{Re}\mu\gg 1, \operatorname{Re}\nu\gg 1$) as

$$\left[\mu\frac{|u|^{2\nu}|s|^{2\mu}|h|^{2\lambda}}{s^{K+1}h}\mathcal{B}_s+\mu\frac{|u|^{2\nu}|s|^{2\mu}|h|^{2\lambda}}{u^K s^{K+1}h}\mathcal{B}_{s,u}(\mu,\nu)\right]_{\lambda=0},$$

extends as a holomorphic function of (μ,ν) to $\{\operatorname{Re}\mu>-\eta,\operatorname{Re}\nu>-\eta\}$ for some $\eta>0$, the value at $\mu=\nu=0$ being equal to

$$\left[\left[\mu\frac{|s|^{2\mu}\,|h|^{2\lambda}}{s^{M+1}h}\mathcal{B}_s\right]_{\lambda=0}\right]_{\mu=0},$$

which is independent of u and annihilated (as a current) by $(\mathcal{I}_{V\cap S\cap U})_{\rm conj}$. This proves that $\overline{\partial}\mathcal{T}$ fulfills conditions 1 and 2 in Definition 2.

Proving the claim clearly amounts to prove that for any positive integers σ,τ, the $(M,M+1)$-current-valued map

$$\{\operatorname{Re}\mu\gg 1,\operatorname{Re}\mu\gg 1\}\longmapsto\left[\mu\frac{|u|^{2\nu}|s|^{2\mu}|h|^{2\lambda}}{u^\tau s^\sigma h}\frac{d\overline{s}}{\overline{s}}\wedge[V]\right]_{\lambda=0}$$

extends as a holomorphic map to $\{\operatorname{Re}\lambda>-\eta,\operatorname{Re}\mu>-\eta\}$ for some $\eta>0$, whose value at $\mu=\nu=0$ is annihilated by $(\mathcal{I}_{V\cap S\cap U})_{\rm conj}$. In order to do that, we need to introduce a smooth log resolution $\mathcal{V}\overset{\pi}{\to}V$ for the closed hypersurface $W=V\cap\{\zeta\,;\,h(\zeta)s(\zeta)u(\zeta)=0\}\subset V$. That is, $\mathcal{V}$ is an $(n-M)$-dimensional complex manifold, and π is a proper surjective holomorphic map such that the closed analytic subset $\mathcal{W}$ (obtained as the union of $\pi^{-1}(W)$ with the set of points in $\mathcal{V}$ about which π is not a local isomorphism) is a closed hypersurface in $\mathcal{V}$ with normal crossings. Such a log resolution can be obtained applying the Hironaka theorem. Let $\iota_V:V\to\mathcal{X}$ be the inclusion embedding. For any $\varphi\in\mathcal{D}^{(n-M,n-M-1)}(U,\mathbb{C})$, we can rewrite, using the properness of π and a (sufficiently refined) partition of unity (V_ι,ρ_ι) subordinated to the support of $(\iota_V\circ\pi)^*[\varphi]$,

$$\left\langle\mu\frac{|u|^{2\nu}|s|^{2\mu}|h|^{2\lambda}}{u^\tau s^\sigma h}\frac{d\overline{s}}{\overline{s}}\wedge[V]_{\rm red},\varphi\right\rangle$$

as a sum of contributions of the form

$$\mu\int_{V_\iota}\frac{|u_\iota\xi_\iota^{\gamma_\iota}|^{2\nu}|s_\iota\xi_\iota^{\beta_\iota}|^{2\mu}|h_\iota\xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota\xi_\iota^{\tau\gamma_\iota+\sigma\beta_\iota+\alpha_\iota}}\left(\frac{d\overline{s}_\iota}{\overline{s}_\iota}+\sum_{j=1}^{n-M}\beta_{\iota,j}\frac{d\overline{\xi}_{\iota,j}}{\overline{\xi}_{\iota,j}}\right)\wedge\rho_\iota(\xi_\iota)(\iota_V\circ\pi)^*[\varphi](\xi_\iota), \tag{12}$$

where $\xi_\iota = (\xi_{\iota,1}, \ldots, \xi_{\iota,n-M})$ denote centered local coordinates in V_ι, $u_\iota, s_\iota, h_\iota$ are invertible functions in V_ι, and $\xi_\iota^{\gamma_\iota}, \xi_\iota^{\beta_\iota}, \xi_\iota^{\alpha_\iota}$ are monomial functions in the centered coordinates $(\xi_{\iota,1}, \ldots, \xi_{\iota,n-M})$ with respective multiexponents $\gamma_\iota, \beta_\iota, \alpha_\iota \in \mathbb{N}^{n-M}$. The function

$$(\lambda, \mu, \nu) \longmapsto \mu \int_{V_\iota} \frac{|u_\iota \xi_\iota^{\gamma_\iota}|^{2\nu} |s_\iota \xi_\iota^{\beta_\iota}|^{2\mu} |h_\iota \xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota \xi_\iota^{\tau\gamma_\iota + \sigma\beta_\iota + \alpha_\iota}} \frac{d\overline{s}_\iota}{\overline{s}_\iota} \wedge \rho_\iota(\xi_\iota)\, (\iota_V \circ \pi)^*[\varphi](\xi_\iota) \quad (13)$$

clearly extends as a holomorphic function of (λ, μ, ν) to a product of half-planes $\{\mathrm{Re}\,\lambda > -\eta, \mathrm{Re}\,\mu > -\eta, \mathrm{Re}\,\nu > -\eta\}$ for some $\eta > 0$, whose value at $\lambda = \mu = \nu = 0$ equals 0. The reason is that the singularities under the integral in (13) are only *holomorphic* singularities. The same remains true if $\varphi = \overline{h}\psi$, where $h \in \mathcal{I}_{V\cap S\cap U}$, since in this case any $\xi_{\iota,j}$, $j = 1, \ldots, n-m$, such that $\beta_{\iota,j} \neq 0$ divides $\pi^* h$, which implies that all *antiholomorphic* singularities in the term under the integral in (12) are thus canceled. It remains to study the meromorphic analytic continuation (as a function of (λ, μ, ν)) of

$$(\lambda, \mu, \nu) \longmapsto \mu \int_{V_\iota} \frac{|u_\iota \xi_\iota^{\gamma_\iota}|^{2\nu} |s_\iota \xi_\iota^{\beta_\iota}|^{2\mu} |h_\iota \xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota \xi_\iota^{\tau\gamma_\iota + \sigma\beta_\iota + \alpha_\iota}} \frac{d\overline{\xi}_{\iota,j}}{\overline{\xi}_{\iota,j}} \wedge \rho_\iota (\iota_V \circ \pi)^*[\varphi] \quad (14)$$

for $j \in \{1, \ldots, n-M\}$ such that $\beta_{\iota,j} > 0$. Using integration by parts, we can rewrite (14) (when $\mathrm{Re}\,\lambda \gg 1$, $\mathrm{Re}\,\mu \gg 1$, $\mathrm{Re}\,\nu \gg 1$) as

$$\frac{\mu}{\alpha_{\iota,j}\lambda + \beta_{\iota,j}\mu + \gamma_{\iota,j}\nu} \times \int_{V_\iota} \frac{|u_\iota \xi_\iota^{\gamma_\iota}|^{2\nu} |s_\iota \xi_\iota^{\beta_\iota}|^{2\mu} |h_\iota \xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota \xi_\iota^{\tau\gamma_\iota + \sigma\beta_\iota + \alpha_\iota}}\, d\overline{\xi}_{\iota,j} \wedge \frac{\partial}{\partial \overline{\xi}_{\iota,j}} \big(\rho_\iota\, (\iota_V \circ \pi)^*[\varphi]\big).$$

We need here to distinguish two more cases.

- If $\gamma_{\iota,j} = 0$, the function

$$\begin{aligned}(\mu, \nu) \quad &\longmapsto \quad \left[\mu \int_{V_\iota} \frac{|u_\iota \xi_\iota^{\gamma_\iota}|^{2\nu} |s_\iota \xi_\iota^{\beta_\iota}|^{2\mu} |h_\iota \xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota \xi_\iota^{\tau\gamma_\iota + \sigma\beta_\iota + \alpha_\iota}} \frac{d\overline{\xi}_{\iota,j}}{\overline{\xi}_{\iota,j}} \wedge \rho_\iota\, (\iota_V \circ \pi)^*[\varphi]\right]_{\lambda=0} \\
&= \Bigg[\frac{\mu}{\alpha_{\iota,j}\lambda + \beta_{\iota,j}\mu} \\
&\quad \times \int_{V_\iota} \frac{|u_\iota \xi_\iota^{\gamma_\iota}|^{2\nu} |s_\iota \xi_\iota^{\beta_\iota}|^{2\mu} |h_\iota \xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota \xi_\iota^{\tau\gamma_\iota + \sigma\beta_\iota + \alpha_\iota}}\, d\overline{\xi}_{\iota,j} \wedge \frac{\partial}{\partial \overline{\xi}_{\iota,j}} \big(\rho_\iota\, (\iota_V \circ \pi)^*[\varphi]\big)\Bigg]_{\lambda=0} \\
&= \frac{1}{\beta_{\iota,j}} \left[\int_{V_\iota} \frac{|u_\iota \xi_\iota^{\gamma_\iota}|^{2\nu} |s_\iota \xi_\iota^{\beta_\iota}|^{2\mu} |h_\iota \xi_\iota^{\alpha_\iota}|^{2\lambda}}{u_\iota^\tau s_\iota^\sigma h_\iota \xi_\iota^{\tau\gamma_\iota + \sigma\beta_\iota + \alpha_\iota}}\, d\overline{\xi}_{\iota,j} \wedge \frac{\partial}{\partial \overline{\xi}_{\iota,j}} \big(\rho_\iota\, (\iota_V \circ \pi)^*[\varphi]\big)\right]_{\lambda=0}\end{aligned}$$

  extends as a holomorphic function to $\{\mathrm{Re}\,\mu > -\eta, \mathrm{Re}\,\nu > -\eta\}$ for some $\eta > 0$.

- If $\gamma_{\iota,j} > 0$, one uses a primitive form of the *Whitney division lemma*, a clever trick introduced by Samuelsson [27]. The hyperplane of coordinates $\{\xi_{\iota,j} = 0\} \cap V_\iota$ lies in the closed analytic set $\{(\iota_V \circ \pi)^*[s] = (\pi \circ \iota_V)^*[u] = 0\} \cap V_\iota$, whose image by π is included in the $(n - M - 2)$-dimensional closed analytic subset of U defined as $\{u = 0\} \cap S \cap V \cap U$. Since any differential form $d\overline{\zeta}_I$, $|I| = n - M - 1$, has a vanishing pullback to $S \cap \{u = 0\} \cap V \cap U$ for dimension reasons, the $(0, n - M - 1)$-differential form $(\iota_V \circ \pi)^*[d\overline{\zeta}_I]$ has a vanishing pullback to $\{\xi_{\iota,j} = 0\} \cap V_\iota$, which means that $(\iota_V \circ \pi)^*[d\overline{\zeta}_I](\xi_\iota) = \overline{\xi}_{\iota,j}\, \overline{\omega}_I(\xi_\iota)$ for some $(0, n - M - 1)$-smooth form $\overline{\omega}_I$ in V_ι. Then $\overline{\xi}_{\iota,j}$ divides $(\iota_V \circ \pi)^*[\varphi]$ in V_ι, which implies that antiholomorphic singularities under the integral in (14) are canceled. Therefore, (14) extends as a holomorphic function of (λ, μ, ν) to a product of half-planes $\{\mathrm{Re}\,\lambda > -\eta, \mathrm{Re}\,\mu > -\eta, \mathrm{Re}\,\nu > -\eta\}$ for some $\eta > 0$.

This completes the proof of the claim and thus of the theorem. □

Proposition 4, together with Proposition 1, implies the following: if $\mathcal{X}, V, E, \Delta, s$ are given as in Proposition 4, then, for all open subsets $U \subset \mathcal{X}$,

$$\overline{\partial}\left(\left[\frac{|s|^{2\mu}}{s} T\right]_{\mu=0}\right) \in \mathrm{CH}_{\mathcal{X}, V^{\mathcal{X}\setminus s^{-1}(0)} \cap s^{-1}(0)}\big(U, \mathcal{O}_{\mathcal{X}}(-\Delta) \otimes E\big) \tag{15}$$

whenever $T \in \mathrm{CH}_{\mathcal{X},V}(U, E)$. Note that $V^{\mathcal{X}\setminus s^{-1}(0)} \cap s^{-1}(0)$ is either purely $(n - M - 1)$-dimensional or empty, in which last case (15) is somehow irrelevant since the current on the left-hand side is 0. We will need in the next section the following result, which is by far more involved, that we formulate here without proof (see [9] for a detailed proof).

Proposition 5 *Let $\mathcal{X}, V, E, \Delta, s$ be as in Proposition 4. Let S be a hypersurface in $\mathcal{X}$ such that $\overline{V \setminus S} = V$ and $\mathcal{T} \in \mathrm{CH}_{\mathcal{X},V}(\mathcal{X}\,; \star S, E)$. The $(0, M+1)$-current-valued map*

$$\nu \in \{\mathrm{Re}\,\nu \gg 1\} \longmapsto \overline{\partial}\left(\frac{|s|^{2\nu}}{s}\right) \wedge \mathcal{T}$$

extends as a holomorphic function to $\{\mathrm{Re}\,\nu > -\eta\}$ for some $\eta > 0$. Moreover, we have

$$\left[\overline{\partial}\left(\frac{|s|^{2\nu}}{s}\right) \wedge \mathcal{T}\right]_{\nu=0} \in \mathrm{CH}_{\mathcal{X}, V^{\mathcal{X}\setminus s^{-1}(0)} \cap s^{-1}(0)}\big(\mathcal{X}; \star\Sigma_{S,s}, \mathcal{O}_{\mathcal{X}}(-\Delta) \otimes E\big),$$

where $\Sigma_{S,s}$ denotes any closed hypersurface in a neighborhood of V in $\mathcal{X}$ such that

$$\overline{\big(V^{\mathcal{X}\setminus s^{-1}(0)} \cap s^{-1}(0)\big) \setminus \Sigma_{S,s}} = V^{\mathcal{X}\setminus s^{-1}(0)} \cap s^{-1}(0)$$

and $\Sigma_{S,s} \supset S^{V\setminus s^{-1}(0)} \cap s^{-1}(0)$, $S^{V\setminus s^{-1}(0)}$ being the union of all components of S whose intersection with V does not lie entirely in $V^{\mathcal{X}\setminus s^{-1}(0)} \cap s^{-1}(0)$.

5 Essential Intersection and Coleff–Herrera Original Construction

Let $\mathcal{X}, V, E, S$ be as in Proposition 5. Let also Δ_1 be a Cartier divisor on $\mathcal{X}$, equipped with a hermitian metric $|\ |$, and s_1 be a holomorphic section of Δ_1. Propositions 4 and 5 imply that any global section $\mathcal{T} \in \mathrm{CH}_{\mathcal{X},V}(\mathcal{X}\ ; \star S, E)$ splits into the sum of an element from $\mathrm{CH}_{\mathcal{X},V^{s_1^{-1}(0)}}(\mathcal{X}; \star S, E)$ and an element in $\mathrm{CH}_{\mathcal{X},V^{\mathcal{X}\setminus s_1^{-1}(0)}}(\mathcal{X}; \star S, E)$. That is,

$$\mathcal{T} = \big[\big(1 - |s_1|^{2\lambda_1}\big)\mathcal{T}\big]_{\lambda_1=0} + \big[|s_1|^{2\lambda_1}\mathcal{T}\big]_{\lambda_1=0} = \mathcal{T}_{|s_1^{-1}(0)} + \mathcal{T}_{\mathcal{X}\setminus s_1^{-1}(0)} \tag{16}$$

(see also [4]). We remark that this splitting is independent of the choice of the metric on Δ_1. To be more specific, suppose that

$$\mathcal{T} = \frac{1}{s} Q^* \left[\frac{[V]}{h}\right],$$

where s is a holomorphic section of a Cartier divisor Δ, h is a holomorphic section of a Cartier divisor D, and $Q \in \mathfrak{D}_{\mathcal{X}}^{n,n-M}(\mathcal{X}, E^* \otimes \mathcal{O}_{\mathcal{X}}(\Delta), D)$ (see Examples 1 and 2). Then, for any test function in $C^\infty_{n,n-M}(\mathcal{X}, E^*)$, we have

$$\begin{aligned}
&\big\langle\big[\big(1 - |s_1|^{2\lambda_1}\big)\mathcal{T}\big]_{\lambda_1=0}, \varphi\big\rangle \\
&\quad = \left[\left[\left[\int_V \frac{|h|^{2\mu}}{h} Q^* \left[\frac{|s|^{2\nu}}{s}\big(1 - |s_1|^{2\lambda_1}\big)\varphi\right]\right]_{\lambda=0}\right]_{\mu=0}\right]_{\lambda_1=0} \\
&\quad = \left[\left[\int_{V^{s_1^{-1}(0)}} \frac{|h|^{2\mu}}{h} Q^* \left[\frac{|s|^{2\nu}}{s}\varphi\right]\right]_{\lambda=0}\right]_{\mu=0}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\left[\bar\partial\left(\frac{|s_1|^{2\lambda_1}}{s_1}\right) \wedge \mathcal{T}\right]_{\lambda_1=0} = \left[\bar\partial\left(\frac{|s_1|^{2\lambda_1}}{s_1}\right) \wedge \mathcal{T}_{|\mathcal{X}\setminus s_1^{-1}(0)}\right]_{\lambda_1=0} \\
&\quad \in \mathrm{CH}_{\mathcal{X},V_1}\big(\mathcal{X}\ ; \star \mathrm{Pol}_1, \mathcal{O}_{\mathcal{X}}(-\Delta_1) \otimes E\big),
\end{aligned} \tag{17}$$

where V_1 stands for the closed analytic set $V^{\mathcal{X}\setminus s_1^{-1}(0)} \cap s_1^{-1}(0)$, and Pol_1 denotes a closed hypersurface in $\mathcal{X}$ satisfying $\overline{V_1 \setminus \mathrm{Pol}_1} = V_1$. Note that (16) can be understood as an analogue (at the level of meromorphic Coleff–Herrera currents) of the *gap sheaf operation* in intersection theory (see, e.g., [23]), namely, the splitting of the cycle $[V]$ corresponding to V as the sum $[V]^{s_1^{-1}(0)}$ of its components whose supports lie completely in the hypersurface $s_1^{-1}(0)$, and the sum $[V]^{\mathcal{X}\setminus s_1^{-1}(0)}$ of the other ones. On the other hand, the wedge product operation (17) can be understood as an analogue (at the level of Coleff–Herrera currents) of the *proper intersection* *product* between two cycles whose corresponding supports $V^{\mathcal{X}\setminus s_1^{-1}(0)}$ and $s_1^{-1}(0)$ intersect properly.

Given an ordered collection $\Delta_1,\dots,\Delta_m$ of Cartier divisors, with $m \le n-M$, together with respective holomorphic sections $s_1,\dots,s_m$, the operation (17) can be iterated because of the iterative process, initiated with $\mathcal{T}_0 = \mathcal{T}$:

$$\mathcal{T}_{j+1} = \left[\overline{\partial}\left(\frac{|s_j|^{2\lambda_j}}{s_j}\right) \wedge \mathcal{T}_j\right]_{\lambda_j=0} = \left[\overline{\partial}\left(\frac{|s_j|^{2\lambda_j}}{s_j}\right) \wedge \mathcal{T}_{j\,|\mathcal{X}\setminus s_j^{-1}(0)}\right]_{\lambda_j=0},\quad 0 \le j < m.$$

When this procedure is carried up to the end, we get

$$\mathcal{T}_m \in \mathrm{CH}_{\mathcal{X},(V\cap s_1^{-1}(0)\cap\cdots\cap s_m^{-1}(0))_{\mathrm{ess}}}\left(\mathcal{X}\,;\star\mathrm{Pol}_m, \bigwedge_1^m \mathcal{O}_{\mathcal{X}}(-\Delta_j)\otimes E\right),$$

where $V\cap s_1^{-1}(0)\cap\cdots\cap s_m^{-1}(0) = V_{\mathrm{ess}}[s]$ stands for the *essential intersection* (see, e.g., [14]) of V respect to the ordered sequence of hypersurfaces $s_1^{-1}(0),\dots,s_m^{-1}(0)$. If $\mathcal{T}_0 = T_0 \in \mathrm{CH}_{\mathcal{X},V}(\mathcal{X},E)$, then the current $\mathcal{T}_m$ is a global section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V_{\mathrm{ess}}[s]}(\cdot,\bigwedge_1^m \mathcal{O}_{\mathcal{X}}(-\Delta_j)\otimes E)$.

