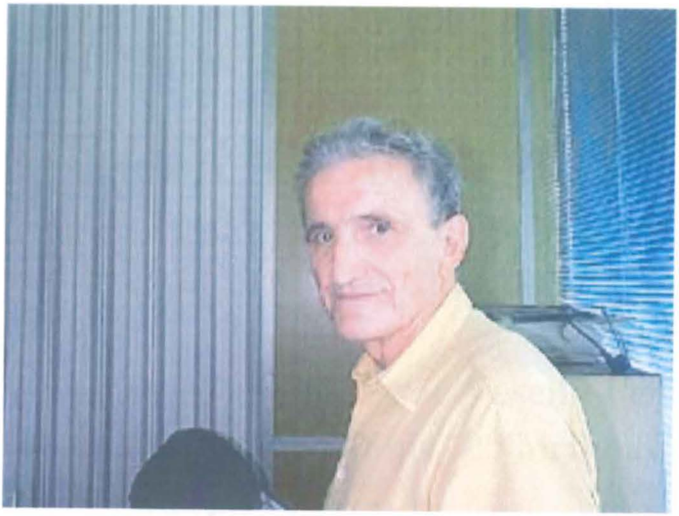

Veselin Perić

SOCIETY OF MATHEMATICIANS OF BANJA LUKA

PROCEEDINGS

OF MATHEMATICAL CONFERENCE BANJA LUKA, DECEMBER 2000

Dedicated to Professor Veselin Perić
on the occasion of his 70^{th} birthday

CONTENTS

BULLETIN OF SOCIETY OF MATHEMATICIANS OF BANJA LUKA

Faculty of Natural Sciences and Mathematics,
Department of Mathematics and Informatics
Banja Luka, 2 Mladen Stojanovic Street, Bosnia and Herzegovina

"**Bulletin of Society of Mathematicians of Banja Luka**" is edited and published by Society of Mathematicians of Banja Luka and the Department of Mathematics and Informatics of Banja Luka University.

"**Bulletin of Society of mathematicians of Banja Luka**" publishes original papers in all fields of mathematics that contain new, substantial and significant results with complete proofs, papers of expository nature (survey papers) and papers of polemical interest. Manuscript should be written in English and published in English.

Manuscripts and correspondences should be send to the following address: **Bulletin**, Society of Mathematicians, Faculty of Natural Sciences and Mathematics, 78000 Banja Luka, 2 Mladen Stojanovic Street, B osnia and Herzegovina

Bull. Soc. Math. Banja Luka **ISSN 0354-5792**

Na osnovu misljenja Ministarstva obrazovanja, nauke i kulture Republike Srpske, broj 02-11711/95 od 17.02.1995. godine ovaj casopis je naucni casopis u smislu clana 17, Stav 1, tacka .11 Zakona o porezu na promet Republike Srpske.

ON RINGS WITH SPECIAL AUTOMORPHISMS, RESP. DERIVATIONS *

VESELIN PERIĆ

ABSTRACT. We give here a survey of some recent results concerning the rings with special automorphisms, resp. derivations.

1. DERIVATION d WITH $d(x)$ ZERO OR INVERTIBLE

Jefrey Bergen, I. N. Herstein and Charles Lanski considered in [4] the following problem:

Let R be a ring with identity 1, and d a derivation on R such that, for every $x \in R$, $d(x)$ is zero or invertible. Does R have a special structure?

They answered this question in the following manner:

Theorem 1.1. *Let R be a ring with identity 1, and d a nonzero derivation on R such that, for all $x \in R$, $d(x)$ is zero or invertible. Then R is*

1. a division ring D or

2. the ring $\mathrm{Mat}(D,2)$ of all matrices of order 2 over a division ring D or

3. the ring $D[x]/(x^2)$, where D is a division ring of characteristic $\mathrm{char}\, D = 2$, $d(D) = 0$ and $d(x) = 1 + ax$ for some $a \in Z$, the center of D.

Moreover, if $2R \neq 0$, then the case $R = \mathrm{Mat}(D,2)$ is possible if and only if D contains all quadratic extensions of Z, i.e. if at least one element of Z is not a square in D.

From the proof of this theorem it follows that, in the case $R = \mathrm{Mat}(D,2)$, derivation d is an inner if $2R \neq 0$, but d need not to be inner if $2R = 0$. Moreover, one see that d cannot be inner in the case $R = D[x]/(x^2)$. We recall that a derivation d on R is an inner, if, for some $a \in R$, $d(x) = [a,x] = ax - xa$ for all $x \in R$.

The authors consider also the case where $d(x)$ is zero or invertible not for all $x \in R$, but for all x from a suitable subset of R. In this way they prove

Theorem 1.2. *Let R be a ring with identity 1, and d a derivation on R such that $d(L) \neq \{0\}$ for some left ideal L of R, and for every $x \in L$, $d(x)$ is zero or invertible. Then, for some division ring D, $R = D$ or $R = \mathrm{Mat}(D,2)$ or $R = D[x]/(x^2)$, where $2R = \{0\}$.*

The authors remark that, for $R = D[x]/(x^2)$, the assumption about d on L, cannot imply any property of d on R. In the case $R = \mathrm{Mat}(D,2)$ they prove than there must not exist a nonzero derivation δ on R such that $\delta(x)$ is zero or invertible for every $x \in R$.

2000 *Mathematics Subject Classification.* Primary: 15W220, 16W25. Secondary: 16N60.

Key words and phrases. Prime rings, centralizing derivation, centralizing automorphism, nilpotent derivation, commutativity.

2. Automorphism ϕ with $\phi(x)$ zero or invertible

If ϕ is an automorphism of a ring R with identity 1, then the mapping δ of R with $\delta(x) = x - \phi(x)$ for all $x \in R$ need not to be a derivation on R, but δ is a special pseudo-derivation on R. We recall that an additive mapping $f : R \to R$ is a pseudo-derivation on R if there exists a function $g : R \to R$ such that

(1) $f(xy) = f(x)g(y) + xf(y) = f(x)y + g(x)f(y)$ $(x, y \in R)$;

(2) $f(g(x)) = g(f(x))$ $(x \in R)$.

It is easy to see, that (1) and (2) are satisfied for $f = \delta$ and $g = \mathrm{id}_R$, the identity mapping of R.

If ϕ is proper, i. e. $\phi \neq \mathrm{id}_R$, then $\delta \neq 0$.

Jefrey Bergen and I. N. Herstein [3] investigate the structure of a ring R with the identity 1 and a automorphism $\phi \neq \mathrm{id}_R$ such that $x - \phi(x)$ is zero or invertible for all $x \in R$, and they prove:

Theorem 2.1. *Let R be a ring with identity 1, and $\phi \neq \mathrm{id}_R$ an automorphism of R such that, for all $x \in R$, $x - \phi(x)$ is zero or invertible. Then*

1. $R = D$ or

2. $R = D \oplus D$ or

3. $R = \mathrm{Mat}(D, 2)$

for some division ring D.

Moreover, the case $R = \mathrm{Mat}(D, 2)$ is possible, for a non-inner automorphism ϕ, if and only if D has a non-inner automorphism ψ, such that $\psi^2(x) = u^{-1}xu$ for all $x \in D$, where $\psi(u) = u$ and $u \neq y\psi(y)$ for all $y \in D$, or, for an inner automorphism ϕ, if and only if D does not contain all square extensions of its own center Z.

The authors remark that, in the case $\mathrm{char}\, R \neq 2$, D does not contain all square extensions of its own center Z, if and only if some $\alpha \in Z$ is not a square in D. In this case, for the automorphism $\psi = \mathrm{id}_D$ surely $\psi^2(x) = x = \alpha^{-1}x\alpha$ and $\alpha \neq y\psi(y)$ for all $y \in D$. Hence, in this case, there is no difference between an inner automorphism ψ and an automorphism ψ which is non-inner. Therefore, for $\mathrm{char}\, R \neq 2$, the above theorem becomes:

If $\mathrm{char}\, R \neq 2$, and R has an automorphism $\phi \neq \mathrm{id}_R$ such that $x - \phi(x)$ is zero or invertible for all $x \in R$, then $R = D$ or $R = D \oplus D$ or $R = \mathrm{Mat}(D, 2)$ for some division ring D; moreover, the case $R = \mathrm{Mat}(D, 2)$ is possible if and only if D has an automorphism ψ for which $\psi^2(x) = u^{-1}xu$, $\psi(u) = u$ and $u \neq y\psi(y)$ for all $y \in D$.

As in [4] for a nonzero derivation d, the authors in [3] consider the problem with $x - \phi(x)$ is zero or invertible not for all $x \in R$, but for all $x \in L$, a left ideal of R. As in the foregoing theorem, they prove that $R = D$ or $R = \mathrm{Mat}(D, 2)$ or $R = D \oplus D$ for some division ring D, but, for this problem, they do not prove the above conditions on D.

3. Centralizing derivations and automorphisms in prim rings

Edward C. Posner [10] considers derivation on prime rings and proves the following two theorems:

Theorem 3.1. *Let R be a prime ring with $\mathrm{char}\, R \neq 2$, and let d_1, d_2 be derivations on R for which the product $d_1 \circ d_2$ is also a derivation. Then one of derivations d_1, d_2 is equal to zero.*

Theorem 3.2. *Let R be a prime ring with centralizing derivation d. Then $d = 0$ or R is commutative.*

We recall that a ring R is a prime ring, if for all $a, b \in R$, $aRb = \{0\}$ implies $a = 0$ or $b = 0$. Thus, a ring R is a prime ring, if for any right ideal A and any left ideal B of R, $AB = \{0\}$ implies $A = \{0\}$ or $B = \{0\}$.

For a function $f : R \to R$ we say to be a centralizing function, if for any $x \in R$, $[x, f(x)] = xf(x) - f(x)x \in Z$, the center of R.

The first theorem is only used in the proof of the second for the case $\operatorname{char} R \neq 2$, and gives no information about R.

Josef H. Mayne [9] inside of a centralizing derivation d considers a centralizing automorphism ϕ of a prime ring R and proves an analogous result:

Theorem 3.3. *If R is a prime ring with a proper centralizing automorphism ϕ, then R is a commutative domain.*

4. Derivations with some commuting properties

I. N. Herstein [6] and Amos Kovacs [7] investigate the relation between a prime ring R and the subset $d(R) = \{d(x) : x \in R\}$, where d is a derivation on R.

In the cited paper Herstein proves the following theorem:

Theorem 4.1. *If for some derivation $d \neq 0$ on a prime ring R, $[d(x), d(y)] = 0$ for all $x, y \in R$, then R is commutative or R is a order in a simple algebra of characteristic 2, which has the dimension 4 as a vector space over the center of the algebra.*

Moreover, Herstein asks the question:

If $d \neq 0$ and if for the standard identity $s_k[x_1, x_2, \ldots, x_k]$, $s_k[d(x_1), d(x_2), \ldots, d(x_k)] = 0$ for all $x_1, x_2, \ldots, x_k \in R$, do we then conclude that R is of some special structure, or does R perhaps satisfy the identity s_k?

In the cited paper, Kovacs answers this question by giving examples which shows:

(a) For any prime number p, there is a prime ring R of the characteristic p with a derivation $d \neq 0$ satisfying $s_{4p+1}[d(x_1), d(x_2), \ldots, d(x_{4p+1})] = 0$ for all $x_1, x_2, \ldots, x_{4p+1} \in R$, such that R satisfies no polynomial identity.

(b) There is a prime ring R of the characteristic 0 with a derivation $d \neq 0$ satisfying $[d(x_1)d(x_2), d(x_3)d(x_4)]d(x_5)[d(x_6)d(x_7), d(x_8)d(x_9)] = 0$ for all $x_1, x_2, \ldots, x_9 \in R$, such that R satisfies no polynomial identity.

In connection with these examples, Jefry Bergen [2] considers the following question:

Let a prime ring R with a derivation $d \neq 0$ be an algebra over a commutative ring A, such that $d(R)$ is contained in a finitely generated submodul of R. Does R satisfy a polynomial identity?

The positive answer to the question is a corollary to the following theorem of Bergen:

Theorem 4.2. *Let R be a prime ring with a pseudo-derivation f. Suppose that R is an algebra over a commutative ring A, such that, for a positive integer n, $d^n(R)$ is contained in a finitely generated submodul of R. If $f^{2n-1} \neq 0$, then R is an order in a simple algebra which is finite dimensional over the center of the algebra.*

Namely, the answer to the above question we get from the foregoing theorem if we take $f = d$ (a derivation) and $n = 1$.

Bergen shows that the condition $f^{2n-1} \neq 0$ cannot be substituted by $f^{2n-2} \neq 0$. Moreover, Bergen remarks, that in the foregoing theorem (for $n = 1$) the constrain on $f(R)$ can be substituted by the constrain on $f(I)$ for any ideal $I \neq \{0\}$ of R.

If ϕ is a homomorphism of R, then $f = \phi - \mathrm{id}_R$ is a pseudo-derivation, and thus (for $n = 1$) the last theorem can be modified for the case of an automorphism $\phi \neq 0$ of a prime ring R.

5. Prime rings with a proper automorphism or a nonzero derivation f satisfying $f([x,y]) - [x,y]$ is zero or invertible for all $x, y \in R$.

V. De Filipps [5] considers prime rings R with a nonzero derivation or a proper automorphism f satisfying

(1) $f([x,y]) - [x,y]$ is zero or invertible for all $x, y \in R$.

The main results of this somewhat longer and very nontrivial paper are contained in the following two theorems:

Theorem 5.1. *Let R be a noncommutative prime ring, I a nonzero ideal of R, and f a proper automorphism of R or a nonzero derivation on R satisfying*

$$f([x,y]) - [x,y] \text{ is zero or invertible for all } x, y \in I.$$

Then $R = D$ or $R = \mathrm{Mat}(D,2)$ for some division ring D.

Theorem 5.2. *Let R be a noncommutative semiprime ring, and f a nonzero derivation on R with the property (1). Then $R = D$ or $R = \mathrm{Mat}(D,2)$ for some division ring D.*

We recall that a ring R is said to be semiprime, if R has no nonzero nilpotent ideals.

6. Nilpotent derivations and commutativity

Lee and Lee [8] proved the following interesting result:

Theorem 6.1. *Let R be a prime ring with center Z, let I be a nonzero ideal of R, and let n be a positive integer. If d is a derivation on R such that $d^n(I) \subseteq Z$, then either $d^n = 0$ or R is commutative.*

At about the same time, Trzepizur [11], as a part of a more general study, proved a related theorem:

Theorem 6.2. *Let n be a nonnegative integer, let R be a prime ring with* $\operatorname{char} R = 0$ *or* $\operatorname{char} R > n+1$, *and let Z be the center of R. If d is a derivation on R and S a subring of R such that $d(S) \subseteq S$ and $d^n(S) \subseteq Z$, then either $d^n(S) = \{0\}$ or $S \subseteq Z$.*

Recently, E. Bell, A. A. Klein and J. Lucier [1] continued similar investigation. First, for the case of special subrings, they prove three following theorems:

Theorem 6.3. *Let R be infinite, and S be a subring of finite index. If d is a derivation on R and $d^n(S) \subseteq Z$ for some positive integer n, then R is commutative or $d^n = 0$.*

Theorem 6.4. *Let n be a positive integer. Let d be a derivation on R, let K be the subring of R generated by $d(R)$, and suppose that $d^n(K) \subseteq Z$. Suppose also that one of the following holds: (i) $n \geq 3$; (ii) $n = 1$ and* $\operatorname{char} R \neq 2$; *(iii) $n = 2$, and $d(Z) \neq \{0\}$. Then either R is commutative or $d^n = 0$.*

Theorem 6.5. *Let H be a commutative subring of R. If d is a derivation on R and $d^n(H) \subseteq Z$ for some positive integer n, then R is commutative or $d^n = 0$.*

Theorem 6.3. follows by Theorem 6.1. since, by a known lamma (Lemma 6.1.) S contains an ideal I of finite index, hence $I \neq \{0\}$, R being infinite.

Lemma 6.1. *Let R be an arbitrary ring and S a subring of R of finite index in R. Then S contains an ideal of R which is of finite index in R.*

In the proof of Theorem 6.4, Theorem 6.1. has been also used together with the known

Lemma 6.2. *Let R be arbitrary ring. If d is a derivation on R such that $d^3 \neq 0$, then the subring generated by $d(R)$ contains a nonzero ideal of R.*

For the case $n = 2$ in Theorem 6.4. the possibility $d^2 = 0$ cannot occur, hence R must be commutative. Moreover, the hypothesis that $d(Z) \neq \{0\}$ cannot be dilated, as we see by letting R be the ring $\mathrm{Mat}(F, 2)$ over a field F of characteristic different from 2 and d be inner derivation induced by the matrix e_{12}.

Theorem 6.5. follows also by Theorem 6.1. using the known

Lemma 6.3. *Let R be noncommutative. Then the commutator subring H contains a nonzero ideal of R.*

Some next results concern prime rings with restricted characteristic.

Theorem 6.6. *Let S be a subring of R. If there exists a derivation d on R such that $\{0\} \neq d(S) \subseteq Z$, then S is commutative. Moreover, if $\operatorname{char} R \neq 2$, then $S \subseteq Z$.*

The authors remark that if $\operatorname{char} R = 2$, then $d(S) \subseteq Z$ does not imply $S \subseteq Z$. Indeed, let R be the ring $\mathrm{Mat}(GF(2), 2)$, let $S = \{0, e_{21}\}$ and let d be the inner derivation determined by e_{12}. Then $d(S) = \{0, 1\} = Z$, but $S \not\subseteq Z$. Among those results the main result is

Theorem 6.7. *Let n be a positive integer, and let $\operatorname{char} R = 0$ or $\operatorname{char} R > n$. If d is a derivation on R and S is a subring of R such that $d(S) \subseteq S$ and $d^n(S) \subseteq Z$, then either S is commutative or $d^n(S) = \{0\}$. Moreover, if $d^n(S) \neq \{0\}$ and $\operatorname{char} R > n + 1$, then $S \subseteq Z$.*

Some results were proved for prime rings of arbitrary characteristic.

For a prime ring R with center $Z \neq \{0\}$, localizing at $Z - \{0\}$ yields a prime ring $\overline{R}$ with center $\overline{Z}$ equal to the quotient field of Z. R is called small, resp. big, if $\overline{R}$ is finite, resp. infinite dimensional as a vector space over $\overline{Z}$.

The authors next prove

Theorem 6.8. *A nonzero left ideal of a big ring R is big.*

The major result on big subrings is the following

Theorem 6.9. *Let R be a big ring with center $Z \neq \{0\}$, and let S be a big subring of R. If there exists a derivation d on R such that $d(S) \subseteq S$ and $d^n(S) \subseteq Z$ for some positive integer n, then either $d^n(S) = \{0\}$ or S is commutative.*

The inductive argument used in the proof of the above theorem gives

Corollary 6.1. *Let R be a big ring with $\operatorname{char} \neq 2$ and center $Z \neq \{0\}$, and let S be a big subring of R. If there exists a derivation d on R such that $d(S) \subseteq S$ and $d^n(S) \subseteq Z$ for some positive integer n, then $d^n(S) = \{0\}$.*

(2) $p = q \Rightarrow xy = yx$,
(3) $p = q \Rightarrow xyy^* = xx^*yy^*$,
(4) $p = q \Leftrightarrow xyy^* = yy^*x$,
(5) $p = q \Leftrightarrow xx^*yy^* = yy^*xx^*$.

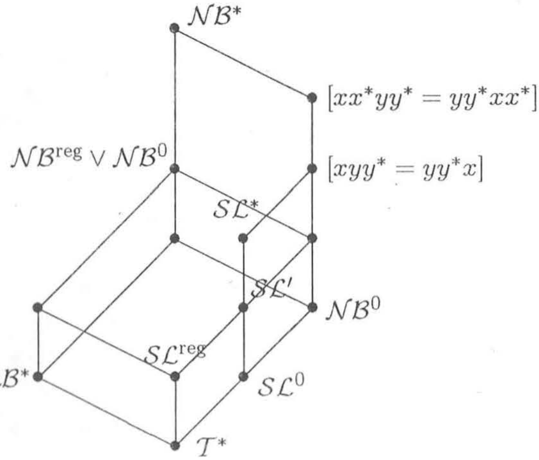

Figure 7. All subvarieties of $\mathcal{NB}^*$

So, no identity of type (1) defines a proper subvariety of $\mathcal{NB}^*$, and hence, such identities are of no importance. If a variety satisfies an identity of type (2), it must be a subvariety of $\mathcal{SL}^*$, when Theorem 1.7.3 applies. On the other hand, from the results of [24] it follows that the identity $xyy^* = xx^*yy^*$ defines the variety $\mathcal{NB}^* \vee \mathcal{NB}^0$ within $\mathcal{NB}^*$; thus, if the equational theory of the considered variety contains an identity of type (3), it is a subvariety of $\mathcal{NB}^* \vee \mathcal{NB}^0$, which is a case taken care of by Theorem 1.7.8. Hence, outside $\mathcal{SL}^*$ and $\mathcal{NB}^* \vee \mathcal{NB}^0$, there are at most two proper subvarieties of $\mathcal{NB}^*$: those defined by $xyy^* = yy^*x$ and $xx^*yy^* = yy^*xx^*$, respectively. It is effectively shown in [20] that these two varieties are different, and so we obtain

Theorem 1.7.10. *The lattice of all varieties of normal bands with involution has the inclusion diagram given in Figure 7.*

Finally, we are going to determine all subvarieties of $\mathcal{J}_4$, thereby answering a question from the beginning of this subsection. To do that, we must employ some more notation and define further notions. The material presented below is published for the first time.

An (involution) semigroup S with zero 0 is said to be *null* (or *constant*) if for all $a, b \in S$ we have $ab = 0$. The variety $\mathcal{N}^{\mathrm{id}}$ of all null semigroups with trivial involution is a minimal one, i.e. it is on the Fajtlowicz's list. It is easy to prove that it is generated by N_2, the two element null involution semigroup with a trivial involution. Further, let $\mathcal{N}^*$ denote the variety of all null involution semigroups, while N_3 is the three-element null involution semigroup in which the involution fixes one of its elements and permutes the other two.

The notion of an *inflation* is familiar in semigroup theory for a long time. Namely, a semigroup V is an inflation of its subsemigroup S if there is a homomorphism $\varphi : V \to S$ such that $\varphi|_S$ is the identity mapping on S and for all $v_1, v_2 \in V$ we have

$$v_1 v_2 = \varphi(v_1)\varphi(v_2).$$

In particular, this means that every product of elements from V lies in S. An inflation of a semigroup S is just a retractive ideal extension of S by the null semigroup $Q \cong V/S$ (see Petrich [85]).

The function φ if often referred to as the *inflation function*.

Now we say that an *involution* semigroup V is a **-inflation* of its involution subsemigroup S if the semigroup reduct of V is an inflation of the semigroup reduct of S, and the corresponding inflation function φ agrees with the star: $\varphi(a^*) = \varphi(a)^*$ for all $a \in S$. Just as in the implication (iii)$\Rightarrow$(iv) of Theorem 1 from Pastijn [79], it is not difficult to prove

Lemma 1.7.11. *Any *-inflation of an involution semigroup S is a subdirect product of S and a null involution semigroup N.*

Now we describe the structure of involution semigroups belonging to the join $\mathcal{V} \vee \mathcal{N}^*$, where $\mathcal{V}$ is an arbitrary involution semigroup variety.

Lemma 1.7.12. *Let $\mathcal{V}$ be an involution semigroup variety. Then $\mathcal{V} \vee \mathcal{N}^*$ consists precisely of all *-inflations of members of $\mathcal{V}$.*

Proof. Clearly, both $\mathcal{V}$ and $\mathcal{N}^*$ are contained in the class of all *-inflations of members of $\mathcal{V}$. On the other hand, by the previous lemma, all involution semigroups from the latter class are contained in $\mathcal{V} \vee \mathcal{N}^*$. Therefore, the proposition will be proved as soon as we show that *-inflations of members of $\mathcal{V}$ constitute a variety.

First of all, for each $i \in I$ (where I is an index set) let V_i be a *-inflation of S_i, with φ_i being the corresponding *-inflation function. Then the *-free reduct of $\prod_{i\in I} V_i$ is an inflation of $\prod_{i\in I} S_i$, the inflation function φ being the target tupling of φ_i's, that is, $\varphi(\langle v_i : i \in I\rangle) = \langle \varphi_i(v_i) : i \in I\rangle$. But

$$\begin{aligned}\varphi(\langle v_i : i \in I\rangle^*) &= \varphi(\langle v_i^* : i \in I\rangle) = \langle \varphi_i(v_i^*) : i \in I\rangle = \\ &= \langle \varphi_i(v_i)^* : i \in I\rangle = \langle \varphi_i(v_i) : i \in I\rangle^*,\end{aligned}$$

thus *-inflations are closed for direct products.

Further, let V be a *-inflation of S (with φ as the *-inflation function), and let T be an involution subsemigroup of V. Then $T \cap S$ is an involution subsemigroup of S (it is easy to see that it cannot be empty), and, moreover, the *-free reduct of T is an inflation of the *-free reduct of $T \cap S$ respect to $\varphi|_T$. Yet, φ agrees with *, and so does $\varphi|_T$. So, T is a *-inflation of $T \cap S \in \mathcal{V}$.

Finally, with the same setting as above, let P be a homomorphic image of V under homomorphism α. Then the *-free reduct of $P = \alpha(V)$ is an inflation of the *-free reduct of $T = \alpha(S)$, and the corresponding inflation function is φ' defined by $\varphi'(p) = t$ if and only if there are $s \in S$, $v \in V$, such that $\alpha(s) = t$, $\alpha(v) = p$ and $\varphi(v) = s$ (one easily shows that such a definition is logically correct). However, α is a *-homomorphism, so $t^* = \alpha(s^*)$ and $p^* = \alpha(v^*)$. Since

$\varphi(v^*) = \varphi(v)^* = s^*$, we have $\varphi'(p^*) = t^* = \varphi'(p)^*$, whence we conclude that φ' agrees with *. □

Since it is easy to calculate that N_2 and N_3 are the only subdirectly irreducibles in $\mathcal{N}^*$ (thus N_3, or any other null involution semigroup with a nonidentical involution, generates $\mathcal{N}^*$), it follows that the list of subdirectly irreducibles of a variety of the form $\mathcal{V} \vee \mathcal{N}^*$ exhausts with the subdirectly irreducibles of $\mathcal{V}$, N_2 and N_3. Therefore, any subvariety of $\mathcal{V} \vee \mathcal{N}^*$ is either of the form $\mathcal{W} \vee \mathcal{N}^{\mathrm{id}}$, or of the form $\mathcal{W} \vee \mathcal{N}^*$, where $\mathcal{W} \subseteq \mathcal{V}$. So, to determine the structure of the lattice of subvarieties of $\mathcal{V} \vee \mathcal{N}^*$, it remains to establish which of the above joins are mutually different. To this end the following auxiliary result will be helpful.

Lemma 1.7.13. *If $\mathcal{W}$ is an involution semigroup variety which does not satisfy $x = x^*$, then $\mathcal{W} \vee \mathcal{N}^{\mathrm{id}} = \mathcal{W} \vee \mathcal{N}^*$.*

Proof. Let $S \in \mathcal{W}$ be an involution semigroup in which $x = x^*$ fails. Denote the elements of N_2 by 0 and 1, and consider the direct product $T = S \times N_2$. Let $P = S \times \{0\}$ and consider the equivalence $\theta = \Delta_{T \setminus P} \cup (P \times P)$ of T (it collapses all pairs whose second coordinate is 0). Obviously, θ is a *-congruence of T, and $N = T/\theta$ is null. As S has elements which are not fixed by *, so has N (because if $a \neq a^*$ for some $a \in S$, then $(a, 1)^* = (a^*, 1) \neq (a, 1)$). Thus, N generates $\mathcal{N}^*$, implying that $\mathcal{N}^* \subseteq \mathcal{W} \vee \mathcal{N}^{\mathrm{id}}$. The lemma now easily follows. □

Our general result (which is related to the main results of Graczyńska [46] and Mel'nik [75]) is now as follows.

Theorem 1.7.14. *Let $\mathcal{V}$ be an involution semigroup variety which does not contain nontrivial null involution semigroups. Let $\mathcal{U}$ be the greatest subvariety of $\mathcal{V}$ satisfying $x = x^*$. Then the lattice of subvarieties of $\mathcal{V} \vee \mathcal{N}^*$ has the structure as depicted in Figure 8, where the interval $[\mathcal{N}^{\mathrm{id}}, \mathcal{U} \vee \mathcal{N}^{\mathrm{id}}]$ is isomorphic to the lattice of subvarieties of $\mathcal{U}$, while the interval $[\mathcal{N}^*, \mathcal{V} \vee \mathcal{N}^*]$ is isomorphic to the lattice of subvarieties of $\mathcal{V}$.*

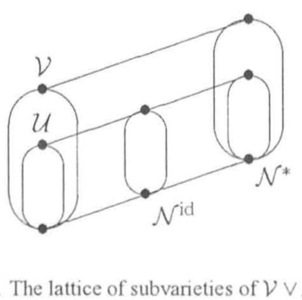

Figure 8. The lattice of subvarieties of $\mathcal{V} \vee \mathcal{N}^*$

Proof. If $\mathcal{W} \subseteq \mathcal{U}$, then $\mathcal{W} \vee \mathcal{N}^{\mathrm{id}}$ satisfies $x = x^*$, and thus it differs from any variety of the form $\mathcal{V}' \vee \mathcal{N}^*$, where $\mathcal{V}' \subseteq \mathcal{V}$. Moreover, from the previous remarks it follows that $\mathcal{W}_1 \vee \mathcal{N}^{\mathrm{id}} = \mathcal{W}_2 \vee \mathcal{N}^{\mathrm{id}}$ implies $\mathcal{W}_1 = \mathcal{W}_2$ for all $\mathcal{W}_1, \mathcal{W}_2 \subseteq \mathcal{U}$. On the other hand, if $\mathcal{W} \not\subseteq \mathcal{U}$, then by Lemma 1.7.13 we have $\mathcal{W} \vee \mathcal{N}^{\mathrm{id}} = \mathcal{W} \vee \mathcal{N}^*$.

Now if $\mathcal{W} \subseteq \mathcal{V}$ is arbitrary, then by listing the subdirectly irreducible members of $\mathcal{W} \vee \mathcal{N}^*$, we obtain that the correspondence $\mathcal{W} \mapsto \mathcal{W} \vee \mathcal{N}^*$ (as well as $\mathcal{W}' \mapsto$

$\mathcal{W}' \cup \mathcal{N}^{\mathrm{id}}$ for $\mathcal{W}' \subseteq \mathcal{U}$) is a bijective one. Thus, to prove the theorem, we need to show that these two correspondences are lattice homomorphisms.

It is immediate to see that these mappings agree with $\vee$, the varietal join operation. For the intersection, i.e. for the equalities

$$(\mathcal{W}_1 \vee \mathcal{N}^{\mathrm{id}}) \cap (\mathcal{W}_2 \vee \mathcal{N}^{\mathrm{id}}) = (\mathcal{W}_1 \cap \mathcal{W}_2) \vee \mathcal{N}^{\mathrm{id}},$$

and

$$(\mathcal{Z}_1 \vee \mathcal{N}^*) \cap (\mathcal{Z}_2 \vee \mathcal{N}^*) = (\mathcal{Z}_1 \cap \mathcal{Z}_2) \vee \mathcal{N}^*,$$

where $\mathcal{W}_1, \mathcal{W}_2 \subseteq \mathcal{U}$ and $\mathcal{Z}_1, \mathcal{Z}_2 \subseteq \mathcal{V}$, it suffices to inspect once more the list of subdirectly irreducibles in corresponding varieties, just as above. The theorem is proved. □

The variety to which we intend to apply the above theorem is

$$\begin{aligned} \mathcal{J}_4 &= \mathcal{RB}^* \vee \mathcal{SL}^{\mathrm{id}} \vee \mathcal{SL}^0 \vee \mathcal{N}^{\mathrm{id}} = \\ &= \mathcal{NB}^{\mathrm{reg}} \vee \mathcal{SL}^0 \vee \mathcal{N}^{\mathrm{id}} = \\ &= (\mathcal{NB}^{\mathrm{reg}} \vee \mathcal{NB}^0) \vee \mathcal{N}^{\mathrm{id}} = \\ &= (\mathcal{NB}^{\mathrm{reg}} \vee \mathcal{NB}^0) \vee \mathcal{N}^*, \end{aligned}$$

as $\mathcal{NB}^{\mathrm{reg}} \vee \mathcal{NB}^0$ does not satisfy $x = x^*$. On the other hand, $\mathcal{SL}^{\mathrm{id}}$ is the *only* nontrivial subvariety of $\mathcal{NB}^{\mathrm{reg}} \vee \mathcal{NB}^0$ equipped with an identical involution, and since $\mathcal{NB}^{\mathrm{reg}} \vee \mathcal{NB}^0$ has 10 subvarieties (as proved by Theorem 1.7.8), it follows from the above theorem that $\mathcal{J}_4$ has exactly 22 subvarieties. The subvarieties missing from Figure 3 are $\mathcal{NB}^0$, $\mathcal{SL}^{\mathrm{id}} \vee \mathcal{NB}^0$ and the joins of these two with $\mathcal{N}^*$.

1.8. **Subdirectly Irreducible Involution Bands.** Subdirectly irreducible algebras are very important building blocks of a variety, determining a great deal its structure and relationships to other varieties (just as it was experienced in the previous considerations). As long as semigroups are concerned, probably the first paper dealing with subdirect decompositions was the one of Thierrin [110]. The main contribution to the topic in the sixties was given by Schein [104], while Gerhard [41] described subdirectly irreducible bands. The characterizations presented in the sequel are just in the style of those given in [41], and they are all due to Dolinka [25].

The first task is certainly to describe subdirectly irreducible involution semilattices. We already met two distinguished semilattices with involution: these are Σ_2 (the two-element semilattice with the identical involution) and Σ_3 (the 0-direct union of a trivial semigroup with its copy). By Σ_4 we denote the involution semilattice obtained from Σ_3 by adjoining an identity element (which is, of course, fixed by the involution). Σ_4 is depicted in the following figure.

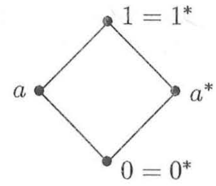

Figure 9. The involution semilattice Σ_4

Theorem 1.8.1. *There are exactly three (nontrivial) subdirectly irreducible involution semilattices: Σ_2, Σ_3 and Σ_4.*

As in the case of bands, it is necessary to distinguish between those involution bands which do or do not contain a zero element. Also similarly to bands, it is much easier to obtain the characterization for involution bands without zero. Recall that if B is an involution band and $a, b \in B$, then $\theta(a, b)$ is a customary notation designed for the *principal* congruence generated by (a, b), that is, for the least congruence containing the indicated pair.

Theorem 1.8.2. *An involution band B without zero is subdirectly irreducible if and only if B is an ideal extension of a rectangular involution band K such that there exist distinct $a, b \in K$ for which $\theta(a, b) \subseteq \theta(c, d)$ holds for all distinct $c, d \in K$, and for all $p, q \in B$, the condition $pk = qk$ and $kp = kq$ for all $k \in K$ implies $p = q$.*

Of course, as one might expect, every subdirectly irreducible involution band has a core — the least non-null *-ideal. So, the above theorem guarantees that in a subdirectly irreducible involution band without zero, its core K is a rectangular band with involution. However, *unlike* ordinary bands, the case when a zero is present splits into two essentially different cases. Namely, it turns out that the core of a subdirectly irreducible can be either a rectangular involution band with zero adjoined (i.e. with structure involution semilattice Σ_2), or of the form $I_0^*(A)$ for some rectangular band A (i.e. with structure involution semilattice Σ_3). The first of these two possibilities is handled easily, while the other is much more involved.

Theorem 1.8.3. *Let B be an involution band with zero. Then it is subdirectly irreducible and has a rectangular involution band with adjoined zero as the core if and only if $B = (B_1)^0$ for some subdirectly irreducible involution band B_1 without zero.*

Theorem 1.8.4. *An involution band B with zero which is not an involution semilattice, whose core has the structure based on Σ_3, is subdirectly irreducible if and only if B is an ideal extension of an involution band of the form $I_0^*(L)$ for some left zero band L, such that there exist distinct $a, b \in L$ for which $\theta(a, b) \subseteq \theta(c, d)$ holds for all distinct $c, d \in L$, and for all $p, q \in B$, the condition $p\ell = q\ell$ and $\ell^* p = \ell^* q$ for all $\ell \in L$ implies $p = q$.*

An interesting special case of the above theorem describes the subdirectly irreducibles in $\mathcal{B}^0$.

Theorem 1.8.5. *An involution band $B \in \mathcal{B}^0$ is subdirectly irreducible if and only if it is of the form $I_0^*(T)$, where T is either a subdirectly irreducible band without zero, or the trivial semigroup.*

In light of Theorem 1.7.8, it follows that all the subdirectly irreducibles of $\mathcal{B}^{\mathrm{reg}} \vee \mathcal{B}^0$ belong either to $\mathcal{B}^{\mathrm{reg}}$, or to $\mathcal{B}^0$.

For regular *-normal bands (i.e. for the variety $\mathcal{NB}^{\mathrm{reg}}$) we can explicitly point out the subdirectly irreducibles. Namely, by Theorem 2.2 of Scheiblich [101], every normal *-regular involution band B can be represented as a spined product of a left normal band L and its anti-isomorphic copy (that is, its dual) R, which is a right normal band, while the involution simply reverses pairs. (Recently, this assertion was generalized to arbitrary involution bands in [26]: if $\varrho^\flat$ denotes the congruence opening of an equivalence ϱ, then every involution band B can be represented as a spined product of the band $B/\mathcal{R}^\flat$ and its dual over $B/\mathcal{D}'$, where $\mathcal{D}' = \mathcal{L}^\flat \circ \mathcal{R}^\flat$, so that the involution is again the reversal of pairs.) It is not difficult to prove that such a band B is subdirectly irreducible as an involution band if and only if L is subdirectly irreducible as a band. But II.2 of [41] lists all left normal subdirectly irreducible bands: these are the trivial semigroup, the two element semilattice, the two element left zero band, and the latter band with adjoined zero. Thus the nontrivial subdirectly irreducible members of $\mathcal{NB}^{\mathrm{reg}}$ are: Σ_2, the 2×2 rectangular involution band RB_2 (which is the only nontrivial subdirectly irreducible rectangular involution band) and RB_2^0. On the other hand, the above theorem implies that the only (nontrivial) subdirectly irreducibles in $\mathcal{NB}^0$ are Σ_3 and $I_0^*(L_2)$, where L_2 denotes the two-element left zero band. In [25], it was proved that the list of all subdirectly irreducible normal bands with involution is completed by Σ_4 and two more normal involution bands, one containing six, and another containing nine elements. This in passing shows that the variety $\mathcal{NB}^*$ is residually < 10.

Theorem 1.8.6. *Aside from those contained in $\mathcal{NB}^{\mathrm{reg}} \vee \mathcal{NB}^0$ and $\mathcal{SL}^*$, there are exactly two more subdirectly irreducible members of $\mathcal{NB}^*$, both with core $I_0^*(L_2)$: one extended by Σ_2 (this one having 6 elements), and one extended by RB_2^0 (thus, having 9 elements), denoted by N_6 and N_9, respectively.*

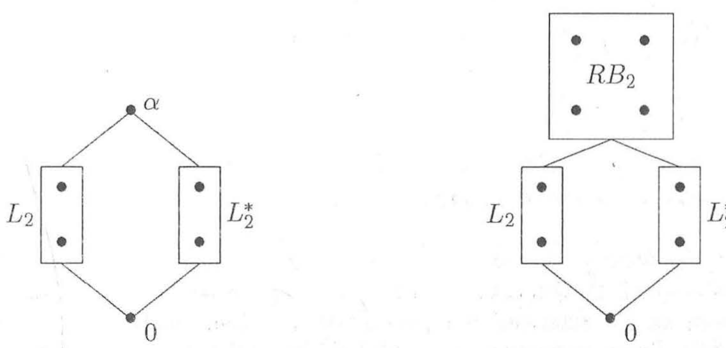

Figure 10. Normal involution bands N_6 and N_9

Since N_9 has noncommuting projections, it must generate the whole $\mathcal{NB}^*$, bearing in mind Theorem 1.7.10. On the other hand, N_6 belongs to the subvariety of $\mathcal{NB}^*$ determined by $xx^*yy^* = yy^*xx^*$ (since $0\alpha = \alpha 0 = 0$), but does not belong to the subvariety given by $xx^*y = yxx^*$, as $\ell_1\alpha = \ell_1 \neq \ell_2 = \ell_2\ell_1 = \alpha\ell_1$, where $L_2 = \{\ell_1, \ell_2\}$. Hence, N_6 generates the former subvariety, whence all members of the latter one turn out to be subdirect products of involution semilattices and normal involution bands from $\mathcal{NB}^0$.

1.9. **Varieties of Regular *-Semigroups with the Amalgamation Property.** Let $\{A_\alpha : \alpha \in I\}$ be a family of universal algebras, sharing a common subalgebra U such that for each $\alpha, \beta \in I$, $\alpha \neq \beta$, we have $A_\alpha \cap A_\beta = U$. Such a family (which is in fact a partial algebra) is called an *amalgam*. It is said to be *weakly embedable* into an algebra B if there exist injective homomorphisms $\varphi_\alpha : A_\alpha \to B$, $\alpha \in I$, agreeing on U ($\varphi_\alpha|_U = \varphi_\beta|_U$ for all $\alpha, \beta \in U$). If, in addition, we have $\varphi_\alpha(A_\alpha) \cap \varphi_\beta(A_\beta) = \varphi_\alpha(U)$ for all different $\alpha, \beta \in I$, then the considered amalgam is *strongly embedded* into B. A variety of algebras $\mathcal{V}$ has the *weak (strong) amalgamation property* if any amalgam of algebras from $\mathcal{V}$ can be weakly (strongly) embedded into an algebra from $\mathcal{V}$.

As known, semigroup amalgams and amalgamation properties in semigroup varieties constitute a well developed and established part of semigroup theory. Yet, there is a major obstacle in completing a number of characterization results which concern amalgams, namely the group varieties. It is still an open question whether there exists a proper nonabelian variety of groups with the weak (strong) amalgamation property (for the strong variant, this is just Problem 6 from [77]). Therefore, it is quite expectable that in considering amalgamation problems for various involution semigroup varieties, groups, and in fact completely simple *-semigroups will remain out of range, so that we obtain descriptions modulo these classes.

For inverse semigroups (recall that they can be considered as regular *-semigroups with the identity $xx^*x^*x = x^*xxx^*$), the following theorem is a result of combined efforts of Hall [47] and Bíró, Kiss and Pálfy [6] (see also [48, 57]).

Theorem 1.9.1. *Aside from the hypothetical proper nonabelian weakly (strongly) amalgamable group varieties, precisely the following inverse semigroup varieties have the weak (strong) amalgamation property:*

(1) *the variety of all inverse semigroups,*
(2) *the variety of all groups,*
(3) *all varieties of commutative inverse semigroups (these are the varieties of semilattices of Abelian groups).*

Later, the focus moved onto *generalized inverse* semigroups — orthodox *-semigroups in which idempotents form a regular *-normal band (of course, inverse semigroups are characterized by the condition that idempotents form a semilattice from $\mathcal{SL}^{\mathrm{id}}$). The investigation along this line was initiated by Imaoka in [58], and the contribution of Hall and Imaoka [50] should be also singled out.

The second author of this survey noted that the results from the last section of [50], when put together with some techniques applied earlier to existence varieties of regular semigroups [49], give a sufficient basis for describing regular *-semigroup varieties with the weak (strong) amalgamation property. In that sense, the paper [28] (where the following result appears) is a continuation of [50]. Note that all varieties listed below are either generalized inverse, or completely simple.

