Mathematica Pannonica
1/1 (1990), 5 – 13

RAINBOW HAMILTONIAN PATHS AND CANONICALLY COLORED SUBGRAPHS IN INFINITE COMPLETE GRAPHS

Paul **Erdős**

Mathematical Institute of the Hungarian Academy of Sciences, 1053 Budapest, Reáltanoda u. 13-15, Hungary.

Zsolt **Tuza***

Computer and Automation Institute of the Hungarian Academy of Sciences, 1111 Budapest, Kende u. 13-17, Hungary.

Received April 1988

AMS Subject Classification: 05 C 55, 05 C 15

Keywords: edge coloring, infinite graph, Hamiltonian path, rainbow subgraph, canonical coloring, 0-1 measure

Abstract: A sufficient condition is given for the existence of a Hamiltonian path all of whose edges have a distinct color, in edge-colored infinite complete graphs. Also, a variant of the Erdős-Rado theorem is presented for canonically colored subgraph.

* Research supported in part by the "AKA" Research Fund of the Hungarian Academy of Sciences.

0. Introduction

In this note we consider complete graphs $K = (X, E)$ with an infinite vertex set X and edge set $E = \{xx' : x, x' \in X\}$. For a given coloring φ of the *edge set*, a subgraph $G \subset K$ is called a ***rainbow subgraph*** of K if $G|_\varphi$, the coloring of G induced by φ, contains no monochromatic pair of edges.

If $Y \subset X$ and G is the complete subgraph induced by Y in K, then we write $Y|_\varphi$ instead of $G|_\varphi$.

Our first aim is to find a condition ensuring the existence of a rainbow Hamiltonian path (i.e., a path visiting all vertices of K) when X is countable. As shown in Theorem 1, it is enough to exclude canonically colored infinite subgraphs (see definition below) from $K|_\varphi$, provided that at each vertex, each color class has a finite or 0-measure infinite degree. This result generalizes a theorem of Hahn and Thomassen [6]. Examples show that the condition in Theorem 1 is nearly the best possible; it would be interesting, however, to see an "if and only if"-type characterization, in terms of forbidden subgraphs (cf. Problem 6).

In the second part of the paper we investigate the question how large canonically colored subgraphs exist in K when X is an ordered set of arbitrary cardinality. We consider a particular class of (so-called "properly ordered") colorings and show that if rainbow triangles are forbidden in $K|_\varphi$ then there can be found a canonically colored complete subgraph on a vertex set of cardinality $|X|$ (Theorem 3). The exclusion of a rainbow K_4, however, is not sufficient, as shown by a suitable coloring for $X = \mathbb{R}$ (the set of real numbers).

1. Rainbow Hamiltonian paths in K_ω

Throughout this section, K denotes the *countable* complete graph with vertex set $X = \{x_1, x_2, \ldots\}$ and edge set $E = \{x_i x_j : i \neq j\}$. We assume there is a 0-1 measure μ on X, i.e., for every $Y \subset X$, $\mu(Y) \in \{0, 1\}, \mu$ is finitely additive, $\mu(X) = 1$, and $\mu(Y) = 0$ for all

finite $Y \subset X$.

For convenience, we denote the colors by integers 1,2,.... Two colorings φ, φ' of a graph G are said to be isomorphic if φ' can be obtained from φ, as well as φ from φ', by renumbering (but not indentifying) the colors. In this sense, two edge-colored graphs G_1, G_2 are isomorphic if for their colorings φ_1, φ_2 we have $G_1|_{\varphi_1} \cong G_2|_{\varphi_2}$, i.e., there is a one-to one mapping between the vertex sets $V(G_1)$ and $V(G_2)$, yielding the isomorphism of φ_1 und φ_2.

Denote by Z^* the complete graph with a countable vertex set $\{z_0, z_1, z_2, \ldots\}$ and having the (canonical) edge coloring in which $z_i z_j$ has color j whenever $i < j$.

Theorem 1. *Suppose $\varphi : E \to \mathbb{N}$ is a coloring of K, such that for each vertex x_i and each color j, the vertices adjacent to x_i by an edge of color j form a set of measure 0. If $K|_\varphi$ contains no subgraph isomorphic to Z^* then K has a one-way infinite and a two-way infinite rainbow Hamiltonian path.*

Proof. We construct a sequence $P_1, P_2, \ldots$ of (finite) rainbow paths with the following properties: $x_i \in P_i$ for all $i \geq 1$, and $P_i \subset P_{i+1}$ in the sense that all edges of P_i are edges of P_{i+1} too. This clearly implies that $\cup P_i$ is a rainbow Hamiltonian path of K.

Let $P_1 = (x_1), P_2 = (x_1 x_2)$. If the Hamiltonian path to be found is one-way infinite then we extend P_i at the end different from x_1; if it should be two-way infinite, we extend P_i at the end being closer to x_1.

Suppose P_i is a rainbow path covering $\{x_1, \ldots, x_i\}$. If $x_{i+1} \in P_i$ define $P_{i+1} = P_i$. Otherwise, denote by y_j the jth vertex of P_i, i.e., $P_i = (y_1 y_2 \ldots y_k)$ where $k = |P_i|$. Set $Y = X \backslash (\{x_{i+l}\} \cup \{y_l, \ldots, y_k\})$.

Delete all vertices y from Y, for which $\varphi(y_k y)$ or $\varphi(x_{i+l} y)$ appears on some edge of P_i. The resulting vertex set Y' has $\mu(Y') = 1$, since each of the $k-1$ colors appearing in P_i defines a neighborhood of x_{i+1} and y_k of measure 0 (and μ is finitely additive). If there is a $y \in Y'$ such that $\varphi(x_{i+1} y) \neq \varphi(y_k y)$ then $P_{i+1} = (y_1 \ldots y_k y x_{i+1})$ is a rainbow path containing x_{i+1}. Otherwise, $\varphi(x_{i+1} y) = \varphi(y_k y)$ for all $y \in Y'$. Let $Y_1 \cup Y_2 \cup \ldots = Y'$ be the partition of Y' in which two vertices y and y' belong to the same class if and only if $\varphi(y_k y) = \varphi(y_k y')$. Then $\mu(Y_m) = 0$ for all $m \geq 1$.

Choose an arbitrary $y' \in Y'$, and delete all y from Y' for which $\varphi(y' y)$ appears in P_i or is identical to $\varphi(y_k y')$. The set of those y

is of measure 0, so that the resulting set Y'' has $\mu(Y'') = 1$. If, for some $y'' \in Y''$, $\varphi(y''y') \neq \varphi(y''y_k)$ then $P_{i+1} = (y_1 \dots y_k y'' y' x_{i+1})$ is a rainbow path containing x_{i+1} and we are home. Otherwise, choose a $y'' \in Y''$ and repeat the same argument. Either a rainbow P_{i+1}, containing x_{i+1}, is found after a finite number of steps, or an infinite sequence y', y'', y''',... of vertices is defined with the property that $\varphi(y^{(p)}y^{(q)} = \varphi(y_k y^{(q)})$ for all $p < q$. In the latter case, however, those vertices would induce a subgraph isomorphic to Z^*, condradicting our assumptions, so that P_i can be extended to a rainbow path P_{i+1}, for all i.

$\diamondsuit$

Corollary 1.1. (Hahn and Thomassen [6]) *If all monochromatic subgraphs are locally finite in a Z^*-free coloring of K, then K contains a rainbow Hamiltonian path.*

$\diamondsuit$

An interesting particular case is when any two edges of the same color in $K|_\varphi$ are vertex-disjoint. Such a φ is called a *proper edge coloring* of K.

Corollary 1.2. *Every proper edge coloring of K contains a rainbow Hamiltonian path.*

$\diamondsuit$

Though Z^* itself contains a rainbow Hamiltonian path, it is very close to being non-Hamiltonian in the following sense. Denote by Z^Δ the graph which is obtained from Z^* by recoloring the edge $z_0 z_1$ to color 2.

Proposition 2. *The graph Z^Δ contains no rainbow Hamiltonian paths.*

Based on a similar idea, the following more general class of examples can be given. Consider an arbitrary complete graph K_n on n vertices, with a coloring φ_n which does not contain a rainbow Hamiltonian path. Suppose φ_n uses colors 1',2',..., none of them appearing among the colors 1,2,.... Replace z_0 by $K_n|_{\varphi_n}$ in Z^*, and define the edge $z_i y$ to have color i, whenever $y \in V(K_n)$ and $i \geq 1$. Denote this edge-colored graph by $Z^*(\varphi_n)$. Now Proposition 2 can be stated in the following stronger form.

Proposition 2'. *If $K_n|_{\varphi_n}$ contains no rainbow Hamiltonian path then neither does $Z^*(\varphi_n)$.*

Proof. Suppose to the contrary that P is a rainbow Hamiltonian path in $Z^*(\varphi_n)$. Then the vertices of K_n induce at least two subpaths P_1, P_2 (both maximal under inclusion) in P. We may assume all vertices between P_1 and P_2 belong to Z^* z_0. Let z_m be the vertex between P_1 and P_2 in P having maximum subscript. Then the two neighbors of z_m in P are adjacent to z_m by edges of color m, contradicting the assumption that P is rainbow.

$\diamond$

In particular, any coloring of K_n with at most n-2 colors satisfies the assumptions on φ_n.

2. Canonically colored subgraphs

In this section we consider infinite complete graphs $K = (X, E)$ with a vertex set X of arbitrary cardinality. We assume there is an ordering $<$ given on X.

Erdős and Rado [2] proved that every coloring φ of K contains an infinite $Y \subset X$ such that $Y|_\varphi$ is rainbow or monochromatic or, $\varphi(yy') = \varphi(yy'')$ either for all $y < y' < y''$ or for all $y'' < y' < y$ $(y, y', y'' \in Y)$.

Call a $Y \subset X$ *cannonically colored* if for all $y, y', y'' \in Y$, $y < y' < y''$, $\varphi(yy') = \varphi(yy'')$. We are interested in the question how large canonically colored complete subgraphs must exist in $K|_\varphi$. The following particular class of colorings will be considered. We say that φ is *properly ordered* if $\varphi(xx') \neq \varphi(xx'')$ whenever $x'' < x' < x (x, x', x'' \in X)$.

Theorem 3. *Let φ be a properly ordered coloring of K, not containing rainbow triangles. Then there is a $Y \subset X$, $|Y| = |X|$, such that $Y|_\varphi$ is canonically colored.*

Proof. For any three elements $x, y, z \in X$, $x < y < z$, either $\varphi(xy) = \varphi(xz)$ or $\varphi(xy) = \varphi(yz)$, since $\varphi(xz) \neq \varphi(yz)$.

If X contains a maximum element x_0 then set $X' = X \backslash \{x_0\}$; otherwise, $X' = X$. Now any two monochromatic edges of X' share a

vertex. Indeed, suppose $\varphi(uv) = \varphi(yz)$. Choose an $x \in X$ such that $x > \max(u, v, y, z)$. Then there is an edge of color $\varphi(uv)$ that joins x to uv and also to yz. Those two edges must coincide, however, since we have a properly ordered coloring.

Thus, each monochromatic subgraph of $X'|_\varphi$ is a star, since monochromatic triangles cannot occur in properly ordered colorings.

Call a monochromatic star non-trivial if it contains at least two edges. Such a star has a (unique) centre, the common vertex of its edges. Observe that every $x \in X'$ is the centre of at most one (non-trivial) star. Otherwise, let $\varphi(xy) = \varphi(xy') \neq \varphi(xz) = \varphi(xz')$. Choose a $w \in X$, $w > \max(x, y, y', z, z')$. Then $\varphi(xy) = \varphi(xw) = \varphi(xz)$ should hold, a contradiction. Since each triangle contains a pair of monochromatic edges, there are at most two vertices x', x'' that are not centres of some star. Set $X'' = X' \backslash \{x', x''\}$.

Thus, each $x \in X''$ is the centre of exactly one non-trivial star S_x. Renumbering the colors, if necessary, we may assume S_x is colored by color x. We define a partition $X_1 \cup X_2 = X''$ as follows: $x \in X_1$ if $y < x$ implies $\varphi(xy) \neq x$; $x \in X_2$ if there is a $y < x$ with $\varphi(xy) = x$. The proof will be done if we show $X_1|_\varphi$ and $X_2|_\varphi$ are both canonically colored.

Suppose $x \in X_2$. If there were a $z > x$ such that $\varphi(xz) \neq x$ then $\varphi(yz) = x$ would follow for any $y, \varphi(xy) = x$, a contradiction as S_y cannot have color x. Hence, X_2 is canonically colored, and $y \in X_1$ whenever $\varphi(xy) \neq y$, $y < x$.

Suppose X_1 is not canonically colored, i.e., there are three elements $x, y, z \in X_1$, $x < y < z$, $\varphi(xy) = a \neq b = \varphi(xz)$. Then $\varphi(yz) = a$ (since φ is properly ordered), so that $y \in X_2$, contrary to our assumption.

$\diamondsuit$

We note that the above argument yields the following result for the finite case.

Theorem 3'. *Every properly ordered coloring of K_n with no rainbow triangle contains a canonicaly colored $K_{\lfloor n/2 \rfloor - 1}$.*

$\diamondsuit$

Instead of K_3, the exclusion of a rainbow K_4 is not sufficient in Theorem 3. This fact can be proved in the following stronger form. ($\mathbb{R}$ denotes the set of real numbers.)

Theorem 4. *For $X = \mathbb{R}$, there exists a properly ordered coloring φ with the following properties:*
(i) *Every canonically colored Y is countable;*
(ii) *$X|_\varphi$ contains no rainbow finite subgraphs of minimum degree greater than 2. (In particular, $X|_\varphi$ is rainbow-K_4-free.)*

Proof. First, consider the properly ordered (canonical) coloring φ^+ defined by $\varphi^+(xy) = x$ for all $x < y$. We modify φ^+ by splitting each color class into two parts, and replacing each color x by two colors x', x''. (Clearly, after any kind of splitting, the obtained coloring remains properly ordered.)