One can consider as well (as in [21]) the $\bigwedge_1^m \mathcal{O}(-\Delta_j)$-valued current $\mathcal{R}^{s_1,\dots,s_m}\wedge[V]$ (which is $\overline{\partial}$-closed) obtained, starting from $\mathcal{R}^{\{\}}\wedge[V]=[V]$, through the inductive procedure

$$\mathcal{R}^{s_1,\dots,s_{j+1}}\wedge[V] = \left[\overline{\partial}\left(\frac{|s_j|^{2\lambda_j}}{s_j}\right)\wedge\mathcal{R}^{s_1,\dots,s_j}\wedge[V]\right]_{\lambda_j=0},\quad 0\le j<m. \tag{18}$$

This point of view was introduced in a slightly different form in [14]. The authors consider there a $(p,0)$-semi-meromorphic form ω on a complex space $(V,\mathcal{O}_V)$, with poles along the union of a finite number of reduced hypersurfaces $S_1,\dots,S_m$ of V (taken in a prescribed order). They construct on $(V,\mathcal{O}_V)$ an (m,p)-residue current $R_{S_1,\dots,S_m}[\omega]$ with support the essential intersection $(S_1\cap\cdots\cap S_m)_{\mathrm{ess}}$. Note that the residual objects defined in [14] are intrinsic with respect to the complex space $(V,\mathcal{O}_V)$, that is, independent of the embedding $\iota: V\to\mathcal{X}$. The construction proposed here and that in [14] are of course related: besides the fact that our currents are treated here as $(M+k,M+p)$-currents, $0\le k\le m$, in the ambient manifold $\mathcal{X}$ instead of (m,p)-currents on the complex analytic space V, the main difference between the two approaches is that the singularities $1/s_j$ in (18) are isolated from local expressions for the denominator of ω.

Consider the particular case where there exist holomorphic bundles $E_1,\dots,E_L$ on $\mathcal{X}$ such that the integration current $[V]$ can be expressed as

$$[V] = \sum_{l=1}^{L} T_{l,0}\wedge\omega_l,$$

where $T_{l,0}\in\mathrm{CH}_{\mathcal{X},V}(\mathcal{X},E_l)$ and $\omega_l\in\Omega_{\mathcal{X}}^{n-M}(\mathcal{X},E_l^*)$ (which occurs, for example, when one restricts $\mathcal{X}$ to some relatively compact open subset. When $\mathcal{X}$ is Stein and

$\mathcal{O}_{\mathcal{X},x}/\mathcal{I}_{V,x}$ is Cohen–Macaulay about each point $x \in V$ (see [3], Example 1), then one can factorize $\mathcal{R}^{s_1,\dots,s_m} \wedge [V]$ as

$$\mathcal{R}^{s_1,\dots,s_m} \wedge [V] = \sum_{l=1}^{L} T_{l,m} \wedge \omega_l, \tag{19}$$

where each $T_{l,m}$ is some $(\bigwedge_1^m \mathcal{O}_{\mathcal{X}}(-\Delta_j)) \otimes E_l$-valued Coleff–Herrera current (with respect to $V_{\mathrm{ess}}[s]$) which is a pole-free Colef–Herrera current. Factorization (19) remains valid in general, but one needs to tolerate then poles in the Coleff–Herrera sections $T_{l,m}$.

In conclusion, we claim that the results presented here (within the robust frame of analytic continuation), together with the geometric formalism of intersection theory (where the role of integration currents on cycles is played by global sections of Coleff–Herrera sheaves), should be a starting point to pursue an approach, introduced in [8], in order to attack division or duality problems with methods inspired by those used in intersection theory.

References

1. Aizenberg, L.A., Yuzhakov, A.P.: Integral Representation and Residues in Multidimensional Complex Analysis. Am. Math. Soc., Providence (1983)
2. Andersson, M.: Uniqueness and factorization of Coleff–Herrera currents. Ann. Fac. Sci. Toulouse Math. Sér. **18**(4), 651–661 (2009)
3. Andersson, M.: Coleff–Herrera currents, duality, and Noetherian operators. Preprint Gothenburg (2009). arXiv:0902.3064
4. Andersson, M., Wulcan, E.: Decomposition of residue currents. J. Reine Angew. Math. **638**, 103–118 (2010)
5. Andersson, M., Wulcan, E.: Residue currents with prescribed annihilator ideals. Ann. Sci. Ec. Norm. Super. **40**(6), 985–1007 (2007)
6. Bernstein, I.N.: The analytic continuation of generalized functions with respect to a parameter. Funct. Anal. Appl. **6**, 273–285 (1972)
7. Aleksandrov, A., Tsikh, A.: Théorie des résidus de Leray et formes de Barlet sur une intersection compléte singuliére. C. R. Acad. Sci. Paris, Sér. I **333**, 1–6 (2001)
8. Berenstein, C.A., Gay, R., Vidras, A., Yger, A.: Residue Currents and Bézout Identities. Progress in Mathematics, vol. 114. Birkhäuser, Basel (1993)
9. Berenstein, C.A., Vidras, A., Yger, A.: Multidimensional residue theory and applications, manuscript in preparation
10. Björk, J.E.: Rings of Differential Operators. North-Holland, Amsterdam (1979)
11. Björk, J.E.: Residue calculus and $\mathcal{D}$-modules on complex manifolds. Preprint, Stockholm University (1996)
12. Björk, J.E.: Residues and $\mathcal{D}$-modules. In: Laudal, O.A., Piene, R. (eds.) The Legacy of Niels Henrik Abel, pp. 605–651. Springer, Berlin (2004)
13. Björk, J.E., Samuelsson, H.: Regularizations of residue currents. J. Reine Angew. Math. **649**, 33–54 (2010)
14. Coleff, N., Herrera, M.: Les courants résiduels associés à une forme méromorphe. Lecture Notes in Math., vol. 633. Springer, Berlin (1978)
15. Dolbeault, P.: On the structure of residual currents. Several complex variables (Stockholm, 1987/1988). Math. Notes **38**, 258–273

16. Ehrenpreis, L.: Fourier Analysis in Several Complex Variables. Wiley-Interscience, New York (1970)
17. El Mir, H.: Sur le prolongement des courants positifs fermés. Acta Math. **153**, 1–45 (1984)
18. Grauert, H., Remmert, R.: Coherent Analytic Sheaves. Grunlehren der math. Wissenschaften, vol. 265. Springer, Berlin (1984)
19. Gyoja, A.: Bernstein–Sato's polynomial for several analytic functions. J. Math. Kyoto Univ. **33**(2), 399–411 (1993)
20. Kashiwara, M.: B-functions and holonomic systems. Invent. Math. **38**(1), 33–53 (1976/77)
21. Lärkäng, R., Samuelsson, H.: Various approaches to products of residue currents. Preprint Gothenburg (2009). arXiv:1005.2056
22. Leray, J.: Le calcul différentiel et intégral sur une variété analytique complexe, Problème de Cauchy III. Bull. Soc. Math. Fr. **87**, 81–180 (1959)
23. Massey, D.B.: Numerical control over complex analytic singularities. Mem. Am. Math. Soc. **163**, 778 (2003)
24. Palamodov, V.P.: Linear Operators with Constant Coefficients. Springer, New York (1970)
25. Poly, J.B.: Sur un théorème de J Leray en théorie des résidus. C. R. Acad. Sci. Paris Ser. A–B **274**, A171–A174 (1972)
26. Sabbah, C.: Proximité évanescente II, Equations fonctionelles pour plusieurs fonctions analytiques. Compos. Math. **64**(2), 213–241 (1987)
27. Samuelsson, H.: Analytic continuation of residue currents. Ark. Mat. **47**(1), 127–141 (2009)

Right Inverses for $P(D)$ in Spaces of Real Analytic Functions

Dietmar Vogt

Abstract In this paper we study the question when a linear partial differential operator $P(D)$ with constant coefficients admits a continuous linear right inverse in the space $A(\mathbb{R}^n)$ of real analytic functions on $\mathbb{R}^n$ (or, more generally, in $A(\Omega)$ where Ω is a open subset of $\mathbb{R}^n$). To obtain a necessary condition, we investigate when $P(D)$ admits solvability "with real analytic parameter" in $A(\Omega)$ and solve it completely for convex Ω, using a different approach from the one used in Domański (Funct. Approx. 44:79–109, 2011). To obtain a sufficient condition, we show that the global real analytic Cauchy problem is solvable if and only if the principal part of $P(D)$ is hyperbolic. In this way we get a complete solution of our main problem for $A(\mathbb{R}^2)$ and, in the homogeneous case, for $A(\Omega)$ where Ω is the open unit ball in $\mathbb{R}^n$.

1 Introduction

Let $\Omega \subset \mathbb{R}^n$ be open, and $P \in \mathbb{C}[z_1, \ldots, z_n]$. We study the linear partial differential operator with constant coefficients $P(D_1, \ldots, D_n)$, with $D_j = -i\frac{\partial}{\partial x_j}$, acting on the space $A(\Omega)$ of real analytic functions on Ω. We want to know when $P(D)$ admits a continuous linear right inverse in $A(\Omega)$.

We recall that $P(D)$ needs not to be surjective in $A(\Omega)$ even for $\Omega = \mathbb{R}^n$. It had been conjectured by De Giorgi and Cattabriga [5] and shown by Piccinini [18, 19] that not every linear differential operator $P(D)$ with constant coefficients is surjective in $A(\mathbb{R}^n)$. Their examples were operators whose principal parts had a mute variable. We study such operators in Sect. 1, and we characterize completely when they are surjective in $A(\Omega \times \mathbb{R})$ for convex Ω (the last variable being assumed to be mute), so extending results of [22]. The same characterization has been obtained in Domański [7]. However methods and proofs there are completely different, so that it appears useful to present our approach. Surjectivity of $P(D)$ in $A(\Omega)$ for convex Ω

Dedicated to the memory of Leon Ehrenpreis.

D. Vogt (✉)
FB Math.-Nat., Bergische Universität Wuppertal, Gauss-Str. 20, 42097 Wuppertal, Germany
e-mail: dvogt@math.uni-wuppertal.de

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_22, © Springer-Verlag Italia 2012

has been characterized by Hörmander [8] in terms of a Phragmén–Lindelöf condition for plurisubharmonic functions on the zero variety of the principal part $P_m(D)$ of $P(D)$. Meise, Taylor, and Vogt [13] characterized, for convex Ω, the polynomials P such that $P(D)$ admits a continuous linear right inverse in $C^\infty(\Omega)$, also in terms of a Phragmén–Lindelöf condition on the zero variety of P. While surjectivity in $A(\Omega)$ depends only on the principal part, the existence of right inverses in $C^\infty(\Omega)$ does not, and the perturbation conditions for lower order parts are unknown up to now. The condition for surjectivity with a mute variable, or surjectivity with parameter dependence, connects both. It depends only on the principal part, which has to admit a continuous linear right inverse in $C^\infty(\Omega)$.

Surjectivity with parameter dependence or surjectivity in $\Omega \times \mathbb{R}$ is, of course, a necessary condition for the existence of a continuous linear right inverse of $P(D)$ in $A(\Omega)$, which yields a necessary condition depending only on the principal part. We do not know whether the existence of a right inverse depends in fact only on the principal part (for the ultradifferentiable case cf. [4]). In some cases, however, the necessary condition turns out to be also sufficient. This is the case for dimension 2 and $\Omega = \mathbb{R}^2$. Here the necessary condition means that the principal part has to be hyperbolic.

We then show, for arbitrary dimension, that hyperbolicity of the principal part is equivalent to unique solvability of the global Cauchy problem in the real analytic functions, for real analytic data. This result is interesting in its own and yields a complete characterization of the $P(D)$ for which there exists a continuous linear solution operator in $A(\mathbb{R}^2)$. Another case where the necessary condition is also sufficient and we get such a characterization is the case of homogeneous operators and Ω a bounded set with C^1-boundary, for instance, the unit ball.

2 Preliminaries

Throughout the paper we denote by $A(\Omega)$ the linear space of real analytic functions on the open set $\Omega \subset \mathbb{R}^n$ equipped with its natural locally convex topology, which is as well of (PDF)-type and ultrabornological (see [12]). This implies by Grothendieck's (or de Wilde's) open mapping theorem that any continuous linear surjective map from $A(\Omega)$ to $A(\Omega)$ is open.

We will use the following condition HPL(Ω, loc) introduced in Hörmander [8]. $\mathscr{K}(\Omega)$ denotes the convex, compact subsets of Ω, and PSH(W) the plurisubharmonic functions on a complex variety W. For any compact convex set $K \subset \mathbb{R}^p$, we denote by $h_K(x) := \sup\{\langle x, \xi\rangle : \xi \in K\}$, $x \in \mathbb{R}^p$, the support function of K.

Let V be the germ of a complex variety at $\xi \in \mathbb{R}^n$. V satisfies HPL(Ω, loc) if there are open sets $U_1 \Subset U_2 \Subset U_3 \Subset \mathbb{C}^n$ with $\xi \in U_1$ such that for each $K \in \mathscr{K}(\Omega)$, there exists $K' \in \mathscr{K}(\Omega)$ and $\delta > 0$ such that each $u \in \text{PSH}(U_3 \cap V)$ satisfying (α) and (β) also satisfies (γ), where

(α) $u(z) \leq h_K(\operatorname{Im} z) + \delta$, $z \in U_3 \cap V$,
(β) $u(z) \leq 0$, $z \in U_2 \cap \mathbb{R}^n \cap V$,
(γ) $u(z) \leq h_{K'}(\operatorname{Im} z)$, $z \in U_1 \cap V$.

For detailed information on this and other related Phragmén–Lindelöf conditions, we refer to [8] and [16]. For unexplained notation and results on partial differential equations, we refer to [9].

3 Solvability with Real Analytic Parameter

We will use the following notation: For $P \in \mathbb{C}[z_1, \dots, z_n]$, we set $P^+ = P$, considered as a polynomial in $\mathbb{C}[z_1, \dots, z_{n+1}]$, and for open $\Omega \subset \mathbb{R}^n$, we consider $P^+(D)$ as acting in $A(\Omega^+)$ where $\Omega^+ = \Omega \times \mathbb{R}$.

We say that $P(D)$ is *solvable in* $A(\Omega)$ *with real analytic parameter* if $P^+(D) : A(\Omega^+) \longrightarrow A(\Omega^+)$ is surjective. Solvability in $A(\Omega)$ with real analytic parameter has been investigated in a different context in [22], and a characterization for $\Omega = \mathbb{R}^n$ was given there. A complete characterization has been given in Domański [7]. Many results of this section, in particular, Theorem 1 can be found also there. However our approach and the methods of proof are entirely different.

Various kinds of parameter dependence in different spaces have also been studied recently in Bonet–Domański [1, 2]. Real analytic parameter-dependence in $\mathscr{D}'(\Omega)$ has been studied and characterized in Domański [6].

The following lemma improves the necessary condition in [22, Proposition 3.1].

Lemma 1 *If* $\Omega \subset \mathbb{R}^n$ *is convex and* $P^+(D)$ *surjective in* $A(\Omega^+)$, *then* $P_m(D)$ *has a right inverse in* $C^\infty(\Omega)$.

Proof Let $V = \{z \in \mathbb{R}^n : P_m(z) = 0\}$ be the zero variety of P_m, and $V^+ = V \times \mathbb{R}$ the same for P_m^+. By [16, Theorem 3.3], it suffices to show that V satisfies HPL(Ω, loc) at zero, and, by [8, Lemma 4.1], we have at our disposal HPL(Ω^+, loc) at any point $\xi^0 \in V^+ \cap \mathbb{R}^{n+1}$ with $|\xi^0| = 1$. We apply it to $\xi^0 = (0, \dots, 0, 1) \in \mathbb{R}^{n+1}$. We find $0 < r_1 < r_2 < r_3$ and for every $K \in \mathscr{K}(\Omega)$, a $K' \in \mathscr{K}(\Omega^+)$ and $\delta > 0$ such that each $u \in \mathrm{PSH}(U_3^+ \cap V^+)$ satisfying (a) and (b) also satisfies (c), where

(a) $u(z) \le h_{K\times\{0\}}(\operatorname{Im} z) + \delta,\ z \in U_3^+ \cap V^+$,
(b) $u(x) \le 0,\ x \in U_2^+ \cap \mathbb{R}^{n+1} \cap V^+$,
(c) $u(z) \le h_{K'}(\operatorname{Im} z),\ \ z \in U_1^+ \cap V^+$.

We have set $U_j^+ = \{z \in \mathbb{C}^{n+1} : |z - \xi^0| < r_j\}$ and put $U_j = \{z \in \mathbb{C}^n : |\zeta| < r_j\}$. We remark that $h_{K\times\{0\}}(x) = h_K(x_1, \dots, x_n)$ for $x \in \mathbb{R}^{n+1}$.

Let now u be a plurisubharmonic function on $U_3 \cap V$, and u^+ be the same function acting on $U_3^+ \cap V^+$. Notice that for $(z_1, \dots, z_{n+1}) \in U_3^+$, we have $(z_1, \dots, z_n) \in U_3$. We assume that

(α) $u(z) \le h_K(\operatorname{Im} z) + \delta,\ z \in U_3 \cap V$,
(β) $u(x) \le 0,\ x \in U_2 \cap \mathbb{R}^n \cap V$.

Then u^+ satisfies (a) and (b) and, hence, (c). For $z \in U_1$, we set $\tilde{z} = (z_1, \dots, z_n, 1)$. Then $\tilde{z} \in U_1^+$, and we have

(γ) $u(z) = u^+(\tilde{z}) \leq h_{K'}(\operatorname{Im} \tilde{z}) = h_{K''}(\operatorname{Im} z)$,

where $K'' = \pi K'$ and $\pi : (x_1, \ldots, x_{n+1}) \mapsto (x_1, \ldots, x_n)$. Clearly $K'' \Subset \Omega$. □

For convex Ω, we obtain a complete characterization of differential polynomials $P(D)$ which admit solvability with a real analytic parameter. The same characterization has been given by a different method in Domański [7, Theorem 6.1].

Theorem 1 *For convex Ω, the following are equivalent*:

1. *$P^+(D)$ is surjective in $A(\Omega^+)$.*
2. *$P_m(D) : C^\infty(\Omega) \longrightarrow C^\infty(\Omega)$ admits a continuous linear right inverse.*

Proof One implication is Lemma 1, the other is [22, Proposition 3.2]. □

If we take into account [16, Corollary 3.14], we get as a special case for $\Omega = \mathbb{R}^n$ [22, Theorem 3.4].

Theorem 2 *For $n > 1$, the following are equivalent*:

1. *$P^+(D)$ is surjective in $A(\mathbb{R}^{n+1})$.*
2. *$P_m(D)$ is surjective in $A(\mathbb{R}^n)$, and P_m has no elliptic factor.*

As an immediate consequence of Theorem 1, we obtain the following:

Corollary 1 *If $P_m(z) = P_m(z_1, \ldots, z_p)$ with $1 \leq p < n$, then the following are equivalent*:

1. *$P(D)$ is surjective in $A(\mathbb{R}^n)$.*
2. *$P_m(D_1, \ldots, D_p) : C^\infty(\mathbb{R}^p) \longrightarrow C^\infty(\mathbb{R}^p)$ admits a continuous linear right inverse.*

Since by a theorem of Grothendieck elliptic $P_m(D_1, \ldots, D_p) : C^\infty(\mathbb{R}^p) \longrightarrow C^\infty(\mathbb{R}^p)$ for $p \geq 2$ never admits a continuous linear right inverse (see Trèves [20, Theorem C.1]), this explains the examples of Di Giorgi, Cattabriga, and Piccinini. A somewhat more general formulation is the following:

Theorem 3 *If there is $N \neq 0$ in $\mathbb{R}^n$ such that $P_m(z + \lambda N)$ does not depend on λ for all $z \in \mathbb{R}^n$, then the following are equivalent*:

1. *$P(D)$ is surjective in $A(\mathbb{R}^n)$.*
2. *$P_m(D) : C^\infty(\mathbb{R}^n) \longrightarrow C^\infty(\mathbb{R}^n)$ admits a continuous linear right inverse.*

Proof By a linear transformation, we may assume that $N = e_n$, and the result follows from Corollary 1. Assertions 2 in both results are then seen to be equivalent, because $P_m(D) = P_m(D_1, \ldots, D_{n-1}) \otimes \mathrm{id}_{C^\infty(\mathbb{R})}$ acting on $C^\infty(\mathbb{R}^n) = C^\infty(\mathbb{R}^{n-1}) \widehat{\otimes} C^\infty(\mathbb{R})$. □

By use of a theorem of Langenbruch [11], for general open $\Omega \subset \mathbb{R}^n$ (conditions for this case see [10]), we obtain the following:

Corollary 2 *If P satisfies the assumptions of Theorem 3, $\Omega \subset \mathbb{R}^n$ is open, and $P(D)$ is surjective in $A(\Omega)$, then $P_m(D) : C^\infty(\mathbb{R}^n) \longrightarrow C^\infty(\mathbb{R}^n)$ admits a continuous linear right inverse.*

Proof By [11] $P(D)$ is surjective in $A(\mathbb{R}^n)$, and we can apply Theorem 3. □

We will need it in the following more special version:

Lemma 2 *If $n > 1$, $\Omega \subset \mathbb{R}^n$ is open, and $P(D)$ is surjective in $A(\Omega^+)$, then $P_m(D)$ has no elliptic factor.*

Proof By [11] $P^+(D)$ is surjective in $A(\mathbb{R}^{n+1})$, and we can apply Theorem 2. □

4 Right Inverses in $A(\Omega)$: Necessary Condition

We begin with a simple observation:

Lemma 3 *If $P(D)$ has a right inverse in $A(\Omega)$, then $P^+(D)$ is surjective in $A(\Omega \times \mathbb{R})$.*

Proof If we identify $A(\Omega \times \mathbb{R}) \cong A(\Omega)\widehat{\otimes}A(\mathbb{R})$, then $P^+(D)$ corresponds to $P(D) \otimes \mathrm{id}_{A(\mathbb{R})}$, which has $R \otimes \mathrm{id}_{A(\mathbb{R})}$ as a right inverse, where R is a continuous linear right inverse for $P(D)$. In particular, $P^+(D)$ is surjective. □

Lemma 3 and Lemma 2 together imply the following:

Proposition 1 *Let $n > 1$ and $\Omega \subset \mathbb{R}^n$ open. If $P(D) : A(\Omega) \longrightarrow A(\Omega)$ admits a continuous linear right inverse, then P_m has no elliptic factor.*

For convex Ω, we can use Theorem 1 to sharpen the necessary criterion.