Theorem 1.9.2. *A regular *-semigroup variety $\mathcal{V}$ has the weak (strong) amalgamation property if and only if one of the following conditions is satisfied:*

(1) *$\mathcal{V}$ is an inverse semigroup variety with the weak (strong) amalgamation property,*
(2) *$\mathcal{V} = \mathcal{U} \vee \mathcal{RB}^*$, where $\mathcal{U}$ is an inverse semigroup variety with the weak (strong) amalgamation property,*
(3) *$\mathcal{V}$ is a completely simple *-semigroup variety with the weak (strong) amalgamation property.*

It is worth mentioning one more ingredient used in obtaining the above result. First of all, note that the Brandt semigroup B_2 can be considered as an inverse semigroup (then it is generated as a regular *-semigroup by a single generator a, subject to the relation $a^2 = 0$). It was proved by Schein [105] (and reproved in [47]) that an inverse semigroup variety consists entirely of semilattices of groups if and only if it omits B_2. This was extended to regular *-semigroup varieties in [28], so that for such a variety $\mathcal{V}$, $B_2 \notin \mathcal{V}$ is equivalent to the fact that $\mathcal{V}$ consists entirely of completely regular *-semigroups, and further, to the fact that $\mathcal{V}$ satisfies an identity of the form $x = ux^2$, where $u = u(x)$ is an involution semigroup word.

However, quite recently it turned out that even the above indicator characterization is just a part of a more general setting. We finish by quoting the main result of [29].

Theorem 1.9.3. *Let $\mathcal{V}$ be an involution semigroup variety. Then the following conditions are equivalent:*

(1) *any member of $\mathcal{V}$ can be decomposed into an involution semilattice of Archimedean semigroups,*
(2) *$\mathcal{V}$ does not contain B_2 and $I_0^*(B_2)$.*

Analogous descriptions for varieties consisting of semilattices of Archimedean semigroups (without involution) were obtained earlier by Sapir and Sukhanov [100] for periodic case, and for the general case by Ćirić and Bogdanović [9].

2. Varieties of Involution Semirings

2.1. The Role of Involution Semirings in Theoretical Computer Science. First of all, we recall that by our definition, a semiring is an algebra with two binary operations, $(S, +, \cdot)$, the first of which is commutative. On the other hand, there are several authors which, while referring to semirings, do not assume the commutativity of $+$, see e.g. [80, 81, 82]. Also, one may often encounter definitions in which $(S, +)$ is required to be a monoid, and its neutral element 0 is then considered as a fundamental constant. However, the latter difference will not cause any major problems: we shall use the semirings with a zero in the present subsection (conforming to the practice in theoretical computer science), and then pass in the subsequent two subsections to the (more general) approach in which the zero is dropped, but the results are always easily transformed from the one variant to another.

Also, in this subsection we shall use another symbol for semiring involutions, namely $^\vee$ instead of the star. There are fairly good reasons for the change of notation. Namely, if Σ is an alphabet, then it is a quite wide-spread notational convention to denote the free monoid on Σ by Σ^*, which consist of all words (finite sequences) over Σ, and free monoids will be important for us in the sequel. Indeed, we may define a semiring with unit

$$L_\Sigma = (\mathcal{P}(\Sigma^*), +, \cdot, \emptyset, \{\lambda\}),$$

where $+$ (for traditional reasons) denotes the set-theoretical union, λ is the empty word, and for $A, B \subseteq \Sigma^*$ we have

$$A \cdot B = \{uv : \ u \in A, \ v \in B\}.$$

The subsets of Σ^* are usually called *languages* (over Σ), and AB is called the *concatenation* of languages A and B. Therefore, we obtain the *language semiring* over Σ. Actually, it is not difficult to see that we can obtain a semiring (with unit) form an arbitrary semigroup (monoid) S, by defining analogous operations of the power set of S,

$$P_S = (\mathcal{P}(S), +, \cdot, \emptyset)$$

(in case S is a monoid, the unit $\{1\}$ is added to the above system). According to the above notation, L_Σ is in fact the same thing as P_{Σ^*}.

Now, one can define a unary operation $A \mapsto A^*$ in P_S (provided S is a monoid) by

$$A^* = \sum_{n \geq 0} A^n$$

(the sum operator denoting the union), where $A^{n+1} = A \cdot A^n$ and by convention, $A^0 = \{1\}$. If we consider the language semiring L_Σ, the above definition introduces the *Kleene star* operation, which is well-known in theoretical computer science, especially in automata theory. By equipping L_Σ with *, we obtain the *language algebra* L_Σ^*. Note that * is here by no means an involution; actually, it satisfies the fixed-point identity $x^{**} = x^*$.

On the other hand, there is an obvious way to define an involution on L_Σ. Namely, if w^R denotes the reverse of the word w, just as in the previous section, for $L \subseteq \Sigma^*$ we may define

$$L^\vee = \{w^R : \ w \in L\}.$$

It is pretty easy to see that $^\vee$ gives L_Σ the structure of a involution semiring with unit, which we denote by $L_\Sigma^\vee$. If both $^\vee$ and * are considered, we obtain the *involution language algebra* $L_\Sigma^{*\vee}$.

Another important examples of involution semirings come from binary relations. If A is an arbitrary set, we define the algebra

$$Rel(A) = (\mathcal{P}(A \times A), \cup, \circ, \emptyset, \Delta_A),$$

where $\circ$ is the relational composition and Δ_A is the diagonal (identity) relation. $Rel(A)$ also turns out to be a semiring with unit, and it can be made into an involution semiring $Rel^{\vee}(A)$ by considering the operation of the *converse* of relations:

$$\varrho^{\vee} = \{(b, a) : (a, b) \in \varrho\}.$$

Similarly as above, we can iterate the relational composition, thus obtaining a unary operation

$$\varrho^{*} = \bigcup_{n \geq 0} \varrho^{n},$$

where $\varrho^{n+1} = \varrho \circ \varrho^{n}$ and $\varrho^{0} = \Delta_A$. The relation ϱ^{*} is actually the reflexive-transitive closure of ϱ. By adding * to (involution) semirings of relations $Rel(A)$ and $Rel^{\vee}(A)$, we obtain *Kleene relation algebras (with involution)* $Rel^{*}(A)$ and $Rel^{*\vee}(A)$, respectively, cf. [59, 70, 22, 23].

Language and relation semirings are just special cases of *complete semirings*, which are of at most importance in the mathematical foundations of computer science, cf. [4, 8, 34, 63, 65, 67]. These are semirings in which an infinite summation operator $\sum_{i \in I}$ is defined, such that if $\{a_i : i \in I\}$ is any family of elements of the considered semiring, we have:

$$\sum_{1 \leq i \leq n} a_i = a_1 + \cdots + a_n,$$

$$\sum_{(i,j) \in I \times J} a_i b_j = \left(\sum_{i \in I} a_i\right)\left(\sum_{j \in J} b_j\right),$$

$$\sum_{i \in I} a_i = \sum_{j \in J} \sum_{i \in I_j} a_i,$$

where I is the disjoint union of the sets I_j, $j \in J$. Of course, the summation is commutative, associative and completely distributive. Further, a complete semiring is *completely additively idempotent* if $\sum_{i \in I} a = a$ holds for any index set I (clearly, each completely additively idempotent semiring is additively idempotent). Note that all the above examples are such. Finally, in any complete semiring one can define the *iteration operation* * by

$$a^{*} = \sum_{n=0}^{\infty} a^{n}.$$

Now we have the following observation.

Lemma 2.1.1. *Every language algebra can be embedded into a Kleene relation algebra. Consequently, every language semiring is isomorphic to a semiring of binary relations.*

Proof. (sketch) Consider the mapping $\xi : \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^* \times \Sigma^*)$ defined for every $A \subseteq \Sigma^*$ by

$$\xi(A) = \{(w, wx) : w \in \Sigma^*, x \in A\}.$$

It is a routine matter to show that ξ is, in fact, an embedding of the algebra L^*_Σ into $Rel^*(\Sigma^*)$. □

Hence, if we denote by $\mathcal{L}$ the variety generated by all language algebras, while $\mathcal{KA}$ denotes the variety of *Kleene algebras*, generated by all Kleene relation algebras, we have $\mathcal{L} \subseteq \mathcal{KA}$, and in particular, all language algebras are Kleene algebras. However, the above inclusion is in fact an equality, $\mathcal{L} = \mathcal{KA}$, because by the Kozen-Németi Theorem (cf. [64, 70]), the free Kleene algebra on Σ is just the subalgebra of L^*_Σ formed by the regular subsets of the free monoid Σ^*. Using this, and knowing the explicit equational axiomatization of Kleene algebras (which is necessarily infinite, cf. [10, 66, 13]), one can easily derive the following result.

Theorem 2.1.2. *Both language semirings and relation semirings generate the variety of idempotent semirings with unit.*

But what is the situation if the involution is present? The above Lemma 2.1.1 is no longer true for the involution case: in fact no involution semiring of the form $L^\vee_\Sigma$ can be embedded in an involution semiring of relations. In other words, if $\mathcal{L}^\vee$ denotes the variety generated by involution language algebras, while $\mathcal{KA}^\vee$ is the variety of *Kleene algebras with involution* generated by all algebras $Rel^{*\vee}(A)$, one can prove that $\mathcal{KA}^\vee \subseteq \mathcal{L}^\vee$, but this inclusion is *proper*. It is just the involution that distinguishes between them, even if we drop the iteration operations and work with involution semirings only. Consider the following identity:

$$x + xx^\vee x = xx^\vee x.$$

It is a routine to see that the above identity is true in binary relations. However, it suffices to consider the one-element alphabet $\Sigma = \{a\}$ and substitute the language $\{a\}$ for x to see that the above identity fails in all involution semirings of languages. In fact, we have a more accurate information concerning this matter.

Theorem 2.1.3. (Bloom, Ésik and Stefanescu, [8]) *The variety $\mathcal{L}^\vee$ is defined by the identities of Kleene algebras, axioms of semiring involution (including $0^\vee = 0$) and*

$$(x^*)^\vee = (x^\vee)^*.$$

Theorem 2.1.4. (Ésik and Bernátsky, [35]) *The variety $\mathcal{KA}^\vee$ is defined as a subvariety of $\mathcal{L}^\vee$ by the identity*

$$x + xx^\vee x = xx^\vee x.$$

¿From these results it is not difficult to obtain

Corollary 2.1.5. *The involution semirings of languages generate the variety of idempotent involution semirings with unit, while the relational involution semirings generate its subvariety determined by $x + xx^\vee x = xx^\vee x$.*

Let us also mention some related results obtained by the authors of this survey and Z.Ésik.

Theorem 2.1.6. (Crvenković, Dolinka and Ésik, [13, 14]) *Varieties $\mathcal{L}^\vee$ and $\mathcal{KA}^\vee$ are both not finitely based. Also, if we drop the union (addition) operation from Kleene relation algebras with involution (resp. involution language algebras), the equational theories of the so obtained varieties consist precisely of those identities of $\mathcal{KA}^\vee$ (resp. $\mathcal{L}^\vee$) which do not contain occurrences of $+$, and these theories are too nonfinitely based.*

The simplest explanation for the second part of the above result is that the interaction between the concatenation and * is from the equational point of view 'too complicated', and exactly this interaction is the origin of all nonfinite axiomatizability results of the above type which concern algebras of formal languages.

It is interesting to remark that there is a 'technical' connection between the first part of the above theorem and Theorem 1.5.2. Namely, there are two ways to prove that $\mathcal{L}^\vee$ and $\mathcal{KA}^\vee$ are not finitely based, knowing that the same holds for $\mathcal{L}$ and $\mathcal{KA}$, respectively, and knowing, of course, Theorems 2.1.3 and 2.1.4. One of these proofs — more syntactical in nature — relies on the same proposition on involutorial identities (proved in [13]), which allowed us to obtain in [15] the result of Theorem 1.5.2. Probably there are some further links between the identities of general algebraic systems with involution and of their involution-free reducts respectively, which are yet to be discovered and explored.

2.2. **Minimal Varieties of Involution Semirings.** Minimal varieties of involution semirings were described by the second author of this survey in [21]. Towards that goal, an important help was the already known list of minimal varieties of ordinary semirings, determined by Polin [92], cf. also [109]. To recall Polin's result and to formulate the main result of [21], we define some binary and unary operations on finite sets $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$ and $4 = \{0, 1, 2, 3\}$.

$\vee$	0	1
0	0	1
1	1	1

$\wedge$	0	1
0	0	0
1	0	1

$\circ$	0	1
0	0	0
1	0	0

$*_\ell$	0	1
0	0	0
1	1	1

$*_r$	0	1
0	0	1
1	0	1

$\wedge_3$	0	1	2
0	0	0	0
1	0	1	0
2	0	0	2

$\circ_3$	0	1	2
0	0	0	0
1	0	0	0
2	0	0	0

$\Diamond$	0	1	2	3
0	0	1	2	3
1	1	1	3	3
2	2	3	2	3
3	3	3	3	3

$\Box$	0	1	2	3
0	0	1	0	1
1	0	1	0	1
2	2	3	2	3
3	2	3	2	3

a	0	1	2
$\bar{a}$	0	2	1

a	0	1	2	3
$\tilde{a}$	0	2	1	3

Theorem 2.2.1. (Polin, [92]) *A variety of semirings is minimal if and only if it is generated by one of the following semirings:*

(1) $(2, \circ, \wedge)$, $(2, \circ, \circ)$, $(2, \vee, \vee)$, $(2, \vee, \wedge)$, $(2, \vee, \circ)$, $(2, \wedge, \circ)$,

(2) $(2, \vee, *_\ell)$, $(2, \vee, *_r)$,

(3) $Z_p = (\{0, 1, \dots, p-1\}, +_p, \cdot_p)$, *where* p *is a prime number, and* $+_p$ *and* $\cdot_p$ *are respectively the addition and the multiplication modulo* p *(i.e.* Z_p *is the finite field with* p *elements),*
(4) $N_p = (\{0, 1, \dots, p-1\}, +_p, \circ_p)$, *where* p *is a prime number, and* $\circ_p$ *is the zero multiplication of the set* $\{0, 1, \dots, p-1\}$.

Note that all varieties of involution semirings having a trivial involution ($x^* = x$) are exhausted by varieties of commutative semirings augmented with the identity mapping, and this conclusion applies to minimal varieties as well. Clearly, (1), (3) and (4) of the above theorem provide all such varieties.

Theorem 2.2.2. (Dolinka, [21]) *A variety of semirings with nontrivial involution is minimal if and only if it is generated by one of:*
(1) $(3, \wedge_3, \wedge_3, ^-)$, $(3, \wedge_3, \circ_3, ^-)$, $(3, \circ_3, \wedge_3, ^-)$,
(2) $(4, \Diamond, \Box, ^-)$.
(3) $(\{0, 1, \dots, p-1\}, +_p, \circ_p, -_p)$, *where* $-_p$ *is the operation of additive inverse modulo a prime number* $p \geq 3$.

It is more or less in the universal algebraic folklore that all of the algebras above generate minimal (equationally complete) varieties. The proof of the other implication, on the other hand, resembles somewhat to the way in which Fajtlowicz obtained the minimal varieties of involution semigroups, because it consists of considering cases according to the properties of Hermitian elements (involution fixed points).

Firstly, one can prove that if an involution semiring which generates a minimal variety $\mathcal{V}$ contains a Hermitian element a which is either not additively idempotent ($a + a \neq a$), or not multiplicatively idempotent ($a^2 \neq a$), then $\mathcal{V}$ consists of commutative involution semirings with a trivial involution, and in that case Theorem 2.2.1 settles the problem. Otherwise, it can be assumed that all Hermitian elements a under consideration satisfy $a + a = a^2 = a$. Now, in any involution semiring S which is not additively idempotent and which belongs to a minimal variety, there is a unique Hermitian element which is:
(1) the multiplicative zero of S,
(2) either the additive zero, or the additive unit of S.
In the latter of the two cases given in (2) above, S must be a ring, $a^\vee = -a$, and, moreover, there is a monogenic subring S' of S and a prime p such that N_p, augmented by the additive inverse modulo p, is a homomorphic image of S'. On the other hand, in the former of the two described cases, S generates the same variety as $(3, \circ_3, \wedge_3, ^-)$ does.

So, it remains to consider minimal varieties generated by additively idempotent involution semirings. If such a variety contains a nontrivial involution semiring with a unique Hermitian element, then it has to contain one of $(3, \wedge_3, \wedge_3, ^-)$, $(3, \wedge_3, \circ_3, ^-)$. Finally, if an involution semiring contains at least two Herimitian elements and generates a minimal variety (even without the condition of the additive idempotency), then it contains an involution subsemiring isomorphic to $(4, \Diamond, \Box, ^-)$, whence our theorem is established.

2.3. **Idempotent Distributive Involution Semirings.** A semiring is *distributive* if it satisfies the dual distributive identity

$$x + yz = (x + y)(x + z)$$

(of course, the above identity and commutativity of $+$ together imply $xy + z = (x + z)(y + z)$). A semiring is *idempotent* if both of its operations are such (we already referred to additive and multiplicative idempotency in semirings). Note that if a semiring S is (additively) idempotent, then $(S, +)$ is a semilattice. If both of the binary reducts $(S, +)$ and $(S, \cdot)$ of S are idempotent and commutative, then S is called a *bisemilattice*. A bisemilattice in which the two operations coincide (i.e. which satisfy $x + y = xy$) is called a *mono-bisemilattice*. Of course, it causes no confusion if we identify (in the notational sense) semilattices and mono-bisemilattices.

Idempotent and distributive semirings are called *ID-semirings* for short. The study of ID-semirings started in the late sixties and continued in the seventies, see e.g. [61, 73, 88], with investigations on distributive bisemilattices. However, the topic gained attention in the early eighties, mainly with contributions of Pastijn and Romanowska [80, 82, 95, 96]. In particular, the lattice of all varieties of ID-semirings (with $+$ commutative) is given in [96]: it is the four-dimensional cube. Recently, Kuřil and Polák [68] found a way to determine all varieties of idempotent semirings (without the requirement of distributivity of $+$ over $\cdot$). On the other hand, Pastijn and Guo [81] described the lattice of all ID-semirings without $+$ being commutative. It is a countably infinite distributive lattice.

Motivated by the result of Romanowska [96], the second author of this survey obtained the lattice of all varieties of ID-semirings with involution. The corresponding result is as follows.

Theorem 2.3.1. (Dolinka, [27]) *There are exactly 64 varieties of ID-semirings with involution, and their lattice coincides with the one depicted in Figure 11.*

As semilattices and mono-bisemilattices can be identified, so can involution semilattices and mono-bisemilattices with involution. Therefore, Σ_2, Σ_3 and Σ_4 will also denote involution semirings in which both operations define the corresponding semilattice with involution. It is easy to see that all of the above algebras are in fact ID-semirings.

It was proved in Theorem 2.1 of [82] that the multiplicative reduct of an ID-semiring must be a normal band. Further, by Theorem 1.6 of the same paper, each ID-semiring is a Płonka sum of a semilattice ordered system of ID-semirings satisfying

$$x + xyx = x.$$

The latter semirings are, in turn, obtained by a special kind of a composition of a distributive lattice ordered system of ID-semirings in which the multiplicative reduct is a rectangular band.

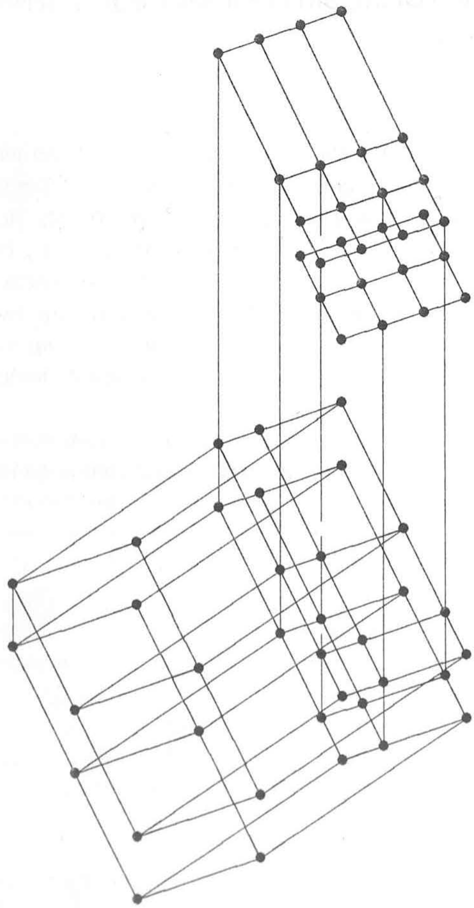

Figure 11. The lattice of all varieties of ID-semirings with involution

All these results can be extended for ID-semirings with involution as well. First of all, one must replace the well-known general algebraic construction of a Płonka sum by the *involutorial Płonka sum of algebras*, introduced in [31]. Here we give the basic definition, restricted to the case of semirings.

Let Y be an involution semilattice. A family of semirings $\{S_i : i \in Y\}$, together with a system of homomorphisms $\{\phi_{i,j} : i,j \in Y,\ i \geq j\}$ and a bijection * on $\bigcup_{i\in Y} S_i$, is called an Y*-ordered system of semirings* if the following conditions are satisfied:

(1) for each $i \in Y$, $\phi_{i,i}$ is the identity mapping on S_i,
(2) for each $i,j,k \in Y$ such that $i \geq j \geq k$ we have

$$\phi_{i,j} \circ \phi_{j,k} = \phi_{i,k},$$

(3) for each $i \in Y$, $^* : S_i \to S_{i^*}$ is a semiring anti-isomorphism,
(4) $\phi_{i^*,j^*}(x) = (\phi_{i,j}(x^*))^*$, for all $i,j \in Y$ such that $i \geq j$ and all $x \in S_{i^*}$.

The *involutorial Płonka sum* of such a system is a semiring with involution S, where $S = \bigcup_{i\in Y} S_i$, with the operations given by

$$\begin{aligned} a+b &= \phi_{i,ij}(a)+\phi_{j,ij}(b), \\ ab &= \phi_{i,ij}(a)\phi_{j,ij}(b), \end{aligned}$$

where $a \in S_i$ and $b \in S_j$.

Theorem 2.3.2. (Dolinka and Vinčić, [31]) *Each ID-semiring with involution can be represented as an involutorial Płonka sum of an involution semilattice-ordered system of ID-semirings satisfying the identity $x + xyx = x$. Conversely, the involutorial Płonka sum of every such system is an ID-semiring with involution.*

As we mentioned above, in [95] Romanowska proved that each ID-semiring satisfying $x + xyx = x$ is the sum of a distributive lattice-ordered m-system of rectangular ID-semirings (i.e. with rectangular multiplicative reduct). This means that we have given a system of disjoint semirings S_i indexed by a distributive lattice $(D, \vee, \wedge)$ (so that $i \in D$), and for each $i, j \in D$ such that $i \geq j$ an embedding $\psi_{i,j} : S_i \to S_j$ such that

(i) $\psi_{i,i}$ is the identity map on S_i for all $i \in D$,

(ii) $\psi_{i,j} \circ \psi_{j,k} = \psi_{i,k}$ for all $i, j, k \in D$ such that $i \geq j \geq k$,

(iii) $\psi_{i,i\wedge j}(S_i) + \psi_{j,i\wedge j}(S_j) \subseteq \psi_{i\vee j,i\wedge j}(S_{i\vee j})$ for all $i, j \in D$.

The sum of this system is defined in such a way that the operations in the resulting semiring $(S, +, \cdot)$ (where $S = \bigcup_{i\in D} S_i$) are given by

$$\begin{aligned} a_i b_j &= \psi_{i,i\wedge j}(a_i)\psi_{j,i\wedge j}(b_j), \\ a_i + b_j &= \psi^{-1}_{i\vee j,i\wedge j}(\psi_{i,i\wedge j}(a_i) + \psi_{j,i\wedge j}(b_j)), \end{aligned}$$

where $a_i \in S_i$ and $b_j \in S_j$.

Now, we are going to call an *m^*-system of semirings* a family of semirings S_i indexed by a distributive lattice with involution $(D, \vee, \wedge, ^*)$, endowed with semiring embeddings $\psi_{i,j}$ for each pair $i \geq j$ and a bijection * on $\bigcup_{i\in D} S_i$ such that the conditions *(i)-(iii)* above are satisfied, as well as the following conditions:

(iv) $^* : S_i \to S_{i^*}$ is a semiring anti-isomorphism for all $i \in D$,

(v) $\psi_{i^*,j^*}(x) = (\psi_{i,j}(x^*))^*$, for all $i, j \in D$ such that $i \geq j$ and all $x \in S_{i^*}$,

which express the compatibility of * with the m-system structure and, respectively, the 'symmetry' of the m-system with respect to the involution.

Theorem 2.3.3. (Dolinka, [27]) *An algebra $(S, +, \cdot, ^*)$ is an ID-semiring with involution satisfying $x + xyx = x$ if and only if it is the sum of an m^*-system of rectangular ID-semirings.*

Finally, it remains to provide some information about rectangular ID-semirings with involution. We recall here a construction which is well-known in universal algebra, called the *matrix power*. Namely, for a universal algebra $(A, \mathcal{F})$ (where $\mathcal{F}$ is a family of finitary operations on A) and $n \in \mathbb{N}$, the n-th matrix power is defined on the set $A^n = A \times \cdots \times A$ such that all fundamental operations from the original algebra are inherited by applying them coordinatewise in A^n, while two operations are added: the n-ary *diagonal operation* d given by

$$d(\mathbf{x}_1,\dots,\mathbf{x}_n) = \langle x_{11}, x_{22}, \dots, x_{nn}\rangle,$$

where $\mathbf{x}_i = \langle x_{i1},\dots,x_{in}\rangle$ for all $1 \leq i \leq n$, and a unary operation p determined by

$$p(\langle x_1, x_2, \dots, x_n\rangle) = \langle x_2, \dots, x_n, x_1\rangle.$$

It is known that for a variety of algebras $\mathcal{V}$ and a given positive integer n, all isomorphic copies of all n-th matrix powers of members of $\mathcal{V}$ also form a variety, denoted by $\mathcal{V}^{[n]}$. Also, it is known that this construction preserves equational completeness, see [71]. For more information about matrix powers and their application in universal algebra, we refer to [53] and [71]. Now we obtain the following theorem, which does not have its non-involutorial analogue.

Theorem 2.3.4. (Dolinka, [27]) *Every rectangular ID-semiring with involution is the matrix square of some semilattice and conversely, every matrix square of a semilattice is a rectangular ID-semiring with involution. In other words, the variety of rectangular ID-semirings is just $\mathcal{SL}^{[2]}$ and thus it has no proper subvarieties (cf. [21]).*

Another nice and in this setting important feature of the paper [31] is that it admits a direct calculation of those involutorial Płonka sums which are subdirectly irreducible, provided that the subdirectly irreducibles are known in the class of (involution) algebras from which the components of the sum are taken. So, the results in [31] generalize the corresponding results on subdirectly irreducible Płonka sums, given in [69]. With a little amount of technical work, one can find the explicit list of subdirectly irreducible ID-semirings, and thereby show that the variety of ID-semirings is — similarly to the variety of normal bands with involution — residually < 10 (in fact, the results presented in the last subsection of the section on involution semigroups can be also derived from the general theorems of [31]). In particular, if an involutorial Płonka sum is subdirectly irreducible, then its structure involution semilattice must be trivial, or it is subdirectly irreducible itself, that is, one of Σ_2, Σ_3 and Σ_4 (by our Theorem 1.8.1).

But first, let L_2 denote the (unique) two-element ID-semirings whose multiplicative reduct is a left zero band. Dually, we have the semiring R_2. These semirings, as well as their direct product $L_2 \times R_2$, are examples of a rectangular ID-semirings. By defining the exchange involution (the reversing of pairs) on the latter one, one obtain a four-element involution semiring, which is isomorphic to the matrix square of the two-element semilattice. This one we denote by RS_2^*.

The two-element and the four-element distributive lattice we denote by D_2 and D_4, respectively. Of course, we can equip the first one by the identity mapping as the involution, thus obtaining the involution lattice D_2^*. In turn, D_4 can be enriched to the involution lattice D_4^* by defining an involution which fixes the top and the bottom element, and exchanges the other two.

Similarly to semigroups, one can adjoin an absorbing element to a semiring (with involution). This is the same as to compose into an involutorial Płonka sum a Σ_2-ordered system consisting of a trivial involution semiring and an arbitrary involution semiring S, and such a construction yields an algebra denoted by S^0. Also, one can perform 0-direct unions by taking a semiring S (without involution), its anti-isomorphic copy $\widetilde{S}$, and a trivial involution semiring (which, considered together, form a Σ_3-ordered system of semirings) and constructing their involutorial Płonka sum. Such a sum is denoted by $I_0^*(S)$.

Finally, assume we are concerned with a Σ_4-ordered system of semirings, where the (involution) semiring assigned to the index 0 is trivial. Further, we have the anti-isomorphic semirings $S = S_a$ and $\widetilde{S} = S_{a^*}$, and the involution semiring S_1, with the structure semiring homomorphism $\phi = \phi_{1,a}$ satisfying the required conditions. The resulting sum we denote by $\diamondsuit_0^*(S, S_1, \phi)$. We omit ϕ if it is (up to an isomorphism of the resulting sum) uniquely determined by the components. Moreover, if S_1 is trivial, then it will be omitted too. The desired key result on subdirectly irreducible ID-semirings is now the following.

Theorem 2.3.5. (Dolinka, [27]) *A nontrivial ID-semiring with involution is subdirectly irreducible if and only if it is isomorphic to one of the following 17 semirings with involution:*

(1) RS_2^*, D_2^*, D_4^*,
(2) Σ_2, $(RS_2^*)^0$, $(D_2^*)^0$, $(D_4^*)^0$,
(3) Σ_3, $I_0^*(L_2)$, $I_0^*(D_2)$,
(4) Σ_4, $\diamondsuit_0^*(L_2)$, $\diamondsuit_0^*(D_2, \phi_0)$, $\diamondsuit_0^*(D_2, \phi_1)$, $\diamondsuit_0^*(L_2, RS_2^*)$, $\diamondsuit_0^*(D_2, D_2^*)$, *and* $\diamondsuit_0^*(D_2, D_4^*)$,

where ϕ_0 maps the only element of the trivial semiring into the lower element of D_2, while ϕ_1 maps into the upper element of D_2.

The above theorem, together with the other structural results presented in this section, are the main ingredients in a lengthy and involved argument, with a number of subtle details, which leads to the result of Theorem 2.3.1. In Figure 11, there are three clearly distinguished intervals of the lattice. The lattice is, of course, intentionally drawn in such a way, because the corresponding proof splits into three separate parts, each producing one of those intervals, starting from the bottom and proceeding to the top.

2.4. **Some Varieties of Involution Rings.** Let $(R, +, \cdot, -, 0)$ be a ring and assume that * is its semiring involution. Then it is very easy to deduce from the ring axioms that for all $r \in R$ we have $(-r)^* = -r^*$ and $0^* = 0$, so that * agrees with the whole ring structure of R. In the way just described, we obtain a *ring with involution* (or a **-ring*).

Involution rings are probably the most important and best studied algebraic structures with involution in mathematics in general. It would take too much space to attempt to give even a shortest account on the results concerning involution rings and their applications. This topic originates back to von Neumann, who considered the adjoint (as an involution) in the algebra of bounded linear operators on a Hilbert space (such an involution algebra is widely used in theoretical physics, especially in quantum mechanics). Classical books on involution rings are e.g. Berberian [3] and Herstein [52].

However, the point of view of considering an involutorial antiautomorphism of a ring as a fundamental operation (and thus, of considering related universal algebraic questions) is somewhat more recent and much easier to review. Such an approach has been taken, for example, in Rowen [97] and in the survey article of Wiegandt [115].

One of the (historically) most important classes of involution rings is the one of *regular *-rings*. Originally, it were *regular rings* which were considered by von Neumann in his fundamental treatise [114] (see also [108]), and which turned out to be the starting point (and the main motivation) for the whole theory of regular semigroups. Regular rings and regular *-rings are in a quite fascinating way strongly related to (orthocomplementded) modular lattices, and thus, in particular, to projective geometries. This link is described by the well-known von Neumann's Full Coordinatization Theorem (which generalizes the classical coordintaization theorems of projective spaces).

Theorem 2.4.1. (von Neumann, [114], Roddy, [94]) *Let M be a (orthocomplemented) modular lattice. Then there is a regular ring (with involution) R whose principal right ideals form a lattice, which is isomorphic to M. Moreover, R can be obtained as a ring of matrices (of a certain finite dimension) over a ring D such that $D \subseteq M$ and the ring operations of D are expressed as polynomials of the lattice M. In the case of ortholattices, the orthocomplementation is uniquely determined by the involution on R.*

It is easy to prove that the condition of a regularity of a *-ring is equivalent to the condition that every principal right ideal is generated by a *projection*, an idempotent fixed by the involution. Therefore, in the orthocomplemented version of the above theorem, one can replace the lattice of principal right ideals of R by the lattice of projections of R with respect to the partial order defined by $e \leq f$ is and only if $ef = e$. Hence, every modular ortholattice can be represented by projections of some regular *-ring.

Further, one can show that the regularity of a *-ring R is equivalent to the implication

$$rr^* = 0 \Rightarrow r = 0,$$

for all $r \in R$. This form of regularity provides an obvious way to equationally define a special ring involution which guarantees the regularity of the underlying ring. Following Yamada [117], we call a *special regular *-ring* an involution ring which satisfies the identity

$$xx^*x = x.$$

Using some results from Nambooripad and Pasijn [76], Yamada first proved that the multiplicative reduct of any special regular *-ring is a semilattice of groups, and moreover, we have $2x = (2x)(2x)^*(2x) = 8xx^*x = 8x$, so $6x = 0$. In light of this, the following result is not so surprising.

Theorem 2.4.2. (Yamada, [117]) *Any special regular *-ring R can be decomposed into a direct sum $R = R_2 \bigoplus R_3$, such that R_2 and R_3 are the *-ideals (ideals closed for *) of R consisting of all the elements of R of order 2 and 3, respectively. Moreover, R_2 satisfies $x^4 = x$, while R_3 satisfies $x^3 = x$, so that R satisfies $x^7 = x$. Consequently (by Jacobson's Theorem), every special regular *-ring is commutative.*

Going in more detail, Yamada in [117] described the subdirectly irreducible special regular *-rings.

Theorem 2.4.3. (Yamada, [117]) *The only subdirectly irreducible special regular *-rings are the finite fields with 2, 3 and 4 elements, with the inverse operation as the involution ($x^* = x^{-1}$ for all $x \neq 0$ and $0^* = 0$).*

Of course, it is well-known that a ring which satisfies the identity $x^{n+1} = x$ for some $n \in \mathbb{N}$ is subdirectly irreducible if and only if it is a field (satisfying the same identity). This fact, and the above theorems of Yamada serve as good inspiration to investigate in general the subdirect decomposition of involution rings obeying an identity of the form $x^{n+1} = x$.

Given a ring R, denote by R^{opp} its opposite ring, i.e. its anti-isomorphic copy. Clearly, the direct sum $R \bigoplus R^{opp}$ is isomorphic to their direct product, and one can define the exchange involution on this sum. The resulting involution ring we denote by $Ex(R)$. Of course, to each ideal I of R it corresponds a *-ideal of $Ex(R)$ obtained as the direct sum of I and I^*. Also, if R is a ring with involution and I is an ideal of the ring reduct of R such that $R = I \bigoplus I^*$, then it follows that $R \cong Ex(I)$.

It is not difficult to analyze all the possible involutions on a finite field $GF(p^k)$. The required involution defines an involutorial automorphism of that field, and it is well-known that every automorphism of the specified finite field is of the form

$$x \mapsto x^{p^m}$$

for some integer $0 \leq m \leq k-1$. Thus, we have

$$x = (x^*)^* = (x^{p^m})^* = x^{p^{2m}}.$$

As the multiplicative group of our field must be cyclic of order $p^k - 1$, we obtain that $(p^k - 1) \mid (p^{2m} - 1)$, that is, $k \mid 2m$. Since $2m < 2k$, this yields two possibilities: $m = 0$, whence the involution is just the identity mapping, and $m = \frac{k}{2}$, provided k is even (otherwise, this case is impossible). The resulting field with involution we denote by $GF(p^k)$ in the former case (abusing slightly the notation), and by $GF^*(p^k)$ in the latter case. Now we have prepared the way for stating our next result.

Theorem 2.4.4. (Crvenković, Dolinka and Vinčić, [16]) *A ring with involution R is subdirectly irreducible and obeys the identity $x^{n+1} = x$ if and only if there is a prime number p and an integer $k \geq 1$ satisfying $(p^k - 1) \mid n$, such that R is isomorphic to one of the following:*

(1) $GF(p^k)$,

(2) *if k is even,* $GF^*(p^k)$,
(3) $Ex(GF(p^k))$.

The key lemma in the course of proving the above theorem is that if R is a ring with involution satisfying the given conditions, then R has an identity element (which is, clearly, fixed by the involution) and R is actually *-simple (meaning that R has no nontrivial *-ideals). The other main ingredient for the proof comes from the paper of Birkenmeier, Groenewald and Heatherly [5] in which the relationships between the ideal and the *-ideal structure of an involution ring were studied. In particular, the result we need is that if R is *-simple, then it is either simple as a ring, or $R \cong Ex(K)$, where K and K^* are the only nontrivial proper ideals of R and $R^2 \neq 0$. From these facts it is possible to derive the previous theorem.

One of the principal applications of the above result is that it helps a lot in determining the lattice $L^{(n)}$ of all subvarieties of the (involution) ring variety $\mathcal{V}^{(n)}$ defined by $x^{n+1} = x$ for a given value of n. Towards this aim, the following observation is very useful. Let $\mathcal{V}_p^{(n)}$ denote the subvariety of $\mathcal{V}^{(n)}$ determined by $px = 0$ (formed by all members of the latter variety of characteristic p), and let $L_p^{(n)}$ be its lattice of subvarieties. Clearly, $\mathcal{V}_p^{(n)}$ is nontrivial if and only if $(p-1) \mid n$. Now if $\{p_1, \dots, p_k\}$ is the set of all prime numbers with this property, then it can be easily shown that the varieties $\mathcal{V}_{p_i}^{(n)}$, $1 \le i \le k$, are *independent*, which means that there is a term $t(x_1, \dots, x_k)$ such that the identity $t(x_1, \dots, x_k) = x_i$ holds in $\mathcal{V}_{p_i}^{(n)}$. If a variety is equal to the join of some of its independent subvarieties, it is usual in universal algebra to say that the variety under consideration decomposes into a *varietal product* of these subvarieties (cf. [74]). In our case, we write $\mathcal{V}^{(n)} = \mathcal{V}_{p_1}^{(n)} \otimes \cdots \otimes \mathcal{V}_{p_k}^{(n)}$. It is well-known that varietal product decompositions induce direct decompositions of the lattice of subvarieties, thus we have

$$L^{(n)} \cong L_{p_1}^{(n)} \times \cdots \times L_{p_k}^{(n)}.$$

Hence, the task of finding the lattice of varieties of rings (with involution) satisfying $x^{n+1} = x$ reduces to the same task in a fixed prime characteristic p, where $(p-1) \mid n$. This is just where Theorem 2.4.4 can be used, for it supplies the corresponding subdirectly irreducibles. It remains then to study their mutual relationships in order to obtain the exact list of varieties they generate.

This is just what have been done in the recent note [30]. Namely, let $\mathcal{F}_p$ denote the set of all finite fields of characteristic p, while $\mathcal{F}_p^*$ denotes the set of all (subdirectly irreducible) involution rings from the above theorem which are of characteristic p. Furthermore, write $R \hookrightarrow S$ if R embeds into S. This relation turns $\mathcal{F}_p$ and $\mathcal{F}_p^*$ into partially ordered sets. Clearly, $(\mathcal{F}_p, \hookrightarrow)$ is isomorphic to the divisibility order of natural numbers (as $GF(p^k)$ embeds into $GF(p^\ell)$ if and only if $k \mid \ell$), but it was shown in [30] that $(\mathcal{F}_p^*, \hookrightarrow)$ can be effectively described as well.

Now let $\mathcal{F}_p(n)$ ($\mathcal{F}_p^*(n)$) denote the set of those $GF(p^k)$ (and $GF^*(p^k)$ and $Ex(GF(p^k))$ in the involutorial case) for which $(p^k - 1) \mid n$. The main result of [30] is as follows.

Theorem 2.4.5. *Let $n \geq 1$ be an integer and p a prime such that $(p-1) \mid n$. Then $L_p^{(n)}$ is isomorphic to the lattice of all ideals of the ordered set $(\mathcal{F}_p(n), \hookrightarrow)$ (resp. $(\mathcal{F}_p^*(n), \hookrightarrow)$).*

By the previous remarks, the finite partial orders from the above theorem turn out to be computable, which establishes an effective algorithm for constructing $L_p^{(n)}$, as required.

Bearing in mind Theorem 2.4.2, let us finish this survey by discussing the case $n = 6$ as an example, so that $(p-1) \mid n$ (where p is a prime) if and only if $p \in \{2, 3, 7\}$.

When we consider ordinary rings, the situation is clear: for $p = 2$ we have two subdirectly irreducibles, $GF(2)$ and $GF(4)$, where $GF(2)$ embeds into $GF(4)$; for $p = 3$ we have $GF(3)$, and for $p = 7$ we have $GF(7)$. Thus, it is easy (using the above theorem) to conclude that there are 12 ring varieties satisfying $x^7 = x$, and that the lattice formed by them is the product of a three-element chain and the square of a two-element chain.

In the case of involution ring varieties, for each of $p = 3, 7$ we have two subdirectly irreducibles, so that $GF(3) \hookrightarrow Ex(GF(3))$ and $GF(7) \hookrightarrow Ex(GF(7))$, and both $L_3^{(6)}$ and $L_7^{(6)}$ are three-element chains. For $p = 2$, a routine calculation shows that $L_2^{(6)}$ is isomorphic to the lattice given in the following figure.

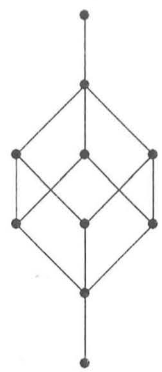

Figure 12. The lattice of all involution ring varieties satisfying $x^7 = x$ and $2x = 0$

Hence, we obtain exactly 90 varieties of involution rings satisfying $x^7 = x$. Only six of them have a special involution, cf. Theorem 2.4.3.

REFERENCES

[1] Adair, C.L., *Varieties of *-Orthodox Semigroups*, Ph.D. thesis, University of South Carolina, 1979.

[2] Adair, C.L., Bands with an involution, *J. Algebra* **75** (1982), 297–314.

[3] Berberian, S.K., *Baer *-Rings*, Springer-Verlag, Berlin, 1972.

[4] Bernátsky, L., Bloom, S.L., Ésik, Z. and Stefanescu, Gh., Equational theories of relations and regular sets, in *Proc. 2nd Int. Colloq. on Words, Languages and Combinatorics* (Kyoto, 1992), pp. 40–48, World Scientific, Singapore, 1994.

[5] Birkenmeier, G.F., Groenewald, N.J. and Heatherly, H.E., Minimal and maximal ideals in rings with involution, *Beiträge zur Algebra und Geometrie* **38** (1997), 217–225.

[6] Bíró, B., Kiss, E.W. and Pálfy, P.P., On the congruence extension property, in *Universal Algebra* (Esztergom, 1977), Colloq. Math. Soc. János Bolyai, Vol. 29, pp. 129–151, North-Holland, Amsterdam, 1982.

[7] Biryukov, A.P., Varieties of idempotent semigroups, *Algebra i Logika* **9** (1970), 255–273 [in Russian].

[8] Bloom, S.L., Ésik, Z. and Stefanescu, Gh., Notes on equational theories of relations, *Algebra Universalis* **33** (1995), 98–128.

[9] Ćirić, M. and Bogdanović, S., Decompositions of semigroups induced by identities, *Semigroup Forum* **46** (1993), 329–346.

[10] Conway, J.H., *Regular Algebra and Finite Machines*, Chapman & Hall, London, 1971.

[11] Crvenković, S., On *-regular semigroups, in *Proc. 3rd Algebraic Conf.* (Beograd, 1982), pp. 51–57, Institute of Mathematics, Novi Sad, 1983.

[12] Crvenković, S. and Dolinka, I., Congruences on *-regular semigroups, *Periodica Math. Hungarica* **45** (2002), 1–13.

[13] Crvenković, S., Dolinka, I. and Ésik, Z., The variety of Kleene algebras with conversion is not finitely based, *Theoret. Comput. Sci.* **230** (2000), 235–245.

[14] Crvenković, S., Dolinka, I. and Ésik, Z., On equations of union-free regular languages, *Inform. Comput.* **164** (2001), 152–172.

[15] Crvenković, S., Dolinka, I. and Vinčić, M., Equational bases for some 0-direct unions of semigroups, *Studia Sci. Math. Hungarica* **36** (2000), 423–431.

[16] Crvenković, S., Dolinka, I. and Vinčić, M., On subdirect decomposition and varieties of some rings with involution, *Beiträge zur Algebra und Geometrie* **43** (2002), 423–432.