The splitting is based on idea due to Sierpiński [5]. Consider a well-ordering $<_L$ of $\mathbb{R}$. For $x < y$, define $\varphi(xy)$ to be x' if $x <_L y$ and to be x'' if $y <_L x$. Let $Y|_\varphi$ be canonically colored, for some $Y \subset X = \mathbb{R}$. We show Y is countable.

Set $E_x = \{xy : x < y \in Y\}$ for $x \in Y$. If Y is canonically colored then each E_x is monochromatic. Divide Y into two (disjoint) parts Y_1, Y_2 as follows: $x \in Y_1$ if E_x has color x' and $x \in Y_2$ if E_x has color x''. By the definition of $<_L$, for each $x \in Y_1$, the set $\{y \in Y_1 : y > x\}$ contains a minimum element y_x. Picking a rational number from the interval $[x, y_x)$, it follows that Y_1 is countable. By a similar argument, considering the sets $\{y \in Y_2 : y < x\}$ and the intervals $(y_x, x]$, it follows that Y_2 is countable.

Let G be a finite rainbow subgraph of $X|_\varphi$, with a vertex set $\{x_1, \ldots, x_n\}$. Then $x = \min x_i$ has degree at most 2, since all edges incident to x in G have color x' or x''.

$\diamond$

3. Concluding remarks

I. Corollary 1.2 is much easier to prove than Theorem 1. As a matter of fact, in a proper edge coloring, P_i can be extended to a suitable P_{i+1} by adding x_{i+1} and at most one extra vertex. The finite version of Corollary 1.2, however, is unknown. A nice construction of Maamoun and Meyniel [3] shows there is a proper edge coloring of the complete graph K_n on $n = 2^k$ vertices (for all $k \geq 2$) not containing a rainbow Hamiltonian path. It would be interesting to see such colorings for all

even n.

On the other hand, Andersen [1] conjectures that every proper edge coloring of K_n contains a rainbow path covering all vertices but one. Some lower bounds on the length of a maximum rainbow path are given by Rödl und Tuza [4]. Here we raise the following related question.

Problem 5. Find the minimum number $f(n)$ of colors, such that every proper edge coloring of K_n by at least $f(n)$ colors contains a rainbow Hamiltonian path.

The examples of [3] show $f(n) \leq n-1$ does not hold in general. It seems to be reasonable to conjecture, however, that $f(n)$ is very close (or, perhaps, equal) to n.

II. All our examples for colorings of a countable complete graph without a rainbow Hamiltonian path have a canonical structure (cf. Proposition 2). Now the following two problems arise.

Problem 6. (a) Find a class $\underline{F}$ of edge-colored countable complete graphs with the following properties:

(i) No $F \in \underline{F}$ contains a rainbow Hamiltonian path.
(ii) All infinite complete subgraphs of $K|_\varphi$ have a rainbow Hamiltonian path if and only if $K|_\varphi$ contains no subgraph isomorphic to any $F \in \underline{F}$.

(b) Do all $f \in \underline{F}$ have a canononical structure?

III. It is easy to show there is a subset $\{a_1, a_2, \ldots\}$ of the natural numbers such that every positive integer occurs exactly once among the numbers and $|a_i - a_j|$, $1 \leq i < j$. In other words, if color $|i-j|$ is assigned to edge $x_i x_j$ then in this coloring of K_ω some rainbow complete subgraph contains all colors. This observation leads to the following questions.

Problem 7. (a) Under what conditions does a countable (or an arbitrary infinite) complete graph K contain a rainbow complete subgraph involving all colors that appear in K?
(b) Find theorems of this type for *finite* complete graphs.
(c) Let $0 < a_1 < a_2 < \ldots < a_k$, and suppose that for each integer $i, 1 \leq i \leq n$, there is exactly one pair $j, m (1 \leq j < m \leq k)$ such that

$a_m - a_j = i$. Find $a(n) = \min a_k$. Also, find the minimum value of $k = k(n)$, for which such a sequence $a_1, \ldots, a_k$ exists.

Note that a greedy argument shows $a(n) \leq 0(n^3)$. In fact, there exists an infinite sequence $a_1, a_2, \ldots$ with $a_n \leq cn^3$ (for some constant c), whose $(2n)^{th}$ slice satisfies the requirements, for all $n \geq 1$.

IV. Concerning Theorem 3, one should ask that, instead of triangles, what sort of rainbow subgraphs F can be excluded so that $X|_\varphi$ must contain a canonically colored subgraph Y of cardinality $|Y| = |X|$. Theorem 4 shows F always has minimum degree at most 2 (when F is finite).

References

[1] ANDERSON, L.D.: Hamilton circuits with many colours in properly edge-coloured complete graphs, *Math. Scandin.* (to appear).

[2] ERDÖS, P. and RADO, R.: A combinatorial problem, *J. London Math. Soc.* **25** (1950), 249-255.

[3] MAAMOUN, M. and MEYNIEL, H.: On a problem of G. Hahn about coloured Hamiltonian paths in K_{2n}, *Discrete Math.* **51** (1984), 213-214.

[4] RÖDL, V. and TUZA, Zs.: Rainbow subraphs in properly edge-colored graphs, (to appear).

[5] SIERPIŃSKI, W.: Sur un problème de la théorie des relations, *Annali R. Scuola Normale Superiore de Pisa*, Ser. 2, **2** (1933), 285-287.

[6] HAHN, G. and THOMASSEN, C.: Path and cycle sub-Ramsey numbers and an edge-colouring conjecture, *Discrete Math.* **62** (1986), 29-33.

Mathematica Pannonica
1/1 (1990), 15 –32

GRÖBNER BASES IN GEOMETRY THEOREM PROVING AND SIMPLEST DEGENERACY CONDITIONS

Franz **Winkler**

Institut für Mathematik and Research Insitute for Symbolic Computation, Johannes Kepler Universität, A-4040 Linz, Österreich.

Received September 1988

AMS Subject Classification: 13 A 15, 13 F 20, 51-04, 68 C 20, 68 G 15

Keywords: geometry theorem proving, polynomial ideals, syzygies, Gröbner bases

Abstract: The method of Gröbner bases has been fruitfully applied to many problems in the theory of polynomial ideals. Recently Gröbner bases have been used in various ways for dealing with the problem of geometry theorem proving as posed by Wu. One approach is centered around the computation of a basis for the module of syzygies of the hypotheses and conclusion of a geometric statement. We elaborate this approach and extend it to a complete decision procedure.

In geometry theorem proving the problem of constructing subsidiary (or degeneracy) conditions arises. Such subsidiary conditions usually are not uniquely determined and obviously one wants to keep them as simple as possible. The question of constructing simplest subsidiary conditions has not been addressed yet. We show that our algorithm is able to construct the simplest subsidiary conditions with respect to certain predefined criteria, such as lowest degree or fewest variables.

0. Introduction

The work of Wu Wen-tsün [Wu 1978], [Wu 1984] has renewed the interest in automated geometry theorem proving. He has developed a decision algorithm for a certain class of geometry problems. The class of problems Wu considers (Wu's geometry, for short) consists, intuitively speaking, of those problems that can be translated into algebraic equations over some ground field K, the number system associated with the geometry. For the relationship between axiomatic geometries and number systems we refer to [Hilbert 1977]. Basically, Wu's geometry allows one to talk about incidence, parallelism, perpendicularity, cocircularity, congruence, etc., but not about "betweenness", because no order predicate is available.

Often a geometric statement is true only in a "generic" sense, i.e. after certain degenerate situations have been ruled out. Such degenerate situations typically occur when triangles collapse to a line segment, circles to a point, etc. and they are usually not explicitly mentioned. An automatic procedure for proving geometry statements has to be able to deal with the problem of such "degeneracy" or "subsidiary" conditions, that means it has to be able to automatically find suitable subsidiary conditions which make the statement a theorem, if such conditions exist at all.

Wu has given a decision procedure for solving the geometry theorem proving problem. His procedure also finds a subsidiary condition, if such a condition exists. Wu's decision algorithm has been partially implemented by himself and by Chou [Chou 1985]. Many interesting theorems have been proved by these implementations, including Simson's theorem, Pascal's theorem, the Butterfly theorem and Feuerbach's theorem. Wu's algorithm is based on the computation of characteristic sets of polynomial ideals, as introduced by Ritt [Ritt 1950].

Different approaches to geometry theorem proving, based on the computation of Gröbner bases [Buchberger 65], [Buchberger 85] for polynomial ideals, have been reported. In [Chou, Schelter 1986] Gröbner bases over the field generated by the independent variables of a geometric construction are employed. Kapur [Kapur 1986a,b] describes

a refutational theorem prover, based on Rabinowitsch's trick for proving Hilbert's Nullstellensatz. Kutzler and Stifter [Kutzler, Stifter 1986a,b] describe various ways of applying Gröbner bases to this problem, one of which is centered on the computation of a basis for the module of syzygies of the geometrical hypotheses and conclusion. This method is not complete. However, we are able to extend it to a complete decision procedure.

As we have mentioned above, an automatic procedure for geometry theorem proving must be able to find subsidiary conditions. Of course it would be of interest to keep the subsidiary condition as simple as possible. Referring to his approach Kapur [Kapur 1986b] claims that *"conditions found using this approach are often simpler and weaker than the ones reported using Wu's method or reported by an earlier version of Kutzler & Stifter's paper as well as Chou & Schelter based on the Gröbner basis method."* However, no algorithm for computing the "simplest" subsidiary condition has been reported up to now. Our algorithm is able to compute the "simplest" subsidiary condition by giving a complete overview of the possible subsidiary conditions. Reasonable criteria for "simplest" might be "of as low a degree as possible" or "involving only certain variables."

The stucture of this paper is as follows. In chapter 1 we give a short introduction to the theory of Gröbner bases, reviewing definitions and basic facts as far as they will be neccessary for the geometry theorem proving problem. In chapter 2 we define the geometry theorem proving problem. We derive a complete decision procedure *GEO*, which is also able to compute the simplest subsidiary condition for a given instance of the geometry theorem proving problem. Finally, in chapter 3 we demonstrate how *GEO* can be applied to concrete geometry problems.

1. The method of Gröbner bases

We define the notion of a Gröbner basis for a polynomial ideal as introduced by Buchberger [Buchberger 1965, 1985].

Let K be a field and $K[x_1, \ldots, x_n]$ (or $K[X]$ for short) the polynomial ring over K in the indeterminates $x_1, \ldots, x_n$. Let $[x_1, \ldots, x_n] = [X]$ denote the monoid of power products in $x_1, \ldots, x_n$. We start by choosing a *term ordering* $\prec$, i.e. a linear ordering on $[X]$ which makes $[X]$ an ordered monoid with $x_1^0 \ldots x_n^0$ as the least element. With respect to $\prec$ every nonzero polynomial $f \in K[X]$ contains a highest power product, which is called the *leading power product of* $f, lpp(f)$. The coefficient of $lpp(f)$ in f is called the *leading coefficient* of $f, lc(f)$. The polynomial which results from f by subtracting the leading power product multiplied by the leading coefficient is called the *reductum* of f, i.e. $red(f) = f - lc(f) \cdot lpp(f)$.

Every nonzero polynomial f gives rise to a *reduction relation* $\rightarrow_f$ on $K[X]$ in the following way: $g_1 \rightarrow_f g_2$ if and only if there is a power product u with a nozero coefficient a in g_1, i.e. $g_1 = au + h$ for some polynomial h which does not contain u, such that $lpp(f)$ divides u, i.e. $u = lpp(f)u'$ for some u', and $g_2 = -\frac{a}{lc(f)}u'red(f) + h$. If F is a set of polynomials, the *reduction relation* modulo F is defined so that $g_1 \rightarrow_F g_2$ if and only if $g_1 \rightarrow_f g_2$ for some $f \in F$. In this case g_1 is *reducible to* g_2 *modulo* F. If there is no such g_2, g_1 is *irreducible modulo* F. For every set of polynomials F the reduction relation $\rightarrow_F$ is Noetherian, i.e. every chain $f_1 \rightarrow_F f_2 \rightarrow_F \cdots$ terminates. We say that g is a *normal form of* f *modulo* F, if f can be reduced to g by a finite number of applications of $\rightarrow_F$, and g is irreducible modulo F. Normal forms are usually not unique.

If F is the basis of a polynomial ideal I, then obviously $f \rightarrow_F 0$ implies $f \in I$. In general, however, the implication in the reverse direction does not hold. A non-zero polynomial f might be irreducible modulo F and still $f \in I$.

Definition 1.1. Let I be an ideal in $K[X]$. A finite set of polynomials G is a *Gröbner basis* for I iff $(G) = I$ (G generates I) and $f \in I \Leftrightarrow f \rightarrow_F 0$, for all $f \in K[X]$. $\diamond$

There are many equivalent definitions for Gröbner bases. The interested reader may confer [Buchberger 1985]. More importantly, however, every ideal I in $K[X]$ has a Gröbner basis and a Gröbner basis for I can always be computed starting with some basis F of I.

Göbner bases are an extremely powerful tools in commutative algebra. We mention some applications, as far as we will need them in the subsequent chapters. For further applications we refer to [Buchberger 1985], [Winkler et al. 1985], [Winkler 1986]. The "main problem" of polynomial ideal theory, namely the question whether $f \in I$ for a polynomial f and a polynomial ideal I, can easily be solved once a Gröbner basis G for I has been computed: reduce f to its unique normal form modulo G and check whether this normal form is 0. The identity $I = J$ for two ideals I and J can be checked algorithmically by computing Gröbner bases G_I and G_J for I and J, respectively, and then checking whether every basis element in G_I is in J and vice versa. The membership problem for the radical of an ideal I (i.e. wheter $f \in \mathit{radical}(I)$) can be solved by computing a Gröbner basis G for $(I, z \cdot f - 1)$, where z is a new variable, and checking whether G contains a constant.