Proposition 2 *Let $n > 1$ and $\Omega \subset \mathbb{R}^n$ open and convex. If $P(D) : A(\Omega) \longrightarrow A(\Omega)$ admits a continuous linear right inverse, then so does $P_m(D) : C^\infty(\Omega) \longrightarrow C^\infty(\Omega)$.*

Homogeneous polynomials admitting a continuous linear right inverse in $C^\infty(\Omega)$ are carefully studied in [13]. In the following case we get a sharp criterion and even a complete characterization:

Proposition 3 *Let $n > 1$, and let $\Omega \subset \mathbb{R}^n$ be open, convex, and bounded with C^1-boundary. If $P(D) : A(\Omega) \longrightarrow A(\Omega)$ admits a continuous linear right inverse, then P_m is, up to a constant factor, a product of real linear forms.*

Proof This follows from Proposition 2 and [13], Theorem 3.8. □

Example 1 If Ω is the unit ball in $\mathbb{R}^d$ and P is homogeneous, then the following are equivalent:

1. $P(D)$ admits a continuous linear right inverse in $A(\Omega)$.
2. $P^+(D)$ is surjective in $A(\Omega^+)$.
3. $P(D)$ admits a continuous linear right inverse in $C^\infty(\Omega)$.
4. P is, up to a constant factor, a product of real linear forms.

The only thing to prove is $4 \Rightarrow 1$. But this is done just by integration. We must construct a right inverse only for P being a real linear form L, which we may assume to be $L(x) = x_1$. Then $f \mapsto \int_0^x f(\xi, x_2, \ldots, x_d)\, d\xi$ is a right inverse.

5 Operators with Hyperbolic Principal Part

The existence of a right inverse of $P(D)$ in $A(\Omega)$ depends in all cases treated up to now only on the principal part $P_m(D)$. Therefore it may be of interest to mention that also "hyperbolicity" in the sense of global existence and uniqueness for the Cauchy problem follows from and is even equivalent to the hyperbolicity of the principal part, and in this case the right inverse can be given in a very explicit way.

We will consider $P(D)$ as acting not only in $A(\Omega)$ but also in $C^\infty(\Omega)$ and in the Gevrey classes $\gamma^{(s)}(\Omega)$ for $s > 1$ defined as follows:

$$\gamma^{(s)}(\Omega) = \left\{ f \in C^\infty(\Omega) : \forall K \Subset \Omega, \varepsilon > 0\ \exists C\ \forall \alpha, x \in K : \left| f^{(\alpha)}(x) \right| \le C \varepsilon^{|\alpha|} \left(|\alpha|! \right)^s \right\}.$$

Here $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ and $|\alpha| = \sum \alpha_j$. $\gamma^{(s)}(\Omega)$ is a Fréchet space equipped with the seminorms

$$\| f \|_n = \sup_{\substack{x \in K_n \\ \alpha}} \left| f^{(\alpha)}(x) \right| \frac{n^{|\alpha|}}{(|\alpha|!)^s},$$

where K_n runs through a compact increasing exhaustion of Ω. For convenience, we set $\gamma^{(+\infty)}(\Omega) = C^\infty(\Omega)$.

Let us remark that $\gamma^{(s)}(\Omega) = \mathscr{E}_{(\omega)}(\Omega)$ with $\omega(t) = t^{1/s}$ in the sense of Braun–Meise–Taylor [3].

We set $\gamma_0^{(s)}(\mathbb{R}^n) = \gamma^{(s)}(\mathbb{R}^n) \cap \mathscr{D}(\mathbb{R}^n)$ equipped with its natural (LF)-topology.

P is called $\gamma^{(m/m-1)}$-hyperbolic with respect to N if there are fundamental solutions $E_\pm \in \gamma_0^{(m/m-1)'}(\mathbb{R}^n)$ of $P(D)$ with support in the cones

$$H_\pm = \left\{ x \mid \langle x, \pm N \rangle > 0 \right\} \cup \{0\}.$$

From [15, Proposition 2.12] and [9, Theorem 12.7.5] we obtain the following:

Proposition 4 *If P_m is hyperbolic with respect to N, then*

1. *P is $\gamma^{(m/m-1)}$-hyperbolic with respect to N.*
2. *The Cauchy problem in Proposition 5 is uniquely solvable in $\gamma^{(m/m-1)}(\mathbb{R}^n)$ for all data $g \in \gamma^{(m/m-1)}(\mathbb{R}^n)$ and $f_0, \ldots, f_{m-1} \in \gamma^{(m/m-1)}(\mathbb{R}^{n-1})$.*

We use it to show the unique solvability of the Cauchy problem in $A(\mathbb{R}^n)$.

Proposition 5 *Let $N = e_1$ and set $x = (x_1, x')$. If P_m is hyperbolic with respect to N, then the Cauchy problem*

$$P(D)f = g, \qquad \frac{\partial^k}{\partial x_1^k} f\bigl(0, x'\bigr) = f_k\bigl(x'\bigr), \quad k = 0, \ldots, m-1,$$

is uniquely solvable for all $g \in A(\mathbb{R}^n)$ and $f_0, \ldots, f_{m-1} \in A(\mathbb{R}^{n-1})$ with $f \in A(\mathbb{R}^n)$.

Proof Since the assumption implies, by [8, Theorem 6.5], that $P(D) : A(\mathbb{R}^n) \to A(\mathbb{R}^n)$ is surjective, it is easily seen that we may assume that $g = 0$. Since $f_0, \ldots, f_{m-1} \in \gamma^{(m/m-1)}(\mathbb{R}^{n-1})$, there is, by assumption and Proposition 4, a unique solution $f \in \gamma^{(m/m-1)}(\mathbb{R}^n)$ of the Cauchy problem.

Due to the Cauchy–Kowalewski theorem, this solution is real analytic in a neighborhood U of $\{0\} \times \mathbb{R}^{n-1}$.

Let $x \in \mathbb{R}^n$, $x_1 > 0$. We choose $\phi \in \gamma^{(m/m-1)}(\mathbb{R}^n)$ such that $\phi \equiv 1$ in a neighborhood of 0 and $\operatorname{supp}(\phi^2 - \phi) \subset x - V$, where $V = U \cap \{\xi \mid |\xi_1| < x_1\}$. Using the fundamental solution $E_+ \in \gamma_0^{(m/m-1)'}(\mathbb{R}^n)$, which exists by Proposition 4, we set

$$T = \phi E_+ \quad \text{and} \quad P(D)T = \delta - S.$$

Then $\operatorname{supp} S \subset x - V$ and $\operatorname{supp} T \subset H_+ \cap \operatorname{supp} \phi$. We obtain

$$0 = \bigl(P(D)f\bigr) * T = f * \bigl(P(D)T\bigr) = f - S * f,$$

i.e., $f(\xi) = S_y(f(\xi - y))$ for all ξ.

For $y \in \operatorname{supp} S$, we have $x - y \in V$, and the same holds for all ξ in a neighborhood of x. Therefore f is real analytic in a neighborhood of x. An analogous argument applies for $x_1 < 0$. □

Theorem 4 *If P_m is hyperbolic, then $P(D)$ admits a continuous linear right inverse in $A(\mathbb{R}^n)$.*

Proof Let P_m be hyperbolic with respect to N. We may assume that $N = e_1$. We set $R(g) := f$ where f is the unique solution of the Cauchy problem in Proposition 5 with $f_0 = \cdots = f_{m-1} = 0$. R is clearly a linear right inverse for $P(D)$; it is continuous because the inverse of the "Cauchy map" χ is continuous (see the proof of Proposition 6). □

Proposition 6 *If the Cauchy problem*

$$P(D)f = 0, \qquad \frac{\partial^k}{\partial x_1^k} f\big(0, x'\big) = f_k\big(x'\big), \quad k = 0, \dots, m-1,$$

is uniquely solvable for all $f_0, \dots, f_{m-1} \in A(\mathbb{R}^{n-1})$ *with* $f \in A(\mathbb{R}^n)$, *then* $P_m(D)$ *is hyperbolic with respect to* e_1.

Proof For $x \in \mathbb{R}^n$, we set again $x = (x_1, x')$. Let χ be the "Cauchy map" $A(\mathbb{R}^n) \longrightarrow A(\mathbb{R}^{n-1})^m$, i.e.,

$$\chi(\varphi) = \big(\varphi\big(0, x'\big), \varphi'\big(0, x'\big), \dots, \varphi^{(m-1)}\big(0, x'\big)\big),$$

where all derivatives are taken with respect to the first variable. By assumption, χ is surjective, hence bijective, and therefore, due to the de Wilde–Grothendieck theorem, a topological isomorphism.

We consider the functions $\varphi_\zeta(x) := e^{ix\zeta}$, $\zeta = \xi + i\eta \in \mathbb{C}^n$, $P(\zeta) = 0$. Then

$$\begin{aligned}\chi(\varphi_\zeta) &= \big(\varphi_\zeta\big(0, x'\big), i\zeta_1\varphi_\zeta\big(0, x'\big), \dots, (i\zeta_1)^{m-1}\varphi_\zeta\big(0, x'\big)\big)\\ &= \varphi_\zeta\big(0, x'\big)\big(1, i\zeta_1, \dots, (i\zeta_1)^{m-1}\big).\end{aligned}$$

Since χ^{-1} is continuous, we have the following estimates for φ_ζ, $P(\zeta) = 0$:

$$\forall r\ \exists R\ \forall \varepsilon > 0\ \exists \delta > 0, C : \|\varphi_\zeta\|_{r,\delta} \le C \big\|\chi(\varphi_\zeta)\big\|_{R,\varepsilon},$$

where

$$\begin{aligned}\big\|\varphi_\zeta(\cdot)\big\|_{r,\delta} &= \sup_{\substack{|x| \le r \\ |y| \le \delta}} e^{-y\xi - x\eta} = e^{r|\eta| + \delta|\xi|},\\ \big\|\varphi_\zeta(0, \cdot)\big\|_{R,\varepsilon} &= \sup_{\substack{|x'| \le r \\ |y'| \le \varepsilon}} e^{-y'\xi' - x'\eta'} = e^{R|\eta'| + \varepsilon|\xi'|}.\end{aligned}$$

Therefore, taking in $A(\mathbb{R}^{n-1})^m$ the maximum of the "norms," we have, for $P(\zeta) = 0$,

$$\big\|\chi(\varphi_\zeta)\big\|_{R,\varepsilon} = \big(1 + |\zeta|\big)^{m-1} e^{R|\eta'| + \varepsilon|\xi'|}.$$

With the quantifiers as above and $c = \log C$, we obtain

$$r|\eta| + \delta|\xi| \le c + (m-1)\log\big(1 + |\zeta_1|\big) + R\big|\eta'\big| + \varepsilon\big|\xi'\big|.$$

Looking for the solutions of $P(\zeta) = 0$ for real ζ', i.e., $\eta' = 0$, and choosing $r = 1$, we obtain, for every $\varepsilon > 1$, a C_ε such that

$$|\eta_1| \le C_\varepsilon + (m-1)\log\big(1 + |\zeta_1|\big) + \varepsilon\big|\xi'\big|,$$

which implies

$$|\eta_1| - (m-1)\log\bigl(1+|\eta_1|\bigr) \le C_\varepsilon + (m-1)\log\bigl(1+|\xi_1|\bigr) + \varepsilon\bigl|\xi'\bigr|$$

and, for large $|\eta_1|$,

$$\frac{1}{2}|\eta_1| \le C_\varepsilon + (m-1)\log\bigl(1+|\xi_1|\bigr) + \varepsilon\bigl|\xi'\bigr|. \tag{1}$$

Assume that there are $\xi \in \mathbb{R}^n$ and $\eta \in \mathbb{R}$ such that $P_m(\xi + i\eta e_1) = 0$. We set $g(z) = P_m(\xi + i(\eta + z)e_1)$ for $z \in \mathbb{C}$. Since $g(0) = 0$ and $g \not\equiv 0$, there is k such that $g(z) = z^k g_0(z)$ and $|g_0(z)| \ge A > 0$ in a neighborhood $U_r(0)$.

Now $t^{-m} P(t(\xi + i(\eta + z)e_1)) = g(z) + h(z)$, where $|h(z)| \le \frac{M}{t}$ for z in $U_r(0)$ and all $t > 0$. We apply the theorem of Rouché for large t to the disc $U_\rho(0)$ with $\rho = Ct^{-1/k}$, where C is chosen such that $AC^k > M$, and obtain z_t with $|z_t| \le Ct^{-1/k}$ such that, for $\zeta_t = t(\xi + i(\eta + z_t)e_1)$, we have $P(\zeta_t) = 0$.

We assume now $\eta \ne 0$ and apply inequality (1) with $\varepsilon > 0$ such that $\varepsilon\,|\xi'| < \frac{1}{2}|\eta|$ to ζ_t for large t. Then instead of η_1, we have $\mathrm{Im}(it(\eta + z_t))$, and we can estimate

$$\bigl|\mathrm{Im}\bigl(it(\eta + z_t)\bigr)\bigr| = t|\eta + \mathrm{Re} z_t| \ge t\bigl(|\eta| - |z_t|\bigr) \ge t\bigl(|\eta| - Ct^{-1/k}\bigr).$$

Instead of ξ_1, we have $t(\xi_1 - \mathrm{Im} z_t)$ and the estimate

$$\bigl|t(\xi_1 - \mathrm{Im} z_t)\bigr| \le t\bigl(|\xi_1| + Ct^{-1/k}\bigr).$$

So (1) takes for all large enough t the form

$$\frac{t}{2}\bigl(|\eta| - Ct^{-1/k}\bigr) \le C_\varepsilon + (m-1)\log\bigl(t\bigl(|\xi_1| + Ct^{-1/k}\bigr)\bigr) + \varepsilon\bigl(\bigl|\xi'\bigr|t\bigr).$$

Dividing by t and letting $t \to +\infty$, we get a contradiction to the choice of ε. Hence, η has to be zero. □

Notice that solvability of the Cauchy problem above for all data implies that e_1 is noncharacteristic, hence the uniqueness of a solution. We obtain the following:

Theorem 5 *The following are equivalent*:

1. *The Cauchy problem in Proposition 5 ("inhomogeneous Cauchy problem") is (uniquely) solvable for all data.*
2. *The Cauchy problem in Proposition 6 ("homogeneous Cauchy problem") is (uniquely) solvable for all data.*
3. *P_m is hyperbolic with respect to e_1.*

6 Case of $n = 2$

We may use this to prove a complete characterization for $n = 2$. We assume Ω to be open in $\mathbb{R}^2$.

Lemma 4 *If $n = 2$ and $P^+(D)$ is surjective in $A(\Omega^+)$, then P_m is, up to a constant factor, the product of real linear forms.*

Proof P_m decomposes into irreducible factors, as follows:

$$P_m(z_1, z_2) = A z_2^{m_1} \prod_{\mu=1}^{m_2} (z_1 + a_\mu z_2),$$

where $A \in \mathbb{C}$, $a_1, \ldots, a_{m_2} \in \mathbb{C}$, and $m_1 + m_2 = m$. If $a_\mu \in \mathbb{C} \setminus \mathbb{R}$, then $D_1 + a_\mu D_2$ is elliptic, which, by Lemma 2, cannot occur. □

We arrive at the theorem:

Theorem 6 *For $n = 2$, the following are equivalent*:

1. *$P(D)$ admits a continuous linear right inverse in $A(\Omega)$ for some open convex Ω.*
2. *$P(D)$ admits a continuous linear right inverse in $A(\mathbb{R}^2)$.*
3. *$P^+(D)$ is surjective in $A(\mathbb{R}^3)$.*
4. *P_m is, up to a constant factor, the product of real linear forms.*

Proof $2 \Rightarrow 1$ is obvious, $1 \Rightarrow 4$ follows from Lemmas 3 and 4, $2 \Rightarrow 3$ is Lemma 3, and $3 \Rightarrow 4$ is Lemma 4. It remains to prove $4 \Rightarrow 2$. We notice that $P_m(D)$ is hyperbolic, even with respect to every noncharacteristic direction. Theorem 4 then gives the result. □

Example 2 Consider the polynomial $P(x,t) = x^2 + it \in \mathbb{C}[x,t]$. Then $P(D_x, D_t) = \partial/\partial t - \partial^2/\partial x^2$ is the heat operator in one space dimension. By Theorem 6 it admits a continuous linear right inverse in $A(\mathbb{R}^2)$, while it does not admit such an inverse in $C^\infty(\mathbb{R}^2)$, since it is hypoelliptic (see [21, Theorem 3.3] or [13, Corollary 2.11]).

7 Case of Convex Ω with Boundary

In this section we return to the case handled in Proposition 2 and Example 1. We collect the information we have up to now in the following theorem:

Theorem 7 *If $\Omega \subset \mathbb{R}^n$, $n > 1$, is a bounded, open, and convex set with C^1-boundary, then the following are equivalent*:

1. *$P^+(D)$ is surjective in $A(\Omega^+)$.*
2. *$P_m(D)$ admits a continuous linear right inverse in $C^\infty(\Omega)$.*
3. *P_m is proportional to a product of real linear forms.*
4. *$P(D)$ is $\gamma^{(m/m-1)}$-hyperbolic in every noncharacteristic direction.*
5. *$P(D)$ admits a continuous linear right inverse in $\gamma^{(m/m-1)}(\Omega)$.*

6. $P^+(D)$ *admits a continuous linear right inverse in* $\gamma^{(m/m-1)}(\Omega^+)$.
   If P *is homogeneous, then the following is also equivalent*:
7. $P(D)$ *admits a continuous linear right inverse in* $A(\Omega)$.

Proof 1 $\Rightarrow$ 2. This is Lemma 1.
2 $\Rightarrow$ 3 follows from [13, Theorem 3.8].
3 $\Rightarrow$ 4. Since P_m is hyperbolic in every noncharacteristic direction, this follows from Proposition 4.
4 $\Rightarrow$ 5 follows from [14, Theorem 4.6].
5 $\Rightarrow$ 6 follows by obvious tensor argument, or also from [14, Theorem 4.6], since 5 implies 3, and 3 implies 3 for P_m^+.
6 $\Rightarrow$ 1 follows from [14, Corollary 5.11].
7 $\Rightarrow$ 1 is always true, as follows from Lemma 3.
Let now P be homogeneous.
3 $\Rightarrow$ 7. We have only to show that a real differential operator of order 1 has a continuous linear right inverse in $A(\Omega)$. We may assume that $P(D) = \partial/\partial x_1$. We act in a similar way as in the proof of Example 1. Set $\omega = \{x' \in \mathbb{R}^{n-1} :$ exists x_1 with $(x_1, x') \in \Omega\}$ and for $x' \in \omega$, let $]\gamma_1(x'), \gamma_2(x')[= \{x_1 : (x_1, x') \in \Omega\}$. We put $\tilde{\gamma} = \frac{1}{2}(\gamma_1 + \gamma_2)$. Then $\tilde{\gamma} \in C(\omega)$. By Whitney's approximation theorem (see [17, Theorem 1.6.5]) we find $\gamma \in A(\omega)$ such that $|\gamma(x') - \tilde{\gamma}(x')| < \frac{1}{2}(\gamma_1(x') - \gamma_2(x'))$ for all $x' \in \omega$. Then $x' \mapsto (\gamma(x'), x')$ is a real analytic section of ω to Ω, and $f \mapsto \int_{\gamma(x')}^{x_1} f(\xi, x')d\xi$ is a continuous linear right inverse. □

The author wants to thank P. Domański for suggesting the use of Whitney's approximation theorem in 3 $\Rightarrow$ 7, which led to a considerable improvement of the author's original statement, where real analyticity of the boundary had been assumed. We have even:

Remark 1 3 $\Rightarrow$ 7 in Theorem 7 holds for any convex, open, and bounded Ω.

We can formulate the last part also in the following way:

Theorem 8 *If* Ω *is bounded and convex with* C^1*-boundary, then the only nonconstant irreducible homogeneous differential operators* $P(D)$ *which admit a continuous linear right inverse are, up to a factor, directional derivatives of order one.*

We end the paper by two more special cases. First we assume that $\Omega \subset \mathbb{R}^n$ and $\omega \subset \mathbb{R}^{n-1}$ are convex and open and $\{0\} \times \omega \subset \Omega \subset \mathbb{R} \times \omega$. So Ω may, for instance, be the open unit ball.

We obtain the following analogue to Proposition 5. For some of the tools we will be using, we refer to the proof of Proposition 5.