[17] Crvenković, S. and Madarász, R.Sz., On Kleene algebras, *Theoret. Comput. Sci.* **108** (1993), 17–24.

[18] Dolinka, I., A charaterization of groups in the class of *-regular semigroups, *Novi Sad J. Math.* **29** (1) (1999), 215–219.

[19] Dolinka, I., Remarks on varieties of involution bands, *Comm. in Algebra* **28** (2000), 2837–2852.

[20] Dolinka, I., All varieties of normal bands with involution, *Periodica Math. Hungarica* **40** (2000), 109–122.

[21] Dolinka, I., Minimal varieties of semirings with involution, *Algebra Universalis* **44** (2000), 143–151.

[22] Dolinka, I., On Kleene algebras of ternary co-relations, *Acta Cybernetica* **14** (2000), 583–595.

[23] Dolinka, I., *On Identities of Algebras of Regular Languages*, Ph.D. thesis, ix+143 pp., University of Novi Sad, 2000 [in Serbian].

[24] Dolinka, I., On the lattice of varieties of involution semigroups, *Semigroup Forum* **62** (2001), 438–459.

[25] Dolinka, I., Subdirectly irreducible bands with involution, *Acta Sci. Math. (Szeged)* **67** (2001), 535–554.

[26] Dolinka, I., Form twists to involution bands, *Filomat (Niš)* **15** (2001), 7–16.

[27] Dolinka, I., Idempotent distributive semirings with involution, *Int. J. Algebra Comput.* **13** (2003), 597–625.

[28] Dolinka, I., Regular *-semigroup varieties with the amalgamation property, *Semigroup Forum* **67** (2003), 419–428.

[29] Dolinka, I., Varieties of involution semilattices of Archimedean semigroups, *submitted for publication.*

[30] Dolinka, I. and Mudrinski, N., On subdirect decomposition and varieties of some rings with involution. II, *submitted for publication.*

[31] Dolinka, I. and Vinčić, M., Involutorial Płonka sums, *Periodica Math. Hungarica* **46** (2003), 17–31.

[32] Drazin, M.P., Regular semigroups with involution, in *Proc. Symp. on Regular Semigroups* (DeKalb, 1979), pp. 29–46, University of Northern Illinois, DeKalb, 1979.

[33] Easdown, D. and Munn, W.D., On semigroups with involution, *Bull. Austral. Math. Soc.* **48** (1993), 93–100.

[34] Eilenberg, S., *Automata, Languages and Machines, Vol. A*, Academic Press, New York, 1974.

[35] Ésik, Z. and Bernátsky, L., Equational properties of Kleene algebras of relations with conversion, *Theoret. Comput. Sci.* **137** (1995), 237–251.

[36] Evans, T., The lattice of semigroup varieties, *Semigroup Forum* **2** (1971), 1–43.

[37] Fajtlowicz, S., Equationally complete semigroups with involution, *Algebra Universalis* **1** (1972), 355–358.

[38] Fennemore, C.F., All varieties of bands. I, *Math. Nachr.* **48** (1971), 237–252. II, *ibid.*, 253–262.

[39] Foulis, D.J., Baer *-semigroups, *Proc. Amer. Math. Soc.* **11** (1960), 648–655.

[40] Gerhard, J.A., The lattice of equational classes of idempotent semigroups, *J. Algebra* **15** (1970), 195–224.

[41] Gerhard, J.A., Subdirectly irreducible idempotent semigroups, *Pacific J. Math.* **39** (1971), 669–676.

[42] Gerhard, J.A., Injectives in equational classes of idempotent semigroups, *Semigroup Forum* **9** (1974), 36–53.

[43] Gerhard, J.A. and Petrich, M., Free involutorial completely simple semigroups, *Canadian J. Math.* **37** (1985), 281–295.

[44] Gerhard, J.A. and Petrich, M., Free bands and free *-bands, *Glasgow J. Math.* **28** (1986), 161–179.

[45] Golan, J.S., *The Theory of Semirings with Applications in Mathematics and Theoretical Computer Science*, Longman Scientific and Techincal, New York, 1993.

[46] Graczyńska, E., On normal and regular identities, *Algebra Universalis* **27** (1990), 387–397.

[47] Hall, T.E., Inverse semigroup varieties with the amalgamation property, *Semigroup Forum* **16** (1978), 37–51.

[48] Hall, T.E., Amalgamation for inverse and generalized inverse semigroups, *Trans. Amer. Math. Soc.* **310** (1988), 313–323.

[49] Hall, T.E., Regular semigroups: amalgamation and the lattice of existence varieties, *Algebra Universalis* **28** (1991), 79–102.

[50] Hall, T.E. and Imaoka, T., Representations and amalgamation of generalized inverse *-semigroups, *Semigroup Forum* **58** (1999), 126–141.

[51] Heatherly, H.E., Lee, E.K.S. and Wiegandt, R., Involutions on universal algebras, in *Nearrings, Nearfields and K-Loops* (Hamburg, 1995), pp. 269–282, Kluwer, Dordrecht, 1997.

[52] Herstein, I.N., *Rings with Involution*, Chicago Lectures in Math., University of Chicago Press, Chicago, London, 1976.

[53] Hobby, D. and McKenzie, R., *The Structure of Finite Algebras*, Contemporary Mathematics Series, Vol. 76, AMS, Providence, 1988.

[54] Hoehnke, H.J., Über antiautomorphe und involutorische primitive Halbgruppen, *Czechoslovak J. Math.* **15 (90)** (1965), 50–63.

[55] Horn, A. and Kimura, N., The category of semilattices, *Algebra Universalis* **1** (1971), 26–38.

[56] Howie, J.M., *Fundamentals of Semigroup Theory*, Oxford University Press & Clarendon Press, Oxford, 1995.

[57] Imaoka, T., Free products with amalgamation of commutative inverse semigroups, *J. Austral Math. Soc. (Ser. A)* **12** (1976), 246–251.

[58] Imaoka, T., Free products and amalgamation of generalized inverse *-semigroups, *Mem. Fac. Sci. Shimane Univ.* **21** (1987), 55–64.

[59] Jónsson, B., The theory of binary relations, in *Algebraic Logic* (Budapest, 1988), Colloq. Math. Soc. János Bolyai, Vol. 54, pp. 245–292, North-Holland, Amsterdam, 1991.

[60] Kalicki, J. and Scott, D., Equational completeness of abstract algebras, *Indag. Math.* **17** (1955), 650–659.

[61] Kalman, J.A., Subdirect decomposition of distributive quasilattices, *Fund. Math.* **71** (1971), 161–163.

[62] Kharlampovich, O.G. and Sapir M.V., Algorithmic problems in varieties, *Int. J. Algebra Comp.* **5** (1995), 379–602.

[63] Kozen, D.C., On Kleene algebras and closed semirings, in *Mathematical Foundations of Computer Science*, Lecture Notes in Comput. Sci., Vol. 452, pp. 26–47, Springer-Verlag, New York, 1990.

[64] Kozen, D.C., A completeness theorem for Kleene algebras and the algebra of regular events, *Inform. Comput.* **110** (1994), 366–390.

[65] Kozen D.C., *Automata and Computability*, Springer-Verlag, 1997.

[66] Krob, D., Complete systems of $\mathcal{B}$-rational identities, *Theoret. Comput. Sci.* **89** (1991), 207–343.

[67] Kuich, W. and Salomaa, A., *Semirings, Automata and Languages*, EATCS Monographs on Theoret. Comput. Sci., Springer-Verlag, New York, 1986.

[68] Kuřil, M. and Polák, L., On varieties of semilattice ordered semigroups, *manuscript, 23pp.*

[69] Lakser, H., Padmanabhan, R. and Platt, C.R., Subdirect decomposition of Płonka sums, *Duke Math. J.* **39** (1972), 485–488.

[70] Madarász, R.Sz. and Crvenković, S., *Relation Algebras*, Mathematical Institute of SANU, Beograd, 1992 [in Serbian].

[71] McKenzie, R., An algebraic version of categorical equivalence for varieties and more general algebraic categories, in *Algebra and Logic* (Pontignano, 1994), Lecture Notes in Pure and Appl. Math., Vol. 180, pp. 211–243, Marcel Dekker, New York, 1996.

[72] McKenzie, R., McNulty, G. and Taylor, W., *Algebras, Lattices, Varieties, Vol. I*, Wadsworth & Brooks/Cole, Monterey, 1987.

[73] McKenzie, R. and Romanowska, A., Varieties of $\cdot$-distributive bisemilattices, in *Contributions to General Algebra* (Klagenfurt, 1978), pp. 213–218, Verlag J. Heyn, Klagenfurt, 1979.

[74] McKenzie, R. and Valeriote, M., *The Structure of Decidable Locally Finite Varieties*, Birkhäuser, Boston, 1989.

[75] Mel'nik, I.I., Nilpotent shifts of varieties, *Mat. Zametki* **14** (1973), 703–712 [in Russian].

[76] Nambooripad, K.S.S. and Pastijn, F.J., Regular involution semigroups, in *Semigroups* (Szeged, 1981), Colloq. Math. Soc. János Bolyai, Vol. 39, pp. 199–249, North-Holland, Amsterdam, 1985.

[77] Neumann, H., *Varieties of Groups*, Springer-Verlag, Berlin, 1963.

[78] Nordahl, T.E. and Scheiblich, H.E., Regular *-semigroups, *Semigroup Forum* **16** (1978), 369–377.

[79] Pastijn, F.J., Constructions of varieties that satisfy the amalgamation property and the congruence extension property, *Studia Sci. Math. Hungarica* **17** (1982), 101–111.

[80] Pastijn, F.J., Idempotent distributive semirings II, *Semigroup Forum* **26** (1983), 151–166.

[81] Pastijn, F.J. and Guo, Y.Q., The lattice of idempotent distributive semiring varieties, *Science in China (Ser. A)* **42** (1999), 785–804.

[82] Pastijn, F.J. and Romanowska, A., Idempotent distributive semirings I, *Acta Sci. Math. (Szeged)* **44** (1982), 239–253.

[83] Penrose, R., A generalized inverse for matrices, *Proc. Cambridge Phil. Soc.* **51** (1955), 406–413

[84] Perkins, P., Bases of equational theories of semigroups, *J. Algebra* **11** (1969), 298–314.

[85] Petrich, M., *Introduction to Semigroups*, Merrill, Columbus, 1973.

[86] Petrich, M., *Inverse Semgroups*, John Wiley & Sons, New York, 1984.

[87] Petrich, M., Certain varieties of completely regular *-semigroups, *Bollettino U.M.I., Sect. B (Ser. VI)* **4** (1985), 343–370.

[88] Płonka, J., On distributive quasi-lattices, *Fund. Math.* **60** (1967), 191–200.

[89] Płonka, J., On a method of construction of abstract algebras, *Fund. Math.* **61** (1967), 183–189.

[90] Płonka, J., On equational classes of abstract algebras defined by regular identities, *Fund. Math.* **64** (1969), 241–247.

[91] Polák, L., A solution of the word problem for free *-regular semigroups, *J. Pure Appl. Algebra* **157** (2001), 107–114.

[92] Polin, S.V., Minimal varieties of semirings, *Mat. Zametki* **27** (1980), 527–537 [in Russian].

[93] Reilly, N.R., A class of regular *-semigroups, Semigroup Forum **18** (1979), 385–386.

[94] Roddy, M.S., On the word problem for orthocomplemented modular lattices, *Canadian J. Math.* **41** (1989), 961–1004.

[95] Romanowska, A., Free idempotent distributive semirings with a semilattice reduct, *Math. Japonica* **27** (1982), 467–481.

[96] Romanowska, A., Idempotent distributive semirings with a semilattice reduct, *Math. Japonica* **27** (1982), 483–493.

[97] Rowen, L.H., *Ring Theory, Vol. I*, Academic Press, London, New York, 1988.

[98] Sapir, M.V., Inherently non-finitely based finite semigroups, *Mat. Sb.* **133** (1987), 154–166 [in Russian].

[99] Sapir, M.V., Identities of finite inverse semigroups, *Int. J. Algebra Comp.* **3** (1993), 115–124.

[100] Sapir, M.V. and Sukhanov, E.V., On varieties of periodic semigroups, *Izv. Vysš. Uč. Zav. Mat.* **4 (227)** (1981), 48–55 [in Russian].

[101] Scheiblich, H.E., Projective and injective bands with involution, *J. Algebra* **109** (1987), 281–291.

[102] Schein, B.M., On the theory of generalized groups, *Dokl. Akad. Nauk SSSR* **153** (1963), 296–299 [in Russian].

[103] Schein, B.M., Atomic semiheaps and involutorial semigroups, *Izv. Vysš. Uč. Zav. Mat.* **3** (1965), 172–184 [in Russian].

[104] Schein, B.M., Homomorphisms and subdirect decopositions of semigroups, *Pacific J. Math.* **17** (1966), 529–547.

[105] Schein, B.M., Completions, translational hulls and ideal extenstions of inverse semigroups, *Czechoslovak Math. J.* **23 (98)** (1973), 575–610.

[106] Schein, B.M., Representation of involuted semigroups by binary relations, *Fund. Math.* **82** (1974), 121–141.

[107] Schein, B.M., Injectives in certain classes of semigroups, *Semigroup Forum* **9** (1974), 159–171.

[108] Skornjakov, L.A., *Complemented Modular Lattices and Regular Rings*, Oliver and Boyd, London, 1964.

[109] Szendrei, Á., A survey on strictly simple algebras and minimal varieties, in *Universal Algebra and Quasigroup Theory*, pp. 209–239, Heldermann Verlag, Berlin, 1992.

[110] Thierrin, G., Sur la structure des demi-groupes, *Publ. Sci. Univ. Alger. (Sér. A3)* **2** (1956), 161–171.

[111] Tiščenko, A.V., The finiteness of a base of identities for five-element monoids, *Semigroup Forum* **20** (1980), 171–186.

[112] Trahtman, A.N., A basis of identities of the five-element Brandt semigroup, *Mat. Zap. Ural. Gos. Univ.* **12** (1981), 147–149 [in Russian].

[113] Vagner, V.V., Generalized heaps and generalized groups with a transitive compatibility relation, *Uč. Zap. Saratov. Univ. Meh.-Mat.* **70** (1961), 25–39 [in Russian].

[114] von Neumann, J., *Continuous Geometry*, Princeton University Press, Princeton, 1960.

[115] Wiegandt, R., On the structure of involution rings with chain condition, *Vietnam J. Math.* **21** (1993), 1–12.

[116] Yamada, M., Finitely generated free *-bands, *Semigroup Forum* **29** (1984), 13–16.

[117] Yamada, M., On the multiplicative semigroups of regular rings with special involution, *Simon Stevin* **59** (1985), 51–57.

DEPARTMENT OF MATHEMATICS AND INFORMATICS, UNIVERSITY OF NOVI SAD, TRG DOSITEJA OBRADOVIĆA 4, 21000 NOVI SAD, SERBIA AND MONTENEGRO

E-mail address: sima@eunet.yu, dockie@im.ns.ac.yu

Difference Schemes for Partial Differential Equations with Generalized Solutions and Singular Coefficients

Boško S. Jovanović

Dedicated to Professor Veselin Perić on the occasion of his 70th birthday

University of Belgrade, Faculty of Mathematics, Studentski trg 16,
11000 Belgrade, Yugoslavia, e–mail: bosko@matf.bg.ac.yu

Abstract. A survey of results concerning the convergence of finite difference schemes for boundary value problems with generalized solutions from Sobolev space is presented. In particular, difference schemes for some problems with singular coefficients are investigated.

Mathematics Subject Classification (1991): 65N15, 65M15

1 Introduction

Finite difference schemes (FDSs) are often used for approximation of boundary value problems (BVPs) with generalized solutions. In such cases it is preferable to have the convergence result for the minimal smoothness of input data. This leads to several problems as: the right hand side of the equation and the solution may be discontinuous functions; small smoothness of the solution requires the convergence rate estimate in the weak norm; coefficients of equation does not belong to standard Sobolev spaces etc. In the case of difference schemes on nonuniform meshes the order of local approximation is usually reduced. The accuracy of the method can be increased by using the approximation of the considered differential equation in some non–mesh points and special averaging operators. In the case of problems with singular coefficients the solution does not belongs to the standard Sobolev spaces. Also there arise nonstandard conjugation conditions.

In this paper we give a survey of techniques for overcoming these problems. Special attention is paid to deriving convergence rate estimates consistent with the smoothness of input data.

2 Poisson Equation

As a model problem we consider the Dirichlet BVP for the Poisson equation in the square $\Omega = (0, 1)^2$:

$$-\Delta u = f(x), \quad x = (x_1, x_2) \in \Omega; \qquad u(x) = 0, \quad x \in \Gamma = \partial\Omega. \qquad (1)$$

We assume that the solution of BVP (1) is sufficiently smooth, that is, the function $f(x)$ satisfies all the necessery conditions for that.

Let $\overline{\omega}_h$ be the uniform mesh in $\overline{\Omega}$ with the step size h, $\omega_h = \overline{\omega}_h \cap \Omega$ and $\gamma_h = \overline{\omega}_h \cap \Gamma$. We define finite differences in the usual way [31]:

$$v_{x_i} = (v^{+i} - v)/h\,, \qquad v_{\bar{x}_i} = (v - v^{-i})/h\,,$$

where $v^{\pm i}(x) = v(x \pm h r_i)$, and r_i denotes the unit vector of the x_i axis. With

$$\|v\|^2_{L_2(\omega_h)} = h^2 \sum_{x \in \omega_h} v^2(x)$$

we denote discrete L_2–norm in ω_h. We also introduce discrete Sobolev norms $\|v\|_{W_2^k(\omega_h)}$ $(k = 1,\, 2,\, ...)$.

We approxsimate (1) with the standard five–point FDS:

$$-\Delta_h v = f\,, \quad x \in \omega_h\,; \qquad v = 0\,, \quad x \in \gamma_h\,. \tag{2}$$

The error $z = u - v$ satisfies the conditions:

$$-\Delta_h z = \psi\,, \quad x \in \omega_h\,; \qquad z = 0\,, \quad x \in \gamma_h\,, \tag{3}$$

where $\psi = \Delta\, u - \Delta_h\, u = \left(\frac{\partial^2 u}{\partial x_1^2} - u_{x_1 \bar{x}_1}\right) + \left(\frac{\partial^2 u}{\partial x_2^2} - u_{x_2 \bar{x}_2}\right) = \psi_1 + \psi_2\,.$

From inequality (see [32]) $\|\Delta_h\, z\|_{L_2(\omega_h)} \geq C_0\, \|z\|_{W_2^2(\omega_h)}$ immediatelly follows a priori estimate

$$\|z\|_{W_2^2(\omega_h)} \leq C\, \|\psi\|_{L_2(\omega_h)}\,. \tag{4}$$

Here C denotes a positive generic constant independent of u and the mesh step–size. In different formulas C may take different values. In such a way, to prove the convergence of FDS (2) we must estimate ψ. From Taylor's formula follows: $\psi_i(x) = \frac{h^2}{12}\,\frac{\partial^4 u(\tilde{x})}{\partial x_i^4}$, where $\tilde{x}$ is some midpoint. From here one immediatelly obtains:

$$\|z\|_{W_2^2(\omega_h)} \leq C\, h^2\, \|u\|_{C^4(\bar{\Omega})}\,.$$

More precise estimate may be obtained using integral representation of residual. We have

$$\psi_1(x) = \frac{1}{h^2} \int_{x_1-h}^{x_1+h} \int_{x_2-h}^{x_2+h} \left(1 - \frac{|x_1' - x_1|}{h}\right) \left(1 - \frac{|x_2' - x_2|}{h}\right) \times$$

$$\times \left(\int_{x_1'}^{x_1} \int_0^{x_1''} \frac{\partial^4 u(x_1''',\, x_2')}{\partial x_1^4}\, dx_1'''\, dx_1'' + \int_{x_1'}^{x_1} \int_{x_2'}^{x_2} \frac{\partial^4 u(x_1'',\, x_2'')}{\partial x_1^3\, \partial x_2}\, dx_2''\, dx_1'' \right) dx_2'\, dx_1'$$

and an analogous formula for ψ_2. Therefrom follows

$$|\psi(x)| \leq C\, h\, \|u\|_{W_2^4(e)} \qquad \text{where} \qquad e = (x_1 - h,\, x_1 + h) \times (x_2 - h,\, x_2 + h)\,.$$

Summing over the mesh ω_h one obtains $\|\psi\|_{L_2(\omega_h)} \le C\,h^2\,\|u\|_{W_2^4(\Omega)}$, wherefrom follows

$$\|z\|_{W_2^2(\omega_h)} \le C\,h^2\,\|u\|_{W_2^4(\Omega)}\,. \tag{5}$$

The estimate (5) can be obtained also by application of the Bramble–Hilbert lemma (see [1], [3]). Moreover, as the value of ψ in the node $x \in \omega$ is a bounded linear functional of $u \in W_2^s(e)$, for $s > 3$, which vanishes on polynomials of third degree, applying the Dupont–Scott lemma [3] we obtain

$$|\psi(x)| \le C\,h^{s-3}\,|u|_{W_s^s(e)}\,, \qquad 3 < s \le 4\,.$$

From here, after summation ofer the mesh ω_h we obtain more general result:

$$\|z\|_{W_2^2(\omega_h)} \le C\,h^{s-2}\,|u|_{W_2^s(\Omega)}\,, \tag{6}$$

for $3 < s \le 4$.

For $s \le 3$ the right hand side of (1) and (2) may be discontinuous function, and consequently, FDS (2) is not well defined. To obtain a well–defined FDS we average $f(x)$ using Steklov averaging operators:

$$T_i f(x) = T_i^- f(x + 0.5\,h\,r_i) = T_i^+ f(x - 0.5\,h\,r_i) = \int_{-1/2}^{1/2} f(x + h\,y\,r_i)\,dy\,.$$

These operators commute and satisfy the following relations

$$T_i^+ T_i^- = T_i^2\,, \quad T_i^- \frac{\partial f}{\partial x_i} = f_{\bar{x}_i}\,, \quad T_i^+ \frac{\partial f}{\partial x_i} = f_{x_i}\,, \quad T_i^2 \frac{\partial^2 f}{\partial x_i^2} = f_{x_i \bar{x}_i}\,.$$

For $s < 2$, the convergence of FDS (2) does not follow from (6). Consequently, the weaker norms must be used to prove the convergence. The following assertion is valid (see [32], [4]).

Lemma 1: *If in (3)* $\psi = \eta_{1,\,\bar{x}_1} + \eta_{2,\,\bar{x}_2}$, *then*

$$\|z\|_{W_2^1(\omega_h)} \le C\left(\|\eta_1\|_{L_2(\omega_h)} + \|\eta_2\|_{L_2(\omega_h)}\right). \tag{7}$$

If $\psi = \zeta_{1,\,x_1\bar{x}_1} + \zeta_{2,\,x_2\bar{x}_2}$ *and* $\zeta_i = 0$ *for* $x_i = 0$ *then*

$$\|z\|_{L_2(\omega_h)} \le C\left(\|\zeta_1\|_{L_2(\omega_h)} + \|\zeta_2\|_{L_2(\omega_h)}\right). \tag{8}$$

Let us consider FDS with averaged right hand side [4]:

$$-\Delta_h\,v = T_1^2 T_2^2 f\,, \quad x \in \omega_h\,; \qquad v = 0\,, \quad x \in \gamma_h\,. \tag{9}$$

The error $z = u - v$ satisfies the conditions (3), where: $\psi = \psi_1 + \psi_2$, $\psi_i = \zeta_{i,\,x_i\bar{x}_i}$, $\zeta_i = T_{3-i}^2 u - u$, $i = 1, 2$. By lemma 1 one obtains a priori estimates (4), (8) and

$$\|z\|_{W_2^1(\omega_h)} \le C\left(\|\zeta_{1,\,x_1}\|_{L_2(\omega_h)} + \|\zeta_{2,\,x_2}\|_{L_2(\omega_h)}\right).$$

Using Dupont–Scott lemma, analogously as in the previous case, one obtains the following convergence rate estimates: (6) for $2 \le s \le 4$,

$$\|z\|_{W_2^1(\omega_h)} \le C\, h^{s-1}\, \|u\|_{W_2^s(\Omega)}, \qquad 1 < s \le 3, \tag{10}$$

and

$$\|z\|_{L_2(\omega_h)} \le C\, h^{s}\, \|u\|_{W_2^s(\Omega)}, \qquad 1 < s \le 2. \tag{11}$$

In the case $0 < s \le 1$ the solution of (1) may be non–continuous function. Let us consider FDS

$$-\Delta_h\, u = T_1^3 T_2^3 f\,, \quad x \in \omega_h\,; \qquad v = 0\,, \quad x \in \gamma_h$$

and define the error in the following manner: $z = T_1 T_2\, u - v$. Similarly as in the previous cases one obtains estimate (11) for $0 < s \le 2$ [4].

Analogous results hold for the FDSs on non–uniform meshes. For example, in [17] for a family of nine–points difference schemes approximating BVP (1) on an arbitrary non–uniform rectangular mesh $\hat{\omega}_h$ are obtained convergence rate estimates

$$\|z\|_{W_2^1(\hat{\omega}_h)} \le C\, h_{max}^2\, \|u\|_{W_2^3(\Omega)}$$

and

$$\|z\|_{L_2(\hat{\omega}_h)} \le C\, h_{max}^2\, \|u\|_{W_2^2(\Omega)}\,.$$

Convergence rate estimates of the form

$$\|z\|_{W_2^k(\omega_h)} \le C\, h^{s-k}\, \|u\|_{W_2^s(\Omega)}, \qquad s \ge k,$$

are called consistent with the smoothness of the solution of BVP (1) (see [26]). For a broad class of FDS such estimates are obtained in [36], [25], [4] etc. A review of results on the convergence of FDSs is given in [12]. An extensive bibliography can be find in [9].

3 Estimates in $\mathbf{W}_p^k$ Norms

Let us consider again the BVP (1) assuming that its solution belongs to Sobolev space $W_p^s(\Omega)$, $1 < p < \infty$. As in previous case, we approximate (1) with the FDS (9). The error $z = u - v$ satisfies the conditions (3), where: $\psi = \psi_1 + \psi_2$, $\psi_i = \zeta_{i,\, x_i \bar{x}_i}$, $\zeta_i = T_{3-i}^2\, u - u$, $i = 1, 2$.

With $\|v\|_{L_p(\omega_h)} = \big(h^2 \sum_{x \in \omega_h} |v(x)|^p\big)^{1/p}$ we denote discrete L_p–norm in ω_h. We also define discrete Sobolev norms $\|v\|_{W_p^k(\omega_h)}$ $(k = 1, 2, ...)$.

The following analog of lemma 1 can be proved using theory of discrete Fourier multipliers [30].

Lemma 2: *FDS (3) satisfies a priori estimate*

$$\|z\|_{W_p^2(\omega_h)} \le C\, \|\psi\|_{L_p(\omega_h)}\,.$$

If $\psi = \eta_{1,\,\bar{x}_1} + \eta_{2,\,\bar{x}_2}$, *then*

$$\|z\|_{W_p^1(\omega_h)} \le C \left(\|\eta_1\|_{L_p(\omega_h)} + \|\eta_2\|_{L_p(\omega_h)} \right).$$

If $\psi = \zeta_{1,\,x_1\bar{x}_1} + \zeta_{2,\,x_2\bar{x}_2}$ *and* $\zeta_i = 0$ *for* $x_i = 0$ *then*

$$\|z\|_{L_p(\omega_h)} \le C \left(\|\zeta_1\|_{L_p(\omega_h)} + \|\zeta_2\|_{L_p(\omega_h)} \right).$$

Estimating $\zeta_{i,\,x_i\bar{x}_i}$, $\zeta_{i,\,x_i}$ and ζ_i by Dupont–Scott lemma one obtains the following convergence rate estimates [33], [2]:

$$\|z\|_{W_p^2(\omega_h)} \le C\, h^{s-2}\, \|u\|_{W_p^s(\Omega)}, \qquad 2 < s \le 4,$$

$$\|z\|_{W_p^1(\omega_h)} \le C\, h^{s-1}\, \|u\|_{W_p^s(\Omega)}, \qquad \max\{1,\, 2/p\} < s \le 3,$$

and

$$\|z\|_{L_p(\omega_h)} \le C\, h^{s}\, \|u\|_{W_p^s(\Omega)}, \qquad 2/p < s \le 2.$$

In the case when $0 < s \le 2/p$ the solution of BVP (1) may be non–continuous function. In this case we may define the error as $z = T_1 T_2\, u - v$ and consider the FDS with stronger averaged right hand side.

4 Technique Based on Interpolation of Hilbert Spaces

As we have been seen, for integer values of smoothness parameter s convergence rate estimates can be constructed "elementary", without the Bramble–Hilbert lemma. Using such estimates and the interpolation theory of Hilbert spaces [27] one easily obtains corresponding estimates for non–integer s.

Let X and Y be two Hilbert spaces and let X be continuously imbeded in Y. Let $0 < \theta < 1$ and let $[X,\, Y]_\theta$ denotes the intermediate space obtained by interpolation [27]. Then $X \subset [X,\, Y]_\theta \subset Y$ and for every $u \in X$ the inequality

$$\|u\|_{[X,\,Y]_\theta} \le C_\theta\, \|u\|_X^{1-\theta}\, \|u\|_Y^{\theta} \tag{12}$$

holds.

Let $W_2^s(\Omega)$ be Sobolev spaces in Ω. Let us introduce also the spaces $W_2^s((0,\,T);\, W_2^r(\Omega))$ and anisotrophic Sobolev spaces in $Q = \Omega \times (0,\,T)$: $W_2^{s,\,r}(Q) = W_2^0((0,\,T);\, W_2^s(\Omega)) \cap W_2^r((0,\,T);\, W_2^0(\Omega))$.

Lemma 3: *Let* $s_1,\, s_2,\, r_1,\, r_2 \ge 0$ *and* $0 < \theta < 1$. *Then*

$$[W_2^{s_1}(\Omega),\, W_2^{s_2}(\Omega)]_\theta = W_2^{(1-\theta)\,s_1 + \theta\,s_2}(\Omega),$$

and

$$
\begin{aligned}
&\left[W_2^{s_1}((0,T);\, W_2^{r_1}(\Omega)),\, W_2^{s_2}((0,T);\, W_2^{r_2}(\Omega))\right]_\theta \\
&\quad = W_2^{(1-\theta)\, s_1+\theta\, s_2}\big((0,T);\, W_2^{(1-\theta)\, r_1+\theta\, r_2}(\Omega)\big)\,.
\end{aligned}
$$

Lemma 4 [27]: *Let A be a bounded linear operator from X_i into Y_i $(i = 0,\, 1)$. Then A is also bounded linear operator from $[X_0,\, X_1]_\theta$ into $[Y_0,\, Y_1]_\theta$ and the following relation holds*

$$
\|A\|_{[X_0,\, X_1]_\theta \to [Y_0,\, Y_1]_\theta} \le C_\theta\, \|A\|^{1-\theta}_{X_0 \to X_1}\, \|A\|^{\theta}_{Y_0 \to Y_1}\,.
$$

Let us consider again FDS (9). Similarly as in the chapter 2, one easily shows that

$$
\|z\|_{W_2^2(\omega_h)} \le C\, h^2\, \|u\|_{W_2^4(\Omega)} \qquad \text{and} \qquad \|z\|_{W_2^2(\omega_h)} \le C\, \|u\|_{W_2^2(\Omega)}\,,
$$

wherefrom, using lemma 4, one immediately obtains estimate (6) for $2 \le s \le 4$. In an analogous manner one obtains convergence rate estimates in other discrete norms (see [10], [11]).

5 Equations with Variable Coefficients

Let us now consider elliptic equation with variable coefficients:

$$
\mathcal{L}\, u \equiv -\sum_{i,\, j=1}^{2} \frac{\partial}{\partial x_i}\left(a_{ij}\, \frac{\partial u}{\partial x_j}\right) = f(x)\,, \quad x \in \Omega \tag{13}
$$

with homogeneous Dirichlet boundary condition. We assume that $u \in W_2^s(\Omega)$ and $f \in W_2^{s-2}(\Omega)$.

Let V and W be two function spaces in the same domain. The space of multipliers $M(V,\, W)$ is defined by: $M(V, W) = \{a(x) \,:\, a(x)\, v(x) \in W,\ \forall\, v(x) \in V\}$, $M(V) = M(V, V)$ (see [29]). It is easy to see that coeficients a_{ij} of equation (13) belong to the space of multipliers $M\big(W_2^{s-1}(\Omega)\big)$.

The following relations are valid [9]:

$$
\begin{aligned}
W_2^{|s-1|}(\Omega) &= M\big(W_2^{s-1}(\Omega)\big)\,, \qquad |s-1| > 1\,, \\
W_{2/|s-1|}^{|s-1|+\varepsilon}(\Omega) &\subset M\big(W_2^{s-1}(\Omega)\big)\,, \quad \varepsilon > 0, \qquad 0 < |s-1| < 1\,, \\
L_\infty(\Omega) = M\big(L_2(\Omega)\big) &= M\big(W_2^{s-1}(\Omega)\big)\,, \qquad s = 1\,.
\end{aligned}
$$

Let us consider FDS

$$
\mathcal{L}_h\, v \equiv -\frac{1}{2} \sum_{i,\, j=1}^{2} \left[\left(a_{ij}\, v_{x_j}\right)_{\bar{x}_i} + \left(a_{ij}\, v_{\bar{x}_j}\right)_{x_i}\right] = T_1^2 T_2^2 f\,, \quad x \in \omega_h \tag{14}
$$

with the previous boundary condition. The error $z = u - v$ satisfies conditions

$$\mathcal{L}_h z = \sum_{i,\,j=1}^{2} \eta_{ij,\,\bar{x}_i}\,, \quad x \in \omega_h\,; \qquad z = 0\,, \quad x \in \gamma_h$$

where $\eta_{ij} = T_i^+ T_{3-i}^2\Big(a_{ij}\,\frac{\partial u}{\partial x_j}\Big) - \frac{1}{2}\,\big(a_{ij}\,u_{x_j} + a_{ij}^{+i}\,u_{\bar{x}_j}^{+i}\big)$. The following a priori estimates hold

$$\|z\|_{W_2^2(\omega_h)} \leq C \sum_{i,\,j=1}^{2} \|\eta_{ij,\,\bar{x}_i}\|_{L_2(\omega_h)} \tag{15}$$

and

$$\|z\|_{W_2^1(\omega_h)} \leq C \sum_{i,\,j=1}^{2} \|\eta_{ij}\|_{L_2(\omega_h)}\,. \tag{16}$$

Using bilinear version of the Bramble–Hilbert lemma or interpolatory properties of bounded bilinear operators from (15) and (16) one obtains convergence rate estimates in the form (see [9], [11], [15])

$$\|z\|_{W_2^2(\omega_h)} \leq C\,h^{s-2} \max_{i,\,j} \|a_{ij}\|_{W_2^{s-1}(\Omega)}\,\|u\|_{W_2^s(\Omega)}\,, \qquad 2 < s \leq 4\,,$$

$$\|z\|_{W_2^1(\omega_h)} \leq C\,h^{s-1} \max_{i,\,j} \|a_{ij}\|_{W_2^{s-1}(\Omega)}\,\|u\|_{W_2^s(\Omega)}\,, \qquad 2 < s \leq 3\,,$$

and

$$\|z\|_{W_2^1(\omega_h)} \leq C\,h^{s-1} \max_{i,\,j} \|a_{ij}\|_{W_{2/(s-1)}^{s-1+\varepsilon}(\Omega)}\,\|u\|_{W_2^s(\Omega)}\,, \qquad 1 < s \leq 2\,.$$

Analogous results for the third boundary value problem are obtained in [20].

In multidimensional case $(n > 2)$ there arise additional problems caused by the discontinuity of right hand side of equation ($f \in W_2^{s-2}(\Omega) \not\subset C(\bar{\Omega})$ for $s \leq 2 + n/2$) or its solution ($u \in W_2^s(\Omega) \not\subset C(\bar{\Omega})$ for $s \leq n/2$). These problems may be resolved by convenient averaging. Note also that $M\big(W_2^{s-1}(\Omega)\big) \neq W_2^{s-1}(\Omega)$ for $s \leq 1 + n/2$.

6 Equations with Singular Coefficients

Interface problems occur in many physical applications. Such problems can be modeled by partial differential equations with singular coefficients. For example, as a model problem let us consider the Dirichlet problem

$$-\Delta\,u + c(x)\,\delta_S(x)\,u = f(x)\,, \quad x \in \Omega\,; \qquad u = 0\,, \quad x \in \Gamma\,, \tag{17}$$

where S is a continuous curve (for example closed curve), $S \subset \Omega$ and $\delta_S(x)$ is Dirac's delta distribution [35] concentrated on S. We suppose that $c(x) \in L_\infty(S)$ and $0 < C_0 \leq c(x) \leq C_1$ almost everywhere on S.

We assume for simplicity that the curve S separates Ω into two regions: $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$, $\Omega_1 \cap \Omega_2 = \emptyset$. Then, at some assumptions for smoothness, the BVP (17) can be rewritten as follows:

$$-\Delta u = f(x)\,, \quad x \in \Omega_1 \cup \Omega_2\,; \quad [u]_S = 0\,, \quad \left[\frac{\partial u}{\partial \nu}\right]_S = c(x)\, u,$$

where $\partial u / \partial \nu$ – is the normal derivative.

We approximate the BVP (17) on the mesh $\overline{\omega}_h$ with the following FDS (see [16])

$$-\Delta_h v + \alpha\, v = T_1^2 T_2^2 f \quad \text{in} \quad \omega_h\,; \qquad v = 0 \quad \text{on} \quad \gamma_h,$$

where

$$\alpha(x) = T_1^2 T_2^2 (c\, \delta_S) = \begin{cases} h^{-2} \displaystyle\int_{S(x)} \kappa(x,\, x')\, c(x')\, dS_{x'}\,, & x \in S_h\,, \\ 0\,, & x \in \omega_h \setminus S_h\,, \end{cases}$$

$\kappa(x,\, x') = \left(1 - \frac{|x_1' - x_1|}{h}\right)\left(1 - \frac{|x_2' - x_2|}{h}\right)$, $S(x) = S \cap e(x)$, $e(x) = (x_1 - h,\, x_1 + h) \times (x_2 - h,\, x_2 + h)$ is the cell attached to the internal node $x \in \omega_h$, and $S_h = \{x \in \omega_h \,:\, S(x) \neq \emptyset\}$.

The error $z = u - v$ satisfies the FDS

$$-\Delta_h z + \alpha\, z = -\psi_{1,\, \bar{x}_1 x_1} - \psi_{2,\, \bar{x}_2 x_2} + \chi \quad \text{in} \quad \omega_h\,; \qquad z = 0 \quad \text{on} \quad \gamma_h$$

where

$$\psi_i = u - T_{3-i}^2 u\,, \quad i = 1,\, 2,$$
$$\chi = \alpha\, u - h^{-2} \int_{S(x)} \kappa(x,\, x')\, c(x')\, u(x')\, dS_{x'}\,, \quad x \in S_h$$
$$\chi = 0\,, \quad x \in \omega_h \setminus S_h\,.$$

The a priori estimate

$$\|z\|_{W_2^1(\omega_h)} \le C \left[\|\psi_{1,\, x_1}\|_{L_2(\omega_h)} + \|\psi_{2,\, x_2}\|_{L_2(\omega_h)} + \left(h^2 \sum_{x \in S_h} \frac{\chi^2}{\alpha}\right)^{1/2}\right] \tag{18}$$

is satisfied. Estimating the terms in the right-hand side of (18) using the Bramble-Hilbert lemma, we obtain the following inequalities

$$\|z\|_{W_2^1(\omega_h)} \le C\, h^{s-1}\, \|u\|_{W_2^s(\Omega)}\,, \qquad 1 < s < 3/2$$

and

$$\|z\|_{W_2^1(\omega_h)} \le C\, h^{s-1} \left(\|u\|_{W_2^s(\Omega_1)} + \|u\|_{W_2^s(\Omega_2)} + \|u\|_{W_2^{s-1/2}(\Omega)}\right), \quad 3/2 \le s < 2\,.$$

In the case when S is a segment parallel to one of the coordinate axes an improved convergence rate estimate holds. Let, for example, S is given by the equation $x_2 = j_0 h$ (j_0 – integer). Then

$$\|z\|_{W_2^1(\omega_h)} \le C\, h^2 \left(\left\| \frac{\partial^3 u}{\partial x_1^2\, \partial x_2} \right\|_{L_2(\Omega)} + \left\| \frac{\partial^3 u}{\partial x_1\, \partial x_2^2} \right\|_{L_2(\Omega_1)} \right.$$

$$\left. + \left\| \frac{\partial^3 u}{\partial x_1\, \partial x_2^2} \right\|_{L_2(\Omega_2)} + \|c\|_{W_2^2(0,1)}\, \|u\|_{W_2^2(S)} \right).$$

7 Parabolic Problems

In parabolic case analogous results hold. Let us consider the folowing initial–boundary value problem (IBVP)

$$\begin{aligned} &\frac{\partial u}{\partial t} + \mathcal{L}\, u = f(x,\, t) \qquad \text{in} \qquad Q = \Omega \times (0,\, T)\,; \\ &u(x,\, 0) = u_0(x)\,; \qquad u(x,\, t) = 0 \quad \text{on} \quad \Gamma \times (0,\, T)\,. \end{aligned} \tag{19}$$

Let us introduce the mesh $Q_{h\tau} = \omega_h \times \omega_\tau$, where ω_τ is uniform mesh with the step size τ in $(0,\, T)$. We also define discrete L_2–norm

$$\|v\|^2_{L_2(Q_{h\tau})} = h^2\, \tau \sum_{(x,t)\in Q_{h\tau}} v^2(x,\, t)\,,$$

and the discrete Sobolev norms $\|v\|_{W_2^{k,\, k/2}(Q_{h\tau})}$.

We consider implicit FDS

$$v_{\bar t} + \mathcal{L}_h\, v = T_1^2 T_2^2 T_t^-\, f\,,$$

with the corresponding boundary and initial conditions, where T_t^- is the Steklov averaging operator on t:

$$T_t^- f(x,\, t) = \frac{1}{\tau} \int_{t-\tau}^{t} f(x,\, t')\, dt'\,.$$

The error $z = u - v$ satisfies the equation

$$z_{\bar t} + \mathcal{L}_h\, z = \varphi_{\bar t} + \sum_{i,\, j=1}^{2} \eta_{ij,\, \bar x_i}$$

and homogeneous boundary and initial conditions. Here

$$\varphi = u - T_1^2 T_2^2 u\,, \qquad \eta_{ij} = T_i^+ T_{3-i}^2 T_t^- \left(a_{ij}\, \frac{\partial u}{\partial x_j} \right) - \frac{1}{2} \left(a_{ij}\, u_{x_j} + a_{ij}^{+i}\, u_{\bar x_j}^{+i} \right).$$

The following a priori estimates are valid

$$\|z\|_{W_2^{2,1}(Q_{h\tau})} \le C\Big(\|\varphi_{\bar t}\|_{L_2(Q_{h\tau})} + \sum_{i,j=1}^{2} \|\eta_{ij,\bar x_i}\|_{L_2(Q_{h\tau})}\Big)$$

and

$$\|z\|_{W_2^{1,1/2}(Q_{h\tau})} \le C\Big([\varphi]_{1/2} + \sum_{i,j=1}^{2} \|\eta_{ij}\|_{L_2(Q_{h\tau})}\Big),$$

where $[\varphi]_{1/2}^2 = h^2 \sum_{x\in\omega_h} \tau^2 \sum_{t,t'\in\bar\omega_\tau,\, t\ne t'} \frac{|\varphi(x,t)-\varphi(x,t')|^2}{|t-t'|^2}$. From here, in a similar manner as in the elliptic case, for $\tau \asymp h^2$, one obtains convergence rate estimates (see [5], [6])

$$\|z\|_{W_2^{2,1}(Q_{h\tau})} \le C\, h^{s-2}\,\big(\max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega)} + 1\big)\,\|u\|_{W_2^{s,s/2}(Q)},$$

for $2 < s \le 4$, and

$$\|z\|_{W_2^{1,1/2}(Q_{h\tau})} \le C\, h^{s-1}\,\Big(\max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega)} + \sqrt{\ln(1/h)}\Big)\,\|u\|_{W_2^{s,s/2}(Q)},$$

for $2 < s \le 3$. For $1 \le s \le 2$ the solution of IBVP (19) may be discontinuous. In that case the error may be defined as: $z = T_1T_2u - v$.