The computation of a Gröbner basis is an important step in solving a system of algebraic equations. The following elimination property of a Gröbner basis with respect to a lexicographic ordering of the variables has been observed by Trinks [Trinks 1978]. It means that the i-th elimination ideal of an ideal I with Gröbner basis G is generated by the basis elements in G that depend only on the first i variables.

Lemma 1.2. *Let I be an ideal in $K[X]$ and G a Gröbner basis for I with respect to the lexicographic ordering $\prec$ with $x_1 \prec x_2 \prec \cdots \prec x_n$. Then, for $1 \leq i \leq n$,*

$$I \cap K[x_1, \ldots, x_i] = (G \cap K[x_1, \ldots, x_i]),$$

where the ideal on the right hand side is formed in $K[x_1, \ldots, x_i]$.

Proof. Obviously the right hand side is contained in the left hand side. On the other hand, assume that $f \in I \cap K[x_1, \ldots, x_i]$. Then f can be reduced to 0 modulo G with respect to the lexicographic ordering $\prec$. All the polynomials occurring in this reduction process depend only on the variables $x_1, \ldots, x_i$, and we get a representation of f as a linear combination of polynomials in G, where all the summands in this representation depend only on $x_1, \ldots, x_i$. $\diamond$

Given bases for the ideals I and J, bases for $(I \cup J)$ and $I \cdot J$ can easily be determined. In general, however, computing bases for $I \cap J$ and $I : J$ is a hard problem.

Lemma 1.3. *Given bases for the ideals I and J in $K[X]$, bases for the following can be computed:*
(a) $I \cap J$,
(b) $I : J$,
(c) $radical(I)$.

Proof. (a) For a new variable z we have

$$I \cap J = ((z-1)I \cup zJ) \cap K[X].$$

From bases for I and J we immediately get a basis for $((z-1)I \cup zJ)$. The intersection with $K[X]$ can be computed by Lemma 1.2.

(b) If $J = (f)$, then compute a basis $\{g_1, \ldots, g_k\}$ of $I \cap (f)$ by (a). $\{g_1/f, \ldots, g_k/f\}$ is a basis for $I : (f)$. In the general case $J = (f_1, \ldots, f_m)$ we have

$$I : J = \bigcap_{i=1}^{m} (I : (f_i)).$$

(c) The zero-dimensional case is treated in [Kalkbrener 1987], [Kobayashi et al. 1988] and the general case in [Kandri-Rody 1984], [Gianni et al. 1988].

$\diamondsuit$

Definition 1.4. Let $< f_1, \ldots, f_m > \in K[X]^m$. $< g_1, \ldots, g_m > \in$ $\in K[X]^m$ is a *syzygy* of $< f_1, \ldots, f_m >$ iff $\sum_{i=1}^{m} f_i g_i = 0$. For a subset M of $K[X]^m$, $< g_1, \ldots, g_m >$ is a *syzygy* of M iff it is a syzygy of every element of M.

$\diamondsuit$

For a finite set $M \subset K[X]$, the syzygies of M are the solutions of a homogeneous system of linear equations with coefficients in M. A (finite) set $M \subset K[X]^m$ generates a module over $K[X]$, and on the other hand, as a consequence of Hilbert's basis theorem, every submodule of $K[X]^m$ has a finite basis. The set of syzygies of a subset M of $K[X]^m$ is equal to the set of syzygies of the module generated by M over $K[X]$, and it forms again a module over $K[X]$. The Gröbner bases algorithm can be used to compute a basis for the module of syzygies of M.

Lemma 1.5. *For every finite subset M of $K[X]^m$ a basis for the module of syzygies of M can be computed.*

Proof. see [Buchberger 1985] for the case $|M| = 1$ and [Winkler 1986] for the general case. An alternative approach via extending the notion of a Gröbner basis to modules is taken in [Galligo 1979] and [Möller, Mora 1986]. ◇

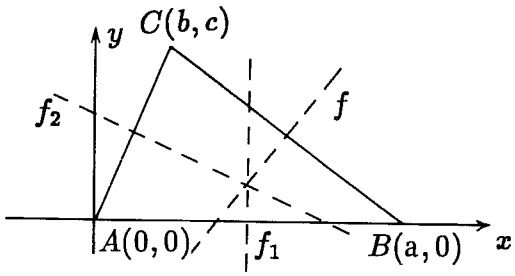
2. Geometry theorem proving: a decision procedure

We consider a geometry whose associated number system is the algebraic closure $\bar{K}$ of a field K, i.e. the geometric objects lie in $\bar{K}^n$ for some $n \in \mathrm{N}$. The statements we allow have to be expressible in the form

$$(2.1) \qquad (\forall x \in \bar{K}^n)[f_1(x) = 0 \wedge \ldots \wedge f_m(x) = 0 \Rightarrow f(x) = 0]$$

for some polynomials $f_1, \ldots, f_m, f$ in $K[x_1, \ldots, x_n] = K[X]$. The $f_1, \ldots, f_m$ are called the *hypothesis polynomials* or *hypotheses* for short, and f is called the *conclusion polynomial* or just the *conclusion.* Basically, this enables us to talk about incidence, parallelism, perpendicularity, cocircularity, congruence, etc., but not about "betweenness", because no order predicate is available.

As an example let us consider the geometric theorem (in $\mathbb{R}^2$) that "*for every triangle ABC the lines orthogonal to the sides of the triangle and passing through the midpoints of the associated sides have a common point of intersection*". Before we can express this theorem algebraically, we have to place the triangle in a two dimensional coordinate system. Without loss of generality we can assume that A is placed at the origin, $A = (0, 0)$, and that the side AB is parallel to the x-axis, $B = (a, 0)$. No restriction is put on $C = (b, c)$.



The equations for f_1, f_2 and f are

$$f_1(x,y) = x - \frac{1}{2}a,$$

$$f_2(x,y) = b(x - \frac{1}{2}b) + c(y - \frac{1}{2}c),$$

$$f(x,y) = (a-b)(x - \frac{1}{2}(a-b)) + c(y - \frac{1}{2}c).$$

In order to prove the theorem, it suffices to show that f vanishes on the variety of $(f_1, f_2) \subset \mathbb{R}(a,b,c)[x,y]$, or in other words that $f \in$ $\in$ *radical*(f_1, f_2). By the method described in Chapter 1 this problem can be decided by computing a Gröbner basis for $(f_1, f_2, z \cdot f - 1)$ in $\mathbb{R}(a,b,c)[x,y]$. The computation can be carried out completely over the field $\mathbb{Q}(a,b,c)$, yielding the Gröbner basis $\{1\}$. So f is indeed in the radical of (f_1, f_2) and the theorem is proved. A geometry theorem prover along these lines is described in [Chou,Schelter 1986].

An important step in this approach is the transition from the question whether a polynomial f vanishes on the variety of an ideal I to the problem whether f is in the radical of I. That is only possible if the varieties are defined over an algebraically closed ground field. So, for instance, one cannot decide geometric statements in real space but only in complex space. Theorems in real geometry can only be confirmed, but not disproved. For actually deciding statements in real geometry one has to consider the theory of elementary algebra and elementary geometry, based on real closed fields.This theory has been

shown to be decidable by Tarski [Tarski 1951] and has become known as Tarski algebra. Tarski's decision procedure has recently been improved in [Collins 1975], [Ben-Or et al. 1984] and [Grigor'ev 1988].

Often a geometric theorem is true only after certain degenerate situations have been ruled out by a nondegeneracy or subsidiary condition. As for the hypotheses and the conclusion, we require that the subsidiary condition be expressible by a polynomial, this time by a polynomial inequation of the form $s(x_1, \ldots, x_n) \neq 0$. So the problem becomes to decide whether for given $f_1, \ldots, f_m, f$ and s in $K[X]$

$$(2.2) \qquad (\forall x \in \bar{K}^n)[f_1(x) = \cdots = f_m(x) = 0 \wedge s(x) \neq 0 \Rightarrow f(x) = 0].$$

Moreover, as we have mentioned above, in a geometry theorem proving setting it is reasonable to require that a subsidiary condition be determined algorithmically.

So we arrive at the following formal specification of the geometry theorem proving problems posed in [Wu 1984]. Let K be a field, $\bar{K}$ the algebraic closure of K.

P_{Wu}:
given: polynomials $f_1, \ldots, f_m, f$ in $K[X]$
decide: does there exist a polynomial $s \in K[X]$ such that
(1) $(\forall x \in \bar{K}^n)(f_1(x) = \ldots = f_m(x) = 0 \wedge s(x) \neq 0 \Rightarrow f(x) = 0)$
and
(2) $(\exists x \in \bar{K}^n)(f_1(x) = \ldots = f_m(x) = 0 \wedge s(x) \neq 0)$?
If so, find such an s.

Part (2) in P_{Wu} guarantees that the subsidiary condition does not exclude all points in the variety of $f_1, \ldots, f_m$. Sometimes it seems natural to use a finite number $s_1, \ldots, s_n$ of subsidiary conditions, replacing $s(x)$ in P_{Wu} by $s_1(x) \neq 0 \wedge \ldots \wedge s_n(x) \neq 0$, thus getting a modified problem. However, it can easily be seen that a single subsidiary condition s is sufficient. The factors of s satisfy the modified problem, and if $s_1, \ldots, s_n$ satisfy the modified problem, then their product $s_1 \cdot \ldots \cdot s_n$ satisfies P_{Wu}.

In [Wu 1984] Wu describes a decision algorithm for P_{Wu}, which has been partially implemented by himself and by Chou [Chou 1985].

Wu's algorithm is based on the computation of characteristic sets of polynomial ideals, as introduced by Ritt [Ritt 1950]. In this paper we solve P_{Wu} by computing a basis for the module of syzygies of the geometrical hypotheses and conclusion, thus getting also a method for computing the simplest subsidiary condition.

Theorem 2.1. *Let $f_1, \ldots, f_m, f$ be the parameters of an instance P of P_{Wu}.*

(i) *Those polynomials $s \in K[X]$, which satisfy part (1) of P, constitute an ideal N_P.*

(ii) *For every $s \in N_P$ there exist $s_1, \ldots, s_m \in K[X]$ and $k \in \mathbb{N}$, such that $< s_1, \ldots, s_m, s^k \cdot f^{k-1} >$ is a syzygy of $< f_1, \ldots, f_m, f >$, i.e. $s_1 \cdot f_1 + \cdots + s_m \cdot f_m + s^k \cdot f^{k-1} \cdot f = 0$.*

(iii) *If $< s_1, \ldots, s_m, s^k \cdot f^{k-1} >$, $k \in \mathbb{N}$, is a syzygy of $< f_1, \ldots, f_m, f >$, then $s \in N_P$.*

(iv) *Set*
$S_P = \{s | < s_1, \ldots, s_m, s >$ is a syzygy of $< f_1, \ldots, f_m, f >$ for some $s_1, \ldots, s_m\}$. Then $N_P = radical(S_P) : (f)$.

Proof. (i) Suppose both s_1 and s_2 solve part (1) of P. Now let t_1, t_2 be arbitrary polynomials, and let $x \in \bar{K}^n$ be such that $f_1(x) = \ldots = f_m(x) = 0$ and $(t_1 s_1 + t_2 s_2)(x) = t_1(x) \cdot s_1(x) + t_2(x) \cdot s_2(x) \neq 0$. Then either $s_1(x) \neq 0$ or $s_2(x) \neq 0$. W.l.o.g. assume that $s_1(x) \neq 0$. But then $f(x) = 0$, since s_1 is a solution of part (1) of P. So $t_1 s_1 + t_2 s_2$ is also a solution of part (1) of P.

(ii) Since $s \in N_P$, we know that $s \cdot f$ vanishes on every common zero of $f_1, \ldots, f_m$ in $\bar{K}$. That, however, means that $s \cdot f$ is in the radical of $(f_1, \ldots, f_m)$, and a power of $s \cdot f$, say $s^k \cdot f^k$, $k \in \mathbb{N}$, is in $(f_1, \ldots, f_m)$. Therefore, for some $s_1, \ldots, s_m \in K[X]$,

$$s_1 \cdot f_1 + \cdots + s_m \cdot f_m + s^k \cdot f^k = 0,$$

i.e. $< s_1, \ldots, s_m, s^k \cdot f^{k-1} >$ is a syzygy of $< f_1, \ldots, f_m, f >$.

(iii) $s_1 \cdot f_1 + \cdots + s_m \cdot f_m + s^k \cdot f^k = 0$, so for every $x \in \bar{K}^n$

$$f_1(x) = \ldots = f_m(x) = 0 \wedge s(x) \neq 0 \Rightarrow f(x) = 0.$$

(iv) Clearly S_P is an ideal in $K[X]$. By (ii) and (iii)

$$N_P = \{s \in K[X] | s^k \cdot f^{k-1} \in S_P \text{ for some } k \geq 1\}$$

If $s \in N_P$, then $s^k f^{k-1} \in S_P$ for some $k \geq 1$. Thus $s^k f^k \in S_P$. This, however, implies $sf \in radical(S_P)$ and therefore $s \in radical(S_P) : (f)$. On the other hand, let $s \in radical(S_P) : (f)$, i.e. $sf \in radical(S_P)$. Then $s^k f^k \in S_P$ for some $k \geq 1$. So $s^{k+1} f^k \in S_P$ and therefore $s \in N_P$. $\diamondsuit$

By Lemma 1.5 a finite basis for the module of syzygies of a sequence of polynomials can be computed. So for every instance P of P_{Wu} one can compute a finite basis for the ideal S_P. From the basis for S_P a basis for N_P can be computed by Lemma 1.3. Hence we have a complete overview of the solutions of part (1) of P_{Wu}. The remaining question is, whether there is a solution of (1), which also satisifies (2).