Proposition 7 *Assume that* $P_m(x) = x_1^m$. *Then the Cauchy problem*

$$P(D)u = 0, \quad \frac{\partial^k}{\partial x_1^k} f(0, x') = f_k(x'), \quad k = 0, \ldots, m-1,$$

is uniquely solvable for all $f_0, \ldots, f_{m-1} \in A(\omega)$ *with* $u \in A(\mathbb{R} \times \omega)$.

Proof Due to the Cauchy–Kowalewski theorem, we find an open neighborhood $W \supset \{0\} \times \omega$ in $\mathbb{R}^n$ and $u_0 \in A(W)$ which solves the Cauchy problem. We choose $\varphi \in \gamma^{(m/m-1)}(\Omega)$ with $\operatorname{supp} \varphi \subset W$ such that $\{x : \varphi(x) = 1\}$ contains a neighborhood of $\{0\} \times \omega$.

Then $w := P(D)(\varphi u_0) \in \gamma^{(m/m-1)}(\mathbb{R} \times \omega)$, $\operatorname{supp} w \subset W$, and $w \equiv 0$ in a neighborhood of $\mathbb{R} \times \omega$. By the assumption and [15, Corollary 2.11] there are fundamental solutions E_+ and E_- with supports in $[0, +\infty) \times \{0\}$ and $(-\infty, 0] \times \{0\}$, respectively. By decomposition of w in an "upper" and a "lower" part, convolution with E_+ or E_-, and putting the results again together, we obtain $v \in \gamma^{(m/m-1)}(\mathbb{R} \times \omega)$ with $P(D)v = w$ and $v \equiv 0$ in a neighborhood of $\{0\} \times \omega$.

We set now $u := u_0 - v$ and obtain a solution in $\gamma^{(m/m-1)}(\mathbb{R} \times \omega)$ of the Cauchy problem such that u is real analytic in a neighborhood of $\{0\} \times \omega$. Now we proceed like in the proof of Proposition 5, using E_+ and E_-. □

In the following we set $N(X) = \{f \in A(X) : P(D)f = 0\}$ for any open subset $X \subset \mathbb{R}^n$.

Theorem 9 *Under the assumptions of Proposition 7, we obtain*:

1. *The restriction map* $N(\mathbb{R} \times \omega) \to N(\Omega)$ *is surjective.*
2. $N(\Omega)$ *is complemented in* $A(\Omega)$.
3. $P(D)$ *has a continuous linear right inverse in* $A(\Omega)$.

Proof $f \to u(f)$, where $u(f)$ is the unique solution of the Cauchy problem of Proposition 7 with $f_k(x') = \frac{\partial^k}{\partial x_1^k} f(0, x')$, is a continuous linear extension operator $N(\Omega) \to N(\mathbb{R} \times \omega)$. This proves 1. Composition with the restriction $A(\mathbb{R}) \to A(\Omega)$ gives the required projection to prove 2. Now $P(D)$ is surjective in $A(\Omega)$, which follows from an easy evaluation of the Phragmén–Lindelöf condition in [8], or from our Remark 1 together with the fact that surjectivity depends only on the principal part (see [8]). Then it is also open (see the Preliminaries). Together with 2, this shows 3. □

Example 3 If $\Omega = \{(t, x) \in \mathbb{R}^2 : t^2 + x^2 < 1\}$ is the open unit ball in $\mathbb{R}^2$, then the heat operator $\partial/\partial t - \partial^2/\partial x^2$ has a continuous linear right inverse in $A(\Omega)$.

Finally, we consider the case of a noncharacteristic half-space.

Theorem 10 *Let* $\Omega = \{x : \langle x, N \rangle < \gamma\}$, *where* $P_m(N) \neq 0$. *Then the following are equivalent*:

1. $P(D)$ *admits a continuous linear right inverse in* $A(\Omega)$.
2. P_m *is hyperbolic with respect to* N.

Proof If 1 is given, then, by Proposition 2, $P_m(D)$ admits a continuous linear right inverse in $C^\infty(\Omega)$, and therefore, by [13, Proposition 3.2], $P_m(D)$ is hyperbolic with respect to N.

To prove the converse, we may assume that $N = e_1$ and $\Omega = \{x : x_1 < 1\}$. The map which assigns to every $f \in A(\Omega)$ the restriction to Ω of the unique solution $u \in A(\mathbb{R}^n)$ of the Cauchy problem

$$P(D)u = 0, \qquad \frac{\partial^k}{\partial x_1^k} u\bigl(0, x'\bigr) = \frac{\partial^k}{\partial x_1^k} f\bigl(0, x'\bigr), \quad k = 0, \dots, m-1,$$

is a continuous projection in $A(\Omega)$ onto $N(\Omega)$. Since, by [13, Proposition 3.2] and [13, Proposition 4.12], $P(D)$ is surjective in $A(\Omega)$, we obtain, like in the proof of Theorem 9, assertion 1. □

References

1. Bonet, J., Domański, P.: Parameter dependence of solutions of differential equations on spaces of distributions and the splitting of short exact sequences. J. Funct. Anal. **230**, 329–381 (2006)
2. Bonet, J., Domański, P.: The splitting of exact sequences of PLS-spaces and smooth dependence of solutions of linear partial differential equations. Adv. Math. **217**, 561–585 (2008)
3. Braun, R., Meise, R., Taylor, B.A.: Ultradifferentiable functions and Fourier analysis. Results Math. **17**, 207–237 (1990)
4. Braun, R., Meise, R., Taylor, B.A.: Perturbation of differential operators admitting a continuous linear right inverse on ultradistributions. Pac. J. Math. **212**, 25–48 (2003)
5. De Giorgi, E., Cattabriga, L.: Una dimostrazione diretta dell'esistenza di solazioni analytiche nel piano reale di equazioni a derivate parziali a coefficienti constanti. Boll. Unione Mat. Ital. **4**, 1015–1027 (1971)
6. Domański, P.: Real analytic parameter dependence of solutions of differential equations. Rev. Mat. Iberoam. **26**, 175–238 (2010)
7. Domański, P.: Real analytic parameter dependence of solutions of differential equations over Roumieu classes. Funct. Approx. **44**, 79–109 (2011)
8. Hörmander, L.: On the existence of real analytic solutions of partial differential equations with constant coefficients. Invent. Math. **21**, 151–183 (1973)
9. Hörmander, L.: The Analysis of Linear Partial Differential Equations II. Springer, Berlin (1983)
10. Langenbruch, M.: Characterization of surjective partial differential operators on spaces of real analytic functions. Studia Math. **162**, 53–96 (2004)
11. Langenbruch, M.: Inheritance of surjectivity for partial differential operators on spaces of real analytic functions. Special issue dedicated to John Horváth. J. Math. Anal. Appl. **297**, 696–719 (2004)
12. Martineau, A.: Sur la topologie des espaces de fonctions holomorphes. Math. Ann. **163**, 62–88 (1966)
13. Meise, R., Taylor, B.A., Vogt, D.: Characterization of the linear partial operators with constant coefficients that admit a continuous linear right inverse. Ann. Inst. Fourier (Grenoble) **40**, 619–655 (1990)
14. Meise, R., Taylor, B.A., Vogt, D.: Continuous linear right inverses for partial differential operators on non-quasianalytic classes and on ultradistributions. Math. Nachr. **180**, 213–242 (1996)
15. Meise, R., Taylor, B.A., Vogt, D.: ω-Hyperbolicity of linear partial differential operators with constant coefficients. In: Complex Analysis, Harmonic Analysis and Applications, Bordeaux, 1995. Pitman Res. Notes Math. Ser., vol. 347, pp. 157–182. Longman, Harlow (1996)
16. Meise, R., Taylor, B.A., Vogt, D.: Phragmén–Lindelöf principles on algebraic varieties. J. Am. Math. Soc. **11**, 1–39 (1998)

17. Narasimhan, R.: Analysis on Real and Complex Manifolds. North-Holland, Amsterdam (1973)
18. Piccinini, L.C.: Non surjectivity of $\partial^2/\partial x^2 + \partial^2/\partial y^2$ as an operator on the space of analytic functions on $\mathbb{R}^3$. In: Lecture Notes of Summer College on Global Analysis, Trieste (1972)
19. Piccinini, L.C.: Non surjectivity of the Cauchy–Riemann operator on the space of the analytic functions on $\mathbb{R}^n$. Generalization to the parabolic operators. Boll. Unione Mat. Ital. **7**, 12–28 (1973)
20. Trèves, F.: Locally Convex Spaces and Linear Partial Differential Equations. Springer, Berlin (1967)
21. Vogt, D.: Some results on continuous linear maps between Fréchet spaces. In: Bierstedt, K.D., Fuchssteiner, B. (eds.) Functional Analysis: Surveys and Recent Results III, pp. 253–264. Elsevier, Amsterdam (1984)
22. Vogt, D.: Real analytic zero solutions of linear partial differential operators with constant coefficients. Bull. Belg. Math. Soc. Simon Stevin **14**, 577–586 (2007)

Averaging Residue Currents and the Stückrad–Vogel Algorithm

Alain Yger

Abstract *Trace formulas* (Lagrange, Jacobi–Kronecker, Bergman–Weil) play a key role in division problems in analytic or algebraic geometry (including arithmetic aspects, see, for example, Berenstein and Yger in Am. J. Math. 121(4):723–796, 1999). Unfortunately, they usually hold within the restricted frame of *complete intersections*. Besides the fact that it allows one to carry local or semi-global analytic problems to a global geometric setting (think about Crofton's formula), averaging the Cauchy kernel (from $\mathbb{C}^n \setminus \{z_1 \dots z_n = 0\} \subset \mathbb{P}^n(\mathbb{C})$), in order to get the Bochner–Martinelli kernel (in $\mathbb{C}^{n+1} \setminus \{0\} \subset \mathbb{P}^{n+1}(\mathbb{C}) = \mathbb{C}^{n+1} \cup \mathbb{P}^n(\mathbb{C})$), leads to the construction of explicit candidates for the realization of Grothendieck's duality, namely *BM residue currents* (Passare et al. in Publ. Mat. 44:85–117, 2000; Andersson in Bull. Sci. Math. 128(6):481–512, 2004; Andersson and Wulcan in Ann. Sci. École Norm. Super. 40:985–1007, 2007), extending thus the cohomological incarnation of duality which appears in the complete intersection or Cohen–Macaulay cases. We will recall here such constructions and, in parallel, suggest how far one could take advantage of the multiplicative inductive construction introduced by Coleff and Herrera (Lecture Notes in Mathematics, vol. 633, Springer, Berlin, 1978), by relating it to the Stückrad–Vogel algorithm developed in Stückrad and Vogel (Queen Pap. Pure Appl. Math. 61:1–32, 1982), Tworzewski (Ann. Pol. Math. 62:177–191, 1995), Andersson et al. (arXiv:1009.2458, 2010) toward improper intersection theory. Results presented here were initiated all along my long-term collaboration with Carlos Berenstein. To both of us, the mathematical work of Leon Ehrenpreis certainly remained a constant and very stimulating source of inspiration. This presentation relies also deeply on my collaboration over the past years with M. Andersson, H. Samuelsson, and E. Wulcan in Göteborg.

A. Yger (✉)
Institut de Mathématiques, Université de Bordeaux, 33405, Talence, France
e-mail: Alain.Yger@math.u-bordeaux1.fr

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*, Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8_23, © Springer-Verlag Italia 2012

1 Coleff–Herrera Residue Currents for Complete Intersections and the Transformation Law

Let $\mathcal{X}$ be an n-dimensional (ambient) complex manifold, and $V \subset \mathcal{X}$ be a closed analytic subset with pure codimension M, equipped with its structure sheaf $\mathcal{O}_V = \mathcal{O}_{\mathcal{X}}/\mathcal{I}_V$, where $\mathcal{I}_{V,x} = \{h_x \in \mathcal{O}_{\mathcal{X},x};\, h_x = 0 \text{ on } V_x\}$. Let $[V]$ be the integration current on the complex subspace $(V, \mathcal{O}_V)$. Let $\Delta_1, \dots, \Delta_m$ be $m \le n - M$ Cartier divisors on $\mathcal{X}$, with respective holomorphic global sections $s_1, \dots, s_m$, such that for all $x \in V \cap \bigcup_{j=1}^m s_j^{-1}(0)$, the germs $(s_{1,x}, \dots, s_{m,x})$ define a regular sequence in $\mathcal{O}_{V,x}$. We denote as $\mathcal{O}_V[s]$ the ideal sheaf of $\mathcal{O}_V$ obtained as the image of $\bigoplus_{j=1}^m \mathcal{O}_V(-\Delta_j)$ by the interior product with $(s_1, \dots, s_m)$. Note that, if the closed hypersurfaces $s_j^{-1}(0)$, $j = 1, \dots, m$, intersect properly on V (that is, $\dim(V \cap \bigcap_{j=1}^m s_j^{-1}(0)) \le n - M - m$) or, equivalently, define a *complete intersection* on V, the required condition holds, provided that the ambient manifold $\mathcal{X}$ is replaced by some convenient neighborhood U of $V \cap s^{-1}(0) := V \cap \bigcap_{j=1}^m s_j^{-1}(0)$. What seems to be today the most robust approach toward the so-called *Coleff–Herrera current* $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$ (originally introduced by Coleff and Herrera in their pioneer work [13]) is the following result (initially obtained in [9] for $m = 2$, then finally extended by Samuelsson [28] for arbitrary m).

Theorem 1 (Robust approach to the Coleff–Herrera residue current attached to a complete intersection) *In the above context, for any choice of C^∞ metrics $|\ |_j$ on the line bundles $\mathcal{O}_{\mathcal{X}}(\Delta_j)$, the holomorphic $\bigoplus_{r=0}^m \bigoplus_{1 \le j_1 < \cdots < j_r \le m} {}'\mathcal{D}^{(M,M+r)}(\mathcal{X}, \bigwedge_{l=1}^r \mathcal{O}_{\mathcal{X}}(-\Delta_{j_l}))$-valued map*

$$(\lambda_1, \dots, \lambda_m) \in \{\lambda;\, \operatorname{Re}\lambda_j > 1,\ j = 1, \dots, m\}$$
$$\longmapsto \left[\bigwedge_{j=m}^{1} \left(1 - |s_j|_j^{2\lambda_j} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_j|_j^{2\lambda_j}}{s_j}\right)\right)\right] \wedge [V] \tag{1}$$

can be analytically continued as a holomorphic map to a product of half-spaces $\{\operatorname{Re}\lambda_j > -\eta\}$ for some $\eta > 0$. Its value at $\lambda = 0$ coincides with its ${}'\mathcal{D}^{M,M+r}(\mathcal{X}, \bigwedge_{j=1}^m \mathcal{O}_{\mathcal{X}}(-\Delta_j))$-component and defines a $\overline{\partial}$-closed bundle-valued current, which is independent of the metrics $|\ |_j$, $j = 1, \dots, m$. This current, supported by $V \cap s^{-1}(0)$, is denoted as $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$. Considered as a $\overline{\partial}$-closed current on the complex space $(V, \mathcal{O}_V)$, it is locally annihilated by any local section of the ideal sheaf $\mathcal{O}_V[s]$. Moreover, when $V = \mathcal{X}$, it realizes the local duality with respect to the ideal sheaf $\mathcal{O}_{\mathcal{X}}[s]$, namely

$$\left(h_x \cdot \bigwedge_{j=1}^m \overline{\partial}(1/s_j) = 0\right) \Longleftrightarrow \left(h_x \in \big(\mathcal{O}_{\mathcal{X}}[s]\big)_x = (s_{1,x}, \dots, s_{m,x})\mathcal{O}_{\mathcal{X},x}\right). \tag{2}$$

The intimate relationship between the sheaf of differential operators with meromorphic coefficients and the local description of such a current arises from the following remark, which refers to the sheaves of Coleff–Herrera currents (here bundle-valued, holomorphic or meromorphic, see [32]) introduced by Björk [11], together with their companion local structure theorems (see [11], or also [32] for a survey in this volume). The current $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$ splits locally about each point of its support $V \cap s^{-1}(0)$ as $\mathcal{T}_x \wedge \omega_x$, where $\mathcal{T}_x$ is a local section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X}, V\cap s^{-1}(0)}(\cdot; \star S_x, \bigwedge_{j=1}^m \mathcal{O}_{\mathcal{X}}(-\Delta_j))$, and ω_x is a local section of $\Omega_{\mathcal{X}}^{M+m}(\cdot, \bigwedge_1^m \mathcal{O}_{\mathcal{X}}(\Delta_j))$. Here S_x denotes a germ of hypersurface at x polar respect to $V_x \cap s_x^{-1}(0)$, i.e., $\overline{(V_x \cap s_x^{-1}(0)) \setminus S_x} = V_x \cap s_x^{-1}(0)$. Moreover, when V is Cohen–Macaulay about $x_0 \in V \cap s^{-1}(0)$ (that is, $\mathcal{O}_{V,x_0}$ is Cohen–Macaulay), the integration current $[V]$ factorizes about x_0 as $\tau_{x_0} \wedge \varpi_{x_0}$, where ϖ_{x_0} is a local section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}$, and ϖ_{x_0} is a section of the sheaf $\Omega_{\mathcal{X}}^M$ (see [4]), which implies in this particular case that one can take about x_0 the current section $\mathcal{T}_{x_0}$ to be a local section T_{x_0} of $\mathrm{CH}_{\mathcal{X}, V\cap s^{-1}(0)}(\cdot, \bigwedge_1^m \mathcal{O}_{\mathcal{X}}(-\Delta_j))$ instead of $\mathrm{CH}_{\mathcal{X}, V\cap s^{-1}(0)}(\cdot; \star S_{x_0}, \bigwedge_1^m \mathcal{O}_{\mathcal{X}}(-\Delta_j))$. This occurs in particular when V is defined as a reduced complete intersection about the point x_0.

Remark 1 Theorem 1 provides a robust approach (via analytic continuation) toward the Coleff–Herrera residue current for complete intersections on a purely dimensional analytic set $(V, \mathcal{O}_V)$, as developed in [13]. It is important here to point out that one may replace in (1) the integration current $[V]$ by any global section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot, E)$, where E denotes a finite-rank holomorphic bundle over $\mathcal{X}$. When $m \leq M$, the first assertion in Theorem 1 remains valid under the complete intersection hypothesis about the s_j's on V, which provides some kind of robustness for multiplicative residue calculus involving Coleff–Herrera currents (see, for example, Proposition 3 below). The proof can be carried in a way similar to that in [28], taking advantage of the wedge product operation introduced in [12] (see also [32]). Be careful however that the definition of the sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot, E)$, whose global sections are $(0, M)$ E-valued currents in the ambient manifold $\mathcal{X}$ with support lying in V, depends on the embedding $V \subset \mathcal{X}$. Nevertheless, Remark 1 will be important with respect to the role of such Coleff–Herrera residue currents in the construction of analytic tools for division theory, in accordance with that played by integration currents in intersection theory (see Sect. 2.2 below).

We considered so far the holomorphic sections $s_1, \dots, s_m$ of the hermitian line bundles $(\mathcal{O}_{\mathcal{X}}(\Delta_j), |\ |_j)$ independently, then introduced the Coleff–Herrera current $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$ through a multiplicative procedure which is reminiscent of the multiplicative operational formalism carried by the Cauchy kernel $dz_1/z_1 \wedge \cdots \wedge dz_m/z_m$. Instead of that, one could interpret $s = s_1 \oplus \cdots \oplus s_m$ as a holomorphic section of the m-holomorphic bundle $\bigoplus_{j=1}^m \mathcal{O}_{\mathcal{X}}(\Delta_j)$, equipped with the metric $\|\ \|^2 = |\ |_1^2 \oplus \cdots \oplus |\ |_m^2$. The robustness of the approach toward the current $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$ (in the complete intersection setting, see Theorem 1) motivates this alternative one. We still use analytic continuation, but this time with respect to a single auxiliary complex parameter λ instead of several. Such an approach was introduced in [9, 27], then extensively developed (after being interpreted within the frame of holomorphic hermitian bundles) in [3, 4, 6, 7]. It is based on an averaging procedure.

Theorem 2 (Bochner–Martinelli approach) *Let $\mathcal{X}$, V, the Δ_j's, and the s_j's be as in the preamble of this section. Let $\| \ \|$ be an arbitrary hermitian metric on the m-dimensional bundle $\bigoplus_{j=1}^m \mathcal{O}_{\mathcal{X}}(\Delta_j) = E_\Delta$, $s = s_1 \oplus \cdots \oplus s_m$, and s^* be the conjugate section of s, that is, $s^*(x)(\xi) = \langle\langle \xi, s(x)\rangle\rangle_x$ for $x \in \mathcal{X}$, $\xi \in E_{\Delta,x}$. The holomorphic $\bigoplus_{r=0}^m {}'\mathcal{D}^{M,M+r}(\mathcal{X}, \bigwedge^r E_\Delta^*)$-valued map*

$$\lambda \longmapsto \left(\left(1 - \|s\|^{2\lambda}\right) + \overline{\partial}\|s\|^{2\lambda} \wedge \left(\sum_{r=1}^m \frac{1}{(2i\pi)^r} \frac{s^* \wedge (\overline{\partial} s^*)^{r-1}}{\|s\|^{2r}}\right)\right) \wedge [V] \qquad (3)$$

(*defined for* $\operatorname{Re}\lambda \gg 1$) *extends as a holomorphic map in a half-plane* $\operatorname{Re}\lambda > -\eta$ *for some* $\eta > 0$. *It coincides at* $\lambda = 0$ *with its* ${}'\mathcal{D}^{M,M+m}(\mathcal{X}, \bigwedge^m E_\Delta^*)$ *component, that is, independently of the choice of the metric* $\| \ \|$ *on* E_Δ, *with the Coleff–Herrera current* $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$.