Problem with coefficients depending on t is considered in [13]. Similar results are obtained for FDS on non–uniform meshes (see [18], [19]).

8 Heat Equation with Concentrated Capacity

Let us consider the IBVP for the heat equation with the presence of concentrated capacity at interior point $x = \xi$ [28]:

$$\begin{gathered}
[c(x) + K\,\delta(x-\xi)]\,\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\Big(a(x)\,\frac{\partial u}{\partial x}\Big) = f(x,t), \quad (x,t)\in Q,\\
u(0,\,t) = 0, \quad u(1,\,t) = 0, \qquad 0 < t < T \\
u(x,\,0) = u_0(x), \qquad x\in(0,\,1),
\end{gathered} \tag{20}$$

where $Q = (0,\,1)\times(0,\,T)$, $K > 0$, $0 < c_1 \le a(x) \le c_2$, $0 < c_3 \le c(x) \le c_4$ and $\delta(x)$ is the Dirac's distribution. The solution of the IBVP (20) satisfies the equation

$$c(x)\,\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\Big(a(x)\,\frac{\partial u}{\partial x}\Big) = f(x,t),$$

for $(x,t)\in Q_1 = (0,\,\xi)\times(0,\,T)$ and $(x,t)\in Q_2 = (\xi,\,1)\times(0,\,T)$, while for $x = \xi$ the conjugation conditions

$$[u]_{x=\xi} \equiv u(\xi+0,t) - u(\xi-0,t) = 0, \qquad \Big[a\,\frac{\partial u}{\partial x}\Big]_{x=\xi} = K\,\frac{\partial u(\xi,t)}{\partial t}$$

are fulfilled.

We introduce the space $\widetilde{L}_2(0,1) = \widetilde{W}_2^0(0,1)$ of functions $w(x) \in L_2(0,1)$ equiped with inner product and norm

$$(u,w)_{\widetilde{L}_2(0,1)} = \int_0^1 u(x)\,w(x)\,dx + u(\xi)\,w(\xi)\,, \quad \|w\|_{\widetilde{L}_2(0,1)} = (u,w)^{1/2}_{\widetilde{L}_2(0,1)}\,.$$

Further we set $\widetilde{W}_2^1(0,1) = \overset{\circ}{W}{}_2^1(0,1)$ and $\widetilde{W}_2^k(0,1) = \overset{\circ}{W}{}_2^1(0,1) \cap W_2^k(0,\xi) \cap W_2^k(\xi,1)$, $k=2,\,3,\,\ldots$. We also define spaces $\widetilde{W}_2^{k,\,k/2}(Q) = L_2(0,T;\,\widetilde{W}_2^k(0,1)) \cap W_2^k(0,T;\,\widetilde{L}_2(0,1))$, $k=0,\,1,\,2,\,\ldots$.

Let ω_h be an uniform mesh in $(0,1)$ with the step–size h. For simplicity we assume that $\xi \in \omega_h$. We approximate the IBVP (20) on the mesh $Q_{h\tau} = \omega_h \times \omega_\tau$ by the implicit FDS with averaged right hand side (see [22], [23])

$$(c + K\,\delta_h)\,v_{\bar{t}} - (\tilde{a}\,v_{\bar{x}})_x = T_x^2 T_t^- f, \qquad (x,t) \in Q_{h\tau},$$

$$v(0,t) = 0, \quad v(1,t) = 0, \qquad t \in \omega_\tau,$$

$$v(x,0) = u_0(x), \qquad x \in \bar{\omega}_h,$$

where $\tilde{a}(x) = \big[a(x-0) + a(x-h+0)\big]/2$, $\delta_h = \delta_h(x-\xi) = \begin{cases} 0, & x \in \omega_h \setminus \{\xi\} \\ 1/h, & x = \xi \end{cases}$ is the mesh Dirac's function and T_x is the Steklov averaging operator on variable x.

Let $(v,w)_h$ be the discrete L_2-inner product in ω_h. Let us set $B_h w = (1+\delta_h)\,w$ and define the energy norms $\|w\|_{B_h} = (B_h w,\,w)_h^{1/2}$ and $\|w\|_{B_h^{-1}} = (B_h^{-1} w,\,w)_h^{1/2}$. We introduce the mesh Sobolev norms with weight operator B_h:

$$\|w\|^2_{\widetilde{L}_2(\omega_h)} = \|w\|^2_{B_h} = \|w\|^2_{L_2(\omega_h)} + w^2(\xi), \quad \|w\|^2_{\widetilde{W}_2^1(\omega_h)} = \|w_{\bar{x}}\|^2_{L_2(\omega_h)} + \|w\|^2_{\widetilde{L}_2(\omega_h)},$$

$$\|w\|^2_{\widetilde{W}_2^2(\omega_h)} = \|w_{\bar{x}x}\|^2_{B_h^{-1}} + \|w\|^2_{\widetilde{W}_2^1(\omega_h)}\,; \quad \|w\|^2_{\widetilde{L}_2(Q_{h\tau})} = \tau \sum_{t \in \omega_\tau} \|w(\,\cdot\,,t)\|^2_{\widetilde{L}_2(\omega_h)}\,,$$

$$\|w\|^2_{\widetilde{W}_2^{1,1/2}(Q_{h\tau})} = \tau \sum_{t \in \bar{\omega}_\tau} \|w(\,\cdot\,,t)\|^2_{\widetilde{W}_2^1(\omega_h)} + \tau^2 \sum_{t \in \bar{\omega}_\tau} \sum_{t' \in \bar{\omega}_\tau,\ t' \neq t} \frac{\|w(\,\cdot\,,t) - w(\,\cdot\,,t')\|^2_{B_h}}{|t-t'|^2}\,,$$

$$\|w\|^2_{\widetilde{W}_2^{2,1}(Q_{h\tau})} = \tau \sum_{t \in \bar{\omega}_\tau} \|w(\,\cdot\,,t)\|^2_{\widetilde{W}_2^2(\omega_h)} + \tau \sum_{t \in \omega_\tau} \|w_{\bar{t}}(\,\cdot\,,t)\|^2_{B_h}\,.$$

The error $z = u - v$ satisfies the following conditions

$$(c + K\,\delta_h)\,z_{\bar{t}} - (\tilde{a}\,z_{\bar{x}})_{\hat{x}} = \varphi, \qquad (x,t) \in \omega_h \times \omega_\tau^+,$$

$$z(0,t) = 0, \quad z(1,t) = 0, \qquad t \in \omega_\tau^+,$$

$$z(x,0) = 0, \qquad x \in \bar{\omega}_h,$$

where $\varphi = \psi_{\bar t} - \chi_x$, $\psi = c\,u - T_x^2(cu)$ and $\chi = \tilde a\, u_{\bar x} - T_x^- T_t^- \left(a\,\frac{\partial u}{\partial x}\right)$. The following a priori estimates hold:

$$\|z\|_{\widetilde W_2^{2,1}(Q_{h\tau})} \le C\left(\tau \sum_{t\in\omega_\tau} \|\varphi(\,\cdot\,,t)\|^2_{B_h^{-1}}\right)^{1/2}, \tag{21}$$

$$\begin{aligned}\|z\|_{\widetilde W_2^{1,1/2}(Q_{h\tau})} \le C\Bigg[\tau^2 \sum_{t\in\bar\omega_\tau}\ \sum_{t'\in\bar\omega_\tau,\ t'\ne t} \frac{\|\psi(\,\cdot\,,t)-\psi(\,\cdot\,,t')\|^2_{B_h^{-1}}}{|t-t'|^2} + \\ +\tau\sum_{t\in\omega_\tau}\left(\frac{1}{t}+\frac{1}{T-t}\right)\|\psi(\,\cdot\,,t)\|^2_{B_h^{-1}} + \tau\sum_{t\in\omega_\tau}\|\chi(\,\cdot\,,t)\|^2_h\Bigg]^{1/2}.\end{aligned} \tag{22}$$

Using integral representations of ψ and χ and the form of corresponding norms, from (21) and (22) we obtain the following convergence rate estimates

$$\|z\|_{\widetilde W_2^{2,1}(Q_{h\tau})} \le C\,(h^2+\tau)\left(\|a\|_{W_2^3(0,\xi)} + \|a\|_{W_2^3(\xi,1)} + \|c\|_{W_2^2(0,1)}\right)\|u\|_{\widetilde W_2^{4,2}(Q)}.$$

$$\begin{aligned}\|z\|_{\widetilde W_2^{1,1/2}(Q_{h\tau})} \le C\left(h^2\sqrt{\ln 1/\tau}+\tau\right)\Big(\|a\|_{W_2^2(0,\xi)} + \|a\|_{W_2^2(\xi,1)} \\ +\|c\|_{W_2^2(0,1)}\Big)\|u\|_{\widetilde W_2^{3,3/2}(Q)}.\end{aligned}$$

Similar estimate can be obtained in the norm $\widetilde L_2(Q_{h\tau})$ using appropriate approximation of initial condition. FDSs on nonuniform meshes are considered in [22].

9 Hyperbolic Problems

Convergence rate estimates for hyperbolic IBVPs, contrary to the case of elliptic and parabolic problems, usually are nonconsistent with the smoothness of data. Let us consider the following IBVP

$$\begin{aligned}&\frac{\partial^2 u}{\partial t^2} + \mathcal{L}\,u = f(x,\,t) \qquad \text{in} \quad Q = \Omega\times(0,T) = (0,1)^2\times(0;T)\,;\\ &u(x,\,0) = u_0(x)\,,\quad \frac{\partial u(x,\,0)}{\partial t} = u_1(x)\,;\quad u(x,\,t) = 0 \ \text{ on } \Gamma\times(0,T)\,.\end{aligned} \tag{23}$$

We introduce the mesh in the same manner as in the Section 7 and define the norm

$$\|v\|_{C_\tau(W_2^1(\omega_h))} = \max_{t\in\omega_\tau}\left[\|v_{\bar t}\|^2_{L_2(\omega_h)} + \sum_{i=1}^{2}\left\|\left(\frac{v+v^-}{2}\right)_{x_i}\right\|^2_{L_2(\omega_h)}\right]^{1/2}$$

Consider FDS

$$v_{t\bar t} + \frac{1}{4}\,\mathcal{L}_h\,(v^+ + 2\,v + v^-) = T_1 T_2 T_t\, f\,, \tag{24}$$

where $v^{\pm} = v(x,\, t\pm\tau)$, with the corresponding initial and boundary conditions. For $\tau \asymp h$ the following convergence rate estimate is valid (see [7], [14])

$$\|z\|_{C_\tau(W_2^1(\omega_h))} \le C\, h^{s-2} \left(\max_{i,\,j} \|a_{ij}\|_{W_2^{s-1}(\Omega)} + 1 \right) \|u\|_{W_2^s(Q)} \tag{25}$$

for $2 < s \le 4$.

In some cases by interpolation technique one can obtains estimates which garantee faster convergence on weaker solutions (see [37]). Let us consider the following model problem

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \quad \text{in} \quad Q = (0,\,1) \times (0,\,T)\,;$$

$$u(x,\,0) = u_0(x)\,, \quad \frac{\partial u(x,\,0)}{\partial t} = 0\,; \quad u(x,\,t) = 0 \ \text{ on } \{0,\,1\} \times (0,\,T)$$

and approximate it by a FDS of the form (24). Using integral representation of the residual, one easily obtains the estimates

$$\|z\|_{C_\tau(W_2^1(\omega_h))} \le C\,(h+\tau)^2\, \|u_0\|_{W_2^4(0,\,1)}$$

and

$$\|z\|_{C_\tau(W_2^1(\omega_h))} \le C\, \|u_0\|_{W_2^1(0,\,1)}\,.$$

From these estimate by interpolation one obtains [8]:

$$\|z\|_{C_\tau(W_2^1(\omega_h))} \le C\,(h+\tau)^{\frac{2}{3}\,(s-1)} \|u_0\|_{W_2^s(0,\,1)}\,, \qquad 1 \le s \le 4\,. \tag{26}$$

Contrary to (25), estimate (26) guaranties convergence even for $1 < s \le 2$.

10 String Equation with Concentrated Mass

Let us consider the first IBVP for the equation of vibrating string with concentrated mass at the interior point $x = \xi$ [34]:

$$\begin{gathered}
\left[c(x) + K\,\delta(x-\xi)\right] \frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial x}\left(a(x)\,\frac{\partial u}{\partial x}\right) = f(x,t), \quad (x,t) \in Q = (0,\,1) \times (0,\,T), \\
u(0,\,t) = 0, \quad u(1,\,t) = 0, \qquad 0 < t < T \\
u(x,\,0) = u_0(x), \quad \frac{\partial u(x,0)}{\partial t} = u_1(x), \qquad x \in (0,\,1),
\end{gathered} \tag{27}$$

where $a(x)$, $c(x)$ and K are the same as in the Section 8. Keeping denotations from the Section 8 we approximate IBVP (27) by symetric weighted difference scheme with arveraged right–hand side (see [21], [24])

$$(c + K\,\delta_h)\, v_{t\bar{t}} - (\tilde{a}\, v_{\bar{x}}^{(\sigma)})_x = T_x^2 T_t^2 f, \quad (x,t) \in Q_{h\tau},$$

$$v(0,t)=0, \quad v(1,t)=0, \quad t\in\bar{\omega}_\tau; \qquad v(x,0)=u_0(x), \quad x\in\omega_h,$$

$$(c+K\,\delta_h)\,v_t(x,0)=T_x^2(cu_1)+K\,\delta_h u_1+\frac{\tau}{2}\,T_x^2\left[\widetilde{T}_t^2 f(x,0)+(a\,u_0'(x))'\right], \quad x\in\omega_h,$$

where $v^{(\sigma)}=\sigma\,v^+ +(1-2\sigma)\,v+\sigma\,v^-$, $\sigma\geq 1/4$ and

$$\widetilde{T}_t^2 f(x,0)=\frac{2}{\tau}\int_0^\tau\left(1-\frac{t'}{\tau}\right)f(x,t')\,dt'.$$

The error $z=u-v$ satisfies a priori estimate

$$\|z\|^2_{C_\tau(\widetilde{L}_2(\omega_h))}\leq C\left[\|\chi\|_h^2+\tau^2\,\|\chi_x\|^2_{B_h^{-1}}+\tau\sum_{t\in\omega_\tau}\|\varphi(\,\cdot\,,t)\|_h^2+\tau\sum_{t\in\omega_\tau}\|\eta_t(\,\cdot\,,t)\|^2_{B_h^{-1}}\right]$$

where

$$\varphi=T_x^- T_t^2\Big(a\,\frac{\partial u}{\partial x}\Big)-\tilde{a}\,u_{\bar{x}}^{(\sigma)}, \qquad \eta=c\,u-T_x^2(c\,u),$$

$$\chi=\frac{\tau}{2}\,T_x^-\left[a\Big(\widetilde{T}_t^2\frac{\partial u}{\partial x}-\frac{du_0}{dx}\Big)\right]\bigg|_{t=0},$$

and

$$\|z\|_{C_\tau(\widetilde{L}_2(\omega_h))}=\max_{t\in\omega_\tau}\|(z+z^-)/2\|_{\widetilde{L}_2(\omega_h)},$$

Estimating the terms χ, φ and η similarly as in previous cases we obtain the following convergence rate estimate

$$\|z\|_{C_\tau(\widetilde{L}_2(\omega_h))}\leq C\,(h^2+\tau^2)\,\Big(\|a\|_{W_2^2(0,\xi)}+\|a\|_{W_2^2(\xi,1)}+\|c\|_{W_2^2(0,1)}\Big)\,\|u\|_{\widetilde{W}_2^3(Q)}\,.$$

Here the space $\widetilde{W}_2^k(Q)$ ($k=0,\,1,\,2,\,\ldots$) is defined as the closure of subset of functions $w\in L_2(Q)$ such that

$$w(0,t)=w(1,t)=0\,,$$

$$\frac{\partial^i w}{\partial t^i}\in L_2(0,T;\,\widetilde{L}_2(0,1))\,, \quad i=0,\,1,\,\ldots,\,k,$$

$$\frac{\partial^i w}{\partial x\,\partial t^{i-1}}\in L_2(Q)\,, \quad i=1,\,2,\,\ldots,\,k,$$

$$\frac{\partial^i w}{\partial x^j\,\partial t^{i-j}}\in L_2(Q_1)\cap L_2(Q_2)\,, \quad 2\leq j\leq k,\ \ i=j,\,\ldots,\,k,$$

in the norm

$$\|w\|^2_{\widetilde{W}_2^k(Q)}=\sum_{i=0}^k\left(\left\|\frac{\partial^i w(\xi,\,\cdot\,)}{\partial t^i}\right\|^2_{L_2(0,T)}+\left\|\frac{\partial^i w}{\partial t^i}\right\|^2_{L_2(Q)}\right)+$$

$$+\sum_{i=1}^{k}\left\|\frac{\partial^i w}{\partial x\,\partial t^{i-1}}\right\|^2_{L_2(Q)}+\sum_{j=2}^{k}\sum_{i=j}^{k}\left(\left\|\frac{\partial^i w}{\partial x^j\,\partial t^{i-j}}\right\|^2_{L_2(Q_1)}+\left\|\frac{\partial^i w}{\partial x^j\,\partial t^{i-j}}\right\|^2_{L_2(Q_2)}\right).$$

Similar result in the norm

$$\|z\|_{C_\tau(\widetilde{W}_2^1(\omega_h))}=\max_{t\in\omega_\tau}\left[\|(z+z^-)/2\|^2_{\widetilde{W}_2^1(\omega_h)}+\|z_t\|^2_{\widetilde{L}_2(\omega_h)}\right]^{1/2}$$

is obtained in [21] and [24].

References

[1] Bramble, J. H.; Hilbert, S. R. (1971): *Bounds for a class of linear functionals with application to Hermite interpolation.* Numer. Math. 16, 362–369.

[2] Drenska, N. T. (1984): *Convergence of finite element scheme for the Poisson equation in L_p-norm.* Vestnik Moskov. Univ. Ser. 15 Vychisl. Mat. Kibernet. 3, 19–22 (Russian).

[3] Dupont, T.; Scott, R. (1980): *Polynomial approximation of functions in Sobolev spaces.* Math. Comput. 34, 441–463.

[4] Jovanović, B. S. (1984): *On the convergence of discrete solutions to generalized solutions of boundary value problems.* In: N. S. Bakhvalov, Yu. A. Kuznetsov (eds.), Variational–difference methods in mathematical physics, Proc. conf. held in Moscow 1983, OVM AN SSSR, Moscow, 120–129 (Russian).

[5] Jovanović, B. S. (1989): *On the convergence of finite difference schemes for parabolic equations with variable coefficients.* Numer. Math. 54, 395–404.

[6] Jovanović, B. S. (1991): *Convergence of finite-difference schemes for parabolic equations with variable coefficients.* Z. Angew. Math. Mech. 71, 647–650.

[7] Jovanović, B. S. (1992): *Convergence of finite-difference schemes for hyperbolic equations with variable coefficients.* Z. Angew. Math. Mech. 72, 493–496.

[8] Jovanović, B. S. (1992): *On the estimates of the convergence rate of the finite difference schemes for the approximation of solutions of hyperbolic problems.* Publ. Inst. Math. 52 (66), 127–135.

[9] Jovanović, B. S. (1993): *The finite difference method for boundary value problems with weak solutions.* Posebna izdanja Mat. Instituta, No 16, Beograd.

[10] Jovanović, B. S. (1994): *Interpolation of function spaces and the convergence rate estimates for the finite difference schemes.* In: 2nd Int. Coll. on Numerical Analysis held in Plovdiv 1993, (D. Bainov and V. Covachev, eds.), VPS, Utrecht, 103–112.

[11] Jovanović, B. S. (1997): *Interpolation technique and convergence rate estimates for finite difference method.* Lect. Notes Comp. Sci. 1196, 200–211.

[12] Jovanović, B. S. (2000): *Difference schemes for partial differential equations with generalized solutions.* In: Proc. of Symp. "Contemporary Mathematics", held in Belgrade 1998, Belgrade, 49–63.

[13] Jovanović, B. S.; Bojović, D. (2000): *Finite difference method for the heat equation with coefficient from anisotropic Sobolev space.* Facta Universitatis, Ser. Math. Inform. 15, 113–122.

[14] Jovanović, B. S.; Ivanović, L. D.; Süli, E. E. (1987): *Convergence of a finite-difference scheme for second-order hyperbolic equations with variable coefficients.* IMA J. Numer. Anal. 7, 39–45.

[15] Jovanović, B. S.; Ivanović, L. D.; Süli, E. E. (1987): *Convergence of finite-difference schemes for elliptic equations with variable coefficients.* IMA J. Numer. Anal. 7, 301–305.

[16] Jovanović, B. S.; Kandilarov, J. D.; Vulkov, L. G. (2001): *Construction and convergence of difference schemes for a model elliptic equation with Dirac's delta function coefficient.* Lect. Notes Comput. Sci. 1988, 431–438.

[17] Jovanovich, B. S.; Matus, P. P. (1999): *Estimation of the convergence rate of finite-difference schemes for elliptic problems.* Zh. vychisl. mat. mat. fiz. 39, 61–69 (Russian).

[18] Jovanovich, B. S.; Matus, P. P.; Shcheglik, V. S. (1998): *Finite-difference schemes on nonuniform meshes for parabolic equation with variable coefficients and weak solutions.* Doklady NAN Belarusi 42, No 6, 38–44 (Russian).

[19] Jovanovich, B. S.; Matus, P. P.; Shcheglik, V. S. (1999): *The rates of convergence of the finite-difference schemes on nonuniform meshes for parabolic equation with variable coefficients and weak solutions.* Sib. zhurn. vychisl. matematiki 2, 123–136 (Russian).

[20] Jovanovich, B. S.; Popović, B. Z. (2001): *Convergence of a finite difference scheme for the third boundary value problem for elliptic equation with variable coefficients.* Comput. Methods Appl. Math. 1, No 4, 356–366.

[21] Jovanović, B. S.; Vulkov, L. G. (2000): *On the convergence of difference schemes for the string equation with concentrated mass.* In: R. Ciegis, A. Samarskiĭ and M. Sapagovas (eds.), Finite–Difference Schemes: Theory and Applications, Proc of 3rd Int. Conf. held in Palanga (Lithuania) 2000, IMI, Vilnius, 107–116.

[22] Jovanović, B. S.; Vulkov, L. G. (2001): *Operator's approach to the problems with concentrated factors.* Lect. Notes Comput. Sci. 1988, 439–450.

[23] Jovanović, B. S.; Vulkov, L. G. (2001): *On the convergence of finite difference schemes for the heat equation with concentrated capacity.* Numer. Math. 89, No 4, 715–734.

[24] Jovanović, B. S.; Vulkov, L. G. (2003): *On the convergence of difference schemes for hyperbolic problems with concentrated data.* SIAM J. Numer. Anal. 41, No 2, 516–538.

[25] Lazarov, R. D. (1981): *On the convergence of finite–difference schemes on generalized solutions of Poisson equation.* Differentsial'nye uravneniya 17, 1285–1294 (Russian).

[26] Lazarov, R. D.; Makarov, V. L.; Samarskiĭ, A. A. (1982): *Application of exact difference schemes for construction and investigation of difference schemes for generalized solutions.* Mat. Sbornik 117, 469–480 (Russian).

[27] Lions, J. L.; Magenes, E. (1968): *Problèmes aux limites non homogènes et applications.* Dunod, Paris.

[28] Lykov, A. V. (1989): *Heat and mass transfer. Spravochnik.* Nauka, Moscow (Russian).

[29] Maz'ya, V. G.; Shaposhnikova, T. O. (1985): *Theory of multipliers in spaces of diferentiable functions.* Monographs and Studies in Mathematics 23, Pitman, Boston, Mass.

[30] Mokin, Yu. I. (1971): *Discrete analog of theorem on multiplier.* Zh. vychisl. mat. mat. fiz. 11, 746–749 (Russian).

[31] Samarskiĭ, A. A. (1983): *Theory of difference schemes.* Nauka, Moscow (Russian).

[32] Samarskiĭ, A. A.; Lazarov, R. D.; Makarov, V. L. (1987): *Finite–difference schemes for difference equations with generalized solutions.* Vysshaya shkola, Moscow (Russian).

[33] Süli, E. E.; Jovanović, B. S.; Ivanović, L. D. (1985): *Finite difference approximations of generalized solutions.* Math. Comput. 45, 319–327.

[34] Tikhonov, A. N.; Samarskiĭ, A. A. (1953): *Equations of mathematical physics*. GITTL, Moscow (Russian).

[35] Vladimirov, V. S. (1998): *Equations of mathematical physics*. Nauka, Moscow (Russian).

[36] Weinelt, W. (1978): *Untersuchungen zur Konvergenzgeschwindigkeit bei Differenzenverfahren*. Zeitschrift der THK 20, 763–769.

[37] Zlotnik, A. A. (1991): *Convergence rate estimates for projection–difference schemes approximating second order hyperbolic equations*. Vychisl. Protsessy Sist. 8, 116–167 (Russian).

SCHLICHT DISCS, BLOCH-BERS SPACE AND HARMONIC MAPS

MIODRAG MATELJEVIĆ

Dedicated to Professor Veselin Perić on the occasion of his 70th birthday

In [3], we characterize Bers space by means of maximal φ-disks:

Theorem 1. *A holomorphic quadratic differential φdz^2 on the unit disc is bounded with respect to the Poincaré metric (i.e. it belongs to Bers Space) if and only if the radii of its maximal φ-discs are uniformly bounded.*

As an application we show that the Hopf differential of a quasiregular harmonic map with respect to strongly negatively curved metric belongs to Bers space. Also we give further sufficient or necessary conditions for a holomorphic function to belong to Bers space.

After writing our paper [3] we realized that Theorem 1 has its roots in known characterizations of Bloch functions.

In this paper we will present the content of our paper (see [3]) and explain links between mentioned Theorem 1 and known characterizations of Bloch functions. For further results related to the subject of this paper we refer the interested reader to author's review papers [29] and [31] (see also [13],[14] and [30]) . In this paper (in section 4) , only a short review of [13] is given .

In section 0 we prove Koebe and Bloch Theorem. The following result is an immediate corollary of Bloch Theorem and Schwarz Lemma.

Theorem SW (Seidel and Walsh). *An analytic function on the unit disc Δ is Bloch function iff schlicht discs in the image surface are uniformly bounded.*

It is clear that Theorem 1 is a generalization of Theorem SW. The proofs of these results are similar except we have some additional difficulties caused by possible zeros of corresponding quadratic differential. Lemma 1.1 (see below) enables us to overcome those difficulties.

0. Bloch's and Koebe's Theorem

We will use the following notation.

If $r > 0$ and a is a complex number

$$B(a;r) = \{z \in \mathbb{C} : |z-a| < r\}$$

is the open circular disc with center at a and radius r. Also we use notation $\Delta_r = B(0,r)$ and $\Delta = \Delta_1$. First, we introduce a particulary interesting class of conformal mappings of the disc, the class S. We denote by S the class of holomorphic functions f in Δ which are injective and satisfy normalization conditions $f(0) = 0$ and $f'(0) = 1$.

Proposition 1. *If $f \in S$ there exists a $g \in S$ such that*

$$g^2(z) = f(z^2),\ z \in \Delta. \tag{1}$$

Proof. Since f belongs to S we can write f in the form $f(z) = z\varphi(z)$, where φ is holomorphic in Δ, $\varphi(0) = 1$ and φ has no zeros in Δ. Hence there exists an analytic function h on Δ with $h(0) = 1$, $h^2(z) = \varphi(z)$. Define g by

$$g(z) = zh(z^2),\ z \in \Delta.$$

Then $g^2(z) = z^2h^2(z^2) = z^2\varphi(z^2) = f(z^2)$.

The following result is well known as the Area Theorem (see for example [1] and [18]).

Theorem A (Area Theorem). *If F is holomorphic in $\Delta \setminus \{0\}$, F is one-to-one in Δ, and*

$$F(z) = \frac{1}{z} + \sum_{k=0}^{\infty} c_k z^k,\ z \in \Delta, \tag{2}$$

then

$$\sum_{k=1}^{\infty} k|c_k|^2 \leq 1.$$

Corollary 1. *Under the same hypothesis, $|c_1| \leq 1$.*

Proposition 2. *If $f \in S$ and $a_2 = \frac{f''(0)}{2}$ is the Taylor coefficient of f then $|a_2| \leq 2$.*

Proof. By Proposition 1 there exists a $g \in S$ so that $g^2(z) = f(z^2)$. If $\Phi = \frac{1}{g}$ then Theorem A applies to Φ, and this will give $|a_2| \leq 2$. Since

$$f(z^2) = z^2(1 + a_2z^2 + ...),$$

we have

$$g(z) = z(1 + \frac{1}{2}a_2z^2 + ...),$$

and hence

$$\Phi(z) = \frac{1}{z}(1 - \frac{1}{2}a_2z^2 + ...) = \frac{1}{z} - \frac{a_2}{2}z + ... \quad .$$

The Corollary 1 shows now that $|a_2| \leq 2$.

Theorem K (Koebe's One-Quarter Theorem). *If $f \in S$ then $f(\Delta) \supset \Delta_{1/4}$.*

Proof. Suppose that $w_0 \notin f(\Delta)$.

Define the auxiliary function

$$A(w) = \frac{w_0 w}{w_0 - w} \text{ and } F = A \circ f.$$

Then $F \in S$ and $A_2 = F''(0)/2 = A''(0)/2 + f''(0)/2 = a_2 + \frac{1}{w_0}$.
Hence, applying Proposition 2 to the function F we obtain

$$|a_2 + \frac{1}{w_0}| \leq 2$$

and since $|a_2| \leq 2$, we finally obtain $|1/w_0| \leq 4$. So

$|w_0| \geq 1/4$ for every $w_0 \notin f(\Delta)$. Thus $f(\Delta) \supset \Delta_{1/4}$.

Lemma 1. *If f is analytic on $B = B(a;r)$ and $\operatorname{Re} f' > 0$ on B, then f is one-to-one on B.*

Proof. Suppose z_1 and z_2 are points in B, $z_1 \neq z_2$, $\gamma(t) = z_1 + t(z_2 - z_1)$ and $I = \int_0^1 f'(\gamma(t))dt$. Then

$$f(z_2) - f(z_1) = \int_{[z_1,z_2]} f' d\zeta = (z_2 - z_1)I.$$

¿From the hypothesis $\operatorname{Re} I > 0$ and therefore $I \neq 0$ so that $f(z_1) \neq f(z_2)$ and f is one-to-one on B.

Lemma 2. *Let f be an analytic nonconstant function on $\overline{B} = \overline{B(a : R)}$ and suppose there exists $q \in (0,1]$ such that*

$$q|f'(z)| \leq |f'(a)|, \; z \in B.$$

Then f is univalent on $B_1 = B(a; R_1)$, where $R_1 = qR$.

A proof of this lemma can be based on the subordination principle.

Theorem B (Bloch's Theorem). *Let f be an analytic function on $\overline{\Delta}$ and $f'(0) \neq 0$. Then there exists a disc $B \subset \Delta$ such that f is univalent on B and $f(B)$ contains a disc of radius greater or equal $\frac{1}{16}|f'(0)|$.* *Proof.* Let $M_0 = \max\{|f'(z)|(1-|z|) : z \in \overline{\Delta}\}$. It is easy to see that there is a point $z_0 \in \Delta$ such that $M_0 = |f'(z_0)|(1-|z_0|)$. If $z \in B_0 = B(z_0; \rho_0)$, where $\rho_0 = \frac{1}{2}(1-|z_0|)$, then

$$|f'(z)| \leq \frac{M_0}{\rho_0} = 2|f'(z_0)|.$$

Hence, by Lemma 2, f is univalent on $B_1 = B(z_0; \rho_1)$, where $\rho_1 = \frac{\rho_0}{2}$. According to Koebe's Theorem, this implies that $f(B_1)$ contains a disc of radius $R = \frac{1}{4}|f'(z_0)|\rho_1 = \frac{1}{16}|f'(z_0)|(1-|z_0|)$.

Since $|f'(z_0)|(1-|z_0|) \geq |f'(0)|$ this gives the result.

We will use notation $W_f = f(\Delta)$. A schlicht disc in W_f is a disc $B \subset W_f$ such that there exists a domain Ω such that $f|_\Omega : \Omega \to B$ is 1-1 and onto.

For $z \in \Delta$ let $d_f(z) = \sup\{r : B(f(z); r)$ is a schlicht disc in $W_f\}$. Set $r_f = \sup_{z\in\Delta} d_f(z)$

Definition 1. A Bloch function is an analytic map $f : \Delta \to \mathbb{C}$ such that

$$||f||_B = \sup_{z\in\Delta}(1-|z|^2)|f'(z)| < +\infty.$$

It follows from the proof of Bloch's Theorem that if r_f is finite then f is a Bloch function. On the other hand if f is an analytic function on Δ then, by Schwarz Lemma,

$$d_f(z) \leq (1-|z|^2)|f'(z)|, \; z \in \Delta.$$

Hence, if f is a Bloch function then r_f is finite.

Thus, we have proved the following result of Seidel and Walsh.

Theorem SW. *For an analytic function on Δ the following conditions are equivalent:*

(a) f is Bloch function

(b) $r_f < \infty$

The following result gives further characterizations of Bloch function.

Theorem SWPo. *For an analytic function on Δ the following conditions are equivalent:*
(a) f is Bloch function
(b) $r_f < \infty$
(c) The family $\{f \circ T - f \circ T(0) : T \in \mathrm{Aut}(\Delta)\}$ is normal
(d) There exists $\alpha > 0$, and a univalent function g on Δ such that $f = \alpha \ln g'$.

Recall that Seidel and Walsh (see Theorem SW above) proved (a) $\Longleftrightarrow$ (b). The equivalence of (a), (c) and (d) was proved by Pommerenke (see [16]).

In [3] we proved a generalization of Theorem SW concerning holomorphic quadratic differentials.

Theorem 1. *A holomorphic quadratic differential φdz^2 on the unit disc is bounded with respect to the Poincaré metric (i.e. it belongs to Bers Space) if and only if the radii of its maximal φ-discs are uniformly bounded.*

Since f is a Bloch function iff holomorphic quadratic differential φdz^2, defined by $\varphi = (f')^2$, belongs to Bers space, then Theorem SW can be consider as a special case of Theorem 1.

After writing the paper [3] we realized that this result has its roots in known characterizations of Bloch functions in terms of their image Riemann surface. As we mentioned Seidel and Walsh and Pommerenke [16] proved that a function f belongs to Bloch space if and only if its image surface W_f contains no large schlicht discs. For relevant definitions related to this result and some generalizations we refer the interested reader to Pommerenke [16], Stegenga and Stephenson [21].

Holomorphic quadratic differentials on a Riemann surface arise in several distinct areas of geometry, for instance in Teichmüller theory and in the theory of harmonic maps (see, for example, Ahlfors [2], Earle and Eells [6], Wolf [27], Jost [10]).

First we give a short review of our results as well as some related ones (see [3]).

In §1, we use a special parameter (natural parameter) in terms of which the differential has a particularly simple representation, along with the theorems of Bloch and Koebe to prove Theorem 1 (just stated above).

Recall that when we work with a natural parameter, we have some additional difficulties caused by possible zeroes of the corresponding quadratic differential. Lemma 1.1 (see below) enables us to overcome those difficulties.

We will mention some recent results, which motivated us.

Wan [26] proved that a harmonic diffeomorphism of the hyperbolic plane $\mathbb{H}^2$ is quasiconformal if and only if its Hopf differential is uniformly bounded with respect to the Poincaré metric. This has also been generalized to hyperbolic Cartan-Hadamard surfaces by Li, Tam and Wang [25].

See Tam and Wan [23], [24] and Han [8] for a general discussion of this area, where this and other questions were discussed.

As an application of Theorem 1 we show that the Hopf differential of a quasiregular harmonic map with respect to a strongly negatively curved metric belongs to Bers space (see below, theorem 2 and 3, §2).

Thus, roughly speaking, we can extend one direction of the above-mentioned characterizations ([26],[25]) of harmonic quasiconformal mappings to harmonic quasiregular mappings.

For a precise definition of quasiregular mapping see §2,

A4. Here we note only that the notion of quasiregular mapping is a natural generalization of the notion of a quasiconformal mapping since one does not require that quasiregular mappings be homeomorphisms.

Our proofs of theorems 2 and 3 are based on the fact that the Bochner formula (see [19], [20], [10]) has a simple form with respect to the natural parameter. This allows us to define a metric by means of the dilatation of the mapping, whose Gaussian curvature is bounded from above by -1, and we use the classical Ahlfors-Schwarz lemma.

In §3 we give further sufficient and necessary conditions for a holomorphic function to belong to Bers space, and show that every quasiregular harmonic mapping is decomposable as a quasiconformal harmonic mapping followed by an analytic function.

For further results and the literature in this growing area we refer the interested reader to [8, 9, 12, 15, 19, 23, 24, 25, 26, 27, 28].

We close our paper with a short discussion concerning some further results and open problems.

Now, we present the complete content of our paper.

1. Maximal φ-disks and Bers space Q

Let φ be an analytic function on the unit disk Δ. Then φ belongs to *Bers space* $Q = Q(\Delta)$ if

$$\operatorname{ess\,sup} \omega(z)^2 |\varphi(z)| < +\infty ,$$

where $\omega(z) = 1 - |z|^2$.

In this section we will give a characterization of Bers space by means of maximal φ-disks (see below theorem 1). First we define maximal φ-disks.

Maximal φ-disk. Let φ be an analytic function on the unit disk Δ and let z_0 be a regular point of φ, i.e. $\varphi(z_0) \neq 0$. Let Φ_0 be a single valued branch of

$$w = \Phi(z) = \int \sqrt{\varphi(z)}\, dz$$

near z_0, $\Phi(z_0) = 0$. There is a neighborhood U of z_0 which is mapped one-to-one conformally onto an open set V in the w-plane. We can assume, by restriction, that V is a disk around $w = 0$. The inverse Φ_0^{-1} is a conformal homeomorphism of V into Δ and evidently there is a largest open disk V_0 around $w = 0$ such that the analytic continuation of Φ_0^{-1} (which is still denoted by Φ_0^{-1}) is homeomorphic, and that $\Phi_0^{-1}(V_0) \subset \Delta$. The image $U_0 = \Phi_0^{-1}(V_0)$ is called the *maximal φ-disk* around z_0; its φ-radius (injectivity radius) $r_0 = R_{z_0}$ is the Euclidean radius of V_0.

Note that if f is a holomorphic function on Δ and $\varphi = (f')^2$, by notation in section 1, $R_z = d_f(z)$, $z \in \Delta$.

For the definition of φ-disks and a discussion of their important role in the theory of holomorphic quadratic differentials we refer the interested reader to Strebel's book [22].

Bloch's and Koebe's theorem. The two following theorems play an important role in the proof of Theorem 1. It is easy to derive from Theorem K and Theorem B (section 0) the following versions of Bloch's and Koebe's theorems.

Theorem B1 (Bloch). *Let $w = f(z)$ be an analytic function on the disk $B = B(z_0, r) = \{z : |z - z_0| < r\}$, $r > 0$, and let $f'(z_0) \neq 0$. Then there is an open disk*

U together with an open set $V \subset B$ such that f restricted to V defines a one-to-one mapping of V onto U and the radius R of U satisfies

$$R \geq C|f'(z_0)|r\,,$$

where C is an absolute constant.

Theorem K1 (Koebe). *Let V be a domain in $\mathbb{C}$ and let f be an analytic and univalent mapping which maps V onto the disk $U = \{w : |w - w_0| < R\}$ and let $z_0 = f^{-1}(w_0)$. Then*

$$dist(z_0, \partial V)|f'(z_0)| \geq \frac{R}{4}\,.$$

The following lemma enables us to use Bloch's theorem. In the proof of this lemma we will use the hyperbolic metric on a disk.

Hyperbolic distance. Let B be the disk with center at z_0 and radius r. Using the conformal automorphisms $\phi_a(z) = \frac{z-a}{1-\bar{a}z}$, $a \in \Delta$, of Δ, one can define pseudo-hyperbolic distance on Δ by

$$\delta(a,b) = |\phi_a(b)|,\ a, b \in \Delta\,.$$

Next, using the conformal map $A(\zeta) = \frac{\zeta - z_0}{r}$ from B onto Δ, one can define pseudo-hyperbolic distance on B by

$$\delta_B(z,w) = \delta(A(z), A(w))$$

and the *hyperbolic metric* on B by

$$\rho(z,w) = \log\frac{1+\delta_B(z,w)}{1-\delta_B(z,w)}$$

for $z, w \in B$.

The following result is well known.

Theorem H. *Let F be an analytic function from a disk B to another disk U. Then F does not increase the corresponding hyperbolic (pseudo-hyperbolic) distances.*

Lemma 1.1. *Let φ be a bounded analytic function on the disk $B = B(z_0, r_0)$ and let $M_0 = \sup\{|\varphi(z)| : z \in B\}$. Suppose that $\varphi(z_0) \neq 0$ and let $r_1 = \frac{r_0}{M_0}|\varphi(z_0)|$. Then φ has no zeroes in the disk $B(z_0, r_1)$.*

Proof. Let $\varphi(z) = 0$ for some $z \in B$. Using the hyperbolic (or pseudo-hyperbolic) distances on B and $B(0, M_0)$, an application of Theorem H to the analytic function φ and the points z and z_0 yields

$$\frac{|z - z_0|}{r_0} \geq \frac{|\varphi(z_0)|}{M_0}\,.$$

□

Let φ be an analytic function on the unit disk Δ. Let $0 < r < 1$ and $\varphi_r(z) = \varphi(rz)\,r^2$, $\psi(z) = |\varphi_r(z)|^{\frac{1}{2}}$ and $\omega(z) = 1 - |z|^2$. Assume that the function $h(z) = \omega(z)\psi(z)$ has the maximum on Δ at the point $z_0 \in \Delta$. Next let $r_0 = \frac{1-|z_0|}{2}$ and let $M_0 = \max\{|\varphi_r(z)| : z \in B(z_0, r_0)\}$. Since

$$h(z_0) \geq h(z)$$

and

$$2\omega(z) \geq \omega(z_0) \text{ for } z \in B(z_0, r_0)$$

then

$$M_0 \leq 4|\varphi_r(z_0)| \tag{1}$$

An application of lemma 1.1 to the disk $B(z_0, r_0)$ shows that φ_r does not have zeroes in the disk $B(z_0, r_1)$, where $r_1 = \frac{r_0}{M_0}|\varphi_r(z_0)|$. Next, by (1)

$$r_1 \geq \frac{r_0}{4} = r_2 \ .$$

Since φ_r does not have zeroes in $B(z_0, r_2) = B$, there is a regular branch of the function $\sqrt{\varphi_r}$ in B, and therefore a regular branch Φ of $\int \sqrt{\varphi_r}$ in B. Since $|\Phi'(z_0)| = \sqrt{|\varphi_r(z_0)|}$, then, by Bloch's theorem, there is a disk V of radius

$$R = R(z_0) \geq C\psi(z_0) \cdot r_2 \ ,$$

where C is an absolute constant, such that Φ^{-1} is univalent on V. Let $R_\infty = \sup\{R_z : z \in \Delta\}$, where R_z is the radius of the maximal φ-disk around z. Suppose that R_∞ is finite. Then

$$R_\infty \geq \frac{C}{8}\psi(z_0)(1 - |z_0|) \geq \frac{C}{16}\psi(z_0)\omega(z_0)$$

When $r \to 1_-$, one can obtain that

$$R_\infty^2 \geq \left(\frac{C}{16}\right)^2 \|\varphi\| \ , \ \|\varphi\| = \sup_{z\in\Delta} \omega^2(z)|\varphi(z)| \tag{2}$$

Lemma 1.2. *Suppose that $\varphi \in Q$. Then R_∞ is finite.*

Proof. Let Δ_z be the maximal φ-disk around $z \in \Delta$ and R_z the euclidean radius of the disk $\Phi(\Delta_z)$, where Φ is the natural parameter. By Koebe's Theorem

$$dist(z, \partial\Delta_z)|\Phi'(z)| \geq \frac{R_z}{4}$$

Since $1 - |z| \geq dist(z, \partial\Delta_z)$ and $|\Phi'(z)| = \sqrt{|\varphi(z)|}$ then

$$\|\varphi\| \geq \frac{R_\infty^2}{16} \tag{3}$$

□

Note that one can use Schwarz Lemma as in the proof of Theorem SW to prove Lemma 1.2.