Theorem 2.2. *Let P be an instance of P_{Wu}, B a finite basis for N_P.*
(i) *If there is a polynomial in N_P which satisfies* (2), *then there is a polynomial in the basis B which satisfies* (2).
(ii) *If B is a Gröbner basis for N_P with respect to the term ordering $\prec$, B' is the set of $b \in B$ which satisfy part* (2) *of P, and $t = \min\{lpp(b) | b \in B'\}$, then for every solution s of P, $lpp(s) \geq t$.*

Proof. (i) Let $f_1, \ldots, f_m, f$ be the parameters of the instance P of P_{Wu} and $B = \{b_1, \ldots, b_r\}$. Assume that no basis polynomial b_i, $1 \leq i \leq r$, satisfies (2), i.e.

$$(\forall x \in \bar{K}^n)(f_1(x) = \ldots = f_m(x) = 0 \Rightarrow b_i(x) = 0) \text{ for all } 1 \leq i \leq r.$$

Then also for every linear combination $s = \sum_{i=1}^{r} h_i b_i$ we have

$$(\forall x \in \bar{K}^n)(f_1(x) = \ldots = f_m(x) = 0 \Rightarrow s(x) = 0),$$

so no $s \in N_P$ satisfies (2).

(ii) Let s be a solution of part (1) of P. $s \in N_P$, so s is reducible to 0 w.r.t. B. Let $C \subseteq B$ be the set of elements of B used in this reduction. Then $lpp(b) \leq lpp(s)$ for every $b \in C$. If no $b \in C$ satisfies part (2) of P, then neither does s. $\diamondsuit$

Theorem 2.2 (ii) establishes that "simplest" subsidiary conditions can be computed by choosing the term ordering $\prec$ appropriately, namely so that s_1 is simpler than s_2 if and only if $lpp(s_1 \prec lpp(s_2)$.

For instance, a Gröbner basis for N_P with respect to a graded ordering contains a solution of lowest degree of P, if any solution exists. A Gröbner basis for N_P with respect to a lexicographic ordering $x_1 \prec \ldots \prec x_m \prec \ldots \prec x_n$ contains a solution depending only on $x_1, \ldots, x_m$, if such a solution exists. The variables $x_1, \ldots, x_m$ could be the "independent" variables (see [Kutzler, Stifter 1986b]) of the geometric construction. So one can ask the question whether there is a nondegeneracy condition depending only on the independent variables. The two orderings can, of course, be combined by ordering the power products in $x_1, \ldots, x_m$ by some ordering $\prec_1$, e.g. according to the degree, and also the power products in $x_{m+1}, \ldots, x_n$ by some ordering $\prec_2$. Then a term ordering $\prec$ can be constructed by

$$u_1 u_2 \prec t_1 t_2 :\Leftrightarrow u_2 \prec_2 t_2 \vee (u_2 = t_2 \wedge u_1 \prec_1 t_1),$$

where u_1, t_1 are power products over $x_1, \ldots, x_m$ and u_2, t_2 power products over $x_{m+1}, \ldots, x_n$. This ordering will lead to a subsidiary condition of lowest degree involving only the independent variables $x_1, \ldots, x_m$.

In their report [Chou,Yang 1986] Chou and Yang consider the problem statement P_{Wu} and claim: "*The algebraic problem in this formulation is well defined. However, the polynomial s sometimes has nothing to do with nondegenerate conditions in geometry. To make things worse, this formulation is unsound from the geometric point of view.*" They go on to stress their point by an example. We will deal with this example and the criticism of P_{Wu} in Chapter 3.

Combining Theorems 2.1 and 2.2 we get the following decision algorithm for P_{Wu}.

Algorithm *GEO* (**in:** polynomials $f_1, \ldots, f_m$, $f \in K[X]$,
out: s, a solution of the instance
$P =< f_1, \ldots, f_m, f >$ of P_{Wu},
if such a solution exists,
or "no");

(1) *Compute a finite basis C for S_P, the ideal generated by the last component of the syzygies of $< f_1, \ldots, f_m, f >$.*

(2) *Compute a basis C' for radical(S_P).*

(3) *Compute a Gröbner basis C'' for $((z-1)C' \cup \{z \cdot f\})$ in $K[X][z]$ with respect to a lexicographic term ordering $x_1 \prec \cdots \prec x_n \prec z$.*

(4) *Set $C''' = C'' \cap K[X]$. C''' is a basis for $radical(S_P) \cap (f)$.*

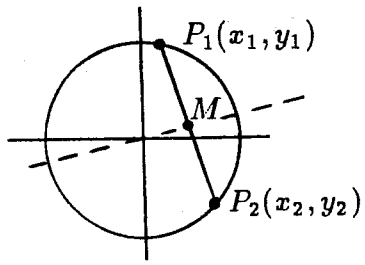
(5) *Set $B = \{h/f | h \in C'''\}$. B is a basis for $radical(S_P) : (f) = N_P$.*

(6) *Check the polynomials b in B for $b \notin radical(I)$, where $I = (f_1, \ldots, f_m)$. If B is a Gröbner basis with respect to the term ordering $\prec$ and b is the element of B with the least leading power product satisfying $b \notin radical(I)$, then b is the simplest subsidiary condition. Set $s = b$ and stop. Otherwise output "no".* ◇

3. Examples

We use the decision algorithm *GEO* to prove that

"if P_1 and P_2 are two points on a circle and M is the midpoint of P_1 and P_2 then the line through M and perpendicular to P_1P_2 contains the center of the circle".



The hypotheses of the given instance P of P_{Wu} are
$f_1 : x_1^2 + y_1^2 - x_2^2 - y_2^2$
(P_1 and P_2 are points on a circle with center (0,0))
$f_2 : a(x_2 - x_1) + b(y_2 - y_1)$
($\binom{a}{b}$ is perpendicular to P_1P_2)

and the conclusion is
$f : a(y_1 + y_2) - b(x_1 + x_2)$
(the line $y = \frac{b}{a}x$ contains M, the midpoint of P_1 and P_2).

First we compute a basis for the ideal S_P, i.e. the third component of the module of syzygies of (f_1, f_2, f). A Gröbner basis for $ideal(f_1, f_2, f)$ in $\mathbb{Q}[a, b, x_1, x_2, y_1, y_2]$ w.r.t. the lexicographic ordering with $a \prec b \prec x_1 \prec x_2 \prec y_1 \prec y_2$ is

$$\{f_1, f_2, f, f_3 = aby_1 - \frac{1}{2}b^2x_2 - \frac{1}{2}a^2x_2 - \frac{1}{2}b^2x_1 + \frac{1}{2}a^2x_1\}.$$

From the Gröbner basis we immediately get a basis for the module of syzygies of $< f_1, f_2, f_3, f >$. By an algorithm described in [Buchberger 1985] this syzygy basis can be transformed to a basis of the syzygies of $< f_1, f_2, f >$:

$$(-b, y_2 + y_1, x_1 - x_2),$$

$$(-a, x_2 + x_1, y_2 - y_1),$$

$$(0, ay_2 + ay_1 - bx_2 - bx_1, -by_2 + by_1 - ax_2 + ax_1),$$

$$(2aby_1 - b^2x_2 - a^2x_2 - b^2x_1 + a^2x_1, ay_2^2 - ay_1^2 + ax_2^2 - ax_1^2,$$
$$-by_2^2 + by_1^2 - bx_2^2 + bx_1^2).$$

Thus $S_P = (x_2 - x_1, y_2 - y_1)$ and $C = \{x_2 - x_1, y_2 - y_1\}$.

S_P is radical, so $C' = C$. Next we determine a Gröbner basis C'' for $((z-1)C' \cup \{zf\})$ in $\mathbb{Q}[a, b, x_1, x_2, y_1, y_2, z]$, getting

$$\begin{aligned}
&x_2z - x_1z - x_2 + x_1,\\
&y_2z - y_1z - y_2 + y_1,\\
&ay_2z + ay_1z - bx_2z - bx_1z,\\
&ay_1z - bx_1z + \frac{1}{2}ay_2 - \frac{1}{2}ay_1 - \frac{1}{2}bx_2 + \frac{1}{2}bx_1,\\
&ax_2y_2 - ax_1y_2 + ax_2y_1 - ax_1y_1 - bx_2^2 + bx_1^2 = (x_2 - x_1)\cdot f,\\
&ay_2^2 - bx_2y_2 - bx_1y_2 - ay_1^2 + bx_2y_1 + bx_1y_1 = (y_2 - y_1)\cdot f.
\end{aligned}$$

Intersecting this basis with $\mathbb{Q}[a, b, x_1, x_2, y_1, y_2]$ and dividing by f we finally get the basis $B = \{x_2 - x_1, y_2 - y_1\}$ for $radical(S_P) : (f) = N_P$.

Neither $x_2 - x_1$ nor $y_2 - y_1$ is in the radical of $ideal(f_1, f_2)$, so both are solutions of the geometric problem instance P, and they are solutions of lowest degree.

That means the theorem holds in $\mathbb{C}^2$ (and therefore also in $\mathbb{R}^2$) if either the x-coordinates or the y-coordinates of the two points P_1 and P_2 differ from one another, i.e. P_1 and P_2 do not collapse to a single point.

For further demonstrating the usefulness of computing a simplest subsidiary condition, we consider an example used in [Chou,Yang 1986]

to support the claim that the polynomial s computed as a solution of P_{Wu} may have nothing to do with a subsidiary condition for the geometric problem.

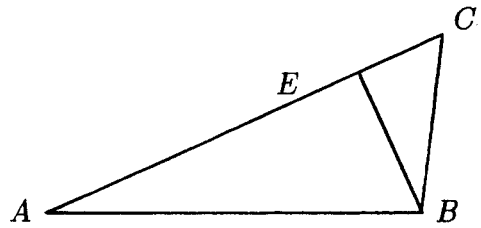
The goal is to prove that *"every triangles is isosceles"*, which of course, is not a theorem in complex geometry. Chou & Yang observe, however, that there is a formulation of this problem as an instance of P_{Wu}, which admits a subsidiary condition s.

The algebraic formulation they use is the following: let ABC be a triangle, and BE the altitude from B. Show that $AB \equiv CB$. As coordinates for the points they choose $A = (0,0)$, $B = (y_1, 0)$, $C = (y_4, y_5)$, and $E = (y_2, y_3)$. Now the hypotheses can be translated into the algebraic equations

$$h_1 = y_3 y_5 + (y_2 - y_1) y_4 = 0 \qquad BE \perp AC$$
$$h_2 = -y_2 y_5 + y_3 y_4 = 0 \qquad E \text{ is on } AC$$

and the conclusion into the equation

$$g = -y_5^2 - y_4^2 + 2 y_1 y_4 = 0 \qquad AB \equiv CB.$$



$s = y_3^2 + y_2^2 - y_1 y_2$ satisfies both conditions in P_{Wu}. In fact, Kapur's theorem prover confirms the "theorem" under the subsidiary condition s. Chou & Yang now state that *"Thus under this formulation we can prove that "every" triangle is isosceles"* and they take this as evidence of their claim that P_{Wu} is *"unsound"*.

In our oppinion, the controversy stems from the fact that the dependent variables y_2, y_3 are not explicitly excluded from the subsidiary condition. If one wants to consider only such subsidiary conditions, which do not involve the dependent variables (which is reasonable from a geometric point of view), then this can be achieved by a suitable

ordering of the power products, e.g. a lexicographic ordering based on

$$\underbrace{y_1 < y_4 < y_5}_{\text{indep. var.}} < \underbrace{y_2 < y_3}_{\text{dep. var.}} .$$

Now the algorithm *GEO* is able to detect that there exists no subsidiary condition involving only the independent variables y_1, y_4, y_5. Actually also Kapur [Kapur 1986b] mentions the possibility of recognizing that there is no such subsidiary condition in a remark following Theorem 2.

Let us apply the algorithm *GEO* to the geometric problem in the formulation above, where h_1, h_2 are the hypotheses and g is the conclusion. We get

$$\begin{aligned}\{b_1 &= y_4 y_3 - y_5 y_2,\\ b_2 &= y_5^2 y_2 + y_4^2 y_2 - y_1 y_4^2,\\ b_3 &= y_3^2 + y_2^2 - y_1 y_2,\\ b_4 &= y_5 y_3 + y_4 y_2 - y_1 y_4\}\end{aligned}$$

as a basis for S_P, the ideal generated by the last component of the syzygies of (h_1, h_2, g). S_P is radical, so we just have to compute the intersection $S_P \cap ideal(g)$ and divide by g. This leads to the basis $\{b_1, b_2, b_3, b_4\}$ for N_P.

Finally in step (6) we detect that $b_3 \notin radical(h_1, h_2)$, but there exists no possible subsidiary condition involving only the independent variables y_1, y_4, y_5.

Acknowledgement. Work reported herein has been supported by the *Österreichische Forschungsgemeinschaft* and by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung* under Project Nr. P.6763.

References

[Ben-Or et al. 1984] BEN-OR, M; KOZEN, D.; REIF, J. : "The Complexity of Elementary Algebra and Geometry", *Proc. 16th ACM Symp. on Theory of Computing*, 457 – 464 (1984).

[Buchberger 1965] BUCHBERGER, B.: Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal, Ph.D. dissertation Univ. Innsbruck, Austria (1965).

[Buchberger 1985] BUCHBERGER, B.: "Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory", in: *Multidimensional Systems Theory*, N.K. Bose, ed., 184 – 232, D. Reidel Publ. Comp. (1985).