Remark 2 In the particular case where all line bundles $\mathcal{O}_{\mathcal{X}}(\Delta_j)$, $j = 1, \ldots, m$, are trivial over $\mathcal{X}$, and the holomorphic sections s_j are holomorphic functions in $\mathcal{X}$ defining a complete intersection on V, one can interpret the current

$$\left[\left(\left(1 - \|s\|^{2\lambda}\right) + \overline{\partial}\|s\|^{2\lambda} \wedge \left(\sum_{r=1}^m \frac{1}{(2i\pi)^r} \frac{s^* \wedge (\overline{\partial} s^*)^{r-1}}{\|s\|^{2r}}\right)\right) \wedge [V]\right]_{\lambda=0}$$

(now $\| \ \|$ is the Euclidean norm on $\mathbb{C}^m$, and $s^*(x)(\xi) = \sum_1^m \overline{s_j(x)}\, \xi_j$ for $\xi \in \mathbb{C}^m$) as an averaged value of the currents

$$\left[\bigwedge_{j=m}^{1} \left(1 - \left|\langle u^j, s\rangle\right|^{2\lambda_j} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|\langle u^j, s\rangle|^{2\lambda_j}}{\langle u^j, s\rangle}\right)\right) \wedge [V]\right]_{\lambda_1 = \cdots = \lambda_m = 0}$$

(where $(u^1, \ldots, u^m) \in (\mathbb{P}^{n-1}(\mathbb{C}))^m$, and $\langle s, u^j\rangle = \sum_{k=0}^m u_k^j s_k$ for $k = 1, \ldots, m$) defined as in (1), for $(u^1, \ldots, u^m)$ generic in $(\mathbb{P}^{m-1}(\mathbb{C}))^m$, with respect to the normalized tensorized Fubini–Study metric on $(\mathbb{P}^{n-1}(\mathbb{C}))^m$.

Averaging Coleff–Herrera currents (realized in a multiplicative form as in Theorem 1) prevents usually from keeping track of the algebraic or arithmetic structure of the data (when there is one). Keeping track of such a structure is indeed better possible through a multiplicative approach such as in Theorem 1. Nevertheless, within the complete intersection frame, the fact that the current $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge [V]$ remains preserved under such averaging (Theorem 2) allows one to reinterpret in geometric terms (and prove) the key (algebraic) operational property of multivariate residue calculus (see, for example, [10, 18, 23]), the so-called *Transformation Law*. In order to formulate such a computational rule (which appears to be the geometric counterpart of Wiebe's theorem [33]) in a rather general form, one needs to introduce two sequences of Cartier divisors on the ambient complex manifold $\mathcal{X}$, namely $(\Delta_1, \dots, \Delta_m, \mathfrak{T}_0, \dots, \mathfrak{T}_{k-1})$ and $(\widetilde{\Delta}_1, \dots, \widetilde{\Delta}_m)$ $(m + k \le n - M)$, together with respective holomorphic sections $(s_1, \dots, s_m, t_0, \dots, t_{k-1}) = (s, t)$, $(\widetilde{s_1}, \dots, \widetilde{s_m}) = \widetilde{s}$; when $k = 0$, one takes $\{\mathfrak{T}_0, \dots, \mathfrak{T}_{k-1}\} = \emptyset$. The geometric hypotheses are:

$$\begin{aligned} &\operatorname{codim}_V \big(V \cap s^{-1}(0)\big) = \operatorname{codim}_V \big(V \cap \widetilde{s}^{-1}(0)\big) = m; \\ &\overline{\big(V \cap s^{-1}(0)\big) \cap t_l^{-1}(0)} = V \cap s^{-1}(0) \quad \forall l = 0, \dots, k-1; \\ &\operatorname{codim}_V \big(V \cap s^{-1}(0) \cap t^{-1}(0)\big) = \operatorname{codim}_V \big(V \cap \widetilde{s}^{-1}(0) \cap t^{-1}(0)\big) = m + k. \end{aligned} \tag{4}$$

Let $E_\Delta = \bigoplus_{j=1}^m \mathcal{O}_{\mathcal{X}}(\Delta_j)$ and $E_{\widetilde{\Delta}} = \bigoplus_{j=1}^m \mathcal{O}_{\mathcal{X}}(\widetilde{\Delta}_j)$. As a pendant algebraic hypothesis, one assumes that there is a meromorphic section H of $\operatorname{Hom}_{\mathbb{C}}(E_\Delta, E_{\widetilde{\Delta}})$, together with positive integers $\nu_0, \dots, \nu_k$, such that $\widetilde{s} = H \cdot s$ and

$$\mathfrak{D} = \det H \otimes \bigotimes_{\kappa=0}^{k-1} t_\kappa^{\otimes \nu_\kappa} \in \mathcal{O}_{\mathcal{X}}\left(\mathcal{X}, \operatorname{Hom}_{\mathbb{C}}(E_\Delta, E_{\widetilde{\Delta}}) \otimes \bigotimes_{\kappa=0}^{k-1} \mathcal{O}_{\mathcal{X}}(\nu_\kappa \mathfrak{T}_\kappa)\right). \tag{5}$$

Theorem 3 (Transformation Law for Coleff–Herrera residue currents) *Under the above geometric and algebraic hypotheses* (4) *and* (5), *one has the following identity between bundle valued currents*:

$$\bigwedge_{j=1}^m \overline{\partial}\left(\frac{1}{s_j}\right) \wedge \bigwedge_{\kappa=0}^{k-1} \overline{\partial}\left(\frac{1}{t_\kappa}\right) \wedge [V] = \left(\bigwedge_{j=1}^m \overline{\partial}\left(\frac{1}{\widetilde{s}_j}\right) \wedge \bigwedge_{\kappa=0}^{k-1} \overline{\partial}\left(\frac{1}{t_\kappa^{\nu_\kappa+1}}\right) \wedge [V]\right) \otimes \mathfrak{D}.$$

Proof In order to get this result, we introduce metrics $|\ |_\kappa$ on the divisors $\mathfrak{T}_\kappa$, $\kappa = 0, \dots, k-1$, and compare the values at $\mu_0 = \cdots = \mu_{k-1}$ (following the analytic continuation) of the two holomorphic maps (for $\operatorname{Re} \mu_\kappa \gg 1$, $\kappa = 0, \dots, k-1$):

$$\begin{aligned} &\mu \longmapsto \bigwedge_{\kappa=k-1}^{0} \overline{\partial}\left(\frac{|t_\kappa|_\kappa^{2\mu_\kappa}}{t_\kappa}\right) \wedge \bigwedge_{j=1}^m \overline{\partial}\left(\frac{1}{s_j}\right) \wedge [V] \\ &\mu \longmapsto \left(\bigwedge_{\kappa=k-1}^{0} \overline{\partial}\left(\frac{|t_\kappa|_\kappa^{2\mu_\kappa}}{t_\kappa}\right) \wedge \bigwedge_{j=1}^m \overline{\partial}\left(\frac{1}{\widetilde{s}_j}\right) \wedge [V]\right) \otimes \frac{\mathfrak{D}}{t_0^{\nu_0} \cdots t_{k-1}^{\nu_{k-1}}}. \end{aligned} \tag{6}$$

These two current-valued maps coincide for $\operatorname{Re} \mu_\kappa \gg 1$, $\kappa = 0, \dots, k-1$, because of the robustness assumption in Theorem 2, together with the holonomy and the Standard Extension Property of sections of the Coleff–Herrera sheaves $\mathrm{CH}_{\mathcal{X}, V \cap s^{-1}(0)}(\cdot; \star S)$ or $\mathrm{CH}_{\mathcal{X}, V \cap \widetilde{s}^{-1}(0)}(\cdot; \star S)$, where $S = \bigcup_{\kappa=0}^{k-1} t_\kappa^{-1}(0)$ (see [32], condition 1 in Definition 2). The fact that they share the same value at $\mu = 0$ about any $x \in V \cap s^{-1}(0) \cap t^{-1}(0)$ follows from the robustness of the approach toward the Coleff–Herrera residue via analytic continuation (Theorem 1). On the other hand, both maps (6) vanish at $\mu = 0$ about any $x \in V \cap (\widetilde{s}^{-1}(0) \setminus s^{-1}(0)) \cap t^{-1}(0)$ because of Cramer's rule, combined with the local duality property (2) in Theorem 1 (here is the crucial algebraic point on which relies Wiebe's theorem in the algebraic context). □

In order to focus on the importance of the Transformation Law, let us conclude this section with few comments inspired by arithmetic considerations. Let $\{p_1, \dots, p_{\mathfrak{m}}\}$ be a collection of polynomials in $\mathbb{Z}[X_1, \dots, X_{\mathfrak{n}}]$ (here $\mathfrak{m} \leq \mathfrak{n}$), with respective degrees $d_1, \dots, d_{\mathfrak{m}}$, whose homogenizations define a complete intersection in $\mathbb{C}^n = \mathbb{P}^n(\mathbb{C}) \setminus \{z_0 = 0\}$. Denote as $T[p_1, \dots, p_{\mathfrak{m}}]$ the $\bigwedge_1^{\mathfrak{m}} \mathcal{O}_{\mathbb{P}^{\mathfrak{n}}(\mathbb{C})}(-d_j)$-valued Coleff–Herrera global current in $\mathbb{P}^{\mathfrak{n}}(\mathbb{C})$ obtained from the current $T[p_1] = \overline{\partial}(1/p_1)$ through the inductive process

$$T[p_1, \dots, p_{l+1}] = \left[\frac{1}{2i\pi}\overline{\partial}\left(\frac{|p_{l+1}|_{l+1}^{2\lambda}}{p_l}\right) \wedge T[p_1, \dots, p_l]\right]_{\lambda=0}, \quad l = 1, \dots, \mathfrak{m} - 1. \tag{7}$$

The metric $|\ |_l$ one takes here on $\mathcal{O}_{\mathbb{P}^{\mathfrak{n}}(\mathbb{C})}(d_{l+1})$ is the Fubini–Study metric, but this is in fact irrelevant, since the result after taking $\lambda = 0$ in (7) is unaffected by such a choice. For $u^1, \dots, u^{\mathfrak{n}-\mathfrak{m}}$ generic in $\mathbb{Z}^{\mathfrak{m}+1} \setminus \{0\}$, the linear subspace

$$\Pi_u := \left\{[z_0 : \dots : z_{\mathfrak{n}}] \in \mathbb{P}^{\mathfrak{n}}(\mathbb{C});\ \langle u^j, z \rangle = 0,\ j = 1, \dots, \mathfrak{m}\right\}$$

is such that $\dim_{\mathbb{C}^{\mathfrak{n}}}(\Pi_u \cap \{p_1 = \dots = p_{\mathfrak{m}} = 0\}) = 0$. Consider the $\bigwedge_1^{\mathfrak{m}} \mathcal{O}_{\mathbb{P}^{\mathfrak{n}}(\mathbb{C})}(-d_j)$-valued $(\mathfrak{n} - \mathfrak{m}, \mathfrak{n})$ $\overline{\partial}$-closed current $T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u]$. For any global smooth $(\mathfrak{m}, 0)$-form $\mathfrak{b}$ in $\mathbb{P}^{\mathfrak{n}}(\mathbb{C})$ with values in $\bigwedge_1^{\mathfrak{m}} \mathcal{O}_{\mathbb{P}^{\mathfrak{n}}(\mathbb{C})}(d_j)$, we have

$$\begin{aligned}
&\langle T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u], \mathfrak{b}\rangle \\
&\quad = \left\langle \left[(1 - |z_0|^{2\lambda}) T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u]\right]_{\lambda=0}, \mathfrak{b}\right\rangle \\
&\qquad + \left\langle \left[|z_0|^{2\lambda} T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u]\right]_{\lambda=0}, \mathfrak{b}\right\rangle \\
&\quad = \left\langle \left(T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u]\right)_{|\{z_0=0\}}, \mathfrak{b}\right\rangle + \left\langle \left(T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u]\right)_{|\mathbb{C}^{\mathfrak{n}}}, \mathfrak{b}\right\rangle.
\end{aligned}$$

An explicit computation of

$$\left\langle \left(T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u]\right)_{|\mathbb{C}^{\mathfrak{n}}}, \mathfrak{b}\right\rangle \tag{8}$$

can be carried thanks to the Transformation Law (such as formulated in Theorem 3 with $k = 0$), following the procedure described in [18]. In particular, if $\mathfrak{b}[q, J]$ is the form expressed in $\mathbb{C}^{\mathfrak{n}}$, in affine coordinates $\zeta_1, \dots, \zeta_{\mathfrak{n}}$, as

$$\mathfrak{b}[q, J] = \frac{q(\zeta)}{(1 + |\zeta|^2)^{(d+\deg q)/2}} \bigwedge_{l=1}^{\mathfrak{m}} d\left(\frac{\zeta_{j_l}}{\sqrt{1 + |\zeta|^2}}\right), \quad d = d_1 + \dots + d_{\mathfrak{m}},$$

for some $q \in \mathbb{Z}[X_1, \dots, X_{\mathfrak{n}}]$ and $1 \leq j_1 < \dots < j_{\mathfrak{m}} \leq \mathfrak{n}$, we get for (8) a somehow explicit rational expression $R_{q,J}[u] \in \mathbb{Q}$. A key point in such a procedure is that it provides a final estimate on the logarithmic height h of the rational function $R_{q,J}[u]$ of the form

$$h\big(R_{q,J}[u]\big) \le h(q) + \kappa_1(n) D(D + \deg q) \max\big(h(p_j), \log \|u^j\|\big), \tag{9}$$

where $D = \prod_1^{\mathfrak{m}}(d_j + 1)$, in accordance with the geometric Bézout theorem and its arithmetic counterpart. If $q\ (T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u])_{\{z_0=0\}}$, we get a rational expression with logarithmic height control (9) for $\langle T[p_1, \dots, p_{\mathfrak{m}}] \wedge [\Pi_u], \mathfrak{b}[q, J]\rangle$.

2 Coleff–Herrera Residue Currents and Ordered Sequences of Sections of Cartier Divisors

Let $\mathcal{X}$ be an n-dimensional complex manifold. Given a purely dimensional cycle $[V] \in Z^{n-M}(\mathcal{X})$ (identified here with the (M, M) associated integration current on it and with support denoted as V) and a closed submanifold $\mathcal{Z} \subset \mathcal{X}$, the splitting operation

$$[V] = [V]^{\mathcal{Z}} + [V]^{\mathcal{X}\setminus\mathcal{Z}}, \tag{10}$$

which consists in separating (as components of $[V]^{\mathcal{Z}}$) the components of $[V]$ with support lying in $\mathcal{Z}$, from the others, is one of the major operational tools in *geometric intersection theory*. Transposed in algebraic terms, it leads to the notion of *gap sheaf* (for an introduction to the subject, see Part I in [26]).

Example 1 ([24, 30, 31]) Let $[V_1], \dots, [V_{\mathfrak{m}}]$ be $\mathfrak{m}$ purely dimensional algebraic cycles in $\mathbb{P}^{\mathfrak{n}}(\mathbb{C})$, $[V]$ be the $(n - M)$-cycle in $\mathcal{X} = \mathbb{P}^n(\mathbb{C})$ ($n = \mathfrak{m}(\mathfrak{n} + 1) - 1$, $M = \sum_{j=1}^{\mathfrak{m}} \operatorname{codim}_{\mathbb{P}^{\mathfrak{n}}(\mathbb{C})} V_j$) corresponding to the *ruled join*

$$J(V_1, \dots, V_{\mathfrak{m}}) = \big\{[Z_1 : \dots : Z_{\mathfrak{m}}] \in \mathbb{P}^{\mathfrak{m}(\mathfrak{n}+1)-1}(\mathbb{C});\ z_j \in V_j\ \forall j = 1, \dots, \mathfrak{m}\big\}$$

(multiplicities been taken into account), and $\mathcal{Z}$ be the diagonal subspace of $\mathbb{P}^n(\mathbb{C})$ defined as a complete intersection as the set of points $[Z_1 : \dots : Z_{\mathfrak{m}}] \in \mathbb{P}^n(\mathbb{C})$ such that $Z^k = Z^1$ for any $k \in \{2, \dots, \mathfrak{m}\}$ (see [19]). Note that, in this particular case, the coherent $\mathcal{O}_{\mathbb{P}^n(\mathbb{C})}$-ideal sheaf $\mathcal{I}_{\mathcal{Z}}$ is globally generated by $(\mathfrak{m} - 1)(\mathfrak{n} + 1)$ holomorphic sections σ_l of the same line bundle $\mathcal{O}_{\mathbb{P}^n(\mathbb{C})}(1)$. Moreover the zero sets $\sigma_l^{-1}(0)$, $l = 1, \dots, (\mathfrak{m} - 1)(\mathfrak{n} + 1)$, define here $\mathcal{Z}$ as a complete intersection in the ambient n-dimensional manifold $\mathcal{X} = \mathbb{P}^n(\mathbb{C})$.

We propose now to transpose, within the frame of global sections of the Coleff–Herrera sheaves $\mathrm{CH}_{\mathcal{X},V}$ (or $\mathrm{CH}_{\mathcal{X},V}(\cdot\,; \star S)$) (instead of the geometric frame of integration currents), the splitting operation (10), together with the construction such an operation generates when applied inductively. We should point out here that such a construction is directly inspired from the ideas introduced by Coleff and Herrera [13] (see, for example, [29]).

2.1 The Coleff–Herrera Product

Let $\mathcal{X}$ and V be as in Sect. 1. Consider also an ordered sequence $\Delta_1, \dots, \Delta_m$ (one drops here the assumption $m \le n - M$) of Cartier divisors on $\mathcal{X}$, equipped respectively with hermitian metrics $|\ |_j$, $j = 1, \dots, m$, and with holomorphic global sections $s_1, \dots, s_m$. The major difference with Sect. 1 is that we forget here the geometric assumption that the closed hypersurfaces $s_j^{-1}(0) \subset \mathcal{X}$, $j = 1, \dots, m$, intersect as a complete intersection on the closed analytic subset V. Let T be a global section over $\mathcal{X}$ of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot, E)$, where E denotes some finite-rank holomorphic bundle over $\mathcal{X}$. Let $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge T$ be the current in

$$\bigoplus_{r=0}^{m} \bigoplus_{1\le j_1<\cdots<j_r\le m} {}'\mathcal{D}^{(0,M+r)}\left(\mathcal{X}, \bigwedge_{l=1}^{r} \mathcal{O}_{\mathcal{X}}(-\Delta_{j_l}) \otimes E\right)$$

defined inductively from

$$\overline{\partial}\left(\frac{1}{s_1}\right) \wedge T = \left[\left(1 - |s_1|_1^{2\lambda} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_1|_1^{2\lambda}}{s_1}\right)\right) \wedge T\right]_{\lambda=0}$$

by the iterative process

$$\bigwedge_{j=1}^{l+1} \overline{\partial}\left(\frac{1}{s_j}\right) \wedge T = \left[\left(1 - |s_{l+1}|_{l+1}^{2\lambda} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_{l+1}|_{l+1}^{2\lambda}}{s_{l+1}}\right)\right) \wedge \bigwedge_{j=1}^{l} \overline{\partial}\left(\frac{1}{s_j}\right) \wedge T\right]_{\lambda=0} \tag{11}$$

for $l = 1, \dots, m-1$. Remark that in spite of the notation $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge T$ (that we use here for simplicity), this current is not in general a $(0, m+M)$-current. The above inductive construction is independent of the choice of the metrics $|\ |_j$ on the line bundles $\mathcal{O}_{\mathcal{X}}(\Delta_j)$ but of course depends on the ordering of the sequence of divisors $\{\Delta_1, \dots, \Delta_m\}$. When the hypersurfaces $s_j^{-1}(0)$, $j = 1, \dots, m$, define a complete intersection on V, this current coincides with its ${}'\mathcal{D}^{(0,M+m)}(\mathcal{X}, \bigwedge_{j=1}^m \mathcal{O}_{\mathcal{X}}(-\Delta_j) \otimes E)$ component and can be recovered in a robust way in a neighborhood of $V \cap s^{-1}(0)$ as the value at $\lambda_1 = \cdots = \lambda_m = 0$ of a holomorphic function in m variables in a product of half planes $\mathrm{Re}\,\lambda_j > -\eta$ for some $\eta > 0$ (see Remark 1 above). The iterative construction is justified by the local structure theorems for sections of the Coleff–Herrera sheaves $\mathrm{CH}_{\mathcal{X},W}(\cdot; \star S, E)$, where W denotes a purely dimensional closed analytic subset of V, and S is a closed hypersurface in a neighborhood of W in $\mathcal{X}$ such that $\overline{W \setminus S} = W$. When $\underline{j} = (j_1, \dots, j_r)$ is an r-uplet ($0 \le r \le m$, $\underline{j} = \emptyset$ if $r = 0$) of strictly increasing integers $1 \le j_1 < \cdots < j_r \le m$, the component $(\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge T)_{\underline{j}}$ in ${}'\mathcal{D}^{(0,M+r)}(\mathcal{X}, \bigwedge_{l=1}^r \mathcal{O}_{\mathcal{X}}(-\Delta_{j_l}) \otimes E)$ is a global section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V_{\underline{j}}[s]}(\cdot; \star S_{\underline{j}}[s], E_{\underline{j}})$, where $E_{\underline{j}} = E_{\underline{j}}[\Delta, E] = \bigwedge_{l=1}^r \mathcal{O}_{\mathcal{X}}(-\Delta_{j_l}) \otimes E$, $V_{\underline{j}}[s]$ is a purely $(n - M - r)$-dimensional closed subset of V, and $S_{\underline{j}}[s]$ is some closed hypersurface in a neighborhood of $V_{\underline{j}}[s]$ such that $\overline{V_{\underline{j}}[s] \setminus S_{\underline{j}}[s]} = V_{\underline{j}}[s]$. The construction of $V_{\underline{j}}[s]$ is carried through the procedure that leads to the construction of the multicycle of contact (see [31]) between a given cycle $[V]$ and a smooth closed submanifold $Z \subset \mathcal{X}$ contained in the intersection of an ordered sequence $H_1, \dots, H_m$ ($m \le \dim V$) of closed hypersurfaces. More precisely, $V_{\underline{j}}[s]$ can be reached through the following iterated splitting operation: when W is a closed purely dimensional subset in $\mathcal{X}$ and S denotes a closed hypersurface in a neighborhood of W, then W is (geometrically) decomposed as