The following result is an immediate corollary of (2) and Lemma 1.2

Theorem 1. *Let φ be an analytic function on Δ. Then $\varphi \in Q$ iff R_∞ is finite.*

2. Harmonic maps and Bers space Q

Harmonic maps play an important role in the parametrization of Teichmüller spaces (see Earle and Eells [6] and Wolf [27]), so it is interesting to understand the relation between universal Teichmüller space and quasiconformal harmonic diffeomorphisms. For further results see Wan [26], Tam and Wan [23], Reich and Strebel [17]). In this direction we have the following result (the terminology will be explained, and the proof given, later in this section).

Theorem 2. *Let ρ be the metric with Gaussian curvature $K \leq -a$ for some constant $a > 0$, and let f be a harmonic quasiregular map from Δ into itself with respect to ρ. Then the Hopf differential φ of f belongs to Q.*

Theorem 2 is an immediate corollary of Theorem 1 and Lemma 2.1. See below for the proof of this lemma and for the definition of a quasiregular function.

Let R and S be two surfaces. Let $\sigma(z)|dz|^2$ and $\rho(w)|dw|^2$ be the metrics with respect to the isothermal coordinate charts on R and S respectively, and let f be a C^2-map from R to S.

It is convenient to use notation in local coordinates $df = p\,dz + q\,d\bar{z}$, where $p = f_z$ and $q = f_{\bar{z}}$. Also we introduce the complex (Beltrami) dilatation

$$\mu_f = Belt[f] = \frac{q}{p}$$

where it is defined.

The energy integral of f is

$$E(f,\rho) = \int_R \rho \circ f(|p|^2 + |q|^2)\, dxdy\,.$$

A critical point of the energy functional is called a harmonic mapping. The Euler-Lagrange equation for the energy functional is

$$\tau(f) = f_{z\bar{z}} + (\log\rho)_w \circ f\, pq = 0\,.$$

Thus, we say that a C^2-map f from R to S is harmonic if f satisfies the above equation. For basic properties of harmonic maps and for further information on the literature we refer to Jost [10] and Schoen-Yau [19].

The following facts and notation are important in our approach:

A1 If f is a harmonic mapping then

$$\varphi dz^2 = \rho \circ f\, p\bar{q}\, dz^2$$

is a quadratic differential on R, and we say that φ is the *Hopf differential* of f and we write $\varphi = $Hopf$(f)$.

A2 The Gaussian curvature on S is given by

$$K_S = -\frac{1}{2}\frac{\Delta \log\rho}{\rho}\,.$$

A3 We will use the following notation $\mu = Belt[f] = \frac{q}{p}$ and $\tau = \log\frac{1}{|\mu|}$ and *Bochner* formula (see [19])

$$\Delta\tau = -K_S|\varphi|\sinh\tau\,.$$

A4 Definition of quasiregular function. Let R and S be two Riemann surfaces and $f : R \to S$ be a C^2-mapping. If P is a point on R, $\tilde{P} = f(P) \in S$, ϕ a local parameter on R defined near P and ψ a local parameter on S defined near $\tilde{P}$, then the map $w = h(z)$ defined by $h = \psi \circ f \circ \phi^{-1}|_V$ (V is a sufficiently small neighborhood of P) is called a local representer of f at P. The map f is called k-quasiregular if there is a constant $k \in (0,1)$ such that for every representer h, at every point of R, $|h_{\overline{z}}| \leq k|h_z|$.

Lemma 2.1. *Let ρ be the metric on Δ with Gaussian curvature K uniformly bonded from above on Δ by the negative constant $-a$, and let f be a harmonic k-quasiregular map from Δ into itself with respect to the metric ρ. If $R = R_z$ is the radius of the maximal φ-disk around z, where $\varphi =$Hopf(f), then R is bounded from above by the constant C which depends only on k and a.*

Proof. Let $R = R_z$ be the radius of the maximal φ-disk $U = U_z$ around $z \in \Delta$. Since f is k-qusiregular then $\tau \geq m$, where $m = \log \frac{1}{k}$. $m > 0$. Let $\zeta = \Phi(z)$ be the natural parameter in U and $\Phi(U) = V = B(0,R)$ With respect to the parameter ζ the Bochner formula takes the simple form

$$\Delta\tau = -K \sinh \tau\,.$$

Since $K \leq -a$ and $\tau \geq m$, we conclude that

$$\Delta\tau \geq \delta e^{\tau} \text{ on } V \tag{4}$$

where $\delta = \frac{a \sinh m}{e^m}$. Let $ds = \lambda(\zeta)|d\zeta|$, where $\lambda(\zeta) = \frac{2R}{R^2 - |\zeta|^2}$ is the hyperbolic metric on V and let $\tilde{\lambda}(\zeta) = \left(\frac{\delta}{2} e^{\tau(\zeta)}\right)^{\frac{1}{2}}$. From(4) we have for the Gaussian curvature of the metric $d\tilde{s} = \tilde{\lambda}(\zeta)|d\zeta|$ on V that $\tilde{K} \leq -1$ and then we can use the Ahlfors-Schwarz Lemma (see [1]) to obtain

$$\frac{\delta}{2k} \leq \tilde{\lambda}^2(\zeta) \leq \lambda^2(\zeta)\,. \tag{5}$$

Setting $\zeta = 0$ in (5) one obtains

$$R^2 \leq \frac{8k}{\delta}\,. \tag{6}$$

□

Let φ be a quadratic differential on a hyperbolic Riemann surface R with Poincaré metric $ds^2 = \rho(z)|dz|^2$. Let $p \in R$ and let z be a local parameter near p. We will define

$$\|\varphi\|(p) = \rho^{-1}(z(p))|\varphi(z(p))|.$$

We say that φ belongs to the *Bers space* of R (notation $Q(R)$) if $\|\varphi\|$ is a uniformly bounded function on R.

Theorem 3. *Let R and S be hyperbolic surfaces with metric densities σ and ρ respectively and let the Gaussian curvature of the metric $ds^2 = \rho(w)|dw|^2$ be uniformly bounded from above on S by the negative constant $-a$.*

If f is a harmonic k-quasiregular map from R into S with Hopf differential φ, then $\varphi \in Q(R)$.

Proof. Let $\tilde{f}$ be the lifting of f which maps Δ into itself and let $\tilde{\varphi}$ be the lifting of the quadratic differential φ. Let $\tilde{\rho}$ be the lifting of the density ρ. Since $\tilde{f}$ is harmonic with respect to the metric $\tilde{\rho}(\tilde{w})|d\tilde{w}|^2$ on Δ and k-quasiregular then, by Theorem 2, $\tilde{\varphi} \in Q(\Delta)$. Hence $\varphi \in Q(R)$. □

3. Further results

In Theorem 4 we will give a characterization of a quasiregular harmonic map.

Theorem 4. *Let f be a k-quasiregular harmonic map from Δ into itself with respect to some metric $ds^2 = \rho(w)|dw|^2$.*

Then $f = F \circ g$, where F is an analytic function from Δ into itself and g is a k-quasiconformal mapping from Δ onto itself, which is harmonic with respect to the metric $d\tilde{s}^2 = \tilde{\rho}(\zeta)|d\zeta|^2$, where $\tilde{\rho} = \rho \circ F\,|F'|^2$

Proof. Since f is harmonic on Δ then $\varphi = \rho \circ f\, p\overline{q}$ is an analytic function on Δ.

Therefore p has isolated zeroes or p is identically 0 on Δ. If $p \equiv 0$ on Δ then $q \equiv 0$ and $f \equiv const$ on Δ and our theorem is trivial. If p has isolated zeroes on Δ then we can define $\mu = \frac{q}{p}$ a.e. on Δ.

It is known that there is a quasiconformal mapping g from Δ onto itself such that g is a solution of *Beltrami* equation

$$g_{\overline{z}} = \mu g_z$$

(see [2], [11]).

Let $F = f \circ g^{-1}$. Then we have for $Belt[F]$ (see [2], [11]) that

$$\mu_F \circ g = \frac{g_z}{\overline{g}_{\overline{z}}} \cdot \frac{\mu_f - \mu_g}{1 - \mu_f \overline{\mu}_g} = 0$$

and we conclude that F is analytic function.

Since f is harmonic with respect to ρ then

$$\varphi(z) = \rho(f(z))\, p\overline{q}$$

is an analytic function in Δ, where $p = f_z$ and $q = f_{\overline{z}}$. Since $p(z) = F'(\zeta)A(z)$ and $q(z) = F'(\zeta)B(z)$, where $A = g_z$, $B = g_{\overline{z}}$ and $\zeta = g(z)$, one can obtain that

$$\varphi(z) = \tilde{\rho}(\zeta)A\overline{B}\,.$$

Since g is quasiconformal $|A| \neq |B|$ a.e. and $\varphi_{\overline{z}} \equiv 0$ on Δ, one can show that $\tau(g) = 0$ (for computation of $\varphi_{\overline{z}}$ see, for example, Jost [10] and Tam-Wan [24]). □

Let D be a hyperbolic domain in $\mathbb{C}$, $z \in D$ and $ds = \rho(z)|dz|$ the corresponding hyperbolic metric on D. Then it is known that (theorem 1.11 [1])

$$\rho(z) \leq \frac{2}{r(z)} \quad , z \in D\,, \tag{7}$$

where $r(z) = dist(z, \partial D)$ and we call $r(z)$ the *distance function* (see [1]).

For an analytic function φ on a domain $D \subset \mathbb{C}$ we say that $\varphi \in \tilde{Q}$ if

$$\operatorname{ess\,sup} dist^2(z, \partial D)|\varphi(z)| < \infty$$

Since the distance function is geometrically simpler than the hyperbolic density, it is reasonable to study the space $\tilde{Q}$.

We say that a domain $D \subset \mathbb{C}$ is *strongly hyperbolic* if it is hyperbolic and diameters of boundary components are uniformly bounded from below by a positive constant.

Theorem 5. *Let D be a strongly hyperbolic and bounded domain in $\mathbb{C}$. Then $\varphi \in \tilde{Q}$ iff $\varphi \in Q$.*

Proof. Because of (7) we have that $Q \subset \tilde{Q}$. Let the diameters of boundary components of D be bounded from below by $d > 0$, and let the diameter of D be equal to M. Now let $z \in D$. We can find a component l of ∂D for which $r(z) = |z - z_0|$, where $z_0 \in l$. Let $\tilde{D} = \overline{\mathbb{C}} \setminus l$ and let ρ and $\tilde{\rho}$ be the corresponding hyperbolic linear densities of D and $\tilde{D}$ respectively. Since $D \subset \tilde{D}$ then $\tilde{\rho}(z) \leq \rho(z)$ for $z \in D$ (see [1]).

Let $\tilde{r}(z) = dist(z, \partial\tilde{D})$, and let $c \in l$ such that $|z_0 - c| = \frac{d}{2}$. The function $\psi(\zeta) = \frac{1}{\zeta - c}$ maps $\tilde{D}$ conformaly onto the domain $G \subset \mathbb{C}$. Since G is conformally equivalent to the unit disk, by the Koebe Theorem

$$\sigma(w) \geq \frac{1}{4|w - w_0|}, \tag{8}$$

where $w = \psi(z)$, $w_0 = \psi(z_0)$ and σ is the linear density of hyperbolic metric on G. ¿From (8) we can conclude that

$$\tilde{\rho}(z) \geq \frac{|z_0 - c|}{|z - c|} \cdot \frac{1}{|z - z_0|}$$

and hence

$$\tilde{\rho}(z) \geq \frac{C}{\tilde{r}(z)}, \tag{9}$$

where $C = \frac{d}{d+2M}$.

Since $\tilde{r}(z) = r(z)$ we finally obtain that

$$\rho(z) \geq \frac{C}{r(z)}.$$

Hence $\tilde{Q} \subset Q$. □

The next example shows that if the boundary of a domain D has a point as a component then the spaces Q and $\tilde{Q}$ are different.

Example 1. Let D be $\Delta \setminus \{0\}$, and $\varphi(z) = \frac{1}{z^2}$. It is obvious that $\varphi \in \tilde{Q}$. The linear density of the hyperbolic metric on D is

$$\rho(z) = \frac{1}{|z| \log \frac{1}{|z|}}.$$

Then

$$\rho^{-2}(z)|\varphi(z)| = \log \frac{1}{|z|}$$

which is not bounded in D, hence $\varphi \notin Q$.

In fact, any function

$$\varphi(z) = \frac{1}{z^2}\psi(z),$$

where ψ is an analytic function in Δ with $\psi(0) \neq 0$, is not in Q.

Example 2. Let D be $\mathbb{C} \setminus [-1, 1]$. Then $\psi(w) = \frac{1}{2}(w + \frac{1}{w})$ is a conformal mapping from $\Delta \setminus \{0\}$ onto D. Let φ be an analytic function on D. It is clear that $\varphi \in Q(D)$ iff $\varphi_1 \in Q(\Delta \setminus \{0\})$, where

$$\varphi_1(w) = \varphi(z)\left(\psi'(w)\right)^2.$$

Let $\varphi(z) = \frac{1}{z^2}$. It is obvious that $\varphi \in \tilde{Q}(D)$. Since

$$\varphi_1(w) = \frac{1}{w^2}\frac{1}{2}\Big(\frac{w^2-1}{w^2+1}\Big)^2$$

then, by Example 1 we conclude that $\varphi_1 \notin Q(\Delta \setminus \{0\})$, hence $\varphi \notin Q(D)$.

Using Koebe's Theorem as in Lemma 1.2 one can prove the following result

Proposition 3.1. Let D be a hyperbolic domain in the complex plane $\mathbb{C}$. If $\varphi \in \tilde{Q}(D)$ then the radii of the maximal φ-disks are uniformly bounded on D.

As we mentioned in the introduction we close with a short discussion of some further results in this area.

Wolf [28] and Minsky [15] have shown that estimates on the dilatation of a harmonic map depend to a great extent on the geometry of the Hopf differential φ (in particular, on the placement of the zeroes of φ and the injectivity radius in the φ-metric).

Han [8] and Han, Tam, Treibergs and Wan [9] have used the Wolf-Minsky type estimates mentioned above to study among other things the images of harmonic diffeomorphisms of $\mathbb{C}$ into the hyperbolic plane $\mathbb{H}$.

We believe that our results can be of use in understanding some parts of this interesting area, as well as being of interest in their own right.

4. A version of Bloch theorem

Also recall ,for further results related to the subject of this paper we refer the interested reader to author's review papers [29] and [31] (see also [13],[14] and [30]) .

For example , in [13] ,using a version of Bloch theorem (see Lemma 1 below) we give a short proof of a Dyakonov's theorem [5]. Also we show that Lemma 1 holds for quasiregular harmonic functions (see Theorem 6 below).

Let U denote the unit disc in the complex plane. If z and w are complex numbers by $\Lambda(z,w)$ we denote the half-line $\Lambda(z,w) = \{z+\rho(w-z) : \rho \geq 0\}$ and $\Lambda(w) = \Lambda(0,w)$.

Lemma 1. *Suppose that f is an analytic function on the unit disc U, $f(0)=0$ and $|f'(0)| \geq 1$. Then there is an absolute constant s such that for every $\theta \in \mathbb{R}$ there exists a point w on the half-line $\Lambda_\theta(0,e^{i\theta}) = \{\rho e^{i\theta} : \rho \geq 0\}$, which belongs to $f(U)$, such that $|w| \geq 2s$.*

Theorem 6. *Suppose that f is a K-quasiregular harmonic mapping on the unit disc U, $f(0)=0$ and $|\operatorname{grad} f(0)| \geq 1$. Then, there exists an absolute constant α such that for every $\theta \in \mathbb{R}$ there exists a point w on the half-line $\Lambda_\theta = \Lambda(0,e^{i\theta}) = \{\rho e^{i\theta} : \rho \geq 0\}$, which belongs to $f(U)$, such that $|w| \geq 2\alpha$.*

Acknowledgement . I wish to thank Professor M. Janjić for inviting me to give lecture at the meeting devoted to academician V. Perić .

References

[1] Ahlfors,L., *Conformal invariants*, McGraw-Hill Book Company, 1973.

[2] Ahlfors,L., *Lectures on Quasiconformal Mappings*, Van Nostrand, 1966.

[3] Anić, I., Marković, V., Mateljević, M., *Uniforly bounded maximal $\varphi-$disks, Bers space and harmonic maps*, Proc. Amer. Math. Soc. **128** (2000), 2947-2956.

[4] Conway, J., *Functions of One Complex Variables*, Springer Verlag (1984).

[5] Dykonov, K., M., *Equivalent norms on Lipschitz-type spaces of holomorphic functions*, Acta Math. **178** (1997), 143–167.

[6] Earle,C.J. and Eells,J., *A Fibre bundle description of Teichmüler theory*, J.Diff.Geom. **3**, (1969) 19-43

[7] Gardiner,F.,P., *Teichmüller Theory and Quadratic Differentials*, New York: Awiley-Interscience Publication, 1987.

[8] Han,Z-C., *Remarks on the geometric behavior of harmonic maps between surfaces*, Elliptic and parabolic methods in geometry. Proceedings of a workshop, Minneapolis, May 23-27, 1994., Wellesley

[9] Han,Z-C., Tam,L-F., Treibergs,A. and Wan,T., *Harmonic maps from the complex plane into surfaces with nonpositive curvature*, Commun.Anal.Geom. **3** (1995) 85-114

[10] Jost,J., *Two-dimensional Geometric Variational Problems*, John Wiley & Sons, 1991.

[11] Lehto,O. and Virtanen,K.I., *Quasiconformal Mappings in the Plane*, Springer-Verlag, 1973.

[12] Marković, V., Mateljević, M., *New version of the main inequality and uniqueness of harmonic maps*, J. d'Analyse Math. **79** (1999) 315-334.

[13] Mateljević, M., *A version of Bloch theorem for quasiregular harmonic mappings*, Proceedings of International Conference on Complex Analysis and Related Topics (IX^{th} Romanian-Finnish Seminar, 2001),Rev Roum Math Pures Appliq (Romanian Journal of Pure and Applied mathematics) 47 (2002) 5-6 , pp 705-707. .

[14] Mateljević, M., *Estimates for the modulus of the derivatives of harmonic univalent mappings*, Proceedings of International Conference on Complex Analysis and Related Topics (IX^{th} Romanian-Finnish Seminar, 2001), Rev Roum Math Pures Appliq (Romanian Journal of Pure and Applied mathematics) 47 (2002) 5-6 , pp 709 -711.

[15] Minsky,Y.,*Harmonic maps, lenght and energy in Teichmüller space*, J.Diff.Geom. **35** (1992), 151-217

[16] Pommerenke,Ch., *On Bloch functions*, J. London Math. Soc. (2),**2** (1970), 689-695

[17] Reich,E. and Strebel,K., *On the Gerstenhaber-Rauch principle*, Israel J.Math. **57**, (1987) 89-100

[18] Rudin, W., *Real and Complex Analysis*, McGraw-Hill Book Co. (1966).

[19] Schoen,R. and Yau,S.,T., *Lectures on Harmonic Maps*, Conf. Proc. and Lect. Not. in Geometry and Topology, Vol.II, Inter. Press, 1997.

[20] Schoen,R. and Yau,S.,T., *On univalent harmonic maps between surfaces*, Invent.Math. **44** (1978), 265-278

[21] Stegenga,D. and Stephenson,K., *A geometric characterization of analytic functions with bounded mean oscillation*, J. London Math. Soc. (2),**24** (1981), 243-254

[22] Strebel,K., *Quadratic Differentials*, Springer-Verlag, 1984.

[23] Tam,L. and Wan,T., *Quasiconformal harmonic diffeomorphism and universal Teichmüler space*, J.Diff.Geom. **42** (1995) 368-410

[24] Tam,L. and Wan,T., *Harmonic diffeomorphisms into Cartan-Hadamard surfaces with prescribed Hopf differentials*, Comm.Anal.Geom.**4** (1994) 593-625

[25] Li,P., Tam,L. and Wang,J., *Harmonic diffeomorphisms between hyperbolic Hadamard manifolds*, to appear Jour.Geom.Anal.

[26] Wan,T., *Conastant mean curvature surface, harmonic maps, and univrsal Teichmüller space*, J.Diff.Geom. **35** (1992) 643-657

[27] Wolf,M., *The Teichmüller theory of harmonic maps*, J.Diff.Geom. **29** (1989) 449-479

[28] Wolf,M., *High-energy degeneration of harmonic maps between surfaces and rays in Teichmüler space*, Topology **30** (1991), 517-540

[29] M . Mateljević , Ahlfors-Schwarz lemma and curvature , Kragujevac Journal of Mathematics (Zbornik radova PMF), Vol 25 ,2003

[30] M. Mateljević ,*The unique extremality II* , Mathematical Reports Vol. **2** (**52**) No. 4 , 2000, 503-525

[31] M.Mateljević , *Dirichlet's principle ,area theorem , uniqueness of harmonic maps and related problems* (accepted for publication in special number of Publication Inst˜Math -Belgrade, editor M.Mateljević)

Faculty of mathematics, Univ. of Belgrade, Studentski Trg 16, Belgrade, YU
E-mail address: miodrag@matf.bg.ac.yu

An essay about geometric combinatorics

Rade T. Živaljević

Dedicated to Professor Veselin Perić on the occasion of his 70th birthday

1 What is geometric combinatorics?

- I believe that we lack another analysis properly geometric or linear which expresses location directly as algebra expresses magnitude.

  G.W. Leibniz (letter to C. Huygens, 1679)

- Poincaré was the first who introduced the idea of *computing with topological objects*, not only with numbers. He did this, ..., by defining the concepts of *homology* and fundamental group.

  J. Dieudonné (History of Algebraic and Differential Topology, 1900–1960)

- Homology theory discovered by Poincaré is perhaps the most profound and far reaching creation in all topology.

  S. Lefschetz

It is believed by many mathematicians that *homology theory*, discovered by Henry Poincaré, provides a direct "analysis" of geometric objects, referred to in the letter of Leibniz to Huygens. One century after its discovery, the homology theory, as an *analysis* and *combinatorics* of topological/geometric objects, remains, together with other related constructions in algebraic geometry and topology, one of central tools for discovering and expressing laws about geometric forms.

Geometric combinatorics is one of the areas of mathematics where the "direct calculus" with geometric objects is one of central themes. It is not an easy task to determine all the themes and driving forces of this field, so the selection in this paper reflects in part the research interest of the Belgrade G-T-A seminar[1]. Formally the article consists of four mathematical études, each composed for a different area of contemporary geometric combinatorics. Since the area of geometric combinatorics is a mixture of fields including combinatorial topology, combinatorial geometry, combinatorics, computational and discrete

[1]The seminar for Geometry, Topology and Algebra (GTA) was founded more than 15 years ago. Perhaps its name should be rightly changed to CGTA to include Combinatorics.

geometry etc. we leave it to the reader to decide what area of mathematics is the most natural environment for a given étude.

Each of the études begins with a short *partiture*. Here, a *partiture* is a short sequence of formulas or statements, most of them related, all of them tied to a given *theme* or *motive*. The rest of the étude consists of *variations* on the main theme. The idea is to follow the partiture, formula after formula, or statement after statement and offer variations, reminiscences, comments, historical details and anything that comes to mind, as in a mathematical jam session.

2 Legacy of Ludwig Schläfli

$$b_0(M) = b_0(\mathbb{R}^n \setminus \cup\mathcal{A}) = \binom{m}{0} + \binom{m}{1} + \ldots + \binom{m}{n} \tag{1}$$

(L. Schläfli, 1901)

$$b_0 = \sum\nolimits_{p \in L_{\mathcal{A}}} |\mu(\hat{0}, p)| \tag{2}$$

(T. Zaslavsky, 1975)

$$P_t(M) = (1+t)(1+2t)\cdots(1+(m-1)t) \tag{3}$$

$M := \{z \in \mathbb{C}^m \mid z_i = z_j \text{ for } i = j\}$

(V.I. Arnold, 1969)

$$P_t(M) = (1+m_1 t)(1+m_2 t)\cdots(1+m_k t) \tag{4}$$

$M := V_{\mathbb{C}} \setminus \mathcal{A}_{\mathbb{C}}$, $\mathcal{A}$ is a reflection arrangement in $V_{\mathbb{R}}$

(E. Brieskorn, 1971)

$$\tilde{H}^k(\mathbb{R}^n \setminus \cup\mathcal{A}; \mathbb{Z}) \cong \oplus_{p \in L} \tilde{H}_{n-k-\dim(p)-2}(\Delta(L)) \tag{5}$$

(M. Goresky – R. MacPherson, 1988)

$$(\cup\mathcal{A})^+ \simeq \bigvee_{p<\hat{1}} \Delta(L_{<p}) * S^{\dim(p)} \tag{6}$$

$$\mathbb{R}^n \setminus \cup\mathcal{A} = S^n \setminus (\cup\mathcal{A})^+ \cong_S \bigvee_{p<\hat{1}} D_{n-\dim(p)-1}(\Delta(L_{<p})) \tag{7}$$

(G. Ziegler – R. Živaljević, 1993)

The first formula is taken from the great posthumous work *Theorie der vielfachen Kontinuität* of the famous Swiss geometer Ludwig Schläfli, published as the volume 38 of Denkschriften der Schwiezerischen naturforschenden Gesellschaft in 1901. The formula appears on page 39 as the answer to the question about the largest possible number of connected components in the complement of m hyperplanes in a n-dimensional affine space $\mathbb{R}^n$.

Schläfli made his reputation as one of the leading geometers of his time by numerous contributions to what today would often be appropriately classified as the field of geometric combinatorics. Perhaps it is not sufficiently widely known that it is Schläfli who completed the classification of all regular polytopes in all dimensions, thus completing the list started in antiquity with the discovery of the five Platonic solids. Around 1852 Schläfli proved that in dimension $n \geq 4$, aside from the obvious examples, n-dimensional regular cube, regular n-simplex and the dual of the cube, the regular, n-dimensional cross polytope (hyperoctahedron), there exist precisely three more regular polytopes, all of them in dimension 4. It is a remarkable fact that facets of these polytopes are respectively 24 octahedra, 120 dodecahedra and 600 tetrahedra.

- As a tribute to Einstein and Schläfli, on the inner wall of central library of the Institute for physics and mathematics in Bern, there is a sentence "*Three quarks for Einstein and Schläfli*"! The reader remembers that Einstein was in Bern for a relatively short period of time on the beginning of his career, where he moved with his wife Mileva Marić-Einstein from Zürich.

The formula (1) is possibly one of the first results in the area of *arrangements of subspaces*, see [20] for a comprehensive account. The formula has been rediscovered in the meantime by other authors and sometimes in different forms. For example, The American Mathematical Monthly published in the issue 50 (1943), p.59 the following problem:

- Show that n cuts can divide a cheese into as many as $\frac{(n+1)(n^2-n+6)}{6}$ pieces (problem E 554).

The formula (1) opened several lines of research. First of all, one can ask about the number of connected components in the complement of an arbitrary hyperplane arrangement. The answer was obtained by an elegant formula of Thomas Zaslavsky, [40], reproduced here as the formula (2). Of course, one needs some additional information about the arrangement $\mathcal{A} = \{H_1, \dots, H_m\}$ of hyperplanes. Such an information is provided by the *Möbius* function of the associated intersection partially ordered set $L = L_{\mathcal{A}}$. Recall that $L_{\mathcal{A}}$ is an abstract poset which has an element p for each intersection of the form $H_{i_1} \cap \dots \cap H_{i_k}$ while $\leq$ records the containment relation between these subspaces. August Ferdinand Möbius introduced his well known arithmetic function $\mu(n)$ in 1832. The extension to general posets was given by Gian-Carlo Rota [24], however see [30] for a more complete historical account. The formula of Philip Hall, identifying the Möbius function as the Euler characteristic of the associated order complex $\Delta(P)$ is reproduced here as the formula (15) in Section 5.

Returning to the main theme, recall that for a given hyperplane arrangement $\mathcal{A}$, the number of components in the complement is equal to the rank of the cohomology group $H^0(\mathbb{R}^n \setminus \cup\mathcal{A}; \mathbb{Q})$. So, in the case of a general, real or complex ($\mathbb{K} = \mathbb{R}, \mathbb{C}$) subspace arrangement $\mathcal{B}$ it is natural to ask the question:

Q: Determine interesting topological invariants of the complement $\mathbb{K}^n \setminus \cup\mathcal{B}$ of the general subspace arrangement $\mathcal{B}$.

Our next example is the celebrated formula (3) of Vladimir Igorevič Arnold. This formula provides an explicit computation of the Poincaré polynomial of the famous *braid arrangement*. Recall that $P_t(X) = b_0 + b_1 t + \ldots + b_k t^k + \ldots$ is the polynomial which has the Betti numbers b_k, i.e. the ranks of the associated (co)homologu groups $H^k(X;\mathbb{Q})$, as the coefficients. The braid arrangement Br_n in $\mathbb{C}^n$ is described as the collection of all hyperplanes $H_{ij} = \{z \in \mathbb{C}^n \mid z_i = z_j\}$ for $1 \leq i \leq j \leq n$. Arnold's motivation for studying the complement of the braid arrangement came actually from another direction. Already as a student at Moscow State University, V. Arnold made a decisive contribution, under guidance of his professor Andrei Kolmogorov, to the solution of the 13 Hilbert problem. A variant of this problem asks if an *algebraic* function can be expressed as a superposition of algebraic functions depending on a fewer number of variables than the original function. Arnold originally proved (see [32]), relying on the results of Dmitrii Fuks on the $\mathbb{Z}/2$-cohomology of $M(Br_n)$, that such superpositions are not always possible.

The braid arrangement is an object ubiquitous in today's mathematics. We refer the reader to [32] for an interesting account from the point of view of discriminants and singular spaces. Perhaps the main reason for its importance is coming from the fact that $M(Br_n)$ is a $K(G,1)$-space where G is the *colored braid group*. There is however yet another, very important reason why the braid arrangement is so popular among mathematicians. Here in the focus are the so called *Knizhnik-Zamolodchikov* equations which arise as the conditions for certain natural connections over $M(Br_n)$ to be flat, i.e. to have the zero curvature, see [29], Chapter 12. The fact that the connection is flat, i.e. the absence of holonomy, implies the existence of a monodromy representation of the group $G = \pi_1(M(Br_n))$. Vladimir Drinfel'd was able to give a detailed analysis, or a categorical description, of this representation. This was the starting point of his celebrated work about Hopf algebras or quantum groups, for which he was awarded a Fields medal.

Egbert Brieskorn was able to extend Arnold's formula to other Coxeter arrangements. He explained the appearance of numbers $1, 2, \ldots, (n-1)$ in Arnold's formula, by identifying them as the *coexponents* of the associated Coxeter arrangement (group), formula (4). All these results still deal with hyperplane, albeit complex, arrangements. A lot of work has been directed towards the understanding the homology of complements of general subspace arrangements, see [20]. This work culminated in the formula (5) of Mark Goresky and Robert MacPherson. This formula appeared at the end of their book on stratified Morse theory, [16], and served as a test example for the powerful general theory they developed. This settled the question of homology invariants of arbitrary subspace arrangements. As in the case of Zaslavsky's formula (2), the homology is expressed in terms of the intersection poset $P = L_{\mathcal{A}}$ of the arrangement $\mathcal{A}$ but with a new ingredient, the dimension function $d : L_{\mathcal{A}} \to \mathbb{N}$, defined by $d(H) = \dim(H)$.

This formula was refined in another direction by Günter Ziegler and Rade Živaljević. The formula (6) which appeared in [41], gave a precise homotopy decomposition for the one point compactification $(\cup\mathcal{A})^+ := (\cup\, \mathcal{A}) \cup \{+\infty\}$ of an arbitrary affine, spherical or even more general arrangement. As a consequence, via the so called S-duality, this formula describes the stable homotopy type of the complement $M(\mathcal{A})$. The last result was independently and by different methods obtained also by Victor Anatol'evich Vassiliev as a part of his general theory of geometric resolutions of discriminants and the topology of their complements. Let us remark that (6), via Alexander duality, provides a new proof of the Goresky and MacPherson formula (5).

This is not the end of the story. The references [13], [15], [27], [37], [32], [38], [39] and others deal with different aspect of the problem of the topology of the complement and the union of the arrangements of subspaces.

3 The story of 3 houses and 3 wells

$$(K_{3,3} \to \mathbb{R}^2) \Rightarrow (2 \to \text{point}) \tag{8}$$

$$(K_{3,3,3} \xrightarrow{s} \mathbb{R}^2) \Rightarrow (3 \to \text{point}) \tag{9}$$

$$(K_{5,5,5} \to \mathbb{R}^3) \Rightarrow (3 \to \text{point}). \tag{10}$$

$$(K_6 \hookrightarrow \mathbb{R}^3) \Rightarrow \text{linking} \tag{11}$$

$$(K_{4,4} \to \mathbb{R}^2) \Rightarrow (4 \to \text{line}) \tag{12}$$

$$(K_{6,6} \to \mathbb{R}^3) \Rightarrow (4 \to \text{line}) \tag{13}$$

The well known Kuratowski nonplanarity criterion implies that the graph $K_{3,3}$ is not embedable in $\mathbb{R}^2$. Recall that $K_{3,3}$ is a complete bipartite graph obtained if three vertices (three "houses") are connected with three other vertices (three "wells") so that each of the houses is connected by a path (an edge in the graph) with each of the wells. Another popular description of $K_{3,3}$ talks about three houses connected with a source of electricity, gas and water. For this reason $K_{3,3}$ is sometimes called the "houses and utilities graph". The graph K_5 is by definition the complete graph on 5 vertices. The statement (8) is one way of expressing the nonplanarity of $K_{3,3}$. It says that for each continuous map $f : K_{3,3} \to \mathbb{R}^2$, there exist 2 points a and b, which belong to 2 *disjoint* edges in $K_{3,3}$, such that $f(a) = f(b)$ (i.e. $2 \to$ point).

An elementary proof that $K_{3,3}$ is not planar is based on the well known Euler relation $f_0 - f_1 + f_2 = 2$. An embedding of $K_{3,3}$ in $\mathbb{R}^2$ automatically yields an embedding in the sphere $S^2 = \mathbb{R}^2 \cup \{\infty\}$. Since there are no cycles of length 3 in the graph $K_{3,3}$, each face must be bounded by at least 4 edges. Hence, $4f_0 \leq 2f_1 = 20$, and therefore, $f_0 - f_1 + f_2 \leq 6 - 9 + 6 = 1$, a contradiction with the Euler relation. A similar proof applies to the graph K_5. There exist however some subtle points that the reader can "ignore in the first reading". Here is a comment on this proof by Allen Hatcher[2].

> Allen Hatcher: There are a couple points in the proof of nonplanarity of the graphs K_5 and $K(3,3)$ that deserve further comment.
>
> (1) It suffices to consider only polygonal embeddings since a topological embedding of a finite graph in the plane can be approximated by a polygonal embedding. This is shown in Bollobas' book "Modern Graph Theory" by a very simple argument.

[2] D. Davis Algebraic Topology discussion group, http://hopf.math.purdue.edu.

(2) In order to apply Euler's formula $f_0 - f_1 + f_2 = 2$, one might think that it is necessary to know that a polygonal embedding of the graph defines a CW structure on the 2-sphere. This is equivalent to knowing the Schoenflies theorem for polygonal simple closed curves in the plane, that such a curve bounds an embedded disk, which is not a trivial theorem. But in fact one can get away with less, just the Jordan curve theorem that such a curve has two complementary components, which is easy to show for polygonal curves. Then interpret the f_2 in Euler's formula to be the number of complementary components of the embedded graph, and Bollobas gives a simple proof that $f_0 - f_1 + f_2 = 2$.

So much about the formula (8). A natural problem is to find generalizations, analogs and relatives of this statement? Note that a consequence of (8) is that for any collection of 3 red and 3 blue points in the plane, there exist two intersecting vertex disjoint line segments with end points of different color. The statement (9) (due to Imre Bárány) claims that from any collection of 3 blue, 3 white and 3 red points in the plane $\mathbb{R}^2$, one can always select three vertex-disjoint, "rainbow" triangles which have a nonempty intersection. A "rainbow" triangle is a triangle having all vertices of different color. Something similar is possible in the 3-space $\mathbb{R}^3$. This time we need at least 5 points of each color in order to guarantee existence of three vertex disjoint, "rainbow" triangles, which have a nonempty intersection. This is a consequence of the statement (10). More formally (10) says that for every continuous map $f : K_{5,5,5} \to \mathbb{R}^3$, where $K_{5,5,5} := [5] * [5] * [5]$ is the 2-complex obtained as the join of three copies of $[5] = \{1, 2, 3, 4, 5\}$, there exist three points in three vertex-disjoint triangles which are mapped to the same point in $\mathbb{R}^3$. The statement (9) is similar except that $f : K_{3,3,3} \to \mathbb{R}^2$ is assumed to be a simplicial map and it is not known if it holds in the case of an arbitrary continuous map $f : K_{3,3,3} \to \mathbb{R}^2$!

Formulas (8) and (10) are special cases of general statements about configurations of "colored" points in $\mathbb{R}^d$, see [34], [42], [43], [48]. A different generalization to the 3-space is provided by the theory of *linkless, windless* etc. embeddings of graphs, [11], [26], [28]. An example from this circle of results is the statement (11) which says that for every embedding of the graph K_6 in $\mathbb{R}^3$, there exist two disjoint circuits C_1, C_2 of K_6 which are linked with a nonzero linking number, [11], [26].

It is shown in [45] that the results listed above can be extended in a systematic way to include higher dimensional statements where the existence of a common point (common 0-dimensional transversal) is replaced by the existence of a common k-dimensional transversal. Recall that a k-dimensional transversal of a family $\mathcal{F} = \{F_j\}_{j=1}^m$ of subsets in $\mathbb{R}^d$ is an affine k-dimensional space $L \subset \mathbb{R}^d$ such that $L \cap F_j = \emptyset$ for all j. For example a simple consequence of the "ham sandwich theorem" is the statement (12) which implies that for any collection of four black and four white points in the plane $\mathbb{R}^2$ there exists a line intersecting four vertex disjoint line segments with end points of different color. Much less trivial is the statement (13) which, in the affine case, says that for every collection of 6 red and 6 blue points in $\mathbb{R}^3$ there exist 4 line segments with end points of different color having a common line transversal. This result can be viewed as a relative of the nonplanarity of $K_{3,3}$. Of course there are higher dimensional complexes which exhibit similar behavior as shown by the following example

$$(\sigma_2^7 \to \mathbb{R}^3) \Rightarrow (4 \to \text{line})$$

where σ_2^7 is the 2-skeleton of a 7-dimensional simplex σ^7. These results are deduced in [45] as corollaries of general statements which could be interpreted as results belonging to the *combinatorial geometry on vector bundles.*

We end this section with an open problem. It is known that aside from planar graphs there exist other topologically defined classes of graphs which admit a combinatorial characterization in terms of "forbidden minors". According to Robertson, Seymour and Thomas, graphs which admit a linkless (windless) embeddings can be characterized as graphs which have no minors in the Petersen family, [28].

Problem: Find a combinatorial characterization in terms of forbidden minors of all graphs K for which the statement (13) is not true. That us characterize all graphs which can be mapped to the 3-space $\mathbb{R}^3$ such that no 4 vertex disjoint edges admit a line transversal.

4 Partitions of masses

M_1 : Let $\mu_1, \mu_2, \ldots, \mu_n$ be a collection of mass distributions in $\mathbb{R}^n$. Then there exists a hyperplane H such that for all $i = 1, \ldots, n$

$$\mu_i(H^+) \geq 1/2\, \mu_i(\mathbb{R}^n) \text{ and } \mu_i(H^-) \geq 1/2\, \mu_i(\mathbb{R}^n)$$

where H^+ and H^- are the closed halfspaces associated to the hyperplane H.

M_2 : Let μ be a mass distribution in $\mathbb{R}^n$. Then there exists a point $x \in \mathbb{R}^n$ so that for every closed halfspace $P \subset \mathbb{R}^d$, if $x \in P$ then

$$\mu(P \cap A) \geq \frac{\mu(\mathbb{R}^n)}{n+1}.$$

M_3 : Let $\mu_0, \mu_1, \ldots, \mu_k$, $0 \leq k \leq n-1$, be a collection of mass distributions in $\mathbb{R}^n$. Then there exists a k-dimensional affine subspace $D \subseteq \mathbb{R}^n$ such that for every closed halfspace $H(v, \alpha) := \{x \in \mathbb{R}^d \mid \langle x, v\rangle \leq \alpha\}$ and each i,

$$D \subseteq H(v,\alpha) \Longrightarrow \mu_i(H(v,\alpha)) \geq \tfrac{1}{n-k+1}\, \mu_i(\mathbb{R}^n).$$

M_4 :

$$\Delta(2,2) = 3 \quad \Delta(1,3) = 3 \quad 4 \leq \Delta(1,4) \leq 5 \quad \Delta(5,2) \leq 9 \quad \Delta(3,3) \leq 9.$$

All the results $M_1 - M_4$ are combinatorial facts about "mass distributions" in $\mathbb{R}^n$. It is perhaps appropriate, before we begin a discussion, to say a few words what is meant by this term. Recall, [7], that a Borel measure μ, defined on a locally compact space X, is a *weak limit* of a sequence of measures μ_n if for each bounded continuous function $f : X \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_X f\, d\mu_n = \int_X f\, d\mu.$$

As a consequence, if μ is a weak limit of (μ_n), then

$$\liminf_{n\to\infty} \mu_n(F) \leq \mu(F) \leq \mu(O) \leq \limsup_{n\to\infty}\, \mu_n(O) \tag{14}$$

for any sets $F \subset O$, where F closed and O open in X.

By a mass distribution on $\mathbb{R}^n$ we mean a measure μ which is a weak limit of a sequence of measures μ_n, absolutely continuous with respect to the Lebesgue measure m. In other words, μ is a weak limit of a sequence (μ_n) such that $d\mu_n = f_n\, dm$ where $f_n : \mathbb{R}^n \to \mathbb{R}$ is an integrable function. Most of interesting measures that appear in combinatorial problems belong to this class. For example all counting measures of finite sets, i.e. measures μ_S defined by $\mu_S(A) := |A \cap S|$ are weak limits of measures absolutely integrable with respect to the Lebesgue measure. The inequalities (14), often permit us to prove a statement for a smaller class of measures, typically measures of the form $d\nu = g\, dm$, where g is everywhere positive, Lebesgue integrable function, and then obtain the result for a general mass distribution by a passage to the limit.

The reader will not have difficulties to recognize in M_1 the well known "Ham sandwich theorem". Indeed, if $A_1, \ldots, A_n$ are measurable sets in $\mathbb{R}^n$, then the hyperplane H is a halving hyperplane for all A_i. In the special case, when A_1, A_2 and A_3 respectively represent the ham, bread and cheese in $\mathbb{R}^3$, the result says that a ham sandwich can always be divided in two equal pieces by a single straight cut of a knife. The result M_2 is also known as the "center point theorem" and it also has a gastronomic reformulation. Namely, suppose you want to split an irregularly shaped pizza with a hungry friend who chooses first and who is supposed to divide the pizza in two pieces by a straight cut of a knife. You are allowed to mark your piece in advance, say by claiming the piece which will contain a particular marking object (say an olive). Then, if you are very careful about marking your piece, you can count on at least one third of the pizza. Note that the pizza is not assumed to be either connected or convex nor homogeneous. So "one third of the pizza" means that there is some measure, made precise in advance, which evaluates the "quality" of different slices of the pizza. Note that if the pizza is convex and homogeneous, a result of Branko Grünbaum, [17], guarantees that the constant 1/3 can be improved to 4/9.

Both the "ham sandwich" and the "center point" theorem have a very interesting history and numerous applications, see [14], [21], [42] for references. The author remembers a conversation with Vrećica Siniša in late 1987, in front of the blackboard in his office, at Mathematics faculty in Belgrade. By accident, two of us mentioned, for different reasons, theorems M_1 and M_2. We instantly observed that M_1 is a statement about n measures in $\mathbb{R}^n$ while M_2 is a statement about a single measure and asked ourselves whether there exists a general statement about k measures, $1 \leq k \leq n$, which reduces to M_1 and M_2 respectively in the boundary cases $k = n$ and $k = 1$. This is how the statement M_3 was born. It took us a little time to detect the correct topological principle which stands behind the proof of this theorem, namely the fact that the cohomology class $(w_k)^{n-k}$ is nonzero where $w_k \in H^k(G_k(\mathbb{R}^n); \mathbb{Z}_2)$ is the k^{th} Stiefel-Whitney characteristic cohomology class of the canonical k-dimensional vector bundle over the Grassmann manifold $G_k(\mathbb{R}^n)$ of all linear, k-dimensional subspaces of $\mathbb{R}^n$. Nevertheless, we both agree that this is an instance of a result which was more difficult to "invent" or contemplate then to find its proof.