[Chou 1985] CHOU, S.-C.: Proving and Discovering Theorems in Elementary Geometry Using Wu's Method, Ph.D. Thesis, Dept. of Mathematics, University of Texas, Austin (1985).

[Chou, Schelter 1986] CHOU, S.-C.; SCHELTER, W.F.: "Proving Geometry Theorems with Rewrite Rules", *J. Automated Reasoning* **2**, 253 – 273 (1986).

[Chou, Yang 1986] CHOU, S.-C.; YANG, J.-G.: "On the Algebraic Formulation of Geometry Theorems", Techn. Rep., Inst. for Computer Science, Univ. of Texas, Austin (1986).

[Collins 1975] COLLINS, G.E.: "Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition", *Proc. 2nd GI Conf. on Automata Theory and Formal Languages*, Springer-Verlag, LNCS 35, 134 – 183, Berlin (1975).

[Galligo 1979] GALLIGO, A.: "Théorème de division et stabilité en géometrie analytique locale", *Ann. Inst. Fourier* **29**, 107 – 184 (1979).

[Gianni et al. 1988] GIANNI, P; TRAGER, B.; ZACHARIAS, G.: "Gröbner Bases and Primary Decomposition of Polynomial Ideals", to appear in *J. of symbolic Computation* (1988).

[Grigor'ev 1988] GRIGOR'EV, D.Yu.: "Complexity of Deciding Tarski Algebra", *J. of Symbolic Computation* **5/1&** **2**, 65 – 108 (1988).

[Hilbert 1977] HILBERT, D.: Grundlagen der Geometrie, Teubner Verlag, Stuttgart (1977).

[Kalkbrener 1987] KALKBRENER, M.: Application of Gröbner Bases: Solution of Algebraic Equations and Decomposition of Radicals, Diplomarbeit, RISC-Linz, J. Kepler Univ. Linz (1987).

[Kandri-Rody 1984] KANDRI-RODY, A.: Effective Methods in the Theory of Polynomial Ideals, Ph. D. thesis, Rensselaer Polytechnic Institute, Troy, NY (1984).

[Kapur 1986a] KAPUR, D.: "Geometry Theorem Proving Using Hilbert's Nullstellensatz", *Proc. SYMSAC'86*, 202 – 208. B.W. Char, ed., ACM (1986).

[Kapur 1986b] KAPUR, D.: "Using Gröbner Bases to Reason About Geometry Problems", *J. of Symbolic Computation* **2/4**, 399 – 408 (1986).

[Kobayashi et al. 1988] KOBAYASHI, H.; MORITSUGU, S. and HOGAN, R.W.: "On Solving Systems of Algebraic Equations", to appear in *J. of Symbolic Computation* (1988).

[Kutzler, Stifter 1986a] KUTZLER, B.; STIFTER, S.: "Automated Geometry Theorem Proving Using Buchberger's Algorithm", *Proc. 1986 Symp. on Symbolic and Algebraic Computation (SYMSAC'86)*, 209 – 214, B.W. Char (ed.), ACM, New York (1986).

[Kutzler, Stifter 1986b] KUTZLER, B.; STIFTER, S.: "On the Application of Buchberger's Algorithm to Automated Geometry Theorem Proving", *J. of Symbolic Computation* **2/4**, 389 – 397 (1986).

[Möller, Mora 1986] MÖLLER, H.M.; MORA, F.: "New Constructive Methods in Classical Ideal Theory", *J. of Algebra* **100/1**, 138 – 178 (1986).

[Ritt 1950] RITT, J.F.: Differential Algebra, AMS Colloquium Publications, New York (1950).

[Tarski 1951] TARSKI, A.: A Decision Method for Elementary Algebra and Geometry, Univ. of California Press (194), 2nd ed. (1951).

[Trinks 1978] TRINKS, W.: Über B. Buchbergers Verfahren, Systeme algebraischer Gleichungen zu lösen", *J. of Number Theory* **10/4**, 475 – 488 (1978).

[Winkler 1986] WINKLER, F.: Solution of Equations I: Polynomial Ideals and Gröbner Bases; lecture notes of the short course on "Symbolic and Algebraic Computation", Conf. "Computers & Mathematics", Stanford Univ. (1986).

[Winkler et al. 1985] WINKLER, F.; BUCHBERGER, B.; LICHTENBERGER, F.; ROLLETSCHEK, H.: "An Algorithm for Constructing Canonical Bases of Polynomial Ideals", *ACM Trans. on Mathematical Software* **11/1**, 66 – 78 (1985).

[Wu 1978] WU, Wen-tsün: "On the Decision Problem and the Mechanization of Theorem Proving in Elementary Geometry", *Scientia Sinica* **21**, 157 – 179 (1978).

[Wu 1984] WU, Wen-tsün: "Basis Principles of Mechanical Theorem-Proving in Elementary Geometry", *J. Syst. Sci.* & Math. Sci. **4/3**, 207 – 235 (1984).

Mathematica Pannonica
1/1 (1990), 33 – 47

ÜBER VERALLGEMEINERTE DIFFERENTIALGLEICHUNGEN VON ACZÉL-JABOTINSKY IM KOMPLEXEN (Parameterabhängigkeit und holomorphe Fortsetzung von Integralen)

Ludwig **Reich**

Institut für Mathematik, Universität, A-8010 Graz, Brandhofgasse 18, Österreich.

Herrn Professor Dr. Helmut Florian zum fünfundsechzigsten Geburtstag am 30. April 1989 gewidmet.

Received April 1989

AMS Subject Classification: 34 A 20, 34 A 10, 30 D 05

Keywords: Differential equations in the complex domain, Continuous depence of solutions on parameters, Functional equations in the complex domain

Abstract: We consider differential equations of the form

(1) $\frac{dw}{dz} = \frac{\delta w^m(1+\delta_1 w+\delta_2 w^2+\ldots)}{\gamma z^n(1+\gamma_1 z+\gamma_2 z^2+\ldots)}$

where $\delta \neq 0, \gamma \neq 0, m, n \geq 1$, and where the power series in the numerator and denominator on the right hand side converge. We are looking for integrals

(2) $w(z) = \rho z + c_2 z^2 + \ldots$

which "enter the singularity (0,0) holomorphically". Firstly, it is proved, that there exist formal integrals of (1) if and only if certain algebraic relations for the coefficients γ_i, δ_j are satisfied, and if $m = n$. But then each formal integral (2) is also convergent, hence a locally analytic solution. Moreover, there exists a continuum of solutions of (1) if this equation is solvable in the

above sense, since the coefficient $c_m(c_1 = \rho)$ if $m = 1)$ may be chosen arbitrarily. It is shown that the functions $(z, c_m) \mapsto w(z, c_m)$ are holomorphic in regions $|c_m| < \eta$, $|z| < \delta(\eta)$. Here $w(z, c_m)$ denotes the unique integral (2) with the given coefficient $c_m, \eta > 0$ is arbitrary, $\delta(\eta) > 0$ depending on η. Applications are given to the third Aczél-Jabotinsky equation in iteration theory and to holomorphic continuation of integrals of (1) into the singularity (0,0).

1. Einleitung

Bei der Beschreibung von Familien vertauschbarer konvergenter und formaler Potenzreihen und der Normalformen solcher Familien gegenüber simultaner Konjugation spielen die sogenannte ("dritte") Differentialgleichung von Aczél-Jabotinsky und eine Verallgemeinerung derselben eine Rolle (vgl. [1], [2], [3], [4], [5], sowie für das System der sogenannten Aczél-Jabotinskyschen Funktionaldifferentialgleichungen im allgemeinen [6], [7], [8]).

Es sind dies (spezielle) Differentialgleichungen der Form

$$\frac{dw}{dz} = \frac{\delta w^m(1 + \delta_1 w + \delta_2 w^2 + \ldots)}{\gamma z^m(1 + \gamma_1 z + \gamma_2 z^2 + \ldots)}, \tag{1}$$

wobei $\gamma, \delta \neq 0$, $m, n, \geq 1$, und die auftretenden Potenzreihen $1 + \delta_1 w + \ldots, 1 + \gamma_1 w + \ldots$ konvergent seien. Bei dem Problem der vertauschbaren Reihen interessiert man sich für konvergente Reihenlösungen von (1) der Form

$$w(z) = \rho z + c_2 z^2 + \ldots, \rho \neq 0. \tag{2}$$

Da auf den Anfangswert $w = 0$ für $z = 0$ bezüglich (1) aber der Existenz- und Eindeutigkeitssatz von Cauchy oder die klassische Methode der Separation der Variablen nicht direkt angewendet werden kann, müssen hier bei Existenzfragen und bei der Untersuchung der Parameterabhängigkeit modifizierte Wege beschritten werden. So wird z.B. in [2] und in [4] die Differentialgleichung (1) auf einen Spezialfall

der Briot-Bouquetschen Differentialgleichung transformiert, der mittels der Majorantenmethode untersucht werden kann.

Wir wollen hier eine Modifikation der Methode der Variablenseparation darstellen, die in der "Singularität" (0,0) Gültigkeit hat. Die Existenz der holomorphen Integrale (2) ergibt sich, wenn überhaupt formale Integrale exisitieren, aus dem Hauptsatz über implizite Funktionen im Komplexen (§2). Die genauere Untersuchung der Abhängigkeit der Lösungen (2) von inneren Parametern ist im Hinblick auf die ursprüngliche, eigangs erwähnte Frage der maximalen Familien vertauschbarer Potenzreihen wichtig und erfordert eine kleine Verschärfung des Satzes über die impliziten Funktionen in §3.

Schließlich werden wir der Frage nachgehen, welche Integrale von (1), die nach dem Satz von Cauchy durch Anfangswerte (z_0, w_0) mit $z_0 \cdot w_0 \neq 0$ eindeutig festgelegt sind, holomorph in die Singularität (0,0) einmünden. Leider kann ich in diesem Punkt kein abschließendes Resultat vorlegen.

2. Die Existenz holomorpher Integrale

Wir wollen in diesem Paragraphen notwendige und hinreichende Bedingungen dafür angeben, daß eine Differentialgleichung (1) Integrale der Form (2) besitzt. Wir beginnen mit formalen (algebraischen) Vorbereitungen, die man hier am bestem im Körper $\mathbb{C} \ll z \gg$ der formalen Laurentreihen mit endlichem Hauptteil, dem Quotientenkörper des Ringes $\mathbb{C}[\![z]\!]$ der formalen Potenzreihen, durchführt. (1) ist, wenn $w(z)$ die Form (2) hat, äquivalent mit folgenden Relationen in $\mathbb{C} \ll z \gg$:

$$\frac{1}{\delta} w^{-m}(1 + \delta_1 w + \delta_2 w^2 + \ldots)^{-1} \frac{dw}{dz} = \frac{1}{\gamma} z^{-n}(1 + \gamma_1 z + \gamma_2 z^2 + \ldots)^{-1},$$

oder nach Berechnung der Reziproken $(1 + \delta_1 w + \ldots)^{-1}$ und $(1 + \gamma_1 z + \ldots)^{-1}$ mittels der geometrischen Reihe und Substitution:

$$\frac{1}{\delta} w^{-m}(1 + P_1(\delta_1)w + P_2(\delta_1, \delta_2)w^2 + \ldots)\frac{dw}{dz} =$$

$$= \frac{1}{\gamma} z^{-n}(1 + P_1(\gamma_1)z + P_2(\gamma_1, \gamma_2)z^2 + \ldots)$$

mit Polynomen $P_1, P_2 \ldots$. Wir setzen $P_0 = 1$. Die Laurentreihen $w^{\nu-m}\frac{dw}{dz}$, $\nu \geq 0$, bilden eine summierbare Familie, da $ord\{w(z)^{\nu-m}\frac{dw}{dz}\} = \nu - m$ und wir finden also die zu (1) äquivalente Relation in $\mathbb{C} \ll z \gg$

$$(3) \qquad \sum_{\nu=0}^{\infty} \frac{1}{\gamma} P_\nu(\delta_1, \ldots, \delta_\nu) w^{\nu-m} \frac{dw}{dz} = \sum_{\nu=0}^{\infty} \frac{1}{\gamma} P_\nu(\gamma_1, \ldots, \gamma_\nu) z^{\nu-n}$$

mit

$$w(z) = \rho z + c_2 z^2 + \ldots, \rho \neq 0.$$

Ist $L(z)$ eine formale Laurentreihe, so ist der Koeffizient von z^{-1} in $\frac{d}{dz}L(z) = L'(z)$ Null, $res\{L'(z)\} = 0$. Nun gilt $w^\lambda \frac{dw}{dz} = \left(\frac{1}{\lambda+1} w^{\lambda+1}\right)$, wenn nur $\lambda \neq -1$ ($\lambda \in \mathbb{Z}$), also

$$res\{w^{\nu-m}\frac{dw}{dz}\} = 0$$

für $\nu \neq m-1$, $\nu \geq 0$, und daher

$$res\left\{\sum_{\nu=0}^{\infty} \frac{1}{\nu} P_\nu(\delta_1, \ldots, \delta_\nu) w^{\nu-m} \frac{dw}{dz}\right\} =$$

$$= \frac{1}{\gamma} P_{m-1}\ (\delta_1, \ldots, \delta_{m-1}) res\left(\frac{1}{w(z)}\frac{dw}{dz}\right).$$

Da wir auf der rechten Seite von (3)

$$res\left\{\sum_{\nu=0}^{\infty} \frac{1}{\delta} P_\nu\ (\gamma_1, \ldots, \gamma_\nu)\ z^{\nu-n}\right\} = \frac{1}{\gamma} P_{n-1}\ (\gamma_1, \ldots, \gamma_{n-1})$$

finden, so ergibt sich

$$\frac{1}{\gamma} P_{m-1}\ (\delta_1, \ldots, \delta_{m-1})\ res\left\{\frac{1}{w}\frac{dw}{dz}\right\} = \frac{1}{\delta} P_{n-1}\ (\gamma_1, \ldots, \gamma_{n-1}).$$

Da

$$w^{-1}\frac{dw}{dz} = (\rho z + c_2 z^2 + \ldots)^{-1}(\rho + 2c_2 z + \ldots) = \frac{1}{z} + Q(z)$$

mit $Q(z) \in \mathbb{C}[\![z]\!]$, so folgt

$$\frac{1}{\gamma} P_{m-1}(\delta_1, \ldots, \delta_{m-1}) = \frac{1}{\delta} P_{n-1}(\delta_1, \ldots, \delta_{n-1}). \tag{4}$$

Wegen $P_0 \equiv 1$ und wegen

$$w^{\nu-m} \frac{dw}{dz} = \rho^{\nu-m+1} z^{\nu-m} + R(z)$$

mit $ord\ R(z) > \nu - m$, folgt durch Vergleich der Ordnungen der linken und rechten Seiten in (3)

$$m = n \tag{5}$$

und somit aus (4)

$$\frac{1}{\gamma} P_{m-1}(\delta_1, \ldots, \delta_{m-1}) = \frac{1}{\delta} P_{m-1}(\gamma_1, \ldots, \gamma_{m-1}). \tag{6}$$

Der Vergleich der niedrigsten Potenzen von z auf den beiden Seiten von (3) ergibt unter Verwendung des Ansatzes (2) und mit $P_0 \equiv 1$

$$\frac{1}{\gamma} \rho^{m-1} = \frac{1}{\delta}, \tag{7}$$

d.h. ρ ist eine, wie wir sehen werden, beliebige, (m-1)-te Wurzel aus γ/δ. Wenn also $m = 1$, so kann $\rho \neq 0$ beliebig sein, es ist dann aber

$$\gamma = \delta \quad (für\ m = 1) \tag{8}$$

notwendigerweise.