$$W = W^S \cup W^{\mathcal{X}\setminus S}, \tag{12}$$

where W^S denotes the union of irreducible components of W lying entirely in S, and $W^{\mathcal{X}\setminus S}$ is the union of the remaining ones. The closed analytic set $V_{\underline{j}}[s]$ appears then as the end term in the inductive sequence:

$$\begin{aligned}
&V_{\underline{j},1}[s] = V^{s_1^{-1}(0)} \text{ if } 1 \notin \{j_1, \dots, j_r\};\ \text{else } V_{\underline{j},1}[s] = V^{\mathcal{X}\setminus s_1^{-1}(0)} \cap s_1^{-1}(0);\\
&\qquad \dots \qquad \dots \qquad \dots\\
&V_{\underline{j},k+1}[s] = \left(V_{\underline{j},k}[s]\right)^{s_{k+1}^{-1}(0)} \text{ if } k+1 \notin \{j_1, \dots, j_r\};\\
&\qquad\qquad\qquad \text{else } V_{\underline{j},k+1}[s] = \left(V_{\underline{j},k}[s]\right)^{\mathcal{X}\setminus s_{k+1}^{-1}(0)} \cap s_{k+1}^{-1}(0);\\
&\qquad \dots \qquad \dots \qquad \dots\\
&V_{\underline{j}}[s] = V_{\underline{j},m}[s] = \left(V_{\underline{j},m-1}[s]\right)^{s_m^{-1}(0)} \text{ if } m \notin \{j_1, \dots, j_r\};\\
&\qquad\qquad\qquad \text{else } V_{\underline{j}}[s] = V_{\underline{j},m}[s] = \left(V_{\underline{j},m-1}[s]\right)^{\mathcal{X}\setminus s_m^{-1}(0)} \cap s_m^{-1}(0).
\end{aligned} \tag{13}$$

If $m \le n - M$ and $r = m$, $V_{\{1,\dots,m\}}[s]$ is the so-called *essential intersection* $(V \cap s_1^{-1}(0) \cap \dots \cap s_m^{-1}(0))_{\text{ess}}$, while, if $r = 0$, $V_{\emptyset}[s]$ is the union of the irreducible components of V which lie entirely in the intersection $s_1^{-1}(0) \cap \dots \cap s_m^{-1}(0)$.

The same iterative procedure allows us also to define the Coleff–Herrera current $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge \mathcal{T}$ when $\mathcal{T}$ is a global section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V}(\cdot;\ \star S, E)$ for some closed hypersurface S in a neighborhood of V such that $\overline{V \setminus S} = V$. As before, each component $(\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge \mathcal{T})_{\underline{j}}$ is then a global section of the Coleff–Herrera sheaf $\mathrm{CH}_{\mathcal{X},V_{\underline{j}}[s]}(\cdot;\ \star S_{\underline{j}}[s,S], E_{\underline{j}})$ for some convenient hypersurface $S_{\underline{j}}[s,S]$ in a neighborhood of $V_{\underline{j}}[s]$ such that $\overline{V_{\underline{j}}[s] \setminus S_{\underline{j}}[s,S]} = V_{\underline{j}}[s]$. We can point out (see [8]) that analytic continuation with respect to a single complex parameter (as in Theorem 2) provides a direct approach to such a current.

Proposition 1 *Let $\mathcal{X}$, V, the Δ_j's, the metrics $|\ |_j$, the holomorphic sections s_j's, and the Coleff–Herrera current $\mathcal{T}$ (with eventual poles) be as above. Let $\gamma_1 > \gamma_2 >$*

$\cdots > \gamma_m \geq 1$ *be* m *positive integers, and* $\epsilon > 0$. *Then the current-valued map*

$$\lambda \in \left\{\lambda \in \mathbb{C};\ \left|\arg_{]\pi,\pi[}(\lambda)\right| < \frac{\pi}{2(\gamma_1+\epsilon)};\ |\lambda| \gg 1\right\}$$
$$\longmapsto \left[\bigwedge_{j=m}^{1}\left(1 - |s_j|_j^{2\lambda^{\gamma_j}} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_j|_j^{2\lambda^{\gamma_j}}}{s_j}\right)\right)\right] \wedge \mathcal{T} \tag{14}$$

extends as a holomorphic map in an open neighborhood of the closed sector $\{|\arg_{]\pi,\pi[}(\lambda)| \leq \pi/(2\gamma_1)\}$. *Its value at* $\lambda = 0$ *equals the current* $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge \mathcal{T}$.

We intend from now on to focus on the advantages (or disadvantages) that carry the idea of averaging (as done in Sect. 1 when stating Theorem 1, then Theorem 2) Coleff–Herrera residues of the form $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge T$ or $\bigwedge_{j=1}^m \overline{\partial}(1/s_j) \wedge \mathcal{T}$, when V is a purely dimension $(n-M)$-closed analytic subset in a complex manifold $\mathcal{X}$, $s_1, \ldots, s_m$ being holomorphic sections of arbitrary Cartier divisors on $\mathcal{X}$.

2.2 Vogel Sequences and Vogel Residue Currents

The first positive point with respect to the averaging idea arises from the theory of improper intersection on $\mathcal{X}$, as developed in the algebraic context in [30] (see also [24]) and in the analytic context in [31]. Let $\mathcal{X}$ and V be as in Sect. 1, and $\mathcal{I}$ be a coherent ideal sheaf in $\mathcal{O}_{\mathcal{X}}$, with $Z(\mathcal{I})$ being the support of the quotient sheaf $\mathcal{O}_{\mathcal{X}}/\mathcal{I}$. We recall that, at the local level, a *local Vogel sequence* (in the geometric sense) at $x \in V$ (for $\mathcal{I}_x$, on the germ of complex analytic space $(V_x, (\mathcal{O}_V)_x)$ is a sequence $(s_{1,x}, \ldots, s_{n-M,x})$ in the ideal $\mathcal{I}_x$ such that there is a neighborhood U_x of x in $\mathcal{X}$, together with representatives $s_{1,x}, \ldots, s_{n-M,x}$ of the germs in this neighborhood, with

$$\operatorname{codim}_V\left[\big((U_x \cap V) \setminus Z(\mathcal{I})\big) \cap \bigcap_{j=1}^{l} s_{j,x}^{-1}(0)\right] = l \quad \text{or } +\infty,\ l = 1, \ldots, n-M. \tag{15}$$

Such geometric conditions (15) (considered in the semi-global setting) are already sufficient in order to imply the following proposition.

Proposition 2 *Let* U *be an open set in* $\mathbb{C}^n$, V *be a purely* M*-codimensional closed analytic subset of* U, I *be an ideal in* $\mathcal{O}_{\mathbb{C}^n}(U)$ (*with zero set* $Z(I)$ *in* U *and generators* $\sigma_0, \ldots, \sigma_L$). *Let* $(s_1, \ldots, s_n)$ *be a sequence of elements in* I *such that*

$$\operatorname{codim}_V\left[\big(V \setminus Z(I)\big) \cap \bigcap_{j=1}^{l} s_j^{-1}(0)\right] = l \quad \text{or } +\infty,\ l = 1, \ldots, n-M. \tag{16}$$

If T is an element in $\mathrm{CH}_{U,V}(U,\mathbb{C})$, *then*

$$\bigwedge_{j=1}^{n-M} \overline{\partial}\left(\frac{1}{s_j}\right) \wedge T = T_{|Z(I)} + \sum_{l=1}^{n-M} T[s_1,\ldots,s_l]_{|Z(I)}, \tag{17}$$

where $T[s_1,\ldots,s_l]$ is defined inductively after $l-1$ operations via

$$T[s_1] = \frac{1}{(2i\pi)}\left[\overline{\partial}\left(\frac{|s_1|^{2\lambda}}{s_1}\right) \wedge T_{|U\setminus s_1^{-1}(0)}\right]_{\lambda=0},$$

$$T[s_1,\ldots,s_{l'+1}] = \frac{1}{(2i\pi)}\left[\overline{\partial}\left(\frac{|s_{l'}|^{2\lambda}}{s_{l'}}\right) \wedge T[s_1,\ldots,s_{l'}]_{|U\setminus s_{l'}^{-1}(0)}\right]_{\lambda=0}.$$

Moreover, all currents $T[s_1,\ldots,s_l]$, $l = 1,\ldots,n-M$, are $\overline{\partial}$-closed and thus are Coleff–Herrera currents. Here $T_{|Z(I)}$ means $[(1-|\sigma|^{2\lambda})T]_{\lambda=0}$, while $T_{|U\setminus Z(I)}$ means $[|\sigma|^{2\lambda}T]_{\lambda=0}$, when T is a Coleff–Herrera current.

Remark 3 Formula (17) remains valid if T is replaced by a meromorphic Coleff–Herrera current $\mathcal{T} \in \mathrm{CH}_{U,V}(U;\star S,\mathbb{C})$ for some closed hypersurface S in U such that $\overline{V\setminus S} = V$. However, the second assertion does not remain true.

Remark 4 All currents $(T[s_1,\ldots,s_l])_{|Z(I)}$ are supported, as the current $T_{|Z(I)}$, by the closed analytic set $V\cap Z(I)$. As noticed in [7], such a $(0,l+M)$-current $(T[s_1,\ldots,s_l])_{|Z(I)}$ vanishes as soon as $l+M < \mathrm{codim}(V\cap Z(I))$, which means that only the terms with index l between $\mathrm{codim}(V\cap Z(I)) - M$ and $n-M$ remain in the development (17). The first term $T_{|Z(I)}$ is of course only present if $\mathrm{codim}\,(V\cap Z(I)) = \mathrm{codim}\,V = M$.

Proof Decompose T as

$$T = \left[\left(1-|s_1|^{2\lambda}\right)T\right]_{\lambda=0} + \left[|s_1|^{2\lambda}\,T\right]_{\lambda=0} = T_{|s_1^{-1}(0)} + T_{|U\setminus s_1^{-1}(0)}. \tag{18}$$

The fact that $T_{|s_1^{-1}(0)}$ has its support included in $V\cap s_1^{-1}(0)$ implies

$$\overline{\partial}\left(\frac{1}{s_1}\right) \wedge T_{|s_1^{-1}(0)} = \left[\left(1-|s_1|^{2\lambda} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_1|^{2\lambda}}{s_1}\right)\right) \wedge T_{s_1^{-1}(0)}\right]_{\lambda=0} = T_{|s_1^{-1}(0)}.$$

Since $T_{U\setminus s_1^{-1}(0)}$ is a Coleff–Herrera current with respect to $V^{U\setminus s_1^{-1}(0)}$, we have

$$\overline{\partial}\left(\frac{1}{s_1}\right) \wedge T_{|U\setminus s_1^{-1}(0)} = \frac{1}{2i\pi}\left[\overline{\partial}\left(\frac{|s_1|^{2\lambda}}{s_1}\right) \wedge T_{U\setminus s_1^{-1}(0)}\right]_{\lambda=0} = T[s_1].$$

We now observe that, since all components in the support $V^{s_1^{-1}(0)}$ of $T_{|s_1^{-1}(0)}$ lie in $V\cap s_1^{-1}(0)$ and have codimension M as V, the geometric condition (16) for $l=2$

implies

$$\bigwedge_{j=2}^{n-M} \overline{\partial}\left(\frac{1}{s_j}\right) \wedge T_{|s_1^{-1}(0)} = T_{|s_1^{-1}(0)} = T_{|Z(I)}.$$

It remains to continue the process to compute

$$\left[\left(1 - |s_2|^{2\lambda} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_2|^{2\lambda}}{s_2}\right)\right) \wedge T[s_1]\right]_{\lambda=0}.$$

In order to do this, we decompose $T[s_1]$ as in (18):

$$T[s_1] = \left[\left(1 - |s_2|^{2\lambda}\right)T[s_1]\right]_{\lambda=0} + \left[|s_2|^{2\lambda}T[s_2]\right]_{\lambda=0} = T[s_1]_{|s_2^{-1}(0)} + T[s_1]_{|U\setminus s_2^{-1}(0)}.$$

The geometric condition (16) for $l = 3$ implies now that

$$\left[\left(1 - |s_2|^{2\lambda} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_2|^{2\lambda}}{s_2}\right)\right) \wedge T[s_1]_{|s_2^{-1}(0)}\right]_{\lambda=0} = T[s_1]_{|Z(I)}.$$

The contribution

$$\left[\left(1 - |s_2|^{2\lambda} + \frac{1}{2i\pi}\overline{\partial}\left(\frac{|s_2|^{2\lambda}}{s_2}\right)\right) \wedge T[s_1]_{|U\setminus s_2^{-1}(0)}\right]_{\lambda=0},$$

equals, as for the first step,

$$\frac{1}{(2i\pi}\left[\overline{\partial}\left(\frac{|s_2|^{2\lambda}}{s_2}\right) \wedge T[s_1]_{|U\setminus s_2^{-1}(0)}\right]_{\lambda=0} = T[s_1, s_2].$$

The procedure can be thus repeated, which leads to (17). In order to prove that all $T[s_1, \dots, s_l]$ are $\overline{\partial}$ closed, it is enough to prove it for $T[s_1]$ (then the proof goes inductively). The result for $T[s_1]$ follows immediately from the fact that

$$T[s_1] = \frac{1}{(2i\pi)}\left[\overline{\partial}\left(\frac{|s_1|^{2\mu}}{s_1}\right) \wedge \left[|s_1|^{2\lambda}T\right]_{\lambda=0}\right]_{\mu=0}, \tag{19}$$

by computing $\overline{\partial}$ of both sides in (19). □

Proposition 3 *Let U be an open set in $\mathbb{C}^n$, $M \le n$, $f_1, \dots, f_M$ be M holomorphic functions in U, and $T = T[f]$ the $(0, M)$-Coleff–Herrera current*

$$T[f] = \left(\bigwedge_{j=1}^{M} \overline{\partial}\left(\frac{1}{f_j}\right)\right)_{(0,M)}$$

(with respect to $(V[f])_{\text{ess}} = (f_1^{-1}(0) \cap \cdots \cap f_M^{-1}(0))_{\text{ess}}$). Let I be an ideal in $\mathcal{O}_{\mathcal{X}}(U)$ such that $\overline{(V[f])_{\text{ess}} \setminus Z(I)} = (V[f])_{\text{ess}}$, with generators $(\sigma_0, \dots, \sigma_L)$. If

$u^1, \ldots, u^{n-M}$, *one after each other, in this order, are generic in* $\mathbb{P}^L(\mathbb{C})$, *such that in particular* $(s_1, \ldots, s_{n-M}) := (\langle u^1, \sigma\rangle, \ldots, \langle u^{n-M}, \sigma\rangle)$ *fulfills conditions* (16) *with* $V = (V[f])_{\mathrm{ess}}$, *then the current*

$$\bigwedge_{j=1}^{n-M} \overline{\partial}\left(\frac{1}{\langle u^j, \sigma\rangle}\right) \wedge T[f] \tag{20}$$

is annihilated as a current by I^{n-M}.

Proof Before presenting a sketch of the proof, let us focus on a simple situation (to which the general case will in fact be reduced). Suppose for the moment that $T \in \mathrm{CH}_{U,V}(U, \mathbb{C})$ and that σ and h are two holomorphic functions in a neighborhood of V such that (σ, h) defines a complete intersection on V. Then, we claim that the current

$$\left[\left[\overline{\partial}\left(\frac{|\sigma h|^{2\lambda}}{\sigma h}\right) \wedge T\right]_{\lambda=0}\right]_{|\sigma^{-1}(0)} = \left[\left(1 - |\sigma|^{2\mu}\right)\left[\overline{\partial}\left(\frac{|\sigma h|^{2\lambda}}{\sigma h}\right) \wedge T\right]_{\lambda=0}\right]_{\mu=0}$$

is annihilated by σ. This follows from the fact that the map

$$(\lambda_1, \lambda_2) \in \{\mathrm{Re}\,\lambda_1 \gg 1, \mathrm{Re}\,\lambda_2 \gg 1\} \longmapsto \frac{|h|^{2\lambda_1}}{h}\, |\sigma|^{\lambda_2}\, T$$

extends as a holomorphic function of two variables to a product of half-spaces $\{\mathrm{Re}\,\lambda_1 > -\eta\} \times \{\mathrm{Re}\,\lambda_2 > -\eta\}$ for some $\eta > 0$. Since $\overline{V \setminus \sigma^{-1}(0)} = \overline{V \setminus h^{-1}(0)} = V$, it follows that

$$\left[\overline{\partial}\left(\frac{|h|^{2\lambda_1}}{h}\, |\sigma|^{2\lambda_2}\, T\right)\right]_{\lambda_1=\lambda_2=0} = \left[\overline{\partial}\left(\frac{|h|^{2\lambda}}{h}\right) \wedge T\right]_{\lambda=0} = \overline{\partial}\left(\frac{T}{h}\right),$$

which is a Coleff–Herrera current on U with respect to $V \cap \{h = 0\}$. Since σ does not vanish identically on any component the set $V \cap \{h = 0\}$, we have

$$\left[\left(1 - |\sigma|^{2\mu}\right) \overline{\partial}\left(\frac{T}{h}\right)\right]_{\mu=0} = 0,$$

which proves the claim.

Consider now the Coleff–Herrera current $T[f]$. Let $\widetilde{U} \xrightarrow{\pi} U$ be the normalized blowup of U along I, and $\mathcal{E}(I)$ be its exceptional divisor (multiplicities being taken into account). Though $\widetilde{U}$ is not smooth, one can (locally) consider an embedding $\widetilde{U} \subset \Omega \subset \mathbb{C}^N$ and treat the current $(\bigwedge_1^M \overline{\partial}(1/\pi^*[f_j]))_{(0,M)}$ as the $\overline{\partial}$-closed $(N-n, M+N-n)$ current $(\bigwedge_1^M \overline{\partial}(1/\pi^*[f_j]))_{(0,M)} \wedge [\widetilde{U}]$ in Ω. It factorizes in Ω as the product of a $(0, M+N-n)$-Coleff–Herrera current with eventual poles (with respect to some closed M-codimensional analytic subset $\widetilde{V}[f]$ of $\widetilde{U}$) times a holomorphic $(N-n, 0)$-form. Moreover, $\overline{\widetilde{V}[f] \setminus \pi^{-1}(Z(I))} = \widetilde{V}[f]$. Actually, one can approach the current $(\bigwedge_1^M \overline{\partial}(1/\pi^*[f_j]))_{(0,M)}$ by $\chi_{\|\pi^*(\sigma)\|\geq\epsilon} \bigwedge_1^M \overline{\partial}(1/\pi^*[f_j])$ as $\epsilon > 0$ tends to 0, where χ denotes a smooth cutoff function as in [12]. The exceptional divisor $\mathcal{E}(I)$ defines then a divisor $\mathcal{E}(I)_{|\widetilde{V}[f]}$ on $\widetilde{V}[f]$. Choose u^1 (generically) such that $\pi^*[s^1]_{|\widetilde{V}[f]}$ vanishes on each irreducible component of the support $\mathcal{E}(I)_{|\widetilde{V}[f]}$ exactly at the multiplicity on this component of the divisor $\mathcal{E}(I)_{|\widetilde{V}[f]}$. This means that $\pi^*[s_1]_{|\widetilde{V}[f]}$ can be expressed locally about a point $x \in \widetilde{V} \cap \pi^{-1}(Z(I))$ as $\pi^*[s_1] = \widetilde{\sigma}_{|\widetilde{V}[f]} h_1$, where $\widetilde{\sigma}$ is the generator at x for the stalk of the principal sheaf ideal $(\pi^*[\sigma_0], \dots, \pi^*[\sigma_L])$, and $\{h_1 = 0\}$ defines with $\widetilde{\sigma}$ on $\widetilde{V}[f]$ (about x) a complete intersection. Then it follows from the case studied at the beginning that the current

$$T[f][s_1]_{|Z(I)} = \left[\left[\overline{\partial}\left(\frac{|s_1|^{2\lambda}}{s_1}\right) \wedge T[f]\right]_{\lambda=0}\right]_{|Z(I)}$$

is annihilated by I. Suppose that $l \geq 2$ and that we know that $T[s_1, \dots, s_{l-1}]$ is annihilated by I^{l-1}, once $s_1, \dots, s_{l-1}$ have been conveniently chosen. Then we have

$$I^{l-1} T[f][s_1, s_2, \dots, s_{l-1}]_{|Z(I)} = 0,$$

which implies that if $h \in I^{l-1}$, the current

$$h\, T[f][s_1, s_2, \dots, s_{l-1}]_{|U\setminus Z(I)} = h\, T[f][s_1, s_2, \dots, s_{l-1}]$$

is a Coleff–Herrera current. Repeating the argument used for $l = 1$, with $T[f]$ replaced by $h\, T[f][s_1, \dots, s_{l-1}]_{|U\setminus Z(I)}$, one can choose s_l in a generic way (genericity depends here on the previous choices of $s_1, \dots, s_{l-1}$ but is independent of h), so that, for any $h \in I^{l-1}$,

$$h\, T[f][s_1, \dots, s_l]_{|Z(I)} = \left[\left[\overline{\partial}\left(\frac{|s_l|^{2\lambda}}{s_l}\right) \wedge h\, T[f, s_1, \dots, s_{l-1}]_{|U\setminus Z(I)}\right]_{\lambda=0}\right]_{|Z(I)}$$

is annihilated by I. It follows then that, with this convenient choice of s_l, the current $(T[f][s_1, \dots, s_l])_{|Z(I)}$ is annihilated by I^l. This concludes the proof. □

It is natural to call the sequence $(\langle u^1, \sigma\rangle, \dots, \langle u^{n-M}, \sigma\rangle)$ a *Vogel sequence for I with respect to the Coleff–Herrera current* $T[f]$ if such a sequence is constructed (in a generic way in terms of the u^j) through the inductive procedure described above step-by-step. The corresponding current $\bigwedge_1^{n-M} \overline{\partial}(1/\langle u^j, \sigma\rangle) \wedge T[f]$ is then called a *Vogel residue current* (*for I with respect to* $T[f]$).