The ham sandwich theorem is a special case of M_3 but this is not the only generalization of this result. In this category is the general problem of studying equipartitions of masses by hyperplanes which was formulated by Branko Grünbaum in [17].

Suppose that

$$\mathcal{M} = \{\mu_1, \mu_2, \ldots, \mu_j\}$$

is a collection of *continuous* mass distributions defined in $\mathbb{R}^d$. If $\mathcal{H} = \{H_i\}_{i=1}^k$ is a collection of k hyperplanes in $\mathbb{R}^d$ in general position, the connected components of the complement $\mathbb{R}^d \setminus \cup \mathcal{H}$ are called (open) k-orthants. A collection $\mathcal{H}$ is an *equipartition*, or more precisely a k-equipartition for $\mathcal{M}$ if

$$\mu_i(O) = \mu_i(\overline{O}) = \frac{1}{2^k}\mu_i(\mathbb{R}^d)$$

for each of the measures $\mu_i \in \mathcal{M}$ and for each k-orthant O associated to $\mathcal{H}$. A triple (d, j, k) of integers is referred to as *admissible* if for any collection $\mathcal{M} = \{\mu_i\}_{i=1}^j$ of j continuous measures in $\mathbb{R}^d$, there exists a collection of k hyperplanes $\mathcal{H} = \{H_i\}_{i=1}^k$ forming an equipartition for all measures in $\mathcal{M}$.

The general problem is to characterize the set $\mathcal{A}$ of all admissible triples. If the emphasis is put on the ambient Euclidean space $\mathbb{R}^d$, the equivalent problem is to determine the smallest dimension $d := \Delta(j,k)$ such that the triple (d,j,k) is admissible. Hugo Hadwiger proved that $\Delta(2,2) = 3$ which also implies $\Delta(1,3) = 3$. The case $k = 1$ is answered by the "ham sandwich theorem" which in the new notation says that $\Delta(d,1) = d$. Edgar Ramos, [23], considerably advanced our knowledge about the function $d = \Delta(j,k)$. He showed for example that $\Delta(1,4) \leq 5$, $\Delta(5,2) \leq 9$, $\Delta(3,3) \leq 9$. Perhaps one of the most interesting open problems in the area is the question of the exact value of $\Delta(1,4)$.

P: Is it true that $\Delta(1,4) = 4$. More explicitly, is it true that for any continuous mass distribution μ on $\mathbb{R}^4$, there exist 4 hyperplanes H_1, H_2, H_3, H_4, which divide $\mathbb{R}^4$ into 16 orthants $\{O_i\}_{i=1}^{16}$, such that for each i = 1,...,16

$$\mu(O_i) = \frac{1}{16}\mu(\mathbb{R}^4).$$

5 Order complexes and Vassilev geometric resolutions

$$\mu(P) = \tilde{\chi}(\Delta(P)) \tag{15}$$

$$\Delta(\widetilde{\Pi}_n) \simeq \bigvee_{i=2}^{n} \Sigma(\Delta(\widetilde{\Pi}_{n-1}^i)) \tag{16}$$

$$\Delta(\widetilde{\mathcal{G}}_n(R)) \simeq S^{n-1} \wedge \Sigma(\Delta(\widetilde{\mathcal{G}}_{n-1}(R))) \tag{17}$$

$$\Delta(\widetilde{\mathcal{G}}_n^+(R)) \simeq (S^{n-1} \vee S^{n-1}) \wedge \Sigma(\Delta(\widetilde{\mathcal{G}}_{n-1}^+(R))) \tag{18}$$

$$\Delta(\exp_n(S^1)) \simeq S^{n-1} \wedge \Delta(\widetilde{\mathcal{B}}_n) \simeq S^{2n-1} \tag{19}$$

We have already met the Möbius function in Section 2. Let us recall that the Möbius function $\mu = \mu_P : P \times P \to \mathbb{Z}$ is an important invariant of a finite poset P which is recursively defined by

$$\mu(p,q) = -\sum_{p \leq z \leq p} \mu(p,z) \text{ and } \mu(p,p) = 1$$

for $p \leq q$ and $\mu(p,q) = 0$ in the opposite case. It follows from the definition that if $Q \subset P$ is a "convex" subposet of P then μ_Q, the Möbius function of Q is the restriction on $Q \times Q$ of the Möbius function μ_P. If $\hat{P} = P \cup \{\hat{0}, \hat{1}\}$ is the poset obtained with a formal addition of the maximal and minimal elements $\hat{1}$ and $\hat{0}$, then $\mu_{\hat{P}}(\hat{0}, \hat{1})$ is the so called Möbius number of P, denoted by $\mu(P)$.

The order complex $\Delta(P)$ of a finite poset P is an abstract simplicial complex (or its geometric realization) where $A \subset P$ is a simplex in $\Delta(P)$ if and only if A is a chain in P. The geometric meaning of $\Delta(P)$ is perhaps best understood if one takes a convex polytope K (alternatively a simplical or regular cell complex) and chooses $P = P_K$ to be the face poset of K. Then it is not difficult to observe that $\Delta(P_K)$ is a simplicial complex isomorphic to the first barycentric subdivision of K.

The relation (15) which identifies the Möbius number of P as the Euler characteristic of the associated order complex $\Delta(P)$ is due to Philip Hall. This equation is not difficult to prove, however it is very appealing and points in the direction of a fantastic possibility that combinatorics and topology (geometry) are possibly just different ways of expressing the same reality!

There are several classes of finite posets which make their appearance throughout mathematics. Among them are

(a) the power set $P([n])$, or the poset of all subspaces of $[n] := \{1, 2, \ldots, n\}$,

(b) the multi set poset of monomials $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_k^{\alpha_k}$ which is alternatively described as the poset $D(n)$ of all divisors of a given integer n,

(c) the poset $L(V)$ of all linear subspaces of a finite dimensional vector space V over a finite field,

(d) the poset Π_n of all partitions of an n-element set.

These are some of the central examples and, following Gian-Carlo Rota, these are test examples for general statements about finite posets.

The *homotopy complementation formula* of A. Björner and J.W. Walker, [8], is a very elegant tool for "computing" the homotopy type of the order complex $\Delta(P)$ such that $\hat{P} = P \cup \{\hat{0}, \hat{1}\}$ is a lattice and for some $c_0 \in P$, the set of all complements $C(c_0)$ of c_0 in $\hat{P}$ is an antichain. In the special case of the partition lattice Π_n, the application of the homotopy complementation formula yields an elegant homotopy recurrence relation (16). From here it is easily deduced by induction that the homotopy type of the lattice Π_n, that is the homotopy type of the order complex $\Delta(\widetilde{\Pi}_n)$, where $\widetilde{\Pi}_n := \Pi_n \setminus \{\hat{0}, \hat{1}\}$, is the wedge of $(n-1)!$ copies of the sphere S^{n-3}.

Most of examples $(a) - (d)$ can be meaningfully extended to the case of general, finite or infinite, topological posets. For example an analog of (c) is the Grassmann poset $\mathcal{G}_n(\mathbb{R})$ of all linear subspaces in $\mathbb{R}^n$. If $G_k(\mathbb{R}^n)$ is the usual Grassmann manifold of all k-dimensional, linear subspaces in $\mathbb{R}^n$, then $\mathcal{G}_n(\mathbb{R}) = \coprod_{i=0}^n G_i(\mathbb{R}^n)$, $\tilde{\mathcal{G}}_n(\mathbb{R}) := \mathcal{G}_n(\mathbb{R}) \setminus \{\hat{0}, \hat{1}\}$ and $\Delta(\tilde{\mathcal{G}}_n(\mathbb{R}))$ is defined as the subspace of the join

$$G_1(\mathbb{R}^n) * G_2(\mathbb{R}^n) * \ldots * G_{n-1}(\mathbb{R}^n)$$

where a simplex $l_1 * l_2 * \ldots * l_{n-1} \subset \Delta(\tilde{\mathcal{G}}_n(\mathbb{R}))$ if and only if $l_1 \subset l_2 \subset \ldots \subset l_{n-1}$, i.e. if and only if $l_1, l_2, \ldots, l_{n-1}$ is a chain.

It turns out that the homotopy complementation formula of Björner and Walker admits an extension to (infinite) topological posets, see [46]. An application of this general result to the Grassmann poset yields the recurrence formula (17). As a consequence, one obtains that $\Delta(\tilde{\mathcal{G}}_n(\mathbb{R}))$ has the homotopy type of the sphere of dimension $\binom{n}{2} + n - 2$, and being a PL-manifold, it is actually homeomorphic o this sphere. This fact was obtained by Vassiliev in [31]. However, it is not a surprise that this mathematical gem or its special cases were found earlier by other mathematicians. Among the predecessors are Borel and Serre [6], N. Kuiper, W. Massey, M.Z. Shapiro, see [32] Section 7.1.5 and [33] for more information. The formulas (18) and (19) are also obtained by applications of the generalized homotopy complementation formulas. Here, $\widetilde{\mathcal{G}}_n^+(R)$ is the Grassmann poset of all oriented, proper subspaces in $\mathbb{R}^n$, while $\exp_n(X)$ is the poset of all (nonempty) k-element subsets of X, for $k \leq n$, topologized by the Vietoris topology or (where applicable) by the Hausdorff metric.

Topological order complexes, just like the finite order complexes are basic structures interesting in itself. Among their most remarkable applications are Vassiliev constructions of *geometric resolutions* of singular spaces. Let us quote give the word to Victor Anatol'evich himself ([33]):

- "If elements of a partially ordered set run over a topological space, then the corresponding order complex admits a natural topology, providing that similar interior points of simplices with close vertices are close to one another. Such *topological order complexes* appear naturally in the *conical resolutions* of many singular algebraic varieties, especially of *discriminant varieties*, i.e. the spaces of singular geometric objects. (...) Using these order complexes we study the cohomology rings of many spaces of nonsingular geometric objects, including the spaces of nondegenerate linear operators in $\mathbb{R}, \mathbb{C} or \mathbb{H}$, or homogeneous functions $\mathbb{R}^2 \to \mathbb{R}$ without roots of high multiplicity in RP^1, of nonsingular hypersurfaces of fixed degree in CP^n, of Hermitian matrices with simple spectra etc." (...).

A general idea behind the *geometric resolution* of a singular variety X is the following. Different points in X are distinguished by different "degrees of singularity". The degrees of singularity form a (topological) partially ordered set $(P, \leq)$. The degree map $D : X \to P$ is assumed to be lower semi-continuous, i.e. the D-inverse images of the lower cones $P_{\leq p}$ are closed sets in X. A geometric resolution $\Gamma(X)$ of X, or more precisely a geometric resolution relative to the function D, is defined by

$$\Gamma(X) = \Gamma_D := \cup_{x \in X} \{x\} \times \Delta(P_{\leq D(p)}) \subset X \times \Delta(P).$$

Here are two examples, the space $\mathcal{P}_n$ of real monic polynomials with multiple (real) roots, and the space $\mathcal{L}_n$ of singular $n \times n$-matrices with real coefficients. The degree maps $D_1 : \mathcal{P}_n \to \exp(\mathbb{R})$ and $D_2 : \mathcal{L}_n \to \tilde{\mathcal{G}}(\mathbb{R}^n)$ are respectively defined by

$$D_1(p) := \{x \in \mathbb{R} \mid p(x) = 0\} \quad \text{and } D_2(A) := \mathrm{Ker}(A).$$

In favorable cases, the projection map $\pi : \Gamma(X) \to X$, from the geometric resolution to the original singular space, is a homotopy equivalence. On the other hand, the topological poset P of "degrees" often comes with a filtration $P_1 \subset P_2 \subset \ldots \subset P$ which induces filtrations both on the order complex $\Delta(P)$ and on the geometric resolution $\Gamma(X)$. For example both $\exp(\mathbb{R})$ and $\tilde{\mathcal{G}}(\mathbb{R}^n)$ have rank functions $r_1 : \exp(\mathbb{R}) \to \mathbb{N}$ and $r_2 : \tilde{\mathcal{G}}(\mathbb{R}^n) \to \mathbb{N}$ defined by

$$r_1(p) = |D_1(p)| \quad \text{and} \quad r_2(A) = \dim(D_2(A)).$$

These rank functions induce filtrations $P_k := \{p \in P \mid \text{rank}(p) \leq k\}$ on the posets which induce filtrations on the associated geometric resolutions $\Gamma(\mathcal{P}_n)$ and $\Gamma(\mathcal{L}_n)$. In both cases the resolutions have the same homotopy type as the original singular spaces. The filtrations induce spectral sequences which can often be used for efficient calculations of the (co)homology of the geometric resolutions and the original singular spaces. For the details and a comprehensive exposition of the general theory, the reader is referred to the book [33].

References

[1] M. Aigner. *Combinatorial Theory,* Springer-Verlag, Berlin 1979.

[2] N. Alon, Some recent combinatorial applications of Borsuk-type theorems, Algebraic, Extremal and Metric Combinatorics, M.M. Deza, P. Frankl, D.G Rosenberg, editors, *Cambridge Univ. Press,* Cambridge 1988, pp. 1–12.

[3] I. Bárány, Geometric and combinatorial applications of Borsuk's theorem, New trends in Discrete and Computational Geometry, János Pach, ed., *Algorithms and Combinatorics 10, Springer-Verlag,* Berlin, 1993.

[4] I. Bárány, J. Matoušek. Simultaneous partitions of measures by k-fans, *Discrete Comput. Geom.* (to appear).

[5] I. Bárány, S.B.Shlosman, A.Szücs. On a topological generalization of a theorem of Tverberg, *J. London Math. Soc. (2)*, vol. 23, 1981, pp. 158–164.

[6] A. Borel, J.-P. Serre. Cohomologie d'immeubles et de groupes S-arithmetiques. *Topology* 15 (1976), 211–231.

[7] P. Billingsley. *Convergence of Probability Measures*, John Wiley & Sons, 1968.

[8] A. Björner, Topological methods, In R. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics.* North-Holland, Amsterdam, 1995.

[9] G.E. Bredon. *Topology and Geometry.* Graduate Texts in Mathematics 139, Springer (1995).

[10] E. Brieskorn. Sur les grupes de tresses. In: Séminaire Bourbaki 1971/72. *Lecture Notes in Math.* 317, Springer 1973, pp. 21–44.

[11] J. Conway and C. Gordon. Knots and links in spatial graphs, *J. Graph Theory*, vol. 7, 1983, pp. 445–453.

[12] C. De Concini, C. Procesi. Wonderful models of subspace arrangements. *Selecta Mathematica*, new Series 1 (1995), 459–494.

[13] P. Deligne, M. Goresky, R. MacPherson. L'algébre de cohomologie du complément dans un espace affine, d'une famille finie de sous-espaces affines, preprint (1999), 23 pp.

[14] J. Eckhoff. Helly, Radon and Carathéodory type theorems, *Handbook of Convex Geometry*, P.M. Gruber and J.M. Wills (eds.), North-Holland, Amsterdam, vol. A, 1993, pp. 389–448.

[15] E.M. Feichtner, G.M. Ziegler. On cohomology algebras of complex subspace arrangements. *Trans. A.M.S.*, 2000.

[16] M. Goresky and R. MacPherson. *Stratified Morse Theory*, volume 14 of *Ergebnisse der Mathematik und ihrer Grenzgebiete.* Springer, Berlin, Heidelberg, New York, 1988.

[17] B. Grünbaum. Partitions of mass–distributions and convex bodies by hyperplanes, *Pacific J. Math.*, 10 (1960), 1257–1261.

[18] H. Hadwiger. Simultane Vierteilung zweier Körper, *Arch. Math.* (Basel), 17 (1966), 274–278.

[19] P. Orlik, L. Solomon. Combinatorics and topology of complements of hyperplanes. *Invent. Math.* 56, 167–189 (1980).

[20] P. Orlik, H. Terao. *Arrangements of Hyperplanes*, 1991.

[21] J. Pach (Ed.). *New Trends in Discrete and Computational Geometry*, Algorithms and Combinatorics 10, Springer (1993).

[22] R. Rado. A theorem on general measure, *J. London Math. Soc.*, vol. 26 (1946), pp. 291–300.

[23] E.A. Ramos. Equipartitions of mass distributions by hyperplanes, *Discrete Comput. Geom.*, vol. 15 (1996), pp. 147–167.

[24] G.C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Z. Wahrscheinlichkeitsrechnung* 2 (1964), 340–368.

[25] G-C. Rota. Ten Mathematics Problems I will never solve. *DMV mitteilungen* 2, 45–52, (1998).

[26] H. Sachs. On spatial representation of finite graphs, in *Finite and Infinite Sets*, Colloq. Janos Bolyai 37 (1981).

[27] C. Schaper. Suspensions of affine arrangements. *Math. Ann.* 309 (1997), 463–473.

[28] P. Seymour. Progress on the four-color theorem. *Proc. Int. Cong. Math. Zürich 1994* Birkhäuser 1995.

[29] S. Shnider, S. Sternberg *Quantum Groups.* Graduate texts in Mathematical Physics, International Press 1993.

[30] R.P. Stanley. *Enumerative Combinatorics*, Vol. 1 and 2. Cambridge University Press 1997 and 1999.

[31] V.A. Vassiliev. Geometric realization of the homology of classical Lie groups and complexes, S–dual to flag manifolds. *St.-Petersburg Math. J.* 3:4, 108–115, (1991).

[32] V.A. Vassiliev *Topology of complements of discriminants* (in Russian). Fazis, Moscow 1997.

[33] V.A. Vassilev Topological order complexes and resolutions of discriminant sets. *Publ. Inst. Math. (Beograd)(N.S.)* 66(88)(1999), 165–185.

[34] S. T. Vrećica and R. T. Živaljević. The ham-sandwich theorem revisited, *Israel J. Math.*, **78** (1992), 21-32.

[35] S. Vrećica, R. Živaljević, New cases of the colored Tverberg theorem, Jerusalem Combinatorics '93, H. Barcelo, G. Kalai (eds.) *Contemporary mathematics,* A.M.S. Providence 1994.

[36] S. T. Vrećica and R. T. Živaljević. Conical equipartitios of mass distributions, *Discrete Comput. Geom.* (to appear).

[37] V. Welker, G.M. Ziegler, R.T. Živaljević. Homotopy colimits – comparison lemmas for combinatorial applications. *J. reine angew. Math.* 509 (1999), 117–149.

[38] S. Yuzvinsky. Small rational models of subspace complements. *Trans. A.M.S.*, (1998).

[39] S. Yuzvinsky. Rational model of subspace complement on atomic complex. *Publ. Inst. Math. (Beograd)(N.S.)* 66(80)(1999), 157–164.

[40] T. Zaslavsky. Facing up to arrangements: Face count formulas for partitions of space by hyperplanes. *Memoirs A.M.S.* 154, 1975.

[41] G. M. Ziegler and R. T. Živaljević. Homotopy types of subspace arrangements via diagrams of spaces. *Math. Ann.*, 295:527–548, 1993.

[42] R. Živaljević. Topological methods, in *CRC handbook of discrete and computational geometry*, J.E. Goodman, J. O'Rourke (eds.), CRC Press, New York (1997).

[43] R. Živaljević. User's guide to equivariant methods in combinatorics, *Publ. Inst. Math. Belgrade*, vol. 59(73) (1996), pp. 114–130.

[44] R. Živaljević. User's guide to equivariant methods in combinatorics II, *Publ. Inst. Math. Belgrade*, vol. 64(78) (1998), pp. 107–132.

[45] R.T. Živaljević, The Tverberg–Vrećica problem and the combinatorial geometry on vector bundles, *Israel J. Math.*, 000–000.

[46] R.T. Živaljević. Combinatorics of topological posets: Homotopy complementation formulas. *Advances in Applied mathematics* 21 (1998), pp. 547–574.

[47] R. Živaljević, S. Vrećica. An extension of the ham sandwich theorem, *Bull. London Math. Soc.*, vol. 22 (1990) pp. 183–186.

[48] R. Živaljević, S. Vrećica. The colored Tverberg's problem and complexes of injective functions, *J. Combin. Theory, Ser. A*, vol. 61 (2) (1992), pp. 309–318.

CONNECTION, METRIC AND CORRESPONDING GEODESIC BALLS AND SPHERES ON ANALYTIC MANIFOLDS

NEDA BOKAN AND MIRJANA DJORIĆ

Dedicated to Professor Veselin Perić on the occasion of his 70th birthday

ABSTRACT. In this survey paper we recall the definitions of geodesic balls on manifolds with different structures. Using the coefficients of power series expansions of their volume, which are locally computable invariants of the structure, the geometric information is obtained and characterizations of some spaces are derived.

§0 INTRODUCTION

There are two basic notions which one can use to develop geometry on a smooth manifold: a metric and a connection. In a very special case, starting from a Riemannian metric, one can construct a uniquely determined connection, called the Levi Civita connection, such that its metric is parallel with respect to this connection and its torsion vanishes. In all other cases, these two basic notions may either be connected with one of the two previously mentioned relations or be without any mutual relations with them. Consequently, one can study the geometry determined either by a metric or by a connection, or by both of them together. A geodesic ball is one interesting object connected with these notions. The main topic of this survey article is to present the development of an idea of a small geodesic ball, its volume function and the corresponding power series expansion and geometry as determined by properties of its coefficients for manifolds with different structures.

Research partially supported by Ministry of Science of Serbia, project MM1646.
2000 *Mathematics Subject Classification*. 53B05, 53A15, 53C40.

The first section deals with geodesic balls of a Riemannian manifold. Geodesic balls for a connection with torsion are considered in the second section, and geodesic balls for torsion-free connections are studied in the third section. We refer to [9], [12], [21], [27] for further details, as well as their references.

Acknowledgements. The authors are grateful to O. Kowalski, V. Miquel, U. Simon, L. Vanhecke and L. Vrancken for valuable discussions concerning this topic.

§1 Geodesic balls and spheres on Riemannian manifolds

Let (M,g) be an n-dimensional analytic Riemannian manifold. More generally, if M is a C^∞ manifold, all considered power series expansions would be defined, but they might not converge. Denote by ∇ its Levi Civita connection and by R the associated Riemannian curvature tensor with components R_{ijkl}, where i,j,k,l are part of an orthonormal basis of the tangent space M_m for some $m \in M$. Further, denote by ρ its Ricci tensor, i.e. $\rho_{ij} = \sum_{k=1}^n R_{ikjk}$, by τ its scalar curvature, i.e. $\tau = \sum_{i,j=1}^n R_{ijij}$ and by Δ the Laplacian. We will always suppose that r is sufficiently small in order to have a diffeomorphic exponential map $\exp_m$ at $m \in M$.

In order to compute the Taylor expansion of the volume function, the general power series expansion of tensor fields in normal coordinates is discussed. Such expansions have been used on several occasions, for example, in the theory of harmonic spaces and in determining the asymptotic expansion for $\sum e^{-\lambda_i t}$, where λ_i are the eigenvalues of the Laplacian of a compact Riemannian manifold.

Let $m \in M$ and let $(x_1,\ldots,x_n)$ be a normal coordinate system defined in a neighborhood of m with $x_1(m) = \cdots = x_n(m) = 0$. In terms of the exponential map, any normal coordinate system of the above type is given by

$$x_j\left(\exp_m\left(\sum_{i=1}^n t_i e_i\right)\right) = t_j, \tag{1.1}$$

where $\{e_1,\ldots,e_n\}$ is an orthonormal basis of M_m ([1], [16]).

If s and σ are the functions defined on neighborhoods of $0 \in M_m$ and $m \in M$ by

$$\begin{aligned} s(x) &= \text{the Euclidean distance from } 0 \text{ to } x, \\ \sigma(p) &= \text{the distance in } M \text{ from } m \text{ to } p, \end{aligned}$$

then

$$\sigma = s \circ \exp_m^{-1}. \tag{1.2}$$

Let $B_0(r)$ be the metric ball of radius r in M_m, i.e.

$$B_0(r) = \{x \in M_m | \|x\| \leq r\}. \tag{1.3}$$

The *geodesic ball* of center m and radius r is the set

$$G_m(r) = \{p \in \mathcal{U} | \sigma(p) \leq r\}.$$

Moreover, by (1.2), we have

$$G_m(r) = \exp_m (B_0(r)).$$

Analogously, the *geodesic sphere* is the set $\{p \in \mathcal{U} | \sigma(p) = r\} = \{\exp_m(x) | x \in M_m, \|x\| = r\}$. Further, let $S_m(r)$ denote the $(n-1)$-dimensional volume of the geodesic sphere (area) and $V_m(r)$ the n-dimensional volume of the corresponding geodesic ball. Since $*ds$ is the volume element of any sphere in M_m and, by the Gauss lemma, $*d\sigma$ is the volume element of any small geodesic sphere in M with center m, it can be proved that

$$S_m(r) = r^{n-1} \int_{S^{n-1}(1)} \omega_{1\ldots n}(\exp_m ru) du,$$

where ω is the standard volume form on M, $\omega_{1\ldots n} = \omega\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right)$ and $(x_1, \ldots, x_n)$ is the system of normal coordinates on M at m. Moreover, we have $V_m(r) = \int_0^r S_m(t)dt$. For the proof and more details we refer to [12] and [13]. Then $V_m(r)$ can be expanded in a power series in r by using a normal coordinate system [11], [13] or with the use of Jacobi vector fields [1], [27]. The coefficients of r^{n+k} vanish provided k is odd and the coefficients of r^{n+k} for even k are given by formulas in the invariants of the curvature operator:

$$V_m(r) = \frac{\alpha_n}{n} r^n \{1 + Ar^2 + Br^4 + Cr^6 + \cdots + O(r^k)\}_m,$$

where $\alpha_n = 2\Gamma\left(\frac{1}{2}\right)^n \Gamma\left(\frac{n}{2}\right)^{-1}$ is the volume of a unit sphere S^n in Euclidean n-space.

The history of power series expansions of volume functions begins in the middle of the 19th century when Bertrand, Diguet and Puiseux [2] computed the first two non-zero terms in the power series expansion for the volume of a geodesic ball at a point m in a surface in $\mathbb{R}^3$:

$$V_m(r) = \pi r^2\{1 - \frac{K}{12}r^2 + O(r^4)\}_m,$$

where K is the Gauss curvature of this surface. Their motivation was to give a new proof of the famous theorema egregium of Gauss, i.e. to prove that the Gauss curvature of a surface in $\mathbb{R}^3$ does not depend on the embedding. The generalization of this formula to Riemannian manifolds was first given in 1917 in a paper by H. Vermeil [26] and then in 1939 in a paper by H. Hotelling [17], and the next term was computed in 1973 by A. Gray [11]:

$$\begin{aligned} A &= -\frac{\tau}{6(n+2)}, \\ B &= \frac{1}{360(n+2)(n+4)}\left(-3\|R\|^2 + 8\|\rho\|^2 + 5\tau^2 - 18\Delta\tau\right). \end{aligned} \tag{1.4}$$

A. Gray derived several consequences of this expansion. For example, he used it in [11] to obtain a local comparison theorem. Namely, for an analytic Riemannian manifold M and $m \in M$, its Ricci scalar curvature is positive if and only if

$$V_m(r) < \frac{\alpha_n}{n} r^n,$$

and its Ricci scalar curvature is negative if and only if

$$V_m(r) > \frac{\alpha_n}{n} r^n,$$

for sufficiently small $r > 0$. Note that for Euclidean space $\mathbb{R}^n$

$$V_m(r) = \frac{\alpha_n}{n} r^n.$$

It is interesting that this result is neither stronger nor weaker than the Bishop-Günther inequalities. On one hand, A. Gray's result holds only for sufficiently small $r > 0$, while the Bishop-Günther inequalities are valid for r up to the first conjugate point. On the other hand, the condition that the scalar curvature be positive at m is weaker than positivity at m of either the sectional curvature or Ricci curvature. See [3] and [4, p. 256] for more details.

The coefficient of r^{n+4} in the expansion of $V_m(r)$ is especially interesting since it is a quadratic invariant of $O(n)$. In the same paper A. Gray compared it with other quadratic invariants which arise from geometrical considerations. Notable among them are the conformal and spectral quadratic invariants and the 4-dimensional Gauss-Bonnet integrand. He also discussed the linear independence among them and the quadratic invariant derived from $V_m(r)$, described above.

A. Gray and L. Vanhecke [13] computed the fourth non-zero term in the expansion of $V_m(r)$:

$$
\begin{aligned}
C = \frac{1}{720(n+2)(n+4)(n+6)} \Big(& -\frac{5}{9}\tau^3 - \frac{8}{3}\tau\|\rho\|^2 + \tau\|R\|^2 + \frac{64}{63}\check{\rho} - \frac{64}{21}\langle \rho\otimes\rho, \bar{R}\rangle \\
& + \frac{22}{7}\langle \rho, \dot{R}\rangle - \frac{110}{63}\check{R} - \frac{200}{63}\check{\bar{R}} + \frac{45}{7}\|\nabla\tau\|^2 + \frac{45}{14}\|\rho\|^2 \\
& + \frac{45}{7}\alpha(\rho) - \frac{45}{14}\|\nabla R\|^2 + 6\tau\Delta\tau + \frac{48}{7}\langle\Delta\rho, \rho\rangle \\
& + \frac{54}{7}\langle\nabla^2\tau, \rho\rangle - \frac{30}{7}\langle\Delta R, R\rangle - \frac{45}{7}\Delta^2\tau \Big) .
\end{aligned}
$$

Since the coefficients of r^{n+4} and r^{n+6} in the expansion of $V_m(r)$ are respective linear combinations of the orders 4 and 6 invariants of the curvature operator, it may be possible to consider various problems related to the volume of geodesic spheres and balls of a Riemannian manifold and its various geometrical and topological properties.

The main purpose of [13] was to study to what extent the functions $V_m(r)$ determine the Riemannian geometry of the ambient space. In particular, the authors were concerned with the following

Conjecture. *Let M be an n-dimensional Riemannian manifold such that for all $m \in M$ and all sufficiently small $r > 0$ we have*

$$V_m(r) = \frac{\alpha_n}{n} r^n,$$

i.e. $V_m(r)$ coincides with the volume of a geodesic ball with radius r in Euclidean space. Then M is locally flat.

In this paper the authors compared the volume of a small geodesic ball with center m and radius r in an arbitrary n-dimensional Riemannian analytic manifold with the volume of a ball of radius r in Euclidean space, using the power series expansion for $V_m(r)$. One may formulate similar conjectures for other two-point homogeneous spaces. This conjecture is true in dimensions 2 and 3 and in some special cases, but is still unresolved in higher dimensions and in general case. We list several more special cases when this Conjecture is true and we refer to [13] for further details:

(1) M has non-positive or non-negative Ricci curvature (in particular if M is Einstein);
(2) M is conformally flat;
(3) M is Bochner flat Kähler manifold;
(4) M is a product of surfaces;
(5) M is a 4- or 5-dimensional manifold with parallel Ricci tensor;
(6) M is compact and the Laplacian of M has the same spectrum on functions as that of a compact flat manifold;
(7) M is a compact, oriented four-dimensional manifold whose Euler characteristic and signature satisfy $\aleph(M) \geq -\dfrac{3}{2}|\tau(M)|$;
(8) M is the product of symmetric spaces of classical type.

Although the foregoing conjecture is still unresolved in higher dimensions, there is a series of results obtained from studying the problem to what extent the expression for the volume of the geodesic ball under certain conditions is characteristic for the manifold. We illustrate these studies with some more results.

Theorem. [5] *Let M be a compact Kähler manifold with complex dimension n, and suppose that for all $m \in M$ and all sufficiently small $r > 0$, $V_m(r)$ is the same as that of an n-dimensional compact Kähler manifold $M(\mu)$ with constant holomorphic sectional curvature μ. Let ω and ω_μ denote the fundamental classes of M and $M(\mu)$. If their generalized Chern numbers satisfy the conditions*

$$\omega^{n-1}c_1(M) = \omega_\mu^{n-1}c_1(M(\mu)),$$
$$\omega^{n-2}c_1^2(M) \geq \omega_\mu^{n-2}c_1^2(M(\mu)),$$

then M has constant holomorphic sectional curvature μ.

Let us introduce *model spaces* to be the flat space E^n and the rank one symmetric spaces. The volume functions $S_m(r)$ and $V_m(r)$ for these spaces can be computed completely by using Jacobi vector fields. See, for example [1], [11], [27], for more details.

Further, another interesting problem is to construct a manifold for which the volume of a geodesic ball at each point approximates the volume of a geodesic ball in a model space. For example, A. Gray and L. Vanhecke [13] constructed interesting examples of non-flat manifolds for which

$$V_m(r) = \frac{\alpha_n}{n} r^n \{1 + O(r^6)\},$$

for all $m \in M$ and sufficiently small $r > 0$. One of these examples is a 4-dimensional positive definite metric which is a generalization of the Schwarzschild metric. Another one is a homogeneous 5-dimensional metric. Moreover, they used a different technique to find a manifold of dimension 734 with

$$V_m(r) = \frac{\alpha_n}{n} r^n \{1 + O(r^8)\}.$$

O. Kowalski [18] developed a method for the construction of homogeneous Riemannian spaces with the property

$$V_m(r) = \frac{\alpha_n}{n} r^n \{1 + O(r^{2k})\}$$

and he constructed a direct product of homogeneous spaces with the property

$$V_m(r) = \frac{\alpha_n}{n} r^n \{1 + O(r^{16})\}.$$

C. Ueda [25] constructed other examples using Kowalski's results [18].

M. Djorić and L. Vanhecke [10] obtained other new characterizations of two-point homogeneous spaces considering the volumes of geodesic spheres, balls and circumscribing tubes. Namely, let $\sigma : [a, b] \longrightarrow (M, g)$ be a smooth embedded geodesic through m and denote by $U_\sigma(r)$ the *tubular neighborhood* of radius r about σ, i.e.

$$U_\sigma(r) = \{p \in M \,|\text{there exists a geodesic } \gamma \text{ of } M \text{ through } p \text{ cutting } \sigma \text{ orthogonally and with length } L(\gamma) \leq r\},$$

where the radius r is smaller than the distance from σ to its nearest focal point. If $\sigma : [-r, r] \longrightarrow (M, g)$ is a unit speed geodesic such that $\sigma(0) = m$, then the set of points of $U_\sigma(r)$ at distance r from σ is called the *circumscribing tube* of the geodesic sphere with center m, axial curve σ and radius r. Such a circumscribing tube generalizes the notion of a circumscribing cylinder of a sphere in Euclidean 3-space. Let $V_\sigma^c(r)$ and $S_\sigma^c(r)$ denote the n-dimensional and $(n-1)$-dimensional volume of the circumscribing tube of the geodesic sphere with σ as axial curve. The power series expansions for these volumes were computed in [14], [15], [12] and [27] by using Fermi coordinates, Fermi vector fields and Jacobi vector fields.

In [10] the authors generalized the old result of Archimedes who proved that the ratio of the area and volume of a sphere and a circumscribing cylinder is constant in three-dimensional Euclidean space. The authors considered several relations between volumes of geodesic spheres, geodesic balls, circumscribing tubes and geodesic disks and derived new local characterizations of two-point homogeneous spaces. Several of these relations are either direct generalizations of classical results in Euclidean geometry or related to some isoparametric inequalities. For example, if one of the ratios $\dfrac{V_\sigma^c}{V_m}$ or $\dfrac{S_\sigma^c}{S_m}$ is constant, for all $m \in M$, all geodesics σ through m and all suficiently small r, then the manifold is locally flat. Hence, the property of Archimedes is characteristic for locally Euclidean geometry. More generally, using the explicit expressions for these ratios for two-point homogeneous spaces, the authors proved that such expressions determine these spaces up to local isometry.

§2 Geodesic balls and spheres on manifolds with metric connection with torsion

Let (M^n, g) be an analytic manifold with a Riemannian metric g and a metric connection D with non-vanishing torsion tensor T. There exist various problems, arising naturally in physics and other areas, which can be studied by using this connection. Let us remark that although the Levi Civita connection for a given metric is uniquely determined, a metric connection with torsion is not unique, for a given metric. However, for a given almost Hermitian manifold (M, g, J), there exists the unique metric connection D called the characteristic conection, which satisfies

$$DJ = 0, \quad T(X,Y) + T(JX, JY) = 0.$$

Using this special connection, V. Miquel studied in [19] the volumes of certain small geodesic balls on almost Hermitian manifolds and expressed the first non-trivial coefficient in the power series expansion of this volume as an almost Hermitian invariant of order two. He proved that, in a certain sense, this coefficient determines some classes (and only these) of almost Hermitian manifolds. Moreover, he obtained a better characterization for the nearly Kähler manifolds. Namely, he proved that if this coefficient is the same as the corresponding one for the Levi Civita connection, then M is a nearly Kähler manifold. Using the Levi Civita connection, the information about the almost complex structure cannot be obtained, since the coefficients in this case are metric invariants. Further, he proved that this coefficient and the spectrum of the complex Laplacian, together, determine the class in which a compact Hermitian manifold lies.

Therefore, let us recall the notion of geodesic ball in a manifold with metric connection with torsion [21]. First, let $\exp_m : \tilde{\mathcal{U}} \to \mathcal{U}$ be the exponential map associated to a metric connection D which is a diffeomorphism on $\tilde{\mathcal{U}}$. For any $p \in \mathcal{U}$ there exists a unique D-geodesic arc joining m and p. Then, if $\delta^D(m,p)$ is the length of this geodesic arc,

$$\delta^D(m,p) = \| \exp^{-1}(p) \|, \tag{2.1}$$

since the velocity vector of a geodesic, for a metric connection, has constant length. Using (2.1) it follows

$$B_r^D(m) = \exp_m(B_0(r)) \tag{2.2}$$

where $B_r^D(m) = \{p \in \mathcal{U} | \delta^D(m,p) \leq r\}$ is the so called *D-geodesic ball* of center m and radius r and $B_0(r)$ is defined by (1.3). Further, since the Gauss lemma fails for general metric connection, V. Miquel used polar coordinates in [20] to obtain an integral formula for the volume $V_m^D(r)$ of D-geodesic ball. To obtain the power series expansion for $V_m^D(r)$, he used the normal coordinates $(x_1, \ldots, x_n)$ defined by (1.1) for the exponential map associated to metric connection D. For an orientable manifold M, choosing normal coordinates such that $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\}$ is a positively oriented local frame, there is a unique volume form ω such that $\omega\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) = 1$ and $D\omega = 0$. Using the general power expansions of tensor fields in normal coordinates and the integral formula for $V_m^D(r)$, V. Miquel in [20] and [21] derived, following the method given in [11], the first non-trivial term in the power series expansion for $V_m^D(r)$:

$$A_D = \frac{1}{n+2}\left(-\frac{1}{6}\tau_D + \frac{1}{3}\sum_{i,j=1}^{n} D_i(T)_{ijj} + \frac{1}{24}\sum_{i,j,k}^{n} T_{ijk}T_{ikj} + \frac{1}{8}\|\overline{T}\|^2\right),$$

where τ_D is the scalar curvature of D at m and $\overline{T}$ is the one form defined by $T_X = \sum_{j=1}^{n} T_{XE_jE_j}$ for any orthonormal frame $\{E_1, \ldots, E_n\}$. He proved in [20] that the metric connection D and the Levi Civita connection ∇ have the same geodesics if and only if $A = A^D$ for any $m \in M$.

Moreover, V. Miquel in [22] computed the first non-trivial term in the power series expansion for the area of a geodesic sphere associated to a metric connection with torsion ($S_m^D(r)$). Since the Gauss Lemma fails for D-geodesic spheres, to overcome these difficulties, he used generalized Jacobi fields (D-Jacobi fields) for connections with torsion and derived the following formula:

$$S_m^D(r) = \alpha_n r^{n-1}\left\{1 - \frac{1}{6n}\left(\tau^D + 2\partial\bar{T} + \frac{1}{6}\|T^1\|^2 - \frac{11+n}{12(n+2)}\|T^2\|^2 - \frac{3n+1}{12}\|T^3\|^2\right) r^2 + O(r^3)\right\}.$$

V. Miquel noticed that although $B_r^D(m) \subset G_m(r)$, where $G_m(r)$ is defined in section 1, for the Levi Civita connection ∇, there is no a priori relation between $S_m(r)$ and $S_m^D(r)$. The consequence of the foregoing formula is that, for r small enough, $S_m(r) \leq S_m^D(r)$.

§3 Geodesic balls on manifolds with torsion-free connection

In this section we consider an analytic structure (M, D, g) where D is a torsion-free and Ricci-symmetric (that means its Ricci tensor is symmetric) connection, which is not necessarily the Levi Civita connection of the metric g. The assumption that Ricci tensor Ric^D is symmetric is equivalent to the existence of D-parallel volume form ω, i.e. $D\omega = 0$.

This structure appears in several situations, for example, statistical manifolds, Codazzi transformations for PDEs, Weyl structures, conjugate triples and hypersurfaces in affine space and in space forms, which motivate a study of a structure (M, D, g). In [7], [8], [9] the authors modified investigations of A. Gray and other authors on small geodesic balls and spheres and considered generalized balls using this structure.

Let us introduce the following notations for the structure (M, D, g): $\nabla(g)$ will denote the Levi Civita connection of g, $\omega(g)$ its oriented Riemannian volume form, and ω an oriented volume form parallel with respect to D, that means $D\omega = 0$. ω is unique modulo a constant non-zero factor. Assume that $\omega(g)$ and ω induce the same orientation; then there exists a positive function μ such that $\omega(g) = \mu\omega$. The torsion-free connections D and $\nabla(g)$ define a $(1,2)$-tensor field $C := D - \nabla(g)$; we denote its trace by $n\hat{T}(X) := \mathrm{tr}\{Y \mapsto C(X,Y)\}$. Using straightforward computations we obtain elementary relations ([9], Lemma 1.1) between the structures of D and g; for example:

$$n\hat{T} = -d \lg \mu,$$

$$D\omega(g) = -n\hat{T} \otimes \omega(g).$$

These relations and the introduction of the difference tensor C allow us to reduce the covariant differentiation to an algebraic computation in terms of the tensor product. Let us remark that these computations fail for a connection which is not Ricci-symmetric. Moreover, from the definition of C and the foregoing formulas we can give geometric interpretations of C and T: C measures the deviation of the connections D and $\nabla(g)$, while T is a measure for the deviation of the volume forms.