Wie schon festgestellt, ist jede Laurentreihe $w^{\nu-m} \frac{dw}{dz}$ für $\nu - m \neq -1$ die Ableitung einer anderen, nämlich der Laurentreihe $\frac{1}{\nu-m+1} w^{\nu-m+1}$. Um den Fall $\nu - m = -1$ zu untersuchen, setzen wir

$$w(z) = \rho z\ (1 + \omega(z)), \tag{9}$$

mit $\omega(z) \in \mathbb{C}[\![z]\!]$, $ord\ \omega \geq 1$. Dann ist

$$w^{-1} \frac{dw}{dz} = \frac{1}{z} + (1 + \omega(z))^{-1} \frac{d\omega}{dz}. \tag{10}$$

Setzen wir (9) und (10) in (3) ein, so finden wir, da sich $\frac{1}{z}$ wegen (5) und (6) weghebt:

$$\sum_{\nu\neq m-1}^{\infty} \frac{1}{\gamma} P_\nu\ (\delta_1,\ldots,\delta_\nu) \left[\frac{1}{\nu-m+1}(\rho z(1+\omega(z))^{\nu-m+1}\right]' +$$

$$+ P_{m-1}(\delta_1,\ldots,\delta_{m-1})\frac{\omega'(z)}{1+\omega(z)} =$$

$$= \sum_{\substack{\nu=0\\ \nu\neq m-1}}^{\infty} \frac{1}{\delta} P_\nu\ (\gamma_1\ldots,\gamma_n u) \left(\frac{z^{\nu-m+1}}{\nu-m+1}\right)'.$$

Es ist aber, bei Verwendung der Logarithmusreihe $ln(1+x)$

$$\frac{\omega'(z)}{1+\omega(z)} = [ln(1+\omega(z))]'.$$

Da $$\left(\tfrac{1}{\nu-m+1}(\rho z(1+\omega(z))^{\nu-m+1}\right)_{\substack{\nu\geq 0\\ \nu\neq m-1}}$$

eine summierbare Familie ist, so sieht man leicht ein, daß sich Summation und Derivation vertauschen lassen, und man findet

$$(11)\quad t + \sum_{\substack{\nu=0\\ \nu\neq m-1}}^{\infty} \frac{1}{\gamma} P_\nu\ (\delta_1,\ldots,\delta_\nu) \left[\frac{1}{\nu-m+1}(\rho z(1+\omega(z))^{\nu-m+1}\right] +$$

$$+P_{m-1}\ (\delta_1,\ldots,\delta_{m-1})ln(1+\omega(z)) = \sum_{\substack{\nu=0\\ \nu\neq m-1}} \frac{1}{\delta} P_\nu\ (\gamma_1,\ldots,\gamma_\nu)\frac{z^{\nu-m+1}}{\nu-m+1}$$

mit einer Integrationskonstanten $t \in \mathbb{C}$.

Es sei zunächst $m = 1$. Die Bedingung (6) besagt hier $\delta = \gamma$. Dann haben wir

$$(12)\quad t + \sum_{\nu=1}^{\infty} \frac{1}{\gamma} P_\nu(\delta_1,\ldots,\delta_\nu) \left[\frac{1}{\nu}\rho^\nu z^\nu (1+\omega(z))^\nu\right] + \frac{1}{\nu} ln(1+\omega(z)) =$$

$$= \sum_{\nu=1}^{\infty} \frac{1}{\gamma} P_\nu(\gamma_1,\ldots,\gamma_\nu)\frac{z^\nu}{\nu},$$

und es ist wegen $ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$ notwendigerweise $t = 0$. Wir erhalten weiters aus (11)

$$\omega(z) = -\sum_{\nu=1}^{\infty} \frac{1}{\gamma} P_\gamma(\delta_1, \ldots, \delta_\nu) \left[\frac{1}{\gamma}\rho^\nu z^\nu (1+\omega(z))^\nu\right] + \tag{13}$$
$$+ \left(\frac{\omega^2(z)}{2} - \frac{\omega^3(z)}{3} + \ldots\right) + \sum_{\nu=1}^{\infty} \frac{1}{\gamma} P_\nu(\gamma_1, \ldots, \gamma_\nu) \frac{z^\nu}{\nu};$$

und der Hauptsatz über implizite Funktionen ergibt die Existenz eines durch (12) eindeutig bestimmten $\omega(z) \in \mathbb{C}[\![z]\!]$ mit $ord\ \omega \geq 1$. Dies ist sogar konvergent, da die Reihen auf der rechten Seite nach Voraussetzung bzw. auf Grund ihrer Erstehung aus konvergenten Reihen konvergent sind.

Sodann sei $m > 1$. Entwickeln wir $(1+\omega(z))^{\nu-m+1}$ für $\nu < m-1$ nach der Binomialreihe, für $\nu \geq m-1$ nach dem binomischen Lehrsatz und setzen wir für $ln(1+\omega(z))$ definitionsgemäß die logarithmische Reihe, und multiplizieren wir (11) mit z^{m-1}, so ergibt sich die Relation

$$tz^{m-1} + \sum_{\substack{\nu=0 \\ \nu\neq m-1}}^{\infty} \tfrac{1}{\gamma} P_\nu(\delta_1, \ldots, \delta_\nu) \left[\tfrac{1}{\nu-m+1}\rho^{\nu-m+1} z^\nu (1+\omega(z))^{\nu-m+1}\right] +$$

$$+\frac{1}{\gamma} P_{m-1}(\delta_1 \ldots, \delta_{m-1}) z^{m-1} ln(1+\omega(z)) = \tag{14}$$

$$= \sum_{\substack{\nu=0 \\ \nu\neq m-1}}^{\infty} \frac{1}{\gamma} P_\nu(\gamma_1, \ldots, \gamma_\nu) \frac{z^\nu}{\nu - m + 1}$$

Wegen $m > 1$ hat das Absolutglied auf der linken Seite von (14) wegen $P_0 = 1$ den Wert $\frac{1}{\gamma}\rho^{-m+1}$, auf der rechten Seite steht $\frac{1}{\delta}$, so daß es sich in dieser Relation weghebt. Der Term $\omega(z)$ tritt nur auf der linken Seite auf, und sein Koeffizient ist $\frac{1}{\gamma}\frac{-m+1}{-m+1}\rho^{-m+1} = \frac{1}{\delta} \neq 0$. Wir können (13) also nach $\omega(z)$ auflösen, und für jedes $t \in \mathbb{C}$ den Hauptsatz über implizite Funktionen anwenden. Da alle auftretenden Reihen konvergieren, ist (bei festem t und ρ) das eindeutig bestimmte $\omega(z)$ mit $ord\omega \geq 1$ auch konvergent. Berechnen wir die Koeffizienten von $\omega(z)$ explizit, so sehen wir, daß der Koeffizient von z^{m-1} in t affin ist, und daher, ebenso wie t, als freier Parameter gewählt werden kann. Dies

trifft daher auch für den Koeffizienten von z^m in $w(z)$ zu. Wir fassen zusammen in

Satz 1.
a) *Eine Differentialgleichung* (1) *mit* $\gamma \neq 0$, $\delta \neq 0, m, n \geq 1$, *besitzt genau dann ein holomorphes Integral* (2),

$$w(x) = \rho z + c_2 z^2 + \ldots + c_m z^m + \ldots,$$

wenn

$$m = n \tag{5}$$

und mit gewissen Polynomen P_{m-1}

$$\frac{1}{\gamma} P_{m-1}(\delta_1, \ldots, \delta_{m-1}) = \frac{1}{\delta} P_{m-1}(\gamma_1, \ldots, \gamma_{m-1}) \tag{6}$$

gilt.
b) *Sind* (5) *und* (6) *erfüllt, so ist, wenn* $m > 1$ *notwendigerweise*

$$\rho = \sqrt[m-1]{\frac{\gamma}{\delta}};$$

wenn $m = 1$, *so ist* $\rho \neq 0$ *beliebig.*
Zu jedem $c_m \in \mathbb{C}(m > 1)$ *bzw.* $\rho \in \mathbb{C}$ $(m = 1)$ *gibt es genau ein Integral*

$$w(z) = \rho z + c_2 z^2 + \ldots + c_m z^m + \ldots$$

von (1). Die Koeffizienten c_ν *sind Polynome in* ρ *und in* c_m.
Speziell sind die Differentialgleichungen von Aczél-Jabotinsky von der Form (1). In ihrem Fall sind die Bedingungen (5) und (6) erfüllt, wie leicht nachzurechnen, und daher ist Satz 1 auf sie anwendbar.

3. Parameterabhängigkeit

Wir wollen nun die Abhängigkeit vom Parameter $c_m(m > 1)$ und vom Parameter $\rho(m = 1)$ genauer untersuchen. Dazu benötigen wir

eine kleine Verschärfung des Satzes über die impliziten Funktionen im Komplexen.

Satz 2. *Vorgelegt sie die Gleichung*

$$w(z,u_1,\dots,u_N) = \sum_{\substack{(\alpha,\beta)\neq(1,0)\\ \alpha+\beta\geq 1}} d_{\alpha\beta\gamma_1,\dots,\gamma_N} w^\alpha z^\beta u_1^{\gamma_1}\dots u_N^{\gamma_N}. \tag{15}$$

Zu jedem $\eta > 0$ gebe es ein $\delta_0(\eta) > 0$, so daß die Potenzreihe auf der rechten Seite von (15) für $|w| < \delta_0(\eta)$, $|z| < \delta_0(\eta)$, $|u_j| < \eta$, $(j = 1,..,N)$ konvergiere. Dann gilt:
Zu jedem $\eta > 0$ gibt es ein $\delta(\eta) > 0$, so daß die gemäß dem Hautpsatz über implizite Funktionen eindeutig bestimmte Lösung $w(z,u_1,\dots,u_N)$ von (15) mit $|u_j| < \eta (j = 1,\dots,N)$ holomorph ist.