Remark 5 When $[V]$ is a purely dimensional cycle in $\mathcal{X}$ (here assimilated to its associated integration current), and $\mathcal{Z} \subset \mathcal{X}$ is a complex submanifold, it is proved in [14] and reinterpreted in algebraic terms in [24] that the (multi)cycle of intersection $[V] \bullet Z$, defined via the Vogel procedure from a prescribed Vogel sequence, is such that its Chow ideal $\mathcal{I}^{\text{Chow}}([V] \bullet [\mathcal{Z}])$ (see [24] for the definition of this notion) lies in the integral closure of the sheaf of ideals $(\mathcal{I}([V]), \mathcal{I}_{\mathcal{Z}})$. Briançon–Skoda's theorem then implies that $(\mathcal{I}^{\text{Chow}}([V] \bullet [\mathcal{Z}]))^n \subset (\mathcal{I}([V]), \mathcal{I}_{\mathcal{Z}})$. Proposition 3 can thus be understood as an analogue result when $[V]$ is replaced by the Coleff–Herrera $T[f]$ and intersection theory is transposed at the (algebraic) level of residue currents instead of the (geometric) one of integration currents.

2.3 Averaging Vogel Residue Currents

Let $\mathcal{X}$ be an n-dimensional complex manifold, U be an open set in $\mathcal{X}$, $M \leq n$, and T be a $(0, M)$ E-valued Coleff–Herrera current in U (or $\mathcal{T}$ a $(0, M)$ E-valued Coleff–Herrera current in U with poles) with respect to a closed analytic subset $V \subset U$ of pure codimension M. Let $\mathcal{I}$ be an ideal sheaf in $\mathcal{O}_{\mathcal{X}}(U)$. Assume that $\sigma_0, \ldots, \sigma_L$ generate $\mathcal{I}(U)$ globally in U. One can think of $(\sigma_0, \ldots, \sigma_L)$ as a holomorphic section of the trivial bundle $E_\sigma = U \times \mathbb{C}^{L+1}$, equipped with its standard hermitian metric. It is natural to propose as an averaged current

$$\int_{(\mathbb{P}^L(\mathbb{C}))^m} \left[\left(\bigwedge_{j=1}^{m} \overline{\partial} \left(\frac{1}{\langle u^j, \sigma \rangle} \right) \right) \wedge (T \text{ or } \mathcal{T}) \right] d\omega_L\big(u^1\big) \otimes \cdots \otimes d\omega_L\big(u^m\big) \qquad (21)$$

(where $m = \min(L + 1, n - M + 1)$, ω_L denotes the Fubini metric on $\mathbb{P}^L(\mathbb{C})$, and $\langle u, \sigma \rangle = \sum_0^L u_l \sigma_l$ for $u = [u_0 : \cdots : u_L] \in \mathbb{P}^L(\mathbb{C})$), the current obtained as the value at 0 of the $\bigoplus_{r=0}^{n-M} {}'\mathcal{D}^{(0,M+r)}(\mathcal{X}, \bigwedge^r E_\sigma^* \otimes E)$ current-valued map

$$\lambda \longmapsto \left(1 - \|\sigma\|^{2\lambda} + \overline{\partial} \|\sigma\|^{2\lambda} \wedge \left(\sum_{r=1}^{\dim V} \frac{1}{(2i\pi)^r} \frac{\sigma^* \wedge (\overline{\partial} \sigma^*)^{r-1}}{\|\sigma\|^{2r}} \right) \right) \wedge (T \text{ or } \mathcal{T}). \quad (22)$$

Here $\sigma^* = \overline{\sigma} = \sum_0^L \overline{\sigma}_l \otimes e_l^*$ and $d\sigma = \sum_0^L d\sigma_l \otimes e_l$ (if $(e_0, \ldots, e_L)$ denotes a standard base of sections for $E_\sigma = U \times \mathbb{C}^{L+1}$). Because of the local structure theorems for sections of the Coleff–Herrera sheaves (see [32]), this map is holomorphic in some half-plane $\operatorname{Re} \lambda > -\eta$ for some $\eta > 0$. It was indeed proved in [8], as a consequence of Crofton's formula, that if $m = \min(L + 1, \dim V + 1)$, the averaging

$$\int_{(\mathbb{P}^L(\mathbb{C}))^m} \left[\left(\bigwedge_{j=1}^{m} \overline{\partial} \left(\frac{1}{\langle u^j, \sigma \rangle} \right) \wedge [V] \right) \wedge \bigwedge_{j=1}^{m} d\langle u^j, \sigma \rangle \right] d\omega_L\big(u^1\big) \otimes \cdots \otimes d\omega_L\big(u^m\big)$$

leads to the current

$$M_V^\sigma := \left[\left(1 - \|\sigma\|^{2\lambda} + \overline{\partial} \|\sigma\|^{2\lambda} \wedge \left(\sum_{r=1}^{\dim V} \frac{1}{(2i\pi)^r} \frac{\sigma^* \wedge (\overline{\partial} \sigma^*)^{r-1}}{\|\sigma\|^{2r}} \right) \right) \wedge [V] \right]_{\lambda=0} (d\sigma, \ldots, d\sigma). \qquad (23)$$

The action on $(d\sigma, \dots, d\sigma)$ corresponds here to the contraction operation between the $\bigwedge^r E_\sigma^*$ and E_σ as follows:

$$\left(\Phi_1^* \wedge \cdots \wedge \Phi_r^*\right)(d\sigma, \dots, d\sigma) = \frac{\bigwedge_{j=1}^r \Phi_j^*(d\sigma)}{r!}, \quad r = 1, \dots, \dim V,$$

for any Φ_j^* of E_σ^*-valued currents. The current (23) is d-closed and positive in this case. Even though it is not an integration current, the vector of the Lelong numbers $(e_0(x), \dots, e_{n-M}(x))$ of its various components (from type $(0,0)$ till type $(n-M, n-M)$), at any point $x \in V \cap Z(\mathcal{I})$, coincides with the vector of *Segre numbers* (see [22]) at x for $\mathcal{I}_x$ with respect to V, which is the minimum (with respect to lexicographic order) of all u-plets $(\deg_x \gamma_{x,0}, \dots, \deg_x \gamma_{x,n-M})$, where γ is a Vogel cycle at x, and $\gamma_{x,k}$, $k = 0, \dots, n-M$, denotes its component at x with codimension k in V (as a subcycle of V_x). Moreover, the relevant part in the Siu decomposition of the averaged current M^σ is expressed as

$$M^\sigma_{V,\text{relevant}} = \sum_{k=0}^{\dim V} \sum_\iota \beta_\iota^k [Z_\iota^k],$$

where the Z_ι^k are the *distinguished varieties* of $\mathcal{I}$ with codimension k (distinguished varieties being defined as the images of the irreducible components of the exceptional divisor $\mathcal{E}$ in the normalized blowup of U along the coherent ideal sheaf $\mathcal{I}$). Here the β_ι^k are positive integer coefficients. This generalized version of H. King's formula is proved in [8]. This shows that, at the semi global level, averaging Coleff–Herrera currents, once they are conveniently multiplied in order to become intersection currents, is an operation that fits well with improper intersection theory, such as implemented in [30, 31]. Note that the Segre numbers remain unchanged if one replaces $s = (\sigma_0, \dots, \sigma_L)$ by some u-plet $(\tilde{\sigma}_0, \dots, \tilde{\sigma}_{\tilde{L}})$ which generates in $\mathcal{O}_\mathcal{X}(U)$ an ideal with, locally about each point, the same integral closure as $\mathcal{I}$. This emphasizes the role of a current M^σ (or, to be more precise, residue currents involved in its factorization) with respect to Briançon–Skoda-type theorems.

In order to extend these ideas to the global setting, one assumes that the coherent sheaf $\mathcal{I}$ is globally generated in $\mathcal{X}$ by holomorphic sections $\sigma_0, \dots, \sigma_L$ of the same line bundle $\mathcal{O}_\mathcal{X}(\Delta)$. Referring to Example 1, the holomorphic sections of $\mathcal{O}_{\mathbb{P}^{m(n+1)-1}(\mathbb{C})}(1)$ defining the diagonal subspace $\mathcal{Z}$ in the join as a complete intersection, provide a useful illustration of such a situation, the coherent sheaf $\mathcal{I}$ being in this case the radical sheaf $\mathcal{I}_\mathcal{Z}$. Let $E_\sigma = \bigoplus^{L+1} \mathcal{O}_\mathcal{X}(\Delta)$. Choose a metric $|\ |$ on $\mathcal{O}_\mathcal{X}(\Delta)$ that induces the metric $\|\ \| = |\ | \oplus \cdots \oplus |\ |$ on E_σ. Let E be a finite-rank holomorphic bundle over $\mathcal{X}$. When T (resp. $\mathcal{T}$) is an element in $\mathrm{CH}_{\mathcal{X},V}(\mathcal{X}, E)$ (resp. in $\mathrm{CH}_{\mathcal{X},V}(\mathcal{X}; \star S, E)$), it is natural to propose, as the averaged current (21), the current obtained as the value at 0 of the $\bigoplus_{r=0}^{n-M} {}'\mathcal{D}^{(0,M+r)}(\mathcal{X}, \bigwedge^r E_\sigma^* \otimes E)$ current-valued map (22), where σ^* denotes the conjugate section of σ with respect to the metric on E_σ. If $\pi : \widetilde{V} \longrightarrow V$ denotes the normalized blowup of the complex space $(V, (\mathcal{O}_\mathcal{X})_{|V})$ along the $\mathcal{O}_V$ ideal sheaf $\mathcal{I}_{|V}$, generated locally by the holomorphic sections $x \mapsto \sigma_l(\iota(x))$ (ι being the embedding $V \subset \mathcal{X}$), $l = 0, \dots, L$, and $\mathcal{E}$ denotes the exceptional divisor in this normalized blowup, then one can check that this averaging leads, from the point of view of intersection theory (that is, when T or $\mathcal{T}$ are replaced by the (M,M)-current $[V]$) to the construction of the *Vogel current*

$$\begin{aligned}\mathrm{Vog}_{V,\sigma,|\ |} &= \Bigg[\Bigg(1-\|\sigma\|^{2\lambda}+\overline{\partial}\|\sigma\|^{2\lambda}\wedge\Bigg(\sum_{r=1}^{\dim V}\frac{1}{(2i\pi)^r}\frac{\sigma^*\wedge(\overline{\partial}\sigma^*)^{r-1}}{\|\sigma\|^{2r}}\Bigg)\Bigg)\\ &\qquad \wedge[V]\Bigg]_{\lambda=0}(d\sigma,\dots,d\sigma)\\ &= \big[V^{\mathcal{Z}(\mathcal{I})}\big]+\sum_{r=1}^{\dim V}\pi_*\Big([\mathcal{E}]\wedge\big(c^1\big(\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\otimes\pi^*[\Delta_{|V}]\big)\big)^{r-1}\Big)\\ &= \sum_{r=0}^{\dim V}\mathrm{Vog}_{(\dots);\ r}. \end{aligned} \tag{24}$$

We denote by $\Delta_{|V}$ the line bundle $\mathcal{O}_V(\Delta)$. Here $\pi^*[\sigma]=e^{[0]}\times\tau$, where τ is a holomorphic nonvanishing u-plet of sections of $\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\otimes\pi^*[\Delta_{|V}]$. The metric on $\mathcal{O}_{\widetilde{V}}(-\mathcal{E})$ is defined by $|e^{[0]}|_{\mathcal{E}}=\|\sigma\circ\pi\|$. It induces a metric on the divisor $\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\otimes\pi^*((\mathcal{O}_{\mathcal{X}}(\Delta))_{|V})$, so that

$$dd^c\log\big|e^{[0]}\big|^2_{\mathcal{E}}=[\mathcal{E}]+c^1\big(\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\big)=[\mathcal{E}]+dd^c\log\|\tau\|^2-\pi^*\big(c^1(\Delta_{|V})\big),$$

that is,

$$\begin{aligned}dd^c\log\|\tau\|^2 &= c^1\big(\pi^*[\Delta_{|V}]\big)+c^1\big(\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\big)\\ &= c^1\big(\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\otimes\pi^*[\Delta_{|V}]\big).\end{aligned}$$

The current $\mathrm{Vog}_{(\dots)}=\sum_{r=0}^{\dim V}\mathrm{Vog}_{(\dots);\ r}$ (that can also be considered as a current on $(V,\mathcal{O}_V)$) is related to the *Segre current*

$$\mathrm{Seg}_{V,\sigma,|\ |}=\big[V^{\mathcal{Z}(\mathcal{I})\cap V}\big]+\sum_{r=1}^{\dim V}\pi_*\Big([\mathcal{E}]\wedge\big(c^1\big(\mathcal{O}_{\widetilde{V}}(-\mathcal{E})\big)\big)^{r-1}\Big)=\sum_{r=0}^{\dim V}\mathrm{Seg}_{V,\sigma,|\ |;r} \tag{25}$$

thanks to the algebraic relations

$$\mathrm{Vog}_{V,\sigma,|\ |;\ r}=\sum_{l=0}^{r-1}\binom{r-1}{l}\,\mathrm{Seg}_{V,\sigma,|\ |;\ r-l}\wedge\big(c_1(\Delta_{|V})\big)^l,\quad r=1,\dots,\dim V.$$

An important particular case occurs where T is a Coleff–Herrera residue of the form $\bigwedge_{j=1}^M\overline{\partial}(1/s_j)$, $s_1,\dots,s_M$ being respective holomorphic sections of Cartier divisors Δ_j, $j = 1, \ldots, M$, on $\mathcal{X}$, such that the $s_j^{-1}(0)$ define a complete intersection. In this particular case, one can consider the $(M + L + 1)$-holomorphic bundle $E_{\Delta,\sigma} = \bigoplus_{j=1}^{M} \mathcal{O}_{\mathcal{X}}(\Delta_j) \oplus E_\sigma$, equipped with the metric $\bigoplus_1^M |\ |_j \oplus \|\ \|$ (where $|\ |_j$ denotes an hermitian metric on $\mathcal{O}_{\mathcal{X}}(\Delta_j)$, $j = 1, \ldots, M$), and propose as an alternative averaged version for all Vogel residue currents

$$\bigwedge_{j=1}^{m} \overline{\partial}\Big(\frac{1}{\langle u^j, \sigma\rangle}\Big) \wedge \bigwedge_{j=1}^{M} \overline{\partial}\Big(\frac{1}{s_j}\Big), \quad u^j \in \mathbb{P}^L(\mathbb{C}),\ 1 \leq j \leq m = \min(L+1, \dim V + 1)$$

the current obtained (still through the analytic continuation process) as the value at $\lambda = 0$ of the $\bigoplus_{r=0}^{n-M} {}'\mathcal{D}^{(0,M+r)}(\mathcal{X}, \bigwedge^r E^*_{\Delta,\sigma})$ current-valued map

$$\lambda \longmapsto 1 - \|s \oplus \sigma\|^{2\lambda} + \overline{\partial}\|s \oplus \sigma\|^{2\lambda} \wedge \left(\sum_{r=1}^{\dim V} \frac{1}{(2i\pi)^r} \frac{(s \oplus \sigma)^* \wedge (\overline{\partial}(s \oplus \sigma)^*)^{r-1}}{\|s \oplus \sigma\|^{2r}}\right), \tag{26}$$

where $s \oplus \sigma = s_1 \oplus \cdots \oplus s_M \oplus \sigma$, and $(s \oplus \sigma)^*$ denotes its conjugate section with respect to the metric which has been chosen on $E_{\Delta,\sigma}$. The two currents, which belong to $\bigoplus_{r=0}^{n-M} {}'\mathcal{D}^{(0,M+r)}(\mathcal{X}, \bigwedge^r E^*_{\Delta,\sigma})$,

$$\left[\left(1 - \|\sigma\|^{2\lambda} + \overline{\partial}\|\sigma\|^{2\lambda} \wedge \left(\sum_{r=1}^{\dim V} \frac{1}{(2i\pi)^r} \frac{\sigma^* \wedge (\overline{\partial}\sigma^*)^{r-1}}{\|\sigma\|^{2r}}\right)\right) \wedge \bigwedge_{j=1}^{M} \overline{\partial}(1/s_j)\right]_{\lambda=0}$$

and

$$\left[1 - \|s \oplus \sigma\|^{2\lambda} + \overline{\partial}\|s \oplus \sigma\|^{2\lambda} \wedge \left(\sum_{r=1}^{\dim V} \frac{1}{(2i\pi)^r} \frac{(s \oplus \sigma)^* \wedge (\overline{\partial}(s \oplus \sigma)^*)^{r-1}}{\|s \oplus \sigma\|^{2r}}\right)\right]_{\lambda=0}, \tag{27}$$

coincide when $\operatorname{codim}(s^{-1}(0) \cap \sigma^{-1}(0)) = M + L + 1$ (see [34]) but differ in general. They can both be used (see [3, 7]) to materialize the residual obstruction for the exactness of the generically exact Koszul complex

$$\begin{aligned} 0 \to \bigwedge^{M+L+1} E^*_{\Delta,\sigma} &\xrightarrow{\rfloor_{s\oplus\sigma}} \bigwedge^{M+L} E^*_{\Delta,\sigma} \longrightarrow \cdots \\ &\xrightarrow{\rfloor_{s\oplus\sigma}} \bigwedge^{l+1} E^*_{\Delta,\sigma} \xrightarrow{\rfloor_{s\oplus\sigma}} \bigwedge^{l} E^*_{\Delta,\sigma} \longrightarrow \cdots \xrightarrow{\rfloor_{s\oplus\sigma}} E^*_{\Delta,\sigma} \xrightarrow{\rfloor_{s\oplus\sigma}} \mathcal{X} \times \mathbb{C}, \end{aligned} \tag{28}$$

since they are both annihilated (as currents) by the operator $2i\pi \rfloor_{s\oplus\sigma} - \overline{\partial}$, where $\rfloor_{s\oplus\sigma}$ denotes the interior multiplication by the holomorphic section $s \oplus \sigma$ of $E_{\Delta,\sigma}$.