Further, let us recall the notion of a geodesic ball on a manifold with torsion-free connection D [8]. For $v \in M_m$, let ${}^D\xi_v$ denote the unique D-geodesic in M with ${}^D\xi_v(0) = m$ and ${}^D\xi_v'(0) = v$. We write ${}^D\exp_m(v) = {}^D\xi_v(1)$, provided that ${}^D\xi_v(t)$ can be defined for $t = 1$. The map ${}^D\exp_m$ may be defined only on a neighborhood of $0 \in M_m$, where it is a diffeomorphism. In case that D is not the Levi Civita connection of g, it follows that the length $\|{}^D\exp_m^{-1}(p)\|$ is not necessarily constant if p varies along a D-geodesic σ through $m \in M$. Thus define a geodesic ball as follows: Let r be small enough so that the map ${}^D\exp_m$ is defined on a ball of radius r in the tangent space M_m. Now, let

$$G_m^D(r) = {}^D\exp_m\left(B_0(r)\right),$$

where $B_0(r)$ is defined by (1.3). We call $G_m^D(r)$ a *D-geodesic g-ball* with center m and radius r. Let $\mathcal{V}_m^D(r)$ denote the volume of $G_m^D(r)$ with respect to the metric g. Then

$$\mathcal{V}_m^D(r) = \int_0^r t^{n-1}\left(\int_{S^{n-1}(1)} \omega(g)({}^D\exp_m(tu))du\right)dt,$$

where u varies on the unit sphere $S^{n-1}(1)$ in M_m. To compute the Taylor expansion of this volume in terms of local invariants of the geometry of (D, g), in particular in terms of the invariants C and $\hat{T}$ and curvature invariants of D, we modify a method described in [11], [12] and [13], and extend the notion of "normal coordinates" to the structure (M, D, g). Let $\{e_1, \ldots, e_n\}$ be a g-orthonormal basis of M_m. We define a real-valued function x_j on a neighborhood of m by $x_j\left({}^D\exp_m\left(\sum t_i e_i\right)\right) = t_j$. Then $(x_1, \ldots, x_n)$ is called the system of *D-normal coordinates* corresponding to $\{e_1, \ldots, e_n\}$. Considering the local Gauss basis $\partial_1, \ldots, \partial_n$ associated to a D-normal coordinate system $(x_1, \ldots, x_n)$, a *D-normal coordinate vector field* at m is a local vector field X of the form $X = \sum a_i \partial_i$, where the a_i's are constants. For many purposes it turns out to be far easier to work with D-normal coordinate vector fields instead of D-normal coordinates. For example, the notion of a D-normal coordinate vector field at m does not depend on the choice of the D-normal coordinate system at m. Following [11],

U. Simon and the authors of this article in [9] expanded $\omega(g)({}^D\exp_m(ru))$ in a power series in r and derived the formula for $\mathcal{V}_m^D(r)$ with the first three non-zero terms:

$$\begin{aligned}\mathcal{V}_m^D(r) = \alpha_n \frac{r^n}{n} \Bigg\{ 1 &+ \frac{1}{2(n+2)} \left(\Delta \lg \mu + 2\|\operatorname{grad} \lg \mu\|^2 - \frac{1}{3}\operatorname{tr}_g Ric^D \right) r^2 \\ &+ \frac{1}{24(n+2)(n+4)} \big[3\Box\Box(\lg \mu) + 12 g(\operatorname{grad} \lg \mu, \operatorname{grad}\Box \lg \mu) \\ &\quad + 6g(Hess^D \lg \mu, Hess^D \lg \mu) + 3(\Box \lg \mu)^2 \\ &\quad + 12 Hess^D(\lg \mu)(\operatorname{grad} \lg \mu, \operatorname{grad} \lg \mu) \\ &\quad + 6\Box \lg \mu \|\operatorname{grad} \lg \mu\|^2 + 3\|\operatorname{grad} \lg \mu\|^4 \\ &\quad - 4Ric^D(\operatorname{grad} \lg \mu, \operatorname{grad} \lg \mu) - 4g(\operatorname{div} Ric^D, \operatorname{grad} \lg \mu) \\ &\quad + 8\sum_{i,j,h} R^h{}_{jij} D_i \lg \mu D_h \lg \mu + 2\sum_{i,j,h} D_j R^h{}_{iji} D_h \lg \mu \\ &\quad + 2\sum_{i,j,h} R^h{}_{iji} D_j D_h \lg \mu - 4R^{ij} D_j D_i (\lg \mu) \\ &\quad - 2(\Box \lg \mu + \|\operatorname{grad} \lg \mu\|^2)\operatorname{tr}_g Ric^D - 2g(\operatorname{grad}\operatorname{tr}_g Ric^D, \operatorname{grad} \lg \mu) \\ &\quad - \frac{3}{5}\Box(\operatorname{tr}_g Ric^D) + \frac{1}{3}(\operatorname{tr}_g Ric^D)^2 - \frac{6}{5}\sum_{i,j} D_j D_i R_{ij} + \frac{2}{3}\|Ric^D\|^2 \\ &\quad - \frac{2}{15}\sum_{i,j,s,h} \left(R^h{}_{isi} R^s{}_{jhj} + R^h{}_{ijs} R^s{}_{jih} + R^h{}_{ijs} R^s{}_{ijh} \right) \big] r^4 \Bigg\}_m + O(r^{n+6}).\end{aligned}$$

Here we use the following notation: For a differentiable function f, its Hessian $Hess^D(f)$ is given by

$$Hess^D(f)(X,Y) := XY(f) - df(D_X Y)$$

and we introduce a Laplace type operator by

$$\Box f := \operatorname{tr}_g(Hess^D(f)),$$

its metric Hessian is denoted by $Hess_g(f)$ and its Laplacian by

$$\Delta f := \operatorname{tr}_g Hess_g(f).$$

After straightforward computation, it follows that for $D = \nabla(g)$ the foregoing formula specializes to (1.4).

Another important special case of our computations is that of a Blaschke structure, see e.g. [24], chapters 4 and 6. In case of our structure (M, D, g) we have the equivalence: $\hat{T} \equiv 0 \iff \omega(g) = \omega$ and therefore $D\omega(g) \equiv 0$. Extending the terminology from affine hypersurface theory to the structure (M, D, g), we call (M, D, g) a Blaschke structure if $\hat{T} \equiv 0$.

If (M, D, g) is a Blaschke structure, for a D-geodesic g-ball, and for $r > 0$ sufficiently small, it is possible to compare its volume with that of a Euclidean sphere $S^n(r)$:

- (i) if D is flat then $\mathcal{V}_m^D(r) = \mathcal{V}(S^n(r)) + O(r^{n+6})$;
- (ii) if D satisfies $\mathrm{tr}_g Ric = 0$ then $\mathcal{V}_m^D(r) = \mathcal{V}(S^n(r)) + O(r^{n+4})$;
- (iii) regarding the expansion up to the order (n+2), the sign of $\mathrm{tr}_g Ric$ determines whether the map ${}^D\exp$ has a decreasing or increasing effect for the volume functions considered.

It is interesting to compare the foregoing result with the corresponding ones in section 1, for Levi Civita connection.

Furhter, consider an $(n+1)$-dimensional affine space $\mathcal{A}^{n+1}$ with associated vector space V, and a determinant form Det fixing an oriented volume. Let $x : M^n \longrightarrow \mathcal{A}^{n+1}$ be an analytic, locally strongly convex, embedded hypersurface with so-called Blaschke structure induced by the unimodular affine normal. The convexity condition implies that the Blaschke metric g on M is Riemannian. The metric g, together with the connection D, induced from the affine normal, and the conormal connection D^*, define a conjugate triple. For Blaschke hypersurfaces, the variational problem for the area functional leads to the Euler-Lagrange equation $H \equiv 0$. Following Calabi, Blaschke hypersurfaces with $H \equiv 0$ are called *affine maximal*. The power series expansion for $\mathcal{V}_m^D(r)$ implies:

- (i) for affine maximal hypersurfaces we have $\mathcal{V}_m^{D^*}(r) = \mathcal{V}(S^n(r)) + O(r^{n+4})$ and $\mathcal{V}_m^D(r) = \mathcal{V}(S^n(r)) + O(r^{n+4})$;
- (ii) for improper affine spheres we have $\mathcal{V}_m^{D^*}(r) = \mathcal{V}(S^n(r)) + O(r^{n+6})$ and $\mathcal{V}_m^D(r) = \mathcal{V}(S^n(r)) + O(r^{n+6})$.

Further, in [9] the authors considered the structure (M, D, g) by studying the projective changes of D and projective flatness of D and the influence of the generalized geodesic ball expansion in such geometric situations. Especially, since it is interesting to study the problem of uniqueness of projectively flat connections, they proved that if g is a metric and D and $D^{\#}$ are projectively flat connections such that

(i) M is diffeomorphic to the sphere S^n;

(ii) the generalized geodesic ball expansions for (D, g) and $(D^{\#}, g)$ coincide up to order $n+2$;

(iii) $\omega(D) = \omega(D^{\#})$;

(iv) the pairs (D, g) and $(D^{\#}, g)$ satisfy Codazzi equations;

then both connections coincide: $D = D^{\#}$.

In the foregoing results two structures on a given hypersurface are compared. However, numerous applications can be obtained in different relative normalizations by comparing two hypersurfaces. For example,

(i) If $x, x^{\#}$ are affine Blaschke hyperspheres and the generalized geodesic ball expansions for $\mathcal{V}_m^{D^*}(r)$ and $\mathcal{V}_m^{D^{*\#}}(r)$ coincide up to order $n+2$, both hyperspheres are of the same type (elliptic, parabolic, hyperbolic).

(ii) If $x, x^{\#}$ are affine 2-spheres with the same metric and the generalized geodesic ball expansions for $\mathcal{V}_m^{D^*}(r)$ and $\mathcal{V}_m^{D^{*\#}}(r)$ coincide up to order $n+2$, then x, $x^{\#}$ are unimodularly congruent.

(iii) Let $x, x^{\#}$ be complete Blaschke hyperspheres such that the generalized geodesic ball expansions for $\mathcal{V}_m^{D^*}(r)$ and $\mathcal{V}_m^{D^{*\#}}(r)$ of order $n+2$ coincide. If x is an elliptic paraboloid, then $x^{\#}$ is an elliptic paraboloid.

Let $x, x^{\#} : M \longrightarrow \mathcal{A}^{n+1}$ be relative hyperovaloids. Assume that

(i) the relative metrics coincide: $h = h^{\#}$;

(ii) for the conormal connections D^* and $D^{*\#}$, the generalized geodesic ball expansions for $\mathcal{V}(D^*, h; m, r)$ and $\mathcal{V}(D^{*\#}, h^{\#}; m, r)$ coincide up to order $n+2$;

(iii) $\omega^* = \omega^{*\#}$.

Then $x, x^{\#}$ are affinely equivalent, which means that there exists an affine transformation $\mathcal{T} : \mathcal{A}^{n+1} \longrightarrow \mathcal{A}^{n+1}$ such that $x^{\#} = \mathcal{T}x$. Moreover, if Blaschke hyperovaloids x,$x^{\#}$ satisfy the conditions (i), (ii) from the previous theorem, then $x, x^{\#}$ are unimodularly equivalent.

Let us remark that that the foregoing assertions hold true if one considers the induced connections D, $D^{\#}$ instead of the conormal connections.

Further, onsider a non-degenerate hypersurface $x : M \longrightarrow \mathbb{R}^{n+1}$ such that its position vector is transversal. Then $y(c) := -x$ is called the *centroaffine normal*. Following Nomizu, we call such a hypersurface together with its centroaffine normalization a *centroaffine hypersurface*. The associated geometry is invariant under the group $GL(n+1,\mathbb{R})$. Then if $x : M \longrightarrow \mathbb{R}^{n+1}$ is a hyperovaloid with centroaffine normalization and the generalized geodesic ball expansions for $\mathcal{V}_m^D(r)$ and $\mathcal{V}_m^{D^*}(r)$ coincide up to order $n+2$, then x is a hyperellipsoid.

For further applications, details and references we refer to [9].

References

[1] M. Berger, *Le spectre des variétés riemanniennes*, Rev. Roumaine Math. Pures Appl. **13** (1968), 915-931.

[2] J. Bertrand, C. F. Diguet and V. Puiseux, *Démonstration d'un théorème de Gauss*, Journal de Mathématiques **13** (1848), 80-90.

[3] R. L. Bishop, *A relation between volume, mean curvature and diameter*, Notices Amer. Math. Soc., **10** (1963), 364 Abstracts #63T-196.

[4] R. L. Bishop and R. J. Crittenden, *Geometry of manifolds*, Academic Press, New York, 1964.

[5] N. Blažić, *The volumes of small geodesic balls and generalized Chern numbers of Kähler manifolds*, Nagoya Math. J. **116** (1989), 181-189.

[6] N. Bokan, *Complex conformal connection of Kähler manifold, its small geodesic balls and generalized Chern numbers*, J. Ramanujan Math. soc. **4 (1)** (1989), 93-108.

[7] N. Bokan and M. Djorić, *On power series expansions of tensor fields for a torsion free connection*, Proc. 10th Yugoslav Congress of Math. (2001), 185-188.

[8] N. Bokan, M. Djorić and U. Simon, *An extension of Gray's investigation on small geodesic balls*, Contemporary Mathematics, The Mathematical Legacy of Alfred Gray, Proceedings of the International Congress on Differential Geometry, September 2000, Bilbao, Spain **288** (2001), 268-272.

[9] N. Bokan, M. Djorić and U. Simon, *Geometric structures as determined by generalized geodesic balls*, Result. Math. **43** (2003), 205-234.

[10] M. Djorić and L. Vanhecke, *A theorem of Archimedes about spheres and cylinders and two-point homogeneous spaces*, Bull. Austral. Math. Soc. **43** (1991), 283-294.

[11] A. Gray, *The volume of a small geodesic ball in a Riemannian manifold*, Michigan Math. J. **20** (1973), 329-344.

[12] A. Gray, *Tubes*, Addison-Wesley, Advanced Book Program, Reading, 1990.

[13] A. Gray and L. Vanhecke, *Riemannian geometry as determined by the volumes of small geodesic balls*, Acta Math. **142** (1979), 157-198.

[14] A. Gray and L. Vanhecke, *The volumes of tubes about curves in a Riemannian manifold*, Proc. London Math. Soc. **44** (1982), 215-243.

[15] A. Gray and L. Vanhecke, *The volumes of tubes in a Riemannian manifold*, Rend. Sem. Mat. Univ. Politec. Torino **39** (1981), 1-50.

[16] N. Hicks, *Notes on differential geometry*, Van Nostrand, New York, 1965.

[17] H. Hotelling, *Tubes and spheres in n-spaces and a class of statistical problems*, Amer. J. Math. **61** (1939), 440-460.

[18] O. Kowalski, *Additive volume invariants of Riemannian manifolds*, Acta Math. **145** (1980), 205-225.

[19] V. Miquel, *Volumes of certain small geodesic balls and almost-Hermitian geometry*, Geom. Dedicata **15** (1984), 261-267.

[20] V. Miquel, *Volumes of small geodesic balls for a metric connection*, Compositio Math. **46** (1982), 121-132.

[21] V. Miquel, *Volúmenes de pequeñas bolas geodeésicas asociades a conexiones métricas con torsion. Aplicaciones*, doct. thesis, Universidad de Valencia, 1979.

[22] V. Miquel, *Volumes of geodesic balls and spheres associated to a metric connection with torsion*, Contemporary Mathematics, The Mathematical Legacy of Alfred Gray, Proceedings of the International Congress on Differential Geometry, September 2000, Bilbao, Spain **288** (2001), 119-128.

[23] V. Miquel and A. M. Naveira, *Sur la relation entre la fonction volume de certaines boules geodésiques et la géometrie d'une variété riemannienne*, C.R. Acad. sc. Paris **290**, Serie A, 379-381.

[24] U. Simon, A. Schwenk-Schellschmidt and H. Viesel, *Introduction to the affine differential geometry of hypersurfaces*, Lecture Notes Science University Tokyo, 1991, ISBN 3798315299.

[25] C. Ueda, *Additive volume invariants of Riemannian manifolds*, Tokyo J. Math. **23 (1)** (2000), 101-112.

[26] H. Vermeil, *Notiz über das mittlere Krümmungsmass einer n-fach ausgedehten Riemanschen Mannigfaltigkeit*, Akad. Wiss. Göttingen Nach. (1917), 334-344.

[27] L. Vanhecke, *Geometry in normal and tubular neighborhoods,*, Rend. Sem. Fac. Sci. Univ. Cagliari Supplemento al **58** (1988), 73-176.

University of Belgrade, Faculty of Mathematics,
Studentski trg 16, p.p. 550, 11001 Belgrade, Serbia and Montenegro
E-mail address: neda@matf.bg.ac.yu, mdjoric@matf.bg.ac.yu,

Algebras of generalized functions-two approaches
Boundary and Cauchy problems with singularities

Stevan Pilipović,
Department of Mathematics and Informatics, University of Novi Sad

Dedicated to Professor Veselin Perić on the occasion of his 70 birthday

Introduction

The aim of this survey article is to explain a general concept of generalized function algebras (Part I) and to illustrate the analysis of equations with singularities within these algebras (Part II). While Part I is simply acceptable, Part II gives just fragments of our approach. We refer to the literature for more information about any of quoted classes of equations.

Colombeau had constructed his well-known algebras by purely algebraic methods. Since then, algebras of Colombeau generalized numbers and functions became a very useful framework for linear problems with singularities and specially for non linear problems [1, 2, 5, 24].

Many linear and nonlinear problems with irregular data or irregular coefficients, have been successfully analyzed by the mean of appropriate approximations through nets of C^∞ functions which fits into Colombeau algebra $\mathcal{G}$ of generalized functions. We extend the references in order to mark a part of large literature related to linear and nonlinear equations in the framework of generalized function algebras.

In Part I, we present a very general construction of generalized function Colombeau type algebras through a purely topological description of Colombeau type algebras. We will show that such algebras fit very well in the general theory of the well known sequence spaces forming appropriate algebras [11]. All these classes of algebras are simply determined by the (locally convex) space E, and a sequence of weights $r : \mathbb{N} \to \mathbb{R}_+$ (or sequence of sequences) which serves to construct an ultrametric on the sequence space $E^{\mathbb{N}}$. The sequence $r = (r_n)_n$ is assumed to be decreasing to zero. This implies that sequence spaces under consideration ($\subset E^{\mathbb{N}}$) contain as a subspace $E \sim \operatorname{diag} E^{\mathbb{N}}$ and that they induce the discrete topology on E. This is well-known for the sharp topology for Colombeau type algebras. But our analysis implies that if one has a Colombeau type algebra containing the Dirac delta distribution δ as an embedded Colombeau generalized function, then the topology induced on the basic space must be discrete. This is an analogous result to the Schwartz's "impossibility result" concerning the product of distributions.construction of Colombeau type algebras.

In Part II we present our method in solving various classes of equations with strong singularities in the framework of generalized function algebras.

Constructions of algebras given in Part I are used

First, we present a quasilinear elliptic equation with Dirichlet's boundary conditions, ([34]) second, wave semilinear ([22]) and, third, heat semilinear ([23]) equations with Cauchy data.

We consider a quasilinear Dirichlet problem for uniformly elliptic equations whose coefficients have lack regularity assumptions and with singular boundary conditions. In our setting of a problem we replace an equation $divA(Du) = 0$ with a net of equations with regular coefficients and a singular boundary condition with an appropriate regularized net of boundary conditions.

As a second illustration we consider a semilinear wave equations in space dimension $n \leq 9$ with singular data and various types of nonlinearities. In general, a nonlinear term is regularized with respect to a small parameter ε such that it becomes globally Lipschitz for each ε. A net of solutions to a net of Cauchy problems obtained in this way determines an element in $\mathcal{G}_{L^2}$, the generalized solution. For certain growth conditions on a nonlinear term the equation is uniquely solved in $\mathcal{G}_{L^2}$ without regularization. Note, in certain cases, a solution to the regularized equation is also a solution to the non-regularized one.

We have studied also the heat equation with singularities, extending the use of semigroups to some classes of PDE's with singular coefficients. The general idea is simple: it lies in the core of a construction of generalized functions. Regularized PDE, in fact a net of equations, is solved with an appropriate net of semigroups. The solution obtained in this way represents a generalized function. The concrete results for the heat equation will not be presented in this paper (cf [23]).

Part I

1 Colombeau type algebras

In this section, we will present traditional approach to Colombeau type algebras and then in the next sections a more abstract and general realization of such algebras.

First, we recall the usual Colombeau type extension of $\mathcal{G}(E)$ ([35]), where E is a vector space on $\mathbb{C}$ with an increasing sequence of seminorms μ_n, $n \in \mathbb{N}$. The space of moderate nets of $\mathcal{E}_M(E)$, respectively, of null nets of $\mathcal{N}(E)$, is constituted by nets $(r_\varepsilon)_{\varepsilon\in(0,1]} \in E^{(0,1]}$ with the properties

$$(\forall n \in \mathbb{N})\ (\exists a \in \mathbb{R})(\mu_n(r_\varepsilon) = \mathcal{O}(\varepsilon^a)), \tag{1.1}$$

$$\text{respectively,} \quad (\forall n \in \mathbb{N})\ (\forall b \in \mathbb{R})(\mu_n(r_\varepsilon) = \mathcal{O}(\varepsilon^b)).$$

($\mathcal{O}$ is the Landau symbol.) The quotient space $\mathcal{G}(E) = \mathcal{E}_M(E)/\mathcal{N}(E)$ with elements $[(f_\varepsilon)_\varepsilon], [(g_\varepsilon)_\varepsilon]$, ..., (equivalence classes are denoted by $[\cdot]$) is called the Colombeau extension of E. Putting $v_n(r_\varepsilon) = \sup\{a;\ \mu_n(r_\varepsilon) = \mathcal{O}(\varepsilon^a)\}$ and $e_n((r_\varepsilon)\varepsilon, (s_\varepsilon)\varepsilon) = \exp(-v_n(r_\varepsilon - s_\varepsilon))$, $n \in \mathbb{N}$, we obtain that $(e_n)_n$ is a sequence of ultra-pseudometrics defining the ultra-metricmetric topology (sharp topology) on $\mathcal{G}(E)$.

If $E = \mathbb{C}$ (or $E = \mathbb{R}$) and the seminorms are equal to the absolute value, then the corresponding spaces are $\mathcal{E}_0$, $\mathcal{N}_0$; $\mathcal{E}_0$ is an algebra and $\mathcal{N}_0$ is an ideal and, as a quotient, one obtains Colombeau algebra of generalized complex numbers $\bar{\mathbb{C}} = \mathcal{E}_0/\mathcal{N}_0$ (or $\bar{\mathbb{R}}$). If a set Ω is open in $\mathbb{R}^n$ and $E = C^\infty(\Omega)$ is endowed with the usual sequence of seminorms (this is Schwartz space $\mathcal{E}(\Omega)$), then the above definition gives Colombeau simplified algebra $\mathcal{G}(\Omega) = \mathcal{E}_M(\Omega)/\mathcal{N}(\Omega)$ ([5], [24]). Its elements are called generalized functions and we keep this name for elements of any spaces or algebras constructed as extensions of some functional space E.

Then the embedding of compactly supported Schwartz distributions (elements of $\mathcal{E}'(\Omega)$) is made through the convolution with a net of mollifiers $h_\varepsilon = \varepsilon^{-n}h(\cdot/\varepsilon)$ constructed by a rapidly decreasing function $h \in \mathcal{S}(\mathbb{R}^n)$ with the properties $\int h(t)dt = 1$, $\int t^m h(t)dt = 0, m \in \mathbb{N}^n$. The embedding is given by

$$f \mapsto [(f * h_\varepsilon|_\Omega)_\varepsilon].$$

By the sheaf properties of $\mathcal{D}'(\Omega)$ and $\mathcal{G}(\Omega)$, this embedding is extended to $\mathcal{D}'(\Omega)$.

Construction needed in applications

Let Ω be a bounded open set in $\mathbb{R}^n$ and $\alpha \in (0,1)$. Recall ([15], p. 94), a domain Ω and its boundary are of $C^{k,\alpha}-$ class $0 \le \alpha \le 1$, if at each point $x_0 \in \partial\Omega$ there is a ball B_{x_0} and a bijection $\psi : B \to D$ such that $\psi(B \cap \Omega) \subset \mathbb{R}^n_+$, $\psi(B \cap \partial\Omega) \subset \partial\mathbb{R}^n_+$, and $\psi \in C^{k,\alpha}(B)$, $\psi^{-1} \in C^{k,\alpha}(D)$. A domain Ω has a boundary portion $T \in \partial\Omega$ of $C^{k,\alpha}-$ class if at each point $x_0 \in T$ there is a ball B_{x_0} in which the above conditions are satisfied and $B \cap \partial\Omega \subset T$.

We will consider the Colombeau extensions in cases $E = C^{k,\alpha}(\bar{\Omega})$, $k \in \mathbb{N}$ and $E = C^\infty(\bar{\Omega})$. We will use the norms

$$|\phi|_{k,\Omega} = \sup\{|f^{(p)}(x)|;\ |p| \le k, x \in \Omega\},$$

$$|f|_{k,\alpha,\Omega} = |f|_{k,\Omega} + [f]_{k,\alpha,\Omega},\ k \in \mathbb{N}_0,$$

where, for $\phi \in C^\infty(\bar\Omega),\ k \in \mathbb{N}_0$,

$$[f]_{k,\alpha,\Omega} = \sup\Big\{\frac{|f^{(p)}(x) - f^{(p)}(y)|}{|x-y|^\alpha};\ (x,y) \in \Omega,\ x \neq y,\ |p| = k\Big\}.$$

The completion of $C^\infty(\bar\Omega)$ with respect to the norm $|\cdot|_{k,\alpha,\Omega}$ defines $E_k = C^{k,\alpha}(\bar\Omega), k \in \mathbb{N}$. Recall, if $k+\alpha < k'+\alpha'$, then the imbedding of $C^{k,\alpha}(\bar\Omega)$ into $C^{k',\alpha'}(\bar\Omega)$ is a compact linear operator.

Note that the sequences of norms $\|\cdot\|_{k,\alpha}, k \in \mathbb{N}$ and $\|\cdot\|_k, k \in \mathbb{N}$ define the same uniform structure on $C^\infty(\bar\Omega)$ as the usual one.

In case $E = C^\infty(\bar\Omega)$, we need one more construction. Let $(g_\varepsilon)_\varepsilon$ be a net in $C^{0,\alpha}(\bar\Omega)$ such that

$$g_\varepsilon \in C^{k,\alpha}(\bar\Omega), \varepsilon < \varepsilon_k, k \in \mathbb{N},$$

where $(\varepsilon_k)_k \in (0,1)^{\mathbb{N}}$ strictly decreases to zero ($(\varepsilon_k)_k \downarrow 0$).

Two such nets are in relation, $(g_\varepsilon)_\varepsilon \sim (r_\varepsilon)_\varepsilon$, if

$$g_\varepsilon = r_\varepsilon, \varepsilon < \varepsilon_0,\ \text{for some } \varepsilon_0 \in (0,1).$$

This is an equivalence relation and with the corresponding classes, elements in $C^{0,\alpha}(\bar\Omega)/\sim$, we define spaces $\mathcal{E}_M[E]$, $\mathcal{N}[E]$ as in (1.1). Then we define the corresponding Colombeau type space $\mathcal{G}[E] = \mathcal{E}_M[E]/\mathcal{N}[E]$.

2 Algebras of weighted sequence spaces

Now we will give another approach to generalized function algebras which is actually the topological description of such algebras.

Consider a semi-normed algebra (E,p) such that $p(a\,b) \le p(a)\,p(b),\ a,b \in E$ and a sequence $r \in \mathbb{R}_+^{\mathbb{N}}$ decreasing to zero.

Define for $f \in E^{\mathbb{N}}$

$$\| f \|_{p,r} := \limsup_{n\to\infty} p(f_n)^{r_n}\ .$$

This is well defined for any $f \in E^{\mathbb{N}}$, with values in $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. With this definition, let

$$\mathcal{F}_{p,r} = \{f \in E^{\mathbb{N}} : \| f \|_{p,r} < \infty\}$$
$$\mathcal{K}_{p,r} = \{f \in E^{\mathbb{N}} : \| f \|_{p,r} = 0\}\ .$$

Then the following holds:

Proposition 1

1. The function

$$d_{p,r}:\ \mathcal{F}_{p,r} \times \mathcal{F}_{p,r} \to \mathbb{R}_+\ ,$$
$$(f,g) \mapsto \| f - g \|_{p,r}$$

is an ultrapseudometric on $\mathcal{F}_{p,r}$.

2. *$\mathcal{F}_{p,r}$ is a subalgebra of $E^{\mathbb{N}}$, and $\mathcal{K}_{p,r}$ is an ideal of $\mathcal{F}_{p,r}$; thus*

$$\mathcal{G}_{p,r} := \mathcal{F}_{p,r}/\mathcal{K}_{p,r}$$

is an algebra.

3.

$$\begin{aligned}\tilde{d}_{p,r} : \mathcal{G}_{p,r} \times \mathcal{G}_{p,r} &\to \mathbb{R}_+ \ , \\ (F,G) &\mapsto d_{p,r}(f,g)\end{aligned}$$

is an ultrametric on $\mathcal{G}_{p,r}$, where $f \in F$, $g \in G$ are any representatives of the classes $F = f + \mathcal{K}_{p,r}$ resp. $G = g + \mathcal{K}_{p,r}$.

4. *$\mathcal{G}_{p,r} = \mathcal{F}_{p,r}/\mathcal{K}_{p,r}$ is a topological algebra, the quotient topology being the same than the topology induced by the ultrametric $\tilde{d}_{p,r}$.*

We give the construction of generalized constants. For this, E will be the underlying field $\mathbb{R}$ or $\mathbb{C}$, and $p = |\cdot|$ the absolute value. As already explained in the introduction, for $r = 1/\log$, we get the ring of Colombeau's numbers $\overline{\mathbb{C}}$. Let $r_n = \frac{1}{\log n}$, $n \geq 2$. Colombeau's algebras of generalized constants represented by sequences with polynomial growth modulo sequences of more than polynomial decrease, because

$$\begin{aligned}\limsup |x_n|^{1/\log n} < \infty &\iff \exists C : \limsup |x_n|^{1/\log n} = C \\ &\iff \exists B, \exists n_0, \forall n > n_0 : |x_n| \leq B^{\log n} = n^{\log B} \\ &\iff \exists \gamma : |x_n| = o(n^\gamma) \ .\end{aligned}$$

If we put, $\limsup = 0$ (for the ideal) then the corresponding C above equals zero and thus $\forall B > 0$ resp. $\forall \gamma$ we have $|x_n| = o(n^\gamma)$.

Consider now Hölder type spaces $E = C^{k,\alpha}(\bar{\Omega})$ (cf. [15]) , $\alpha \in (0,1]$ and $k \in \mathbb{N}_0$. With $|\cdot|_{k,\alpha}$–norm It is a Banach space and we can apply the same construction with $p = \|\cdot\|_{k,\alpha}$.

The corresponding Colombeau type algebra is defined by $\mathcal{G}_{C^{k,\alpha}} :\equiv \mathcal{F}/\mathcal{K}$ where

$$\begin{aligned}\mathcal{F} &:= \left\{u \in (C^{k,\alpha}(\bar{\Omega}))^{\mathbb{N}} \mid \limsup \|u_n\|_{s,\infty}^{\frac{1}{\log n}} < \infty\right\} \ , \\ \mathcal{K} &:= \left\{u \in (C^{k,\alpha}(\bar{\Omega}))^{\mathbb{N}} \mid \limsup \|u_n\|_{s,\infty}^{\frac{1}{\log n}} = 0\right\} \ .\end{aligned}$$

This algebra is already described in Section 1; it will be used for the analysis of elliptic equation in Part II.

Constructions with locally convex vector spaces

Consider now an algebra E which is a locally convex vector space on $\mathbb{C}$, equipped with an arbitrary set of seminorms $p \in \mathcal{P}$ determining its locally convex structure. Assume that

$$\forall p \in \mathcal{P}, \ \exists \bar{p} \in \mathcal{P}, \ C \in \mathbb{R}_+ : \forall x, y \in E : p(x\,y) \leq C\,\bar{p}(x)\,\bar{p}(y) \ .$$

Let
$$\mathcal{F}_{\mathcal{P},r} = \{\, f \in E^{\mathbb{N}} \mid \forall p \in \mathcal{P} : \| f \|_{p,r} < \infty \,\}$$
and
$$\mathcal{K}_{\mathcal{P},r} = \{\, f \in E^{\mathbb{N}} \mid \forall p \in \mathcal{P} : \| f \|_{p,r} = 0 \,\} \ .$$
Then the following holds:

Proposition 2 *1. For every $p \in \mathcal{P}$,*
$$d_{p,r} : \ E^{\mathbb{N}} \times E^{\mathbb{N}} \to \overline{\mathbb{R}}_+ \ , \quad (f,g) \mapsto \| f - g \|_{p,r}$$
is an ultrapseudometric on $\mathcal{F}_{\mathcal{P},r}$.

2. *$\mathcal{F}_{\mathcal{P},r}$ is a (sub-)algebra of $E^{\mathbb{N}}$, and $\mathcal{K}_{\mathcal{P},r}$ is an ideal of $\mathcal{F}_{\mathcal{P},r}$.*

3. *$\mathcal{G}_{\mathcal{P},r} := \mathcal{F}_{\mathcal{P},r}/\mathcal{K}_{\mathcal{P},r}$ is an algebra.*

4. *For every $p \in \mathcal{P}$,*
$$\widetilde{d}_{p,r} : \ \mathcal{G}_{\mathcal{P},r} \times \mathcal{G}_{\mathcal{P},r} \to \overline{\mathbb{R}}_+ \ , \quad (F,G) \mapsto d_{p,r}(f,g)$$
is an ultrametric on $\mathcal{G}_{\mathcal{P},r}$, where f,g are any representatives of the classes $F = f + \mathcal{K}_{\mathcal{P},r}$ resp. $G = g + \mathcal{K}_{\mathcal{P},r}$.

5. *$\mathcal{G}_{\mathcal{P},r} := \mathcal{F}_{\mathcal{P},r}/\mathcal{K}_{\mathcal{P},r}$ is a topological algebra, the quotient topology being the same than the topology induced by the family of ultrametrics $\left\{\widetilde{d}_{p,r}\right\}_{p\in\mathcal{P}}$.*

Let $E = \mathcal{C}^\infty(\Omega)$, $\mathcal{P} = \{p_\nu\}_{\nu\in\mathbb{N}}$ with
$$p_\nu(f) := \sup_{|\alpha|\le\nu,\, |x|\le\nu} |f^{(\alpha)}(x)| \ ,$$
and $r = \frac{1}{\log}$. Then, $\mathcal{G}_{\mathcal{P},r} = \mathcal{F}_{\mathcal{P},r}/\mathcal{K}_{\mathcal{P},r}$ with
$$\mathcal{F}_{\mathcal{P},r} = \left\{ (f_n)_n \in \mathcal{C}^\infty(\Omega)^{\mathbb{N}} \;\middle|\; \forall \nu \in \mathbb{N} : \limsup_{n\to\infty} p_\nu(f_n)^{1/\log n} < \infty \right\} \ ,$$
$$\mathcal{K}_{\mathcal{P},r} = \left\{ (f_n)_n \in \mathcal{C}^\infty(\Omega)^{\mathbb{N}} \;\middle|\; \forall \nu \in \mathbb{N} : \limsup_{n\to\infty} p_\nu(f_n)^{1/\log n} = 0 \right\} \ .$$
we obtain the simplified Colombeau algebra $\mathcal{G}_s$.

So called full Colombeau algebra $\mathcal{G}$ is related to a more delicate procedure and it is omitted. We only note that the embedding of Schwartz distributions and of smooth functions into $\mathcal{G}$ is well-known. Also it is well-known that the multiplication of smooth function embedded into $\mathcal{G}$ is the usual multiplication.

The following example is also of interest. Take $E = \mathcal{D}_{L^p}(\Omega)$, $p > 1$, $\mathcal{P} = \{p_\nu\}_{\nu\in\mathbb{N}}$ with
$$p_\nu(f) := \sup_{|\alpha|\le\nu} ||f^{(\alpha)}||_{L^p} \ ,$$

and $r = \frac{1}{\log}$. Then, $\mathcal{G}_{L^p} = \mathcal{F}_{\mathcal{P},r}/\mathcal{K}_{\mathcal{P},r}$ with

$$\mathcal{F}_{\mathcal{P},r} = \left\{ (f_n)_n \in \mathcal{D}_{L^p}(\Omega)^{\mathbb{N}} \;\middle|\; \forall \nu \in \mathbb{N} : \limsup_{n\to\infty} p_\nu(f_n)^{1/\log n} < \infty \right\} ,$$

$$\mathcal{K}_{\mathcal{P},r} = \left\{ (f_n)_n \in \mathcal{D}_{L^p}(\Omega)^{\mathbb{N}} \;\middle|\; \forall \nu \in \mathbb{N} : \limsup_{n\to\infty} p_\nu(f_n)^{1/\log n} = 0 \right\} .$$

is Colombeau type algebra used for the investigations of wave and heat equation.

We will consider later $\mathcal{G}_{L^2}(\mathbf{R}^n)$.

Projective and inductive limits

Projective limit Let $(E^\mu_\nu, p^\mu_\nu)_{\mu,\nu\in\mathbb{N}}$ be a family of semi-normed algebras over $\mathbb{C}$, such that

$$\forall \mu, \nu \in \mathbb{N} : \; E^\mu_{\nu+1} \hookrightarrow E^\mu_\nu \;, \quad E^{\mu+1}_\nu \hookrightarrow E^\mu_\nu \;,$$

where $\hookrightarrow$ means continuously embedded. This implies that there exist constants C^μ_ν, $\tilde{C}^\mu_\nu \in \mathbb{R}_+$ such that

$$\forall \mu, \nu \in \mathbb{N} : p^\mu_\nu \le C^\mu_\nu \, p^\mu_{\nu+1} \;, \quad p^\mu_\nu \le \tilde{C}^\mu_\nu \, p^{\mu+1}_\nu \;,$$

but without loss of generality one can take C^μ_ν, $\tilde{C}^\mu_\nu = 1$, $\forall \mu, \nu \in \mathbb{N}$.
Then let

$$\overleftarrow{E} := \operatorname*{proj\,lim}_{\mu\to\infty} \overleftarrow{E}^\mu = \operatorname*{proj\,lim}_{\mu\to\infty} \operatorname*{proj\,lim}_{\nu\to\infty} E^\mu_\nu = \operatorname*{proj\,lim}_{\nu\to\infty} E^\nu_\nu \;.$$

Define

$$\overleftarrow{\mathcal{F}}_{p,r} = \left\{ f \in \overleftarrow{E}^{\mathbb{N}} \mid \forall \mu, \nu \in \mathbb{N} : \| f \|_{p^\mu_\nu, r} < \infty \right\} ,$$

$$\overleftarrow{\mathcal{K}}_{p,r} = \left\{ f \in \overleftarrow{E}^{\mathbb{N}} \mid \forall \mu, \nu \in \mathbb{N} : \| f \|_{p^\mu_\nu, r} = 0 \right\} .$$

(Here $p \equiv ((p^\mu_\nu)_\nu)^\mu$ stands (on the l.h.s.) for the whole family of seminorms.)
Then Proposition 2 holds, with the slight changes of notations introduced above.

Inductive limit Consider now a family $(E^\mu_\nu, p^\mu_\nu)_{\mu,\nu\in\mathbb{N}}$ of semi-normed spaces over $\mathbb{C}$, such that

$$\forall \mu, \nu \in \mathbb{N} : \; E^\mu_\nu \hookrightarrow E^\mu_{\nu+1} \;, \quad E^{\mu+1}_\nu \hookrightarrow E^\mu_\nu \;.$$

This implies that there exist constants C^μ_ν, $\tilde{C}^\mu_\nu \in \mathbb{R}_+$ such that

$$\forall \mu, \nu \in \mathbb{N} : p^\mu_{\nu+1} \le C^\mu_\nu \, p^\mu_\nu \;, \quad p^\mu_\nu \le \tilde{C}^\mu_\nu \, p^{\mu+1}_\nu \;,$$

but again one can assume C^μ_ν, $\tilde{C}^\mu_\nu = 1$, $\forall \mu, \nu \in \mathbb{N}$. Now let

$$\forall \mu \in \mathbb{N} : \; \overrightarrow{E}^\mu = \operatorname*{ind\,lim}_{\nu\to\infty} E^\mu_\nu \;.$$

Assume that for every $\mu, \nu', \nu'' \in \mathbb{N}$ there exist $\nu \in \mathbb{N}$ and $C > 0$ such that

$$p^\mu_\nu(fg) \le C p^\mu_{\nu'}(f) p^\mu_{\nu''}(g), \; f \in E^\mu_{\nu'}, \; g \in E^\mu_{\nu''} \;.$$

We have

$$\forall \mu \in \mathbb{N} : \overrightarrow{E}^{\mu+1} \hookrightarrow \overrightarrow{E}^\mu \;.$$

Now let

$$\overrightarrow{E} := \operatorname{proj}\lim_{\mu\to\infty} \overrightarrow{E}^{\mu} = \operatorname{proj}\lim_{\mu\to\infty} \operatorname{ind}\lim_{\nu\to\infty} E^{\mu}_{\nu} ,$$

and define

$$\overrightarrow{\mathcal{F}}_{p,r} := \left\{ f \in \overrightarrow{E}^{\mathbb{N}} \mid \forall \mu \in \mathbb{N}, \exists \nu \in \mathbb{N} : f \in (E^{\mu}_{\nu})^{\mathbb{N}} \wedge \| f \|_{p^{\mu}_{\nu},r} < \infty \right\} ,$$

$$\overrightarrow{\mathcal{K}}_{p,r} := \left\{ f \in \overrightarrow{E}^{\mathbb{N}} \mid \forall \mu \in \mathbb{N}, \exists \nu \in \mathbb{N} : f \in (E^{\mu}_{\nu})^{\mathbb{N}} \wedge \| f \|_{p^{\mu}_{\nu},r} = 0 \right\} .$$

Proposition 3 :

(i) Writing $\overleftrightarrow{\cdot}$ *for both,* $\overleftarrow{\cdot}$ *or* $\overrightarrow{\cdot}$*, we have that* $\overleftrightarrow{\mathcal{F}}_{p,r}$ *is an algebra and* $\overleftrightarrow{\mathcal{K}}_{p,r}$ *is an ideal of* $\overleftrightarrow{\mathcal{F}}_{p,r}$*; thus,* $\overleftrightarrow{\mathcal{G}}_{p,r} := \overleftrightarrow{\mathcal{F}}_{p,r} / \overleftrightarrow{\mathcal{K}}_{p,r}$ *is an algebra.*

(ii) For every $\mu, \nu \in \mathbb{N}$, $d_{p^{\mu}_{\nu}} : (E^{\mu}_{\nu})^{\mathbb{N}} \times (E^{\mu}_{\nu})^{\mathbb{N}} \to \overline{\mathbb{R}}_{+}$ *defined by* $d_{p^{\mu}_{\nu}}(f,g) = \| f - g \|_{p^{\mu}_{\nu},r}$ *is an ultrapseudometric on* $(E^{\mu}_{\nu})^{\mathbb{N}}$*. Moreover* $(d_{p^{\mu}_{\nu}})_{\mu,\nu}$ *induces a topological algebra structure on* $\overleftarrow{\mathcal{F}}_{p,r}$ *(since* $d_{p^{\mu}_{\nu}}(0, f \cdot g) \leq d_{p^{\mu}_{\nu}}(0,f) d_{p^{\mu}_{\nu}}(0,g)$*) such that the intersection of neighborhoods of zero equals* $\overleftarrow{\mathcal{K}}_{p,r}$*.*

(iii) From (ii), $\overleftarrow{\mathcal{G}}_{p,r} = \overleftarrow{\mathcal{F}}_{p,r} / \overleftarrow{\mathcal{K}}_{p,r}$ *becomes a topological algebra which topology can be defined by the family of ultrametrics* $(\tilde{d}_{p^{\mu}_{\nu}})_{\mu,\nu}$ *where* $\tilde{d}_{p^{\mu}_{\nu}}([f],[g]) = d_{p^{\mu}_{\nu}}(f,g)$*,* $[h]$ *standing for the class of* h*.*

(iv) If τ_{μ} *denotes the inductive limit topology on* $\mathcal{F}^{\mu}_{p,r} = \bigcup_{\nu \in \mathbb{N}} ((\tilde{E^{\mu}_{\nu}})^{\mathbb{N}}, d_{\mu,\nu})$*,* $\mu \in \mathbb{N}$*, then* $\overrightarrow{\mathcal{F}}_{p,r}$ *is a topological algebra for the projective limit topology of the family* $(\mathcal{F}^{\mu}_{p,r}, \tau_{\mu})_{\mu}$*.*

($(\tilde{E^{\mu}_{\nu}})^{\mathbb{N}}$, consists of elements $f \in (E^{\mu}_{\nu})^{\mathbb{N}}$ with finite $d_{\mu,\nu}(f)$.)

Without assuming completeness of E, it holds:

Proposition 4

(i) $\overleftarrow{\mathcal{F}}_{p,r}$ *is complete.*

(ii) If for all $\mu \in \mathbb{N}$*, a subset of* $\overrightarrow{\mathcal{F}}^{\mu}_{p,r}$ *is bounded iff it is a bounded subset of* $(E^{\mu}_{\nu})^{\mathbb{N}}$ *for some* $\nu \in \mathbb{N}$*, then* $\overrightarrow{\mathcal{F}}_{p,r}$ *is sequentially complete.*

Comments on the Schwartz' impossibility result

In the definition of sequence spaces $\overrightarrow{\mathcal{F}}_{p,r}$ resp. $\overleftarrow{\mathcal{F}}_{p,r}$, we assumed $r_n \searrow 0$ as $n \to \infty$. Clearly, one could consider sequence spaces of the same type with r_n only bounded, or even $r_n \to \infty$. In the former case (r_n bounded), the space $\overleftrightarrow{\mathcal{F}}_{p,r}$ (where $\overleftrightarrow{\cdot}$ stands for $\overleftarrow{\cdot}$ or $\overrightarrow{\cdot}$) contains $\operatorname{diag} E^{\mathbb{N}}$ topology, via the embedding $\overleftrightarrow{E} \ni f \mapsto (f)_n \in \overleftrightarrow{E}^{\mathbb{N}}$. In the second case (when $r_n \to \infty$), this embedding is not possible.