Beweis. Der Hauptsatz über implizite Funktionen im Komplexen behauptet die Existenz genau einer formalen und sogar konvergenten Potenzreihe

$$w(z,u_1,\dots,u_N) = \sum_{\alpha\geq 1} c_{\alpha\gamma_1,\dots,\gamma_N} z^\alpha u_1^{\gamma_1}\dots u_N^{\gamma_N},$$

welche (15) erfüllt. Die Berechnung dieser (formalen) Reihe $w(z,u_1,\dots,u_N)$ kann auf zwei Arten erfolgen. Man kann erstens die Koeffizienten $c_{\alpha\gamma_1\dots\gamma_N}$ aus einer mehrstufigen Rekursion (z.B. nach der lexikographischen Ordnung der Multiindizes) berechnen. Man kann aber zweitens die Reihe auf der rechten Seite von (15)

$$\sum_{\substack{(\alpha,\beta)\neq(1,0)\\ \alpha+\beta\geq 1}} d_{\alpha\beta\gamma_1\dots\gamma_N} w^\alpha z^\beta u_1^{\gamma_1}\dots u_N^{\gamma_N} =$$

$$= \sum_{\alpha,\beta}\left(\sum_{\gamma} d_{\alpha\beta\gamma_1\dots\gamma_N} u_1^{\gamma_1}\dots u_N^{\gamma_N}\right) w^\alpha z^\beta := \sum_{\substack{(\alpha,\beta)\neq(1,0)\\ \beta\geq 1}} F_{\alpha\beta}(u) w^\alpha z^\beta$$

mit $F_{\alpha\beta}(u) \in \mathbb{C}[\![u_1\dots,u_N]\!]$ und

$$w(z) = \sum_{\delta\geq 1} c_\delta(u) z^\delta, c_\delta(u) \in \mathbb{C}[\![u_1,\dots u_N]\!],$$

schreiben und die Koeffizienten $c_\delta(u)$ rekursiv bestimmen. Man sieht leicht ein, daß die $c_\delta(u)$ Polynome mit positiven Koeffizienten in den formalen Reihen $F_{\alpha\beta}(u)$ mit $\alpha+\beta \leq \delta$ sind:

$$c_\delta(u) = Q_\delta(F_{\alpha\beta}(u)),\ \delta = 1, 2, \ldots. \tag{16}$$

Diese Umordnung ist auch im Sinne der Analysis durchführbar, da die auftretenden Reihen innerhalb gewisser Polyzylinder absolut konvergieren. Nach Voraussetzung sind die Reihen $F_{\alpha\beta}(u)$ für $|u_\gamma| < \eta$ mit beliebigem $\eta > 0$ konvergent, wir wählen $\eta = \eta_0 > 0$ fest. Dies gilt auch für die "kleinsten" Majoranten

$$F^*_{\alpha\beta}(u) = \sum_\gamma |d_{\alpha\beta\gamma_1\ldots\gamma_N}|u_1^{\gamma_1}\ldots u_N^{\gamma_N}.$$

Betrachten wir das majorante implizite Problem

$$W(z,u) = \sum F^*_{\alpha\beta}(u) w^\alpha z^\beta \tag{17}$$

mit $W(0,u) \equiv 0$, so ist

$$W(z,u) = \sum_{\delta\geq 1} C^*_\delta(u) z^\delta$$

ebenfalls konvergent, und $W(z,u)$ ist eine Majorante von $w(z,u)$, ebenso wie $C^*_\delta(u)$ von $c_\delta(u)$, für $\delta = 1, 2, \ldots$, was aus der Relation

$$C^*_\delta(u) = Q_\delta(F^*_{\alpha\beta}(u)) = \sum c^*_{\delta\gamma_1\ldots\gamma_N} u_1^{\gamma_1}\ldots u_N^{\gamma_N} \tag{18}$$

folgt, da die Q_δ positive Koeffizienten haben. Schließlich untersuchen wir auch noch das "spezialisierte" implizite Problem. Wir setzen $u_1^0 = \ldots = u_N^0 = \eta_0$, und lösen

$$\phi(z) = \sum F^*_{\alpha\beta}(u^0)\phi^\alpha(z) z^\beta \phi(z) = \sum_{\nu\geq 1} \varphi_\nu z^\nu. \tag{19}$$

Nach dem Satz über implizite Funktionen ist $\phi(z)$ wieder konvergent, etwa für $|z| < \sigma$, und es gilt, auch hier wegen der absoluten Konvergenz

$$\varphi_\delta = Q_\nu(F^*_{\alpha\beta}(u)) = \sum c^*_{\delta\gamma_1\ldots\gamma_N}(u_1^0)^{\gamma_1}\ldots(u_N^0)^{\gamma_N}.$$

Wähle x_0 mit $o < x_0 < \sigma$ beliebig. Dann ist

$$\phi(x_0) = \sum_{\nu \geq 1} Q_\nu(F^*_{\alpha\beta}(u^0))x_0^\nu$$

eine konvergente Reihe mit nichtnegativen Gliedern, und daher nach dem Umordnungssatz

$$\phi(x_0) = \sum_{\nu,\gamma} c^*_{\nu\gamma_1\ldots\gamma_N}(u_1^0)^{\gamma_1}\ldots(u_N^0)^{\gamma_N}x_0^\nu,$$

weil die gleichen Umordnungen wie bei den formalen Reihen möglich sind. Da

$$|c_{\delta\gamma_1\ldots\gamma_N}| \leq c^*_{\delta\gamma_1\ldots\gamma_N},$$

so ist für alle komplexen $v_1, \ldots v_N$, z_0 mit $|v_\gamma| < |u_\gamma^0| = \eta_0$, $|z_0| < \sigma$:

$$w(z_0, v) = \sum_{\delta,\gamma} c_{\delta\gamma_1\ldots\gamma_N} v_1^{\gamma_1}\ldots v_N^{\gamma_N} z_0^\delta$$

absolut konvergent. Da hier $\eta_0 > 0$ beliebig war und σ (in Abhängigkeit von η_0) passend gewählt werden konnte, ist damit der Satz bewiesen.
Wir kommen nun zur Anwendung dieses Satzes auf die holomorphen Integrale der Differentialgleichungen (1). Eine solche Differentialgleichung ist, wenn $m = 1$, zu (13) äquivalent; wenn $m > 1$ so ist sie äquivalent mit (14). Bei (13) und (14) tritt die Bedingung $ord_z\omega(z) \geq 1$ hinzu. Im Falle $m = 1$ ist $\rho \in \mathbb{C}$ beliebig wählbar, hingegen, wenn $m > 1$, $t \in \mathbb{C}$ beliebig, und ρ ist eine der endlich vielen Bestimmungen von $\left(\frac{\gamma}{\delta}\right)^{\frac{1}{m-1}}$. Wenden wir uns (14) zu. Da t nur im Koeffizienten von z^{m-1} auftritt, so gilt offensichtlich: Ist $\eta > 0$ beliebig, so gibt es ein $\delta_0(\eta)$, sodaß alle in (14) auftretenden Reihen in z und ω für $|t| < \eta$, $|z|$, $|\omega| < \delta_0(\eta)$ konvergieren, in diesem Fall ist sogar δ_0 unabhängig von η wählbar. Somit ist Satz 2 anwendbar, und es ergibt sich, daß die Lösung $\omega(t, z)$ mit $\omega(t, 0) = 0$ für $|t| < \eta$, $|z| < \delta(\eta)$ holomorph sind, wobei $\eta > 0$ beliebig, $\delta(\eta) > 0$ in Abhängigkeit von η gewählt werden kann.
Sodann sei $m = 1$, d.h. (1) ist zu (13) äquivalent. Hier tritt der Parameter $\rho \in \mathbb{C}$ stets in Potenzen $(\rho z)^\nu (\nu \geq 1)$ auf. Ist daher $\eta > 0$ beliebig gegeben, so gibt es ein $\delta_0(\eta) > 0$, sodaß die in (13) auftretenden

Reihen in z und ω für $|\rho| < \eta$, $|z|$, $|\omega| < \delta(\eta)$ konvergieren, wobei im allgemeinen $\delta_0(\eta)$ von η abhängen, ja für $\eta \to \infty$ gegen 0 gehen wird. Satz 2 ist daher auch hier anwendbar, und es folgt, daß die Integrale $\omega(\rho, z)$ mit $\omega(0, \rho) \equiv 0$ für $|\rho| < \eta$, $|z| < \delta(\eta)$ mit einem geeigneten $\delta(\eta) > 0$ holomorph sind.

Zusammenfassend ergibt sich

Satz 3. a) *Hat eine Differentialgleichung* (1) *mit* $m > 1$ *Integrale der Form*

$$w(z, c_m) = \rho z + \ldots + c_m z^m + \ldots,$$

so gilt (bei festem $\rho \in \mathbb{C}$*): Zu jedem* $\eta > 0$ *gibt ein* $\delta(\eta) > 0$*, sodaß die Abbildung*

$$(z, c_m) \;\to\; w(z, c_m)$$

für $|z| < \delta(\eta)$, $|c_m| < \eta$ *holomorph ist.*

b) *Hat eine Differentialgleichung* (1) *mit* $m = 1$ *Integrale der Form*

$$w(z, \rho) = \rho z + c_2 z^2 + \ldots \quad (\rho \in \mathbb{C}),$$

so gilt: Zu jedem $\eta > 0$ *gibt es ein* $\delta(\eta) > 0$*, sodaß die Abbildung*

$$(z, \rho) \mapsto w(z, \rho)$$

für $|z| < \delta(\eta)$, $|\rho| < \eta$ *holomorph ist.*

Speziell gilt dieser Satz also für die Differentialgleichungen von Aczél-Jabotinsky, die in [1], [2], [3], [4], [5] untersucht und auf die Gruppe der biholomorphen Transformationen einer autonomen Differentialgleichung in sich sowie auf Familien formaler und konvergenter vertauschbarer Potenzreihen angewendet wurden. Die formalen Untersuchungen in [1] lassen sich ebenfalls auf die impliziten Probleme (13) und (14) zurückführen, da wir diese Beziehungen hier mit rein algebraischen Überlegungen hergeleitet haben.

Analoge Resultate zu den Sätzen 1 und 3 gelten, wie in [2] und [5] ausgeführt wurde, auch für diejenigen Briot-Bouquetschen Differentialgleichungen, welche ein Kontinuum von Integralen besitzen, die, wie es L. Bieberbach ausdrückt, holomorph in die Singularität (0,0) einmünden.

4. Holomorphe Fortsetzung von Integralen in die Singularität der Differentialgleichung

Die in $z = 0$ holomorphen Integrale von (1) oder der vorhin genannten Briot-Bouquetschen Differentialgleichungen können für $z_0 \neq 0$, aber in hinreichend kleiner Umgebung von 0, auch nach dem Existenz- und Eindeutigkeitssatz von Cauchy durch ihren Anfangswert $y(z_0)$ festgelegt werden, da die rechte Seite in $(z_0, y(z_0))$ holomorph ist. Man kann nun umgekehrt fragen, für welche Paare (z_0, y_0), $(z_0 \neq 0)$ an denen die rechte Seite von (1) holomorph ist, gemäß dem Satz von Cauchy Integrale $y(z; z_0, y_0)$ mit $y(z_0; z_0, y_0) = y_0$ bestimmt werden, die sich nach $z = 0$ holomorph fortsetzen lassen. Der Satz 3 über die Abhängigkeit vom Parameter c_m (bzw. ρ) liefert das Resultat, daß dies, für hinreichend kleines z_0, bei einer sehr großen, nämlich nichtleeren offenen Menge von Anfangswerten y_0 der Fall ist. Es gilt

Satz 4. *Gegeben sei eine Differentialgleichung* (1) *unter den Voraussetzungen von Satz 1. Dann gilt für alle $z_0 \neq 0$, die hinreichend nahe an $z = 0$ liegen: Es gibt ein Gebiet G_{z_0}, sodaß für alle $y_0 \in G_{z_0}$ die durch $y(z_0; z_0, y_0) = y_0$ gemäß dem Satz von Cauchy festgelegten Integrale $y(z; z_0, y_0)$ von* (1) *als Lösung von* (1) *holomorph nach $z = 0$ fortsetzbar sind. Es gibt für $m > 1$ genau ein c_m bzw. für $m = 1$ genau ein ρ, sodaß in der Bezeichnungsweise von Satz 3*

$$y(z; z_0, y_0) = w(z, c_m) \text{ für } m > 1$$

bzw.

$$y(z, z_0, y_0) = w(z, \rho) \text{ für } m = 1.$$

Beweis. Wir betrachten hier der Kürze halber nur den Fall $m > 1$, da $m = 1$ ganz ähnlich untersucht werden kann. Es sei $\eta > 0$ beliebig und $\delta(\eta) > 0$ gemäß Satz 3 gewählt, es sei $z_0 \neq 0$ und $|z_0| < \delta(\eta)$; ferner $|c_m| < \eta$. Nach Satz 3 ist die Abbildung

$$\Theta_{z_0} : c_m \;\rightarrow\; y_0 = \rho z_0 + \ldots + c_m z_0^m + \sum_{\nu \geq m+1} Q_\nu(c_m) z_0^\nu$$

für $|c_m| < \eta$ holomorph. Sie ist nicht konstant. Denn wäre für u, v mit $|u|, |v| < \eta$ und $u \neq v \Theta_{z_0}(u) = \Theta_{z_0}(v)$, so hätten die Integrale $w(z, u)$

und $w(z,v)$ von (1) an z_0 den gleichen Wert $w(z_0,u) = w(z_0,v) = y_0$, es ist aber $w(z,u) \neq w(z,v)$, da die Taylorreihen dieser Funktionen bei z^m verschiedene Koeffizienten u und v haben. Andererseits existiert nach dem Satz vom Cauchy genau ein in z_0 holomorphes Integral $y(z,z_0,y_0)$ von (1) mit dem Anfangswert y_0; also ist in einer Umgebung von z_0 $w(z,u) = w(z,v) = y(z;z_0,y_0)$, dann nach dem Identitätssatz und durch holomorphe Fortsetzung auch $w(z,u) = w(z,v)$ in Umgebung von $z = 0$, im Widerspruch zu $u \neq v$; also ist Θ_{z_0} injektiv und nicht konstant. Nach dem Satz von der Gebietstreue ist also $\Theta_{z_0}(\{c_m|\ |c_m| < \eta\})$ ein Gebiet G_{z_0}. Für alle $y_1 \in G_{z_0}$ gibt es genau ein v_1, für das $|v_1| < \eta$ und $\Theta_{z_0}(v_1) = y_1$, d.h.

$$y_1 = \rho z_0 + \ldots + v_1 z_0^m + \sum_{\nu \geq m+1} Q_\nu(v_1) z_0^\nu .$$

$w(z,v_1) = \rho z + \ldots + v_1 z^m + \sum_{\nu \geq m+1} Q_\nu(v_1) z^\nu$ ist ein in $z = 0$ holomorphes Integral von (1) mit $w(z_0,v_1) = y_1$. Das lokale Integral $y(z;z_0,y_1)$ stimmt in Umgebung von z_0 mit $w(z,v_1)$ überein, und ist daher nach $z = 0$ als Lösung von (1) holomorph fortsetzbar. Damit ist Satz 4 bewiesen.

Satz 4 gilt auch für Briot-Bouquetsche Differentialgleichungen mit unendlich vielen in die Singularität mündenden Integralen.

Literatur

[1] REICH, L.: On a differential equation arising in iteration theory in rings of formal power series in one variable. In: Iteration Theory and its Functional Equations (Eds. R. Liedl, L. Reich, Gy. Targonski), *Lecture Notes in Math.*, **1163**, 135 – 148, Springer-Verlag Berlin 1986.