3 About a Result by M. Hickel, M. Andersson, and E. Götmark

In order to emphasize what averaged Vogel residue currents introduced in Sect. 2 could be useful for, and, in parallel, to illustrate how far using their use reveals to be successful, let us focus on the effective geometric global formulation of Briançon–Skoda's theorem in $\mathbb{C}[X_1, \dots, X_\mathfrak{n}]$, as obtained first by Hickel [20] and then reformulated (and thus reproved), using the frame developed above, by Andersson and Götmark [5]. Before stating the result, one needs to recall basic facts about Lojasiewicz exponents at infinity in $\mathbb{C}^\mathfrak{n} \subset \mathbb{P}^\mathfrak{n}(\mathbb{C}) = \mathbb{C}^\mathfrak{n} \cup \{z_0 = 0\}$. Let $p_1, \dots, p_\mathfrak{m}$ be $\mathfrak{m}$ polynomials in $\mathbb{C}[X_1, \dots, X_\mathfrak{n}]$, together with their homogenizations $P_1, \dots, P_\mathfrak{m}$, considered as respective holomorphic sections of the Cartier divisors $\mathcal{O}_{\mathbb{P}^\mathfrak{n}(\mathbb{C})}(\deg p_j)$. Let $\mathcal{E} = \bigcup_\iota \mathcal{E}_\iota$ be the (reduced) exceptional divisor of the normalized blowup $\widetilde{\mathbb{P}}^\mathfrak{n}(\mathbb{C}) \xrightarrow{\pi} \mathbb{P}^\mathfrak{n}(\mathbb{C})$ of $\mathbb{P}^\mathfrak{n}(\mathbb{C})$ along the coherent ideal sheaf $\mathcal{I}(P)$, generated locally by the holomorphic sections $P_1, \dots, P_\mathfrak{m}$. Let $\mu_\iota(P)$ be the multiplicity of $\mathcal{I}(P) \cdot \mathcal{O}_{\widetilde{\mathbb{P}}^\mathfrak{n}(\mathbb{C})}$, and $\mu_\iota(z_0)$ be the multiplicity of $\mathcal{I}_{\{z_0=0\}} \cdot \mathcal{O}_{\widetilde{\mathbb{P}}^\mathfrak{n}(\mathbb{C})}$ along the same component. The Lojaciewicz exponent $\nu_\infty(P)$ is defined as

$$\nu_\infty(P) = \sup_\iota \left(\frac{\mu_\iota(P)}{\mu_\iota(z_0)} \right).$$

In the special case where $\mathfrak{m} = \mathfrak{n}$ and $\dim_{\mathbb{C}^\mathfrak{n}} p^{-1}(0) \le 0$, one has the Lojasiewicz inequality at infinity in $\mathbb{C}^\mathfrak{n} \subset \mathbb{P}^\mathfrak{n}(\mathbb{C})$, namely

$$\sum_{j=1}^{\mathfrak{n}} \frac{|p_j(\zeta)|}{|\zeta|^{\deg p_j}} \ge \frac{\kappa}{|\zeta|^{\nu_\infty(P)}}, \quad |\zeta| \gg 1, \tag{29}$$

for some strictly positive constant κ. In particular, for p to be in this case a proper map from $\mathbb{C}^\mathfrak{n}$ into $\mathbb{C}^\mathfrak{n}$, it is enough that $\nu_\infty(P) < \deg p_j$, $j = 1, \dots, \mathfrak{m}$. Notice that, when $\mathfrak{m} = \mathfrak{n}$ and $\dim_{\mathbb{C}^\mathfrak{n}} p^{-1}(0) \le 0$, the current

$$\left[\bigwedge_1^{\mathfrak{n}} \bar\partial \left(\frac{1}{P_j} \right) \right]_{|\{z_0=0\}}$$

is annihilated by $z_0^{\mathfrak{n}\nu_\infty(P)}$. This can be seen using (as in the proof of Proposition 3) the normalized blowup of $\mathbb{P}^\mathfrak{n}(\mathbb{C})$ along the coherent ideal sheaf $\mathcal{I}_{\{z_0=0\}}$. On the other hand, the current

$$\left[\bigwedge_1^{\mathfrak{n}} \bar\partial \left(\frac{1}{P_j} \right) \right]_{|\mathbb{C}^n}$$

is annihilated by any homogeneous polynomial of the form Q such that the map $\zeta \mapsto Q(1, \zeta)/\|p(\zeta)\|^\mathfrak{n}$ is locally bounded in $\mathbb{C}^\mathfrak{n}$ (see [5]).

M. Hickel's result ([20], revisited with the analytic methods presented above in [5]) can be stated as follows (we just mention here the result when $\mathfrak{m} \le \mathfrak{n}$).

Theorem 4 ([5, 20]) *Assume that $\mathfrak{m} \le \mathfrak{n}$ and let $p_1, \dots, p_\mathfrak{m}$ be $\mathfrak{m}$ polynomials in $\mathbb{C}[X_1, \dots, X_\mathfrak{n}]$. Let q be a polynomial in $\mathbb{C}[X_1, \dots, X_\mathfrak{n}]$ such that the function $|q|/|p|^\mathfrak{m}$ is locally bounded in $\mathbb{C}^\mathfrak{n}$. Then, there exist polynomials $a_1, \dots, a_\mathfrak{m} \in \mathbb{C}[X_1, \dots, X_\mathfrak{n}]$ such that $q \equiv \sum_{j=1}^{\mathfrak{m}} a_j p_j$ with $\deg a_j p_j \le \deg q + [\mathfrak{m}\nu_\infty(p)] + 1$, where $[\gamma]$ denotes the integer part of the rational number γ.*

As a consequence of this theorem, it appears that the effective realization of the global Briançon–Skoda theorem (q being in the ideal $(p_1,\dots,p_{\mathfrak{m}})$ in $\mathbb{C}[X_1,\dots,X_{\mathfrak{n}}]$ if the germ q_ζ, for each $\zeta\in p^{-1}(0)$, lies in the $\mathfrak{m}$-power of the integral closure of the ideal $(p_{1,\zeta},\dots,p_{,\zeta})$ in the local ring $\mathcal{O}_{\mathbb{C}^{\mathfrak{n}},\zeta}$) can be achieved with degree estimates for the quotients a_j controlled by the geometric Bézout theorem, which means that the effectivity of the problem is governed by geometric intersection theory.

The proof of Theorem 4 relies on the use of the averaged Bochner–Martinelli version (22) ($E_\sigma=\bigoplus_1^{\mathfrak{m}}\mathcal{O}_{\mathbb{P}^{\mathfrak{n}}(\mathbb{C})}(d_j)$ with the Fubini–Study metric on each component, $\sigma=P$, $V=\mathbb{P}^{\mathfrak{n}}(\mathbb{C})=\mathcal{X}$), of all currents

$$\bigwedge_1^m \overline{\partial}\Big(\frac{1}{\langle u^j,P_j\rangle}\Big),$$

$u^1,\dots,u^m\in\mathbb{P}^{\mathfrak{m}-1}(\mathbb{C})$, $m=\min(\dim(\{P=0\})+1,\mathfrak{m})$, with respect to the tensorized Fubini–Study metric on $\mathbb{P}^{\mathfrak{m}-1}(\mathbb{C})\otimes\cdots\otimes\mathbb{P}^{\mathfrak{m}-1}(\mathbb{C})$ (m times). As pointed out in [24], when dealing with intersection problems involving arithmetic aspects, it is more adequate to use a specific Vogel cycle of intersection instead of what could be understood as an averaged version, for example, the intersection cycle constructed in [31] and reinterpreted in [8] in the currential setting. The methods presented here (end of Sects. 1 and 2.2) aim precisely to give some support to the following conjecture:

Conjecture 1 (Global arithmetic Briançon–Skoda) Let $p_1,\dots,p_{\mathfrak{m}}$ be $\mathfrak{m}$ polynomials in $\mathbb{Z}[X_1,\dots,X_{\mathfrak{n}}]$ with degrees bounded by d and logarithmic sizes bounded by h. Let $q\in\mathbb{Z}[X_1,\dots,X_{\mathfrak{n}}]$ be such that the function $q/|p|^{\min(\mathfrak{m},\mathfrak{n})}$ is locally bounded in $\mathbb{C}^{\mathfrak{n}}$. Then, one can find $\delta\in\mathbb{N}^*$ and $a_1,\dots,a_{\mathfrak{m}}\in\mathbb{Z}[X_1,\dots,X_{\mathfrak{n}}]$ such that

$$\begin{aligned}&\delta\, q^{\kappa(\mathfrak{n})}=\sum_{j=1}^{\mathfrak{m}}a_j p_j,\qquad \max_j\big(\deg(a_jp_j)\big)\le\kappa(\mathfrak{n})\deg q+\kappa_0(\mathfrak{n})\,d^{\gamma_0\min(\mathfrak{n},\mathfrak{m})},\\&\max_j\big(h(\delta),h(a_jp_j)\big)\le\kappa(\mathfrak{n})\,h(q)+\kappa_1(\mathfrak{n})\,(h+\log\mathfrak{m})\,d^{\gamma_1\min(\mathfrak{n}+1,\mathfrak{m})},\end{aligned}\tag{30}$$

where $\kappa(\mathfrak{n}),\kappa_0(\mathfrak{n}),\kappa_1(\mathfrak{n})$ are numerical constants depending only on the number $\mathfrak{n}$ of variables, γ_0,γ_1 being universal constants (idealistically $\kappa(\mathfrak{n})=1$, $\kappa_0(\mathfrak{n})=\mathfrak{n}$, $\gamma_0=\gamma_1=1$).

This result was obtained by Elkadi [16] when $\dim(p^{-1}(0))=0$. On the other hand, the arithmetic membership problem can be solved with such bounds (30) when $\mathfrak{m}\le\mathfrak{n}$ and $(p_1,\dots,p_{\mathfrak{m}})$ defines a complete intersection in $\mathbb{C}^{\mathfrak{n}}$ (see [17]), which includes in particular the case of the arithmetic nullstellensatz [10, 15, 25]. This fits well with the ideas that govern the construction of Vogel sequences. Nevertheless, it is known now that methods based on multidimensional residue calculus (relying essentially on Cauchy–Weil integral formula and associated Bergman–Weil developments) do not provide the sharpest bounds for the arithmetic nullstellensatz, which is truly a problem related to arithmetic intersection theory (see the recent approach in [15], based on a arithmetic version of O. Perron's theorem used for the algebraic nullstellensatz in [21]). As a consequence, this makes more clear that such methods were in fact more in the spirit of Conjecture 1. Using the Stückrad–Vogel approach [30], combined with a precise analysis of the Vogel residue currents involved (for a description of their annihilators and explicit computations of the restrictions to $\mathbb{C}^n$ of the auxiliary Coleff–Herrera currents involved in their expansion, see Propositions 2 and 3 above, together with the concluding comments in Sect. 1), seems to be a natural way to tackle such a conjecture. It is indeed necessary to overcome the difficulty which is inherent to the fact that averaging such Vogel residue currents in order to get suitable Bochner–Martinelli currents (for control of the degree in effectivity questions) does not preserve the arithmetic structure of the data (which would be necessary in order to get in parallel control on the heights). It seems also opportune to mention that the initial approach to Theorem 4 by Amoroso [2] relies on the Northcott–Rees notion of *superficial elements* in ideals that is also present in the construction of Vogel sequences (more specifically of *filtered sequences*, see [1]).

This presentation of the Coleff–Herrera machinery, combined in Sect. 2.2 with that of the Stückrad–Vogel approach [8, 30, 31], transposed to the context of residue currents instead of integration currents on cycles, intends to be a modest invitation toward such an approach to effectivity questions in arithmetic polynomial geometry, when they require operational tools related to duality, such as Briançon–Skoda's theorem or multidimensional operational residue calculus.

References

1. Achilles, R., Rams, S.: Intersection numbers, Segre numbers and generalized Samuel multiplicities. Arch. Math. **77**, 391–398 (2001)
2. Amoroso, F.: On a conjecture of C. Berenstein and A. Yger. In: Gonzalez Vega, L., Recio, T. (eds.) Algorithms in Algebraic Geometry and Applications (Santander 1994). Progress in Mathematics, vol. 143, pp. 17–28. Birkhäuser, Basel (1996)
3. Andersson, M.: Residue currents and ideals of holomorphic functions. Bull. Sci. Math. **128**(6), 481–512 (2004)
4. Andersson, M.: Coleff–Herrera currents, duality, and Noetherian operators. Bull. Soc. Math. Fr. **139**(4), 535–554 (2011)
5. Andersson, M., Götmark, E.: Explicit representation of membership in polynomial ideals. Math. Ann. **349**(2), 345–365 (2011)
6. Andersson, M., Wulcan, E.: Residue currents with prescribed annihilator ideals. Ann. Sci. Ec. Norm. Super. **40**, 985–1007 (2007)
7. Andersson, M., Wulcan, E.: Decomposition of residue currents. J. Reine Angew. Math. **638**, 103–118 (2010)
8. Andersson, M., Samuelsson, H., Wulcan, E., Yger, A.: Nonproper intersection theory and positive currents I, local aspects. Preprint (2010). arXiv:1009.2458
9. Berenstein, C.A., Gay, R., Vidras, A., Yger, A.: Residue Currents and Bézout Identities. Progress in Mathematics, vol. 114. Birkhäuser, Basel (1993)

10. Berenstein, C.A., Yger, A.: Residue calculus and effective Nullstellensatz. Am. J. Math. **121**(4), 723–796 (1999)
11. Björk, J.E.: Residues and $\mathcal{D}$-modules. In: Laudal, O.A., Piene, R. (eds.) The Legacy of Niels Henrik Abel, pp. 605–651. Springer, Berlin (2004)
12. Björk, J.E., Samuelsson, H.: Regularizations of residue currents. J. Reine Angew. Math. **649**, 33–54 (2010)
13. Coleff, N., Herrera, M.: Les courants résiduels associés à une forme méromorphe. Lecture Notes in Math., vol. 633. Springer, Berlin (1978)
14. Cygan, E.: Intersection theory and separation exponent in complex analytic geometry. Ann. Pol. Math. **69**(3), 287–299 (1998)
15. D'Andrea, C., Krick, T., Sombra, M.: Heights of varieties in multiprojective spaces and arithmetic nullstellensatz. Preprint (2011). arXiv:1103.4561
16. Elkadi, M.: Une version effective du théorème de Briançon–Skoda dans le cas algébrique discret. Acta Arith. **66**, 201–220 (1994)
17. Elkadi, M.: Bornes pour les degrés et les hauteurs dans le problème de division. Mich. Math. J. **40**, 609–618 (1993)
18. Elkadi, M., Yger, A.: Residue calculus and applications. Publ. Res. Inst. Math. Sci. **43**(1), 55–73 (2007)
19. Flenner, H., Carroll, L.O., Vogel, W.: Joins and Intersections. Springer Monographs in Mathematics. Springer, Berlin (1999)
20. Hickel, M.: Solution d'une conjecture de C. Berenstein et A. Yger et invariants de contact à l'infini. Ann. Inst. Fourier (Grenoble) **51**(3), 707–744 (2001)
21. Jelonek, Z.: On the effective Nullstellensatz. Invent. Math. **162**(1), 1–17 (2005)
22. Gaffney, T., Gassler, R.: Segre numbers and hypersurface singularities. J. Algebr. Geom. **8**, 695–736 (1999)
23. Griffiths, P., Harris, J.: Principles of Algebraic Geometry. Wiley-Interscience, New York (1978)
24. Kollár, J.: Effective Nullstellensatz for arbitrary ideals. J. Eur. Math. Soc. **1**(3), 313–337 (1999)
25. Krick, T., Pardo, L.M., Sombra, M.: Sharp estimates for the arithmetic Nullstellensatz. Duke Math. J. **109**(3), 521–598 (2001)
26. Massey, D.B.: Numerical control over Complex Analytic Singularities. Memoirs of the American Mathematical Society, vol. 163, no. 778 (2003)
27. Passare, M., Tsikh, A., Yger, A.: Residue currents of the Bochner–Martinelli type. Publ. Mat. **44**, 85–117 (2000)
28. Samuelsson, H.: Analytic continuation of residue currents. Ark. Mat. **47**(1), 127–141 (2009)
29. Solomin, J.E.: Le résidu logarithmique dans les intersections non complètes. C. R. Acad. Sci. Paris Sér. A–B **284**(17), A1061–A1064 (1977)
30. Stückrad, J., Vogel, W.: An algebraic approach to the intersection theory. Queen Pap. Pure Appl. Math. **61**, 1–32 (1982)
31. Tworzewski, P.: Intersection theory in complex analytic geometry. Ann. Pol. Math. **62**, 177–191 (1995)
32. Vidras, A., Yger, A.: Coleff–Herrera currents revisited. In: Sabadini, I., Struppa, D.C. (eds.) The Mathematical Legacy of Leon Ehrenpreis, 1930–2010. Springer, Berlin (2011)
33. Wiebe, H.: Über homologische Invarianten lokaler Ringe. Math. Ann. **179**, 257–274 (1969)
34. Wulcan, E.: Products of residue currents of the Cauchy–Fantappié–Leray type. Ark. Math. **45**, 157–178 (2007)

ERRATUM

Erratum to: Analyticity on Curves

Mark Agranovsky and Lawrence Zalcman

Erratum to: I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16, pp. 36–37
DOI 10.1007/978-88-470-1947-8_4,
© Springer-Verlag Italia 2012

In the original version of this chapter some notations were not correct. The correct notations are given below

1. On page 36, lines 7, 8 from the top should read "… there exists a (smooth) 2-chain $L \subset \Phi(\overline{\Delta} \times S^1)$, whose boundary lies outside of $\Phi(\Delta \times S^1)$, which intersects …".
2. On page 37, line 6 from the top should read "Let L dual 2-cocycle …" (i.e. 2-cocycle instead of 1-cocycle).
3. On page 37, line 8 from the top should read "chains L" (i.e. "chains" instead of "curves").

The online version of the original chapter can be found under doi: 10.1007/978-88-470-1947-8_4.

M. Agranovsky (✉) · L. Zalcman
Bar-Ilan University, Ramat-Gan 52900, Israel
e-mail: agranovs@macs.biu.ac.il

L. Zalcman
e-mail: zalcman@macs.biu.ac.il

A Letter of Leon Ehrenpreis to Bernard Malgrange

The letter whose scan appears in the next page was written by Leon Ehrenpreis to Bernard Malgrange in June 1960 during the period to which Professor Malgrange makes reference in his introductory article. The letter is now in possession of Professor Struppa, through the kind generosity of Professor Malgrange, and is enclosed in this volume, as it offers a direct insight into Ehrenpreis' style and thinking.

The editors

I. Sabadini, D.C. Struppa (eds.), *The Mathematical Legacy of Leon Ehrenpreis*,
Springer Proceedings in Mathematics 16,
DOI 10.1007/978-88-470-1947-8, © Springer-Verlag Italia 2012

ב"ה
Blessed be the Lord

(1)

~ Juin 1960

Rome

Dear Malgrange,

I guess I didn't understand correctly your problem: Is $\delta + \mu = S$ invertible if support μ does not meet the origin? The following more general result is quite simple: Let $S = T + \mu$, T invertible, support $T \cap$ support $\mu = \phi$. Then S is invertible.

Proof. Use my criterion that we only have to verify: $f \to 0$ in $\mathcal{D}$ if $S * f \to 0$ in $\mathcal{D}$, support f small enough; the proof is clear since support $T * f \cap$ support $\mu * f = \phi$.

Actually, it is of interest to prove directly that $\hat{S}$ is slowly decreasing; this is not trivial! In case $n = 1$, the proof is as follows (using only the fact that $\hat{T}$ is slowly decreasing): Cover support T with intervals I_j such that for each j, k we have

$$\text{distance } (\text{support } \mu, I_j) \geq 8\epsilon$$

$$\text{distance } [I_j, I_k] \geq 8\epsilon.$$

Then Construct
$$\begin{cases} f^j_M = 1 \text{ on } I_j \\ f^j_M(x) \equiv 0 \quad \text{if distance } [x, I_j] \geq \epsilon. \end{cases}$$

also
$$\int_{|y| \geq M} |F^j_M(y)|\, dy \leq \epsilon_1^{-M}$$

Where ϵ_1 is some constant; F^j_M is the Fourier transform of f^j_M.

3

Now, in Division IV I show we may assume $\hat{S}, \hat{T}, \hat{\mu}$ bounded (say by 1) on R. Then we have

$$T = \sum_j f_M^j S$$

$$\hat{T} = \sum_j F_M^j * \hat{S}$$

$$\hat{T}(y_0) = \sum \int f_M^j (y_0 - y)\, \hat{S}(y)\, dy$$

$$= \sum \int_{|y_0 - y| \leq M} + \int_{|y_0 - y| \geq M}$$

The proof now follows easily by reductio ad absurdum, choosing $M \sim$ ~~const~~ $c \log(1 + |y_0|)$ for suitable c.

To construct f_M^j, ~~we start with the Fourier transform of~~ roughly: ~~let G~~ (assume there is only ~~on~~ one I_j; $I = [-\ell, \ell]$. let $G_M(y) = \left[\frac{\sin \epsilon M y}{\epsilon M y}\right]^M$

call $h_M(x) = \int_{-\infty}^{x} g_M(x)$ and set

$$f_M(x) = h_M(x + \ell)\, h_M(-x + \ell).$$

The case $n > 1$ is more complicated. One needs a suitable partition of unity by products of these f_M and I have not written the details in all cases.

Please tell me if I have done something silly.

I have thought a little about functions analytic at the origin and it seems certain that the fundamental principle works. Also, your conjecture about $\| \mu * f \|_{L^2} \geq c_K \| f \|_{L^2}$ for $f \in \mathcal{D}_K$, $\mu \in \mathcal{E}'$ may not be true, though I have no counter example

(2)

4

and one is surely very hard to find.

It seems fairly certain that if you combine my ideas on the fundamental principle with the Malgrange-Lojasiewicz method, you can get very good ideas for real analytic manifolds.

Sincerely,

Leon Ehrenpreis

P. S. In case you have anything of interest, my address for about 6+ weeks is

c/o Neuberg
7 Shilo St.
Tel-Aviv, ISREAL