In the case we consider ($r_n \to 0$), the induced topology on $\overleftrightarrow{E}$ is a discrete topology. But this is necessarily so, since we want to include "divergent" sequences in $\overleftrightarrow{\mathcal{F}}_{p,r}$.

In order to have an appropriate topological algebra containing "δ", we must have that our generalized topological algebra induces a discrete topology on the original algebra $\overleftrightarrow{E}$.
This conclusion is in analogy to Schwartz' impossibility statement for multiplication of distributions.

General remarks on embedding of duals

Under mild assumptions on $\overleftrightarrow{E}$, we can show that our algebras of (classes of) sequences contains elements of the strong dual space $\overleftrightarrow{E}'$.

Let $\mathcal{C}^0(\mathbb{R}^s)$ be the space of continuous functions with projective topology given by sup norms on the balls of radius $\nu \in \mathbb{N}^*$, $p_\nu(f) = \sup\{|f(x)|; |x| \le \nu\}$.

We shall assume in the sequel that $\overleftrightarrow{E}$ is a dense subspace of $\mathcal{C}^0(\mathbb{R}^s)$ and the inclusion mapping $\overleftrightarrow{E} \to \mathcal{C}^0(\mathbb{R}^s)$ is continuous.

Then, we have the following

Proposition 5 *(i)*

$$\delta : \overleftrightarrow{E} \to \mathbb{C}\ , \quad \delta(\phi) := \phi(0)$$

is an element of $\overleftrightarrow{E}'$.

ii) *Let* $\overleftrightarrow{E}$ *be sequentially weakly dense in* $\overleftrightarrow{E}'$. *Then, a sequence* $(\delta_n)_n \in E \cap (\mathcal{C}^0)'$ *with the property* $\exists \eta, \theta > 0 : \forall n \in \mathbb{N} : \sup\limits_{|x|>\theta} |\delta_n(x)| < \eta$, *converging weakly to* δ, *cannot be bounded in* $\overleftrightarrow{E}$.

Thus, the appropriate choice of the sequence r appeared to be important to have at least δ embedded into the corresponding algebra. It can be chosen such that:

In $\overleftarrow{E}$ case, for every $\mu, \nu \in \mathbb{N}$

$$\limsup_{n\to\infty} p_\nu^\mu (\delta_n)^{r_n} = A_\nu^\mu \ \text{ and } \ \exists \mu_0, \nu_0 : A_{\nu_0}^{\mu_0} \neq 0\ .$$

In $\overrightarrow{E}$ case, for every $\mu \in \mathbb{N}$ exists $\nu \in \mathbb{N}$ such that the above limit holds.

3 Association

The notion of a weak limit or of a weak solutions is transferred to generalized function algebras to various notions of associations. Thus their importance is underlined through the applications to nonlinear equations or linear one with singularities.

General concept: $\mathcal{J} - X$–association The $\mathcal{J} - X$-association of elements $F, G \in \mathcal{G} = \mathcal{F}/\mathcal{K}$ is defined in terms of an additive subgroup $\mathcal{J}$ of $\mathcal{F}$ containing the ideal $\mathcal{K}$, and a set X of generalized numbers, by

$$F \underset{\mathcal{J},X}{\approx} G \iff \forall x \in X : x \cdot (F - G) \in \mathcal{J}/\mathcal{K}\ .$$

As $\mathcal{J}$ is not an ideal, the association is not compatible with the multiplication in $\mathcal{F}$ (not even by generalized numbers, only by elements of E). However, in the case of differential algebras, $\mathcal{J}$ is usually chosen such that $\underset{\mathcal{J}}{\approx}$ is stable under differentiation.

If the set X contains only number 1, then we simply write $F \underset{\mathcal{J}}{\approx} G \iff F - G \in \mathcal{J}/\mathcal{K}$.

For example, consider $N = \{x \in \mathbb{C}^{\mathbb{N}} \mid \lim x_n = 0\}$, the set of null sequences. This gives usual association of generalized numbers,

$$[x] \sim [y] \iff [x] \underset{N}{\approx} [y] \iff x_n - y_n \to 0$$

which is well defined because all elements of the ideal tend to zero.

Strong s–association is defined for $s \in \mathbb{R}_+$ by $F \overset{s}{\simeq} G \iff F \underset{\mathcal{J}^{(s)}_{\mathcal{P},r}}{\approx} G$ with $\mathcal{J}^{(s)}_{\mathcal{P},r} = \{f \in \mathcal{F} \mid \forall p \in \mathcal{P} : \| f \|_{p,r} < e^{-s}\}$.
For $s = 0$, we write $F \simeq G$ and simply call them strongly associated.
On the other hand, $F \overset{s}{\simeq} G$ for all $s \geq 0$ implies $F = G$.

Weak associations. The following types of associations are defined in terms of a duality product[1] $\langle\cdot,\cdot\rangle : \overleftrightarrow{E} \times D \to \mathbb{C}$, and

$$\mathcal{J} = \mathcal{J}_M = \left\{ f \in \overleftrightarrow{E}^{\mathbb{N}} \mid \forall\psi \in \boldsymbol{D} : (\langle f_n, \psi\rangle)_n \in M \right\} .$$

where M is some additive subgroup of $\mathbb{C}^{\mathbb{N}}$.

$s - D'$**–association** is defined by $F \overset{s}{\approx} G \iff F \underset{\mathcal{J}_N, X_s}{\approx} G$ with $X_s = \left\{ \left[\left(e^{s/r_n}\right)_n \right] \right\}$ for $s \in \mathbb{R}$.

Example 6 *In the case of Colombeau's algebra this has already been considered (with $\boldsymbol{D} = \mathcal{D}$): For $s = 0$ we get the so-called weak association $[f] \approx [g] \iff f_n - g_n \to 0$ in $\mathcal{D}'$. For $s \neq 0$, $[f] \overset{s}{\approx} [g] \iff n^s(f_n - g_n) \to 0$ in $\mathcal{D}'$. In the case of ultradistributions, we take $\boldsymbol{D} = \mathcal{D}^{(m)}$ and $e^{s/r_n} = \exp[s\, n^{\frac{1}{m'-1}}]$ for Beurling case, and analogous definitions in the Roumieu case.* **Weak s–association** *is defined by $F \overset{s}{\approx} G \iff F \underset{\mathcal{J}_I}{\approx} G$ where $I = \mathcal{J}_{|\cdot|,r,s}$ for any $s \in \mathbb{R}$.*
For $s = 0$, we write $F \overset{\mathrm{sw}}{\approx} G$ and call F and G **strong–weak associated**.

[1] D stands for a test function space such that $E \hookrightarrow D'$.

Part II

4 Quasilinear elliptic equation

Let $(\mathcal{Q}_\varepsilon)_\varepsilon$ be a net of elliptic nonlinear operators of divergent type of the form

$$\mathcal{Q}_\varepsilon(u) = div\mathbf{A}_\varepsilon(Du) = a_\varepsilon^{i,j}(Du)D_{i,j}u, \varepsilon < 1, \tag{4.2}$$

where $a_\varepsilon^{i,j}(p) = D_{p_i}A_\varepsilon^j(p)$, or, in case $n = 2$, let $(\mathcal{Q}_\varepsilon)_\varepsilon$ be a net of elliptic nonlinear operators of the form

$$\mathcal{Q}_\varepsilon(u) = a_\varepsilon^{i,j}(x, u, Du)D_{i,j}u, \ u \in C^\infty(\bar{\Omega}). \tag{4.3}$$

We assume that $a_\varepsilon^{i,j}, \varepsilon \in (0,1)$ are smooth functions on Ω. If λ_ε and Λ_ε denote respectively the minimum and maximum eigenvalues, then we have

$$0 < \lambda_\varepsilon(x,t,p)|\xi|^2 \leq a_\varepsilon^{i,j}(x,t,p)\xi_i\xi_j \leq \Lambda_\varepsilon(x,t,p)|\xi|^2, \tag{4.4}$$

$$p \in \mathbb{R}^n, \ \xi \in \mathbb{R}^n \setminus \{0\}, \ x \in \Omega, \ t \in \mathbb{R}, \ \varepsilon < \varepsilon_0.$$

Assume additionally:

$$(\forall d \in \mathbb{N}_0^n)(\exists l_d \in \mathbb{R})(\exists a_d \in \mathbb{R}) \tag{4.5}$$

$$\sup\left\{\frac{|\partial_x^d a_\varepsilon^{i,j}(x,t,p)|}{(1+|t|+|p|)^{a_d}}; x \in \bar{\Omega}, t \in \mathbb{R}, p \in \mathbb{R}^n\right\} = \mathcal{O}(\varepsilon^{l_d}).$$

$$(\exists C > 0)(\exists \mu > 0)(\exists b \in \mathbb{R}) \tag{4.6}$$

$$\frac{\varepsilon^\mu}{C}(1+|t|+|p|)^b \leq \lambda_\varepsilon(x,t,p)) \leq \Lambda_\varepsilon(x,t,p) \leq \frac{C}{\varepsilon^\mu}(1+|t|+|p|)^b,$$

$$p \in \mathbb{R}^n, x \in \bar{\Omega}, t \in \mathbb{R}, \varepsilon < \varepsilon_0.$$

In the case when the net $(\mathcal{Q}_\varepsilon)_\varepsilon$ is of the form (4.3) then $n = 2$, and if it is of the divergent form (4.2), then we exclude variables x and t in the conditions given above.

Note that condition (4.6) implies

$$\Lambda_\varepsilon/\lambda_\varepsilon \leq C^2/\varepsilon^{2\mu}, \varepsilon < \varepsilon_0. \tag{4.7}$$

With the given properties $(\mathcal{Q}_\varepsilon)_\varepsilon$ is called the net of uniformly elliptic moderate continuous operators.

Example

Consider in $\mathbb{R}^3$ the operator

$$Q(x,u,Du) = (1 + \sum_{i=1}^{3} \delta(D_i))\Delta u \ (\delta \text{ is the delta distribution}).$$

With the regularization of δ, we have

$$Q_\varepsilon(x,u,Du) = \Big(\frac{1}{\varepsilon}\psi\Big(\frac{D_1 u}{\varepsilon}\Big) + \frac{1}{\varepsilon}\psi\Big(\frac{D_2 u}{\varepsilon}\Big)\frac{1}{\varepsilon}\psi\Big(\frac{D_3 u}{\varepsilon}\Big) + 1\Big)\Delta u.$$

(ψ is a compactly supported smooth function with the integral equals 1.)

Then, $\lambda_\varepsilon = 1$ and $\Lambda_\varepsilon = (\frac{1}{\varepsilon}\psi(\frac{p_1}{\varepsilon}) + \frac{1}{\varepsilon}\psi(\frac{p_2}{\varepsilon})\frac{1}{\varepsilon}\psi(\frac{p_3}{\varepsilon}) + 1)$.

This operator is of the form (4.2) for which all the assumptions given above hold. We need a "slope condition" adapted to the setting of Colombeau theory.

Definition 7 *Let $E = C^{k,\alpha}(\bar{\Omega})$ for some $k \in \mathbb{N}$ (cf. 1.1 and $\mathcal{G}_{C^{k,\alpha}}$), $(\phi_\varepsilon)_\varepsilon \in \mathcal{F} = \mathcal{E}_{C^{k,\alpha}}$ and $\Gamma_\varepsilon = \{(x, z_\varepsilon), x \in \partial\Omega, z_\varepsilon = \phi_\varepsilon(x)\}$. Then $(\phi_\varepsilon)_\varepsilon$ and the boundary $\partial\Omega$ satisfies a moderate slope condition if for any $P_\varepsilon \in \Gamma_\varepsilon$ there exist hyperplanes $\pi^+_{\varepsilon,P_\varepsilon}$ and $\pi^-_{\varepsilon,P_\varepsilon}$ defined by $z_\varepsilon = \pi^+_{\varepsilon,P_\varepsilon}(x)$ and $z_\varepsilon = \pi^-_{\varepsilon,P_\varepsilon}(x)$ such that*

$$\pi^-_{\varepsilon,P_\varepsilon}(x) \leq \phi_\varepsilon(x) \leq \pi^+_{\varepsilon,P_\varepsilon}(x), x \in \partial\Omega, \varepsilon < \varepsilon_0$$

and such that for some $K > 0$ and some $m \in \mathbb{R}$,

$$\sup\{|D\pi^+_{\varepsilon,P_\varepsilon}(x)|, |D\pi^-_{\varepsilon,P_\varepsilon}(x)|; x \in \partial\Omega, P_\varepsilon \in \Gamma_\varepsilon\} \leq K\varepsilon^m, \varepsilon < \varepsilon_0.$$

With all the definition given above and by the use a generalized version of the Leray-Schauder fixed point theorem we are able to solve a quasilinear equation¿

Proposition 8 *Let $(\mathcal{Q}_\varepsilon)_\varepsilon$ be a net of uniformly elliptic operators of the form (4.2) or (4.3) with $a^{i,j}_\varepsilon \in C^{k+1}(\bar{\Omega})$ ($k \in \mathbb{N}$)satisfying (4.5) with $d \leq k+1$ and (4.6). Let $E = C^{k+2,\alpha}(\bar{\Omega})$ $(\phi_\varepsilon)_\varepsilon \in \mathcal{E}_{C^{k+2,\alpha}}$ where $\partial\Omega$ is of $C^{k+2,\alpha}$ class and it satisfies a moderate slope condition with $(\phi_\varepsilon)_\varepsilon$. Then, there exists $(u_\varepsilon)_\varepsilon \in \mathcal{E}_{C^{k+2,\alpha}}$ such that*

$$\mathcal{Q}_\varepsilon(u_\varepsilon) = 0, u_\varepsilon\Big|_{\partial\Omega} = \phi_\varepsilon, \quad in\ \varepsilon < 1. \tag{4.8}$$

This theorem implies the solvability in $\mathcal{G}_{C^{k,\alpha}}$.

Remark that the process of regularization of equation $divA(Du) = 0,\ u_{|\partial\Omega} = \phi$ with singular coefficients and singular data leads to the approximated net of solutions by the mean of previous theorem.

5 Semilinear wave equation

In this setting we connect two areas: the L^2-theory for the nonlinear wave equation

$$\partial_t^2 u - \triangle u + g(u) = 0,\ g(0) = 0,\ u = u(x,t),\ x \in \mathbf{R}^n,\ t \geq 0, \tag{5.9}$$
$$u(x,0) = a(x), ; u_t(x,) = b(x), ; x \in \mathbf{R}^n,$$

involving energy estimates and the theory of generalized functions where nonlinear operations makes sense for a large collection of singular objects.

Concerning g, if it is not globally Lipshitz, then it is substituted by a net of globally Lipschitz functions $g_\varepsilon(u)$. Then the obtained net of equations, called regularized equation, is solved for each fixed ε.

In some cases g is not regularized and the growth conditions on g are involved for the existence and unicity of a solution similarly as in the classical theory.

We use here the algebra $\mathcal{G}_{L^2}([0,T)\times\mathbf{R}^n)$ (cf. Example 5 with simple modifications). Also we use the notation $\mathcal{F}=\mathcal{E}_{L^2}([0,T)\times\mathbf{R}^n)$.

Consider a family of equations in $\mathcal{E}_{L^2}([0,T)\times\mathbf{R}^n)$

$$(\partial_t^2-\triangle)G_\varepsilon=-g(G_\varepsilon),\ G_\varepsilon|_{t=0}=A_\varepsilon, \partial_t G_\varepsilon|_{t=0}=B_\varepsilon,\ \varepsilon\in(0,1), \tag{5.10}$$

where $A_\varepsilon, B_\varepsilon\in\mathcal{E}_{L^2}(\mathbf{R}^n)$ and $g:\mathbf{R}^n\to\mathbf{R}$ is smooth, polynomially bounded together with all its derivatives and $g(0)=0$.

Equation (5.10), with the regularization g_ε instead of g is called the regularized equation for (5.9).

Proposition 9 *a) Let $n\leq 5$. Then there exists a regularized net g_ε such that for every $T>0$ there exists a unique solution to 5.10 in $\mathcal{G}_{L^2}([0,T)\times\mathbf{R}^n)$.*

b) Let $n=6$ and let $\|A_\varepsilon\|_{H^{3,2}}$ and $\|B_\varepsilon\|_{H^{2,2}}$ be bounded by $(\log(\log(\varepsilon^{-1})))^s$, as $\varepsilon\to 0$, where $s<1$. Then there exists a regularized net g_ε such that for every $T>0$ there exists a unique solution to 5.10 in $\mathcal{G}_{L^2}([0,T)\times\mathbf{R}^n)$.

Remark 10 *Let $n=7$. In order to obtain the existence of a unique solution with the moderate growth of all its derivatives, we need that $H^{3,2}$-norms of initial data are bounded by $\underbrace{\log(\log\ldots(\log}_{}q\varepsilon^{-1})\ldots)^s$ with respect to ε for some s and q. This follows from [27], Theorem 4.8. Cases $n=8,9$ can be handled out using the procedure and Lemmas 2.1-2.20 in the same paper as well as a composition of the logarithmic function sufficiently many times.*

The proof of quoted theorem for $n=3$ implies the next corollary.

Corollary 11 *Let $n=3$, $g(y)$ be globally Lipschitz and its first derivative be polynomially bounded. Then for every $T>0$ there exists a solution to (5.10) in $\mathcal{G}_{L^2}([0,T)\times\mathbf{R}^n)$.*

Remark 12 *If $g(y)$ is globally Lipschitz, for $n=4,5,6$, we need to assume appropriate conditions for the first and second derivatives of g. If $n=7,8,9$, then the assumptions of corollary are more complicated.*

Especially, we have

Proposition 13 *Equation*

$$(\partial_t^2-\Delta)G=-G^3,\ G|_{t=0}=A,\ \partial_t G|_{t=0}=B,$$

where $A,B\in\mathcal{G}_{L^2(\mathbf{R}^3)}$, has a unique solution in $\mathcal{G}_{L^2}([0,T)\times\mathbf{R}^3)$ for every $T>0$ if there exist representatives of initial data such that

$$\|(\nabla^2 A_\varepsilon,\nabla B_\varepsilon)\|_{L^2}=o((\log\varepsilon^{-1})^{1/2}).$$

References

[1] H A Biagioni, *A Nonlinear Theory of Generalized Functions*, Springer-Verlag, Berlin-Heidelberg-New York, 1990.

[2] H A BIAGIONI, J F COLOMBEAU, *New Generalized Functions and C^∞ Functions with Values in Generalized Complex Numbers*, J. London Math. Soc. (2) 33, 1.(1986), 169–179.

[3] H A BIAGIONI, M OBERGUGGENBERGER, *Generalized Solutions to Burgers's Equation*, J. Differential Equations, 97(1992), 263-287.

[4] H A BIAGIONI, M OBERGUGGENBERGER, *Generalized Solutions to the Korteweg-de Vries and the Regularized Long-wave Equations*, SIAM J. Math. Anal. 23(1992), 923-940.

[5] J-F COLOMBEAU, *New Generalized Functions and Multiplication of the Distributions*, North Holland, 1983.

[6] J-F COLOMBEAU, *Multiplication of Distributions*, Lect. Not. Math. 1532, Springer, Berlin, 1992.

[7] J F COLOMBEAU, M LANGLAIS, *Existence and Uniqueness of Solutions of Nonlinear Parabolic Equations with Cauchy Data Distributions*, J. Math. Anal. Appl. 145(1990), 186-196.

[8] J F COLOMBEAU, A Y LE ROUX, *Multiplications of Distributions in Elasticity and Hydrodynamics*, J. Math. Phys. 29(1988), 315-319.

[9] J F COLOMBEAU, M OBERGUGGENBERGER, *On a Hyperbolic System with a Compatible Quadratic Term: Generalized Solutions*, Delta Waves, and Multiplication of Distributions, Comm. Part. Diff. Eq. 15(1990), 905-938.

[10] J F COLOMBEAU, A HEIBIG, M OBERGUGGENBERGER, *Le Probleme de Cauchy dans un Espace de Fonctions Généralisées*, I. C. R. Acad. Sci. Paris 317(1993), 851-855.

[11] A DELCROIX, M HASLER, S PILIPOVIĆ, V VALMORIN, *Sequence space representation of Colombeau type algebras,*, In Proc. AMS, part I, in Math. Proc. Camb. Phil. Soc., part II, to appear.

[12] A DELCROIX, D SCARPALEZOS, *Sharp topologies on asymptotic algebras"*, Mh Math **129** (2000) 1–14.

[13] N DJAPIĆ, S PILIPOVIĆ, D SCARPALEZOS, *Intrinsic Microlocal Characterization of Colombeau's Generalized Functions*, J. Anal. Math. **75** (1998), 51–66.

[14] A EIDA, S PILIPOVIĆ, *On the Microlocal Decomposition of some Classes of Hyperfunctions*, Math. Proc. Camb. Phil. Soc., 125 (1999), 455–461.

[15] D GILBARG, N S TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag Berlin Heidelberg New York Tokyo, 1983.

[16] T GRAMCHEV, *Nonlinear Maps in Spaces of Distributions*, Math. Zeitschr. 209 (1992), 101–114.

[17] H KOMATSU, *Ultradistributions, I – III*, J. Fac. Sci. Univ. Tokyo, Sect. IA 20(1973), 25–105; 24(1977), 607–628; 29(1982), 653–717.

[18] H KOMATSU, *Microlocal Analysis in Gevrey Classes and in Convex Domains*, Springer, Lec. Not. Math. 1726(1989), 426–493.

[19] D KOVAČEVIĆ, S PILIPOVIĆ, *Structural Properties of the Space of Tempered Ultradistributions, Complex Analysis and Generalized Functions*, (1993), Proc. Conf. "Comp. Analysis and Applications '91 with Symposium on Generalized Functions", Varna 1991, 169–184.

[20] J-A MARTI, *Fundamental structures and asymptotic microlocalization in sheaves of generalized functions*, Integral Transforms and Special Functions, 6(1998), 223–228.

[21] J-A MARTI, *$(\mathcal{C}, \mathcal{E}, \mathcal{P})$-sheaf structures and applications*, in Michael Grosser *et al.*, Chapman & Hall/Crc Research Notes in Mathematics, Nonlinear Theory of Generalized Functions, p 175–186, 1999.

[22] M NEDELJKOV, M OBERGUGGENBERGER, S PILIPOVIĆ, *Generalized Solution to a Semilinear Wave Equation*, preprint.

[23] M NEDELJKOV, S PILIPOVIĆ, D RAJTER, *Semigroups in generalized functions algebras. Heat equation with singular potential and singular data,*, Preprint.

[24] M OBERGUGGENBERGER, *Multiplications of Distributions and Applications to Partial Differential Equations*, Longman, 1992.

[25] M OBERGUGGENBERGER, *Case Study of a Nonlinear, Nonconservative, Non-strictly Hyperbolic System*, Nonlinear Anal. Theory and Application, 19(1992), 53-79.

[26] M OBERGUGGENBERGER, Y G WANG, *Generalized Solutions to Conservation Laws*, Zietschr. Anal. Anw. 13(1994), 7-18.

[27] H PECHER, *Ein nichtlinearer Interpolationssatz und seine Anwendung auf nichtlineare Wellengleichungen*, Math. Z., 161(1978), 9-40.

[28] H J PETZSCHE, *Generalized Functions and the Boundary Values of Holomorphic Functions*, J. Fac. Sci. Univ. Tokyo, Sect. IA Math. 31 (1984), 391-431.

[29] S PILIPOVIĆ, *Microlocal Analysis of Ultradistributions*, Proc. AMS, 126(1998), 105-113.

[30] S PILIPOVIĆ, *Colombeau's Generalized Functions and Pseudodifferential Operators*, University of Tokio, Lecture Notes Series, 1994.

[31] S PILIPOVIĆ, M STOJANOVIĆ, *Generalized Solutions to Nonlinear Volterra Integral Equations with non-Lipschitz Nonlinearity*, Theory, Methods & Applications, 37(1999), 319-335.

[32] S PILIPOVIĆ, D SCARPALÉZOS, *Differential Operators with Generalized Constant Coefficients in Colombeau Algebra*, Portugal. Math. 53(1996), 305-324.

[33] S PILIPOVIC, D SCARPALEZOS, *Colombeau ultradistributions*, Math Proc. Camb. Phil. Soc., 2001.

[34] S PILIPOVIC, D SCARPALEZOS, *Nonlinear analysis in the Colombeau extension of a locally convex space*, Preprint, 2001.

[35] D SCARPALÉZOS, *Some Remarks on Functoriality of Colombeau's Construction: Topological and Microlocal Aspects and Applications*, Integral Transforms and Special Functions, 6 (1998), 295–307.

[36] V VALMORIN, *On the multiplication of periodic hyperfunctions of one variable*, in Michael Grosser *et al.*, Chapman & Hall/Crc Research Notes in Mathematics, Nonlinear Theory of Generalized Functions, p 219–228, 1999.

Some new results in theory of operators

Milutin Dostanić

Dedicated to Professor Veselin Perić on the occasion of his 70th birthday

University of Belgrade, Faculty of Mathematics,

Studentski Trg 16, 11000 Belgrade, Yugoslavia

In this short exposition will be stated some new results concerning spectral properties of certain important singular linear operators that are often met in Analysis.

Namely, we shall consider operators acting on the space $L^2(\Omega)$ ($\Omega \subset \mathbb{C}$ - simple connected domain in $\mathbb{C}$) that are defined in the following way:

$$\mathcal{C}f(z) = -\tfrac{1}{\pi}\int\limits_{\Omega} \frac{f(\xi)}{\xi - z} dA(\xi)$$

(Cauchy's operator)

$$\mathcal{L}f(z) = -\tfrac{1}{2\pi}\int\limits_{\Omega} \ln|z-\xi| dA(\xi)$$

(Operator of the logarithmic potential type)

$$\mathcal{R}f(z) = \pi^{\alpha-1}\frac{\Gamma(1-\frac{\alpha}{2})}{\Gamma(\frac{\alpha}{2})}\int\limits_{\Omega} \frac{f(\xi)}{|\xi-z|^{2-\alpha}} dA(\xi),\ (0 < \alpha < 2)$$

(Riesz' type operator).

For $z = x + iy$ we denote by $dA(z) = dxdy$ the Lebesque measure on Ω.

It is well-known fact that $\mathcal{C}, \mathcal{L}$ and $\mathcal{R}$ are compact operators on $L^2(\Omega)$. It is also known that $\mathcal{R}$, in the case $0 < \alpha < 2$, is positive. It follows immediately by applying the Fourier's transformation.

H.Widam [9] has shown that for the eigenvalues of $\mathcal{R}$ (denoted by $\lambda_n(\mathcal{R})$) holds

$$\lim_{n\to\infty} n^{\frac{\alpha}{2}}\lambda_n(\mathcal{R}) = \pi^{\frac{\alpha}{2}}|\Omega|^{\frac{\alpha}{2}}, \tag{1}$$

where $|\Omega|$ is the measure (area) of Ω. Therefore, the spectrum of the operator $\mathcal{R}$ determines geometrical property (area) of the domain Ω.
Some related questions concerning the operators $\mathcal{C}$ and $\mathcal{L}$ are also interesting.

The operator $\mathcal{L}$ is self-adjoint and it is proved in [7] that for its eigenvalues holds

$$\lim_{n\to\infty} n\lambda_n(\mathcal{L}) = \frac{|\Omega|}{4\pi}. \tag{2}$$

The operator $\mathcal{C}$ is not self-adjoint so we shall investigate asymptotic behavior of its singular values $s_n(\mathcal{C})$, that is, the eigenvalues of the positive operator $(\mathcal{C}^*\mathcal{C})^{\frac{1}{2}}$.

It is shown in [5] that holds

$$\lim_{n\to\infty} n^{\frac{1}{2}} s_n(\mathcal{C}) = \sqrt{\frac{|\Omega|}{\pi}}. \tag{3}$$

From (1), (2) and (3) we conclude that all the operators $\mathcal{C}$, $\mathcal{L}$ and $\mathcal{R}$ have the property that their spectral characteristics detect the area of the domain Ω on which these operators act.

Paper [2] is devoted to the investigation of the spectrum of the operator $\mathcal{C}^*\mathcal{C}$ in the case when $\Omega = D$ is the unit disc. In that case singular values are completely described as well as the vectors which are included in the singular expansion in terms of the Bessel's functions.
Specially, we have

$$||\mathcal{C}|| = \frac{2}{j_0},$$

where j_0 is the smallest positive root of the Bessel's function

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(k!)^2}\left(\frac{x}{2}\right)^{2k}.$$

It is conjectured that for an arbitrary domain $\Omega \subset \mathbb{C}$ holds

$$||\mathcal{C}|| = \frac{2}{\sqrt{\lambda_1}}, \tag{4}$$

where λ_1 is the smallest eigenvalue of the Dirichlet boundary-value problem

$$\begin{array}{l} -\Delta u = \lambda u, \\ u|_{\partial\Omega} = 0, \end{array} . \tag{5}$$

It is proved in [5] that this hypothesis is true. The proof is based on the following lemma which is also interesting itself.

Lemma 1 ([5]) *Given $f \in L^2(\Omega)$, Ω being a bounded domain in $\mathbb{C}$. Take $\hat{f}(x) = \frac{1}{2\pi}\int_\Omega e^{-iux_1 - ivx_2} f(u,v)dudv$. Then for all $0 < \alpha \le \frac{1}{2}$ holds*

$$\int_{R^2} \frac{|\hat{f}|^2}{|x|^{2\alpha}} dx \le \lambda_1{}^{-\alpha} \int_D |f(x)|^2 dx,$$

where λ_1 is the smallest eigenvalue of the boundary-value problem (5) *and* $x = (x_1, x_2)$, $|x| = \sqrt{x_1^2 + x_2^2}$, $dx = dx_1 dx_2$.

By applying lemma 1 and the Cauchy-Green's formula it is possible to obtain

Theorem 1 ([4]) *Let G and D be simple connected domains in $\mathbb{R}^2$ with piecewise smooth boundaries and $G \subset D$. If $u \in C^1(\bar{D})$ and $u|_{\partial D} = 0$, then the following inequality holds*

$$\int_G |u|^2 dx \leq \frac{1}{\sqrt{\lambda_1(G)\lambda_1(D)}} \int_D |\nabla u|^2 dx, \tag{6}$$

where $\lambda_1(D)$ and $\lambda_1(G)$ are the smallest eigenvalues of the boundary-value problems

$$\begin{matrix} -\Delta u = \lambda u & & -\Delta v = \lambda v \\ & \textit{and} & \\ u|_{\partial D} = 0 & & v|_{\partial G} = 0, \end{matrix},$$

respectively.

In the case $G = D$ the inequality (6) becomes the well-known Friedrich's inequality (or Poincaré's or Nirenberg's).

Using the following Faber-Krahn's inequality from [3]

$$\lambda_1(G) \geq \frac{\pi j_0^2}{|G|}, \qquad \lambda_1(D) \geq \frac{\pi j_0^2}{|D|},$$

we obtain in (6) a weaker but more useful inequality

$$\int_G |u|^2 dA \leq \frac{\sqrt{|G| \cdot |D|}}{\pi j_0^2} \int_D |\nabla u|^2 dA.$$

Denote by $L_a^2(\Omega)$ the space of analytic functions on Ω such that

$$\int_\Omega |f|^2 dA < \infty.$$

Then $L_a^2(\Omega)$ is a Hilbert's subspace of $L^2(\Omega)$ and it is called Bergmann's subspace. Next denote by P the orthogonal projector from $L^2(\Omega)$ onto $L_a^2(\Omega)$ which is called Bergmann's projection. It is estimated in [1] the order of the growth of the singular values for the operators $\mathcal{C}\mathcal{P}$ and $\mathcal{L}\mathcal{P}$. It is shown that

$$s_n(\mathcal{C}\mathcal{P}) \sim n^{-1}, \qquad s_n(\mathcal{L}\mathcal{P}) \sim n^{-2}, \tag{7}$$

hold.

The authors remarked that they have no explanation for a double acceleration of the decrease of the singular values for the operators $\mathcal{C}$ and $\mathcal{L}$ multiplied by P. It is also remained open question of the exact values for the constants in the asymptotic formulae (7).

In the papers [6] and [8] these problems are investigated. The following theorems are proved.

Theorem 2 *Let Ω be a bounded, simple connected domain in $\mathbb{C}$ with an analytical boundary. Then for the operators $\mathcal{C}$ and $\mathcal{L}$ holds*

$$\lim_{n\to\infty} n \cdot s_n(\mathcal{PC}) = \frac{|\partial\Omega|}{2\pi},$$

$$\lim_{n^2\to\infty} n \cdot s_n(\mathcal{PL}) = \frac{|\partial\Omega|^2}{16\pi^2}.$$

Theorem 3 *Let Ω be a bounded, simple connected domain in $\mathbb{C}$ with an analytical boundary. Then we have*

$$\lim_{n\to\infty} n^\alpha \cdot s_n(\mathcal{RP}) = \left(\frac{|\partial\Omega|}{2\pi}\right)^\alpha c(\alpha)\sin\frac{\alpha\pi}{2} \cdot \sqrt{\Gamma(1+2\alpha)\cdot d(\alpha)},$$

where

$$d(\alpha) = \int_0^\infty \int_0^\infty \frac{x^{\frac{\alpha}{2}-1} y^{\frac{\alpha}{2}-1}(1+x)^{\frac{\alpha}{2}}(1+y)^{\frac{\alpha}{2}}}{(1+x+y)^{1+2\alpha}} dxdy.$$

Here we denote by $|\partial\Omega|$ the length of the boundary of the domain Ω.

Note that multiplication of the operators $\mathcal{C}$, $\mathcal{L}$ and $\mathcal{R}$ by the Bergmann's projection implies a double acceleration of the decrease of the singular values. Spectral characteristics of these products detects the length of the boundary of the domain.

References

[1] J. Arazi and D. Khavinson: *Spectral Estimates of Cauchy's Transform in $L^2(\Omega)$*, Integral Equation and Operator Theory, Vol. 15, (1992), 901-919

[2] J. M. Anderson, D. Khavinson and L. Lomonosov: *Spectral Properties of Some Integral Operators Arising in Potential Theory*, Quart. J. Math. Oxford (2), 43 (1992), 387-407

[3] C. Bandle: *Isoperimetric Inequalities and Applications*, Pitman, London 1980, M. R. 81e: 35095

[4] M. Dostanić: *On an Inequality of Friedrich's Type*, Proc. of the Amer. Math. Soc., vol. 125, N^o 7, (1997), 2115-2118

[5] M. Dostanić: *The Properties of the Cauchy Transform on a Bounded Domain*, J. Operathor Theory 36 (1996), 233-247

[6] M. Dostanić: *Spectral Properties of the Cauchy Operator and its Product with Bergman's Projection on a Bounded Domain*, Proc. London. Math. Soc. (3), 76 (1998), 667-684

[7] M. Dostanić: *Asymptotic Behaviour of Eigenvalues of Certain Integral Operator*, Publ. Inst. Math. (Beograd) (N.S.) 59(73), (1996), 95-113

[8] M. P. Достанич: *Спектральные свойства оператора типа потенциала Рииса и его произведения на проекцио Бергмана в ограниченоий области*, Математический Сборник, T.191 (2000), 23-42

[9] H. Widam: *Asymptotic Behaviour of the Eigenvalues of Certain Integral Equations*, Trans. Amer. Math. Soc. V. 109 (1963), 278-295

Commutativity of Rings With Constraints on Finite Sets

Milan Janjić

Dedicated to Professor Veselin Perić on the occasion of his 70th birthday

Abstract

In this paper we prove three commutativity results for rings which extends some known results about commuting powers.

For a ring R we will consider the following conditions

$$x^s[x^k, y^k] = 0, \tag{1}$$

where $x, y \in R$ and $k \geq 1, s \geq 0$ are integers.

For each $x, y \in R$ there exists a regular element $r = r(x, y)$ such that

$$[r, x] = [r, y] = 0. \tag{2}$$

We shall prove the following commutativity results.

Theorem 1 *Let R be a ring which satisfies* (2) *and let for each finite subset F of R there exists a set $M = M(F)$ of positive integers such that there is no a prime number which divides each element of M and let* (1) *holds for each $x, y \in F$, each $k \in M$ and some integer $s = s(f) \geq 0$. If $a, b \in N$, then $[a, b] = 0$. Additionally, if R is a prime ring with no non-zero nil ideals, then R is commutative.*

Theorem 2 *Let R be a ring with unit element and let for each finite subset F of R there exists a subset $M = M(F)$ of positive integers such that there is no a prime number which divides each element of M. If further there exists an integer $s = s(F) \geq 0$ such that* (1) *holds for each $x, y \in F$, each $k \in M$, then R is commutative.*

Theorem 3 *Let R be a ring with unit element and let for each subset F of R, consisting of four elements, there exist relatively prime integers $m = m(F) \geq 1$, $n = n(F) \geq 1$ and an integer $s = s(F) \geq 0$ such that* (1) *holds for each $x, y \in F$, $k = m$, $k = n$, then R is commutative.*

First we shall prove a lemma which is a slight improvement of some well known results.

Lemma 1 1° *Let R be a ring and for $x \in R$ there exists a regular element $r \in R$ such that $[r,x]=0$. If for $y \in R$ there is an integer $n \geq 1$ such that*

$$x^n y = (x+r)^n y = 0$$

then $y=0$.

2° *Let R be a ring and $a, x \in R$ with $[x,a]=0$. If for $b \in R$ there exists an integer $k \geq 0$ such that $x^k[a,[a,b]]=0$ then*

$$x^k[a^n, b] = nx^k a^{n-1}[a,b].$$

Proof of the Lemma. 1°. Expanding $(x+r)^n$ we get

$$0 = (x+r)^n y = x^n + (nx^{n-1}r + \cdots + r^n)y = (nx^{n-1}r + \cdots + nxr^{n-1})y + r^n y.$$

If we denote $-(nx^{n-2}r + \cdots + nr^{n-1})$ by a we obtain $r^n y = axy$. This implies

$$r^{n^2} y = a^n x^n y = 0$$

i.e. $y=0$ since r is regular.

Proof of 2° follows easily by induction on n.

Proof of Theorem 1. If $a \in R$ is niloptent, t its index of nilpotency, then

$$(r - ra)\cdot(r + ra + ra^2 + \cdots + ra^{n-1}) = r^2,$$

which implies that $r - ra$ and so $r + ra$ are regular.

For $a, b \in N$ take the set $F = \{r+ra, r+ra^2, \ldots, r+rb, r+rb^2, \ldots\}$ which is obviously finite. Let $M(F)$ and $s = s(F)$ be as in Theorem 1. Since all elements of F are regular we have

$$[x^k, y^k] = 0, \text{ for all } x, y \in F \text{ and for all } k \in M(F). \tag{3}$$

Let $k \in M(F)$ and $y \in F$ be arbitrary. Since a is nilpotent there exists an integer $p \geq 1$ such that

$$k[a^i, y^k] = 0,$$

holds for every $i \geq p$. Suppose p_0 is minimal with this property. If $p_0 > 1$ by (3) we have

$$k[a^{p_0-1}, y^k] = 0 = [(r + ra^{p_0-1})^k, y^k],$$

which is a contradiction to the choice of p_0. Thus p_0 must be equal 1 and we have

$$k[a, y^k] = 0, \ y \in F, \ k \in M(F). \tag{4}$$

Since b is nilpotent in the same way from (4) we get

$$k^2[a,b] = 0, \ k \in M(F)$$

and by the property of $M(F)$ we conclude that $[a,b]=0$.

Suppose now that R is a prime ring with non-zero nil ideals. Let $a \in N$ and $x \in R$ be arbitrary and take the set $F = \{r + ra, r + ra^2, \ldots, ax, xa\}$ which is finite. Let $s = s(F)$ and $M(F)$ be as in Theorem 1. In the same way as above we prove that $k[a, (ax)^k] = 0$, for all $k \in M(F)$. This implies

$$ka^i(ax)^k \left(x(ax)^{k-1}\right)^{i-1} = k(ax)^{ik},$$

for all integers $i \geq 1$. Specially, if p is the index of nilpotency of a, we get

$$k(ax)^{(p-1)k} = 0,$$

for all $k \in M(F)$. By the property of $M(F)$ from this we easily conclude that $(ax)^N = 0$, for some integer $N \geq 1$ which means that ax is nilpotent. In the same way we get that xa is nilpotent. This, with the fact that nilpotent elements mutually commute, means that N is a nil ideal of R and so N must be equal $\{0\}$. Since R is prime and has no non-zero nilpotent element it also has no non-trivial zero divisors. The condition (1) is now reduced to the condition $[x^k, y^k] = 0$ and R is commutative, by a result of Herstein [2].

Proof of Theorem 2. Since R has unit element 1 we may take $r = 1$ in (2) so that R satisfies the condition of Theorem 1. From this theorem we have

$$C(R) \subset N, \; N^2 \subset Z(R). \tag{5}$$

Now we will prove that $N \subset Z(R)$. For $a \in N$ and $x \in R$ take $F = \{1 + a, x, x + a, x + xa\}$. Let m, n and s be as in our Theorem. From $[(1+a)^m, x^m] = [(1+a)^n, x^n] = 0$ and (5) we easily conclude that

$$m[a, x^m] = n[a, x^n] = 0. \tag{6}$$

Conditions

$$x^s[x^k, (x + a)^k] = x^s[x^k, (x + xa)^k] = 0,$$

for $k = m$ and $k = n$ and (5) implies

$$x^s[x^m, [x^m, a]] = x^s[x^n, \; [x^n, a]] = 0.$$

By Lemma 2° we get

$$x^s[x^{m^2}, a] = mx^{s+m(m-1)}[x^m, a], \; x^s[x^{n^2}, a] = nx^{s+n(n-1)}[x^n, a].$$

Using (6) we obtain $x^s[x^{m^2}, a] = x^s[x^{n^2}, a] = 0$ and since m^2 and n^2 are relatively prime we easily obtain $x^N[x, a] = 0$, for some $N \geq 1$. If we take $F = \{1 + a, y, y + a, y + ya\}$ where $y = x + 1$ then repeating the same argument as above to get $(x + 1)^M[x, a] = 0$, for some $M \geq 1$. Now by Lemma 1° we conclude that $[x, a] = 0$, which proves that $N \subset Z(R)$.

Let $x, y \in R$ be arbitrary. Take $F = \{x, y\}$ and m, n and s as in Theorem 2. By the fact that $N \subset Z(R)$ we may use Lemma 2° to obtain

$$0 = x^s[x^m, y^m] = m^2x^{s+m-1}y^{m-1}[x, y],$$

and

$$0 = x^s[x^n, y^n] = n^2 x^{s+n-1} y^{n-1}[x, y].$$

Since m^2 and n^2 are relatively prime we get

$$x^t y^t [x, y] = 0,$$

for some integer $t \geq 0$. Repeating the same argument for $x + 1$ instead of x and later for $y + 1$ instead of y using Lemma 1° we get $[x, y] = 0$ which completes the proof of Theorem 2.

Proof of Theorem 3. An inspection of the proof of preceding theorems shows that only for the proof of (5) we have used finite subsets with eventually more then four elements. We must show that this part of the proof may be derive using only sets with four elements.

Invertible elements of R forms a multiplicative group which is commutative by a result in [1]. If $a \in R$ is nilpotent then $1 + a$ is invertible. This implies that nilpotent elements of R mutually commute. To show that nilpotent elements form an commutative ideal of R it is sufficient to prove that R is commutative, under the additional condition that R is a prime ring without non-zero nil ideals. If $u \in R$ is invertible and $x \in R$ arbitrary then there exist relatively prime integers $m \geq 1$, $n \geq 1$ such that $[u^m, x^m] = [u^n, x^n] = 0$ holds. It follows that there exist integers $N \geq 1$ and $k \geq 1$ such that

$$[u^N, x^k] = [u^{N+1}, x^k] = 0,$$

holds. From that we easily conclude that $[u, x^k] = 0$, which means that u lies in the hypercenter of R. Herstein's hypercenter theorem shows that u belongs to the center of R. We thus obtain that $N \subset Z(R)$. It follows that $N = \{0\}$ and the condition (1) again becomes $[x^k, y^k] = 0$ and R is commutative as above. From this we conclude that (5) holds.

References

[1] M. Hongan, *A commutativity theorem for s-unital rings II*, Math. J. Okayama Univ. 25(1983), 19-22

[2] I.N. Herstein, *A commutativity theorem*, J. Algebra 38(1976), 112-118

Milan Janjić,
Prirodno-matematički fakultet Banja Luka
M. Stojanovića 2, Banja Luka