[2] REICH, L.: Holomorphe Lösungen der Differentialgleichung von E. Jabotinsky. *Sitzungsberichte der Österr. Akademie der Wissenschaften, Mathem.-Naturw. Klasse, Abt. 11,* **145**, 157 – 166 (1986).

[3] REICH, L.: On families of commuting formal power series. In: Selected Topics on Functional Equations, *Berichte der Mathematisch-statistischen Sektion in der Forschungsgesellschaft Joanneum-Graz, Bericht* **294**, 1 – 18 (1988).

[4] PERKO, R.: Some remarks on a generalization of the Jabotinsky equation. In: Selected Topics on Functional equations, *Berichte der Mathematisch-statistischen Sektion in der Forschungsgesellschaft Joanneum-Graz, Bericht* **293**, 1 – 10 (1988).

[5] REICH, L.: Die Differentialgleichungen von Aczél-Jabotinsky, von Briot-Bouquet und maximale Familien konvergenter vertauschbarer Potenzreihen. In: Ausgewählte Beiträge zur Komplexen Analysis (Hsg. C. Withalm) Akademie-Verlag Berlin (In Vorbereitung).

[6] ACZÉL, J. and GRONAU, D.: Iteration, Translation, Commuting and Differential Equations. In: Selected topics in Functional Equations, *Berichte der Mathematisch-statistischen Sektion in der Forschungsgesellschaft Joanneum-Graz, Bericht* . **285**, 1 – 6 (1988).

[7] ACZÉL, J. and GRONAU, D.: Some differential equations related to iteration theory. *Can. J. Math.* **40** (1988), 695 – 717.

[8] TARGONSKI, Gy.: New directions and open problems in iteration theory. *Berichte der Mathematisch-statistischen Sektion im Forschungszentrum Graz,* **229** (1984).

Mathematica Pannonica
1/1 (1990), 49 – 56

AN EMBEDDING THEOREM FOR FREE ASSOCIATIVE ALGEBRAS

P.M. **COHN**

Department of Mathematics, University College, London WC1E 6BT, Gower Street, England.

For Hiroyuki Tachikava on his 60th birthday.

Received September 1988

AMS Subject Classification: 16 A 06

Keywords: Free algebra, inert embedding, factorization, universal field of fractions, derivation, specialization.

Abstract: Embeddings of a free algebras of countable rank in free algebras of rank two are studied, which possess special properties, such as inertia or honesty. Their existence has the consequence that the embedding can be extended to one of their universal fields of fractions.

1. Introduction

It is well known, and easily verified, that a free (associative) algebra F of countable rank can be embedded in a free algebra G of rank 2; thus in $k < x, y >$ the elements $y^n x$ $(n = 0, 1, \ldots)$ freely generate a free subalgebra. But frequently one needs embeddings with special properties, and here the above example is usually insufficient. Thus one may want embeddings $F \to G$ with one or more of the following

properties:

1. 1-inert embeddings. This means that if $c \in F$ has a factorization $c = ab$ in G, then there exists a unit u in G such that $au, u^{-1}b \in F$ (for simplicity we have here identified F with its image in G.)

2. More generally, an n-inert embedding is the case where the matrix ring $\mathcal{M}_n(F)$ is 1-inertly embedded in $\mathcal{M}_n(G)$.

3. Honest embeddings. Their definition will be given below in §2; it amounts to requiring the universal field of fractions of F to be embedded in that of G.

The existence of 1-inert embeddings of F in G was conjectured by G.M. Bergmann in [1] and later proved (though not published) by him (cf. Th. 4.5.3, p. 217 of [5]). Our aim in this note is to construct an honest embedding of F in G, in §2, and to give an illustration, in §3. Our construction also provides an example of a 1-inert embedding which is simpler than that in [5]; whether it is n-inert for $n > 1$ remains open.

2. An honest embedding

Throughout, all rings are associative with 1, which is preserved by homomorphisms and inherited by subrings. Fields need not be comutative; sometimes the prefix 'skew' is used for emphasis.

A matrix C over a ring R is said to be *full* if it is square, say $n \times n$ and cannot be written in the form $C = PQ$, where P is $n \times r$, Q is $r \times n$ and $r < n$. A homomorphism of rings is called *honest* if it keeps full matrices full. Since every non-zero element, as 1×1 matrix, is full, an honest homomorphism is necessarily injective. To give an example, in the embedding $F \to G$ described in §1, with $z_n \mapsto y^n x$ say,

$$\begin{pmatrix} z_0 & z_1 \\ z_2 & z_3 \end{pmatrix} \text{ maps to } \begin{pmatrix} x & yx \\ y^2x & y^3x \end{pmatrix} = \begin{pmatrix} 1 \\ y^2 \end{pmatrix} (x \ yx),$$

and this shows that the mapping is not honest.

The importance of full matrices is that certain classes of rings such as semifirs have the property that each can be embedded in a skew field over which every full matrix from the ring becomes invertible. This

is called the *universal field of fractions* of the ring and for a ring R it will be denoted by $U(R)$ (cf. [5], Ch. 7 for details). Thus if we have a homomorphism of semifirs $\beta : F \to G$, we can form a commutative square as shown precisely when β is honest, and one method of establishing that β is honest is to prove the existence of such a commutative square.

$$\begin{tikzcd} F \arrow[r, "\beta"] \arrow[d] & G \arrow[d] \\ U(F) \arrow[r, "\beta^*"'] & U(G) \end{tikzcd}$$

Let D be a skew field and K a subfield. By the *tensor-D-ring* over K on a set X we understand the ring generated by D (as ring) and X, with the defining relations

$$\alpha x = x\alpha \quad \text{for all} \quad x \in X,\ \alpha \in K. \tag{1}$$

This ring will be denoted by $D_K < X >$; when $D = K$, it reduces to the free K-ring $K < X >$ (called a *free K-algebra* when K is commutative), and the free D-ring over K may be expressed as a free product (coproduct)

$$D_K < X > = D_K^* \, K < X > .$$

Frequently it is assumed that K is contained in the centre of D, but we shall not make this assumption here. We must then bear in mind that we cannot substitute arbitrary elements of D for X (because the relations (1) might then be violated), but we have to restrict X to values in the centralizer of K.

We shall write $F = K < Z >$, where $Z = \{z_0, z_1, \ldots\}$, $G = k < x, y >$; as indicated above, to find an honest embedding of F in G we only need an embedding of $U(F)$ in $U(G)$. Such an embedding was described in [2] (cf. [3], p. 120), but under that mapping the image of F was not confined to G. It was obtained by finding an automorphism permuting Z transitively, and realizing this automorphism by conjugation in G. We shall find that the same purpose can be achieved by a derivation, and this time the image of F stays within G. We first describe the derivation.

Lemma 2.1. *The free K-ring $K < Z >$, where $Z = \{z_0, z_1, \ldots\}$, has a derivation δ over K such that $z_i^\delta = z_{i+1}$, and δ extends to a unique derivation of $U(F)$.*

Proof. The mapping

$$\delta : z_i \mapsto z_{i+1} \quad (i = 0, 1, \ldots) \tag{2}$$

extends to a unique derivation of F because F is free (Prop. 1 of 3.3, p. 67 of [4]). Thus we have a derivation δ of F satisfying (2). We can write this as a homomorphism form F to $M_2(F)$:

$$\Delta : a \mapsto \begin{pmatrix} a & a^\delta \\ 0 & a \end{pmatrix}; \tag{3}$$

it induces a homomorphism from $\mathcal{M}_n(F)$ to $\mathcal{M}_{2n}(F)$ such that every full matrix over F maps to an invertible matrix over $U(F)$. For if A is full over F, then it is invertible over $U(F)$, hence

$$\begin{pmatrix} A & A^\delta \\ 0 & A \end{pmatrix} \quad \text{has the invers} \quad \begin{pmatrix} A^{-1} & -A^{-1}A^\delta A^{-1} \\ 0 & A^{-1} \end{pmatrix}.$$

Therefore the homomorphism Δ can be extended to a unique homomorphism from $U(F)$ to $\mathcal{M}_2(U(F))$, again denoted by Δ. Clearly it has again the form (3) and the (1,2)-entry is a derivation of $U(F)$ extending δ; it is unique because the extension of Δ was unique.
We can now obtain the desired embedding by realizing δ as an inner derivation. As usual we write $[a, b] = ab - ba$.

Theorem 2.2. *Let $G = K < x, y >$, $F = K < Z >$, where $Z = \{z_0, z_1, \ldots\}$ and K is a skew field. Then there is an embedding $\beta_0 : F \mapsto G$ defined by*

$$z_0 \mapsto y, \; z_1 \mapsto [y, x], \; z_2 \mapsto [[y, x], x], \; \ldots \tag{4}$$

If the image of F is denoted by F_0, then $G = \oplus_{n=0}^{\infty} F_0 x^n$. Moreover, the embedding is 1-inert and honest.

Proof. Let δ be the derivation of F defined as in Lemma 2.1, (1), and denote by H the skew polynomial ring $H = F[x; 1, \delta]$ with the commutation rule

$$ax = xa + a^\delta \quad \text{for all} \quad a \in F. \tag{5}$$

Then we have by (2) and (5), $z_{i+1} = z_i^\delta = z_i x - x z_i = [z_i, x]$. We claim that H is the free K-ring on x, z_0. For it is clearly generated by x and z_0 over K; to show that x, z_0 are free generators, we establish a homomorphism $\beta : H \to G$ such that $x \mapsto x$, $z_0 \mapsto y$. We begin by defining $\beta : Z \to G$ by

$$\beta : z_n \mapsto [\ldots[y, x], \ldots, x] \text{ with } n \text{ factors } x.$$

Since F is free on Z, this mapping extends to a homomorphism $\beta_0 : F \to G$. Moreover, we have $z_n^\delta \beta_0 = z_{n+1}\beta_0 = [\ldots[y, x], \ldots, x] = [z_n\beta_0, x]$ (where there are $n+1$ factors x). Hence if δ_x is the inner derivation defined by x in G, we have $\delta\beta_0 = \beta_0\delta_x$. Now the defining relations of H in terms of F are just the equations (5), which may be written $\delta = \delta_x$. Hence on H we have $\delta_x\beta_0 = \beta_0\delta_x$; thus the defining relations of H are preserved by β_0 and so β_0 may be extended to a homomorphism β of H into G. Since G is free on x, y, this shows H to be free on x, z_0, as claimed. Moreover, we see that β is surjective, hence it is an isomorphism between H and G, and the structure of H as skew polynomial ring over F shows that $G = \oplus_{n=0}^{\infty} F_0 x^n$, where F_0 is the image of F under β_0.

We now repeat the construction with $U(F)$ instead of F, thus we form the ring $U(F)[x; 1, \delta]$; this is justified by the fact that δ is defined on $U(F)$. This gives us a skew polynomial ring over a field, and we can form its field of fractions $L = U(F)\ (x; 1, \delta)$. Since $H = F[x; 1, \delta]$ is generated over K by x, z_0, it follows that L is generated by x, z_0 over K. Now the homomorphism $F \to G$ extends to a specialization from $U(F)$ to $U(G)$ and this extends to a specialization of L as H-field to $U(G)$ (cf. Ch. 7 of [5]). But G is free on x, y, so the specialization must be an isomorphism, by the universality of $U(G)$, and we find that $L \cong U(G)$. This provides an embedding of $U(F)$ in $U(G)$; in particular, any full matrix over F is invertible over $U(F)$, hence also over $U(G)$ and so is full over G. Thus we have shown that the mapping $\beta_0 : F \to G$ is honest.

It remains to show that β_0 is 1-inert. Given $c \in F$, suppose that in H we have $c = ab$, $a, b \in H$. We can write $a = a_0 x^r + \ldots$, $b = b_0 x^s + \ldots$, where $a_0, b_0 \in F$ and dots denote terms of lower degree in x. Then $c = ab = a_0 b_0 x^{r+s} + \ldots$; by uniqueness, $r + s = 0$, hence $r = s = 0$ and $a, b \in F$. This shows F to be 1-inert in H, hence the mapping β_0 is 1-inert, as we wished to show.

We can extend the scope of this result as follows.

Proposition 2.3. *Let F, G be semifirs that are K-rings, with an honest embedding $\lambda : F \to G$, and let D be a skew field containing K. Then the induced embedding $D_K^* F \to D_K^* G$ is honest.*

Proof. Our aim will be to show that $\lambda : F \to G$ induces an embedding $U(D^*F) \to U(D^*G)$. We begin by showing that

$$U(D^*F) \cong U(D^*U(F)). \tag{6}$$

On the left we have the universal field of fractions of the semifir D^*F. On the right we have a field generated by the subring D^*F, hence a specialization of the left-hand side. It is a proper specialization precisely if some full matrix over D^*F is not invertible over the right-hand side. But $D^*U(F)$ is a localization of D^*F, so the embedding $D^*F \to D^*U(F)$ is honest, and every full matrix over D^*F is full over $D^*U(F)$, hence invertible over the right-hand side of (6). Hence the specialization is improper, i.e. (6) is an isomorphism.

It now remains to show that there is a natural embedding of $U(D^*U(F))$ in $U(D^*U(G))$. Let us write $U(F) = L$, $U(G) = M$; we have an embedding $L \to M$ and we shall show that the embedding

$$D^*L \to D^*M \tag{7}$$

is honest; this will complete the proof.

Write $U_1 = U(D^*L)$; this is a field containing D and L and we have a natural homomorphism

$$D_K^* M \to U_{1L}^* M, \tag{8}$$

which reduces to the indentity on D and M. Moreover, it is an epimorphism, since the right-hand side is contained in $U(U_1^* M)$, which is generated, as a field, by D and M. Let A be a full matrix over D^*L; then A is invertible over $U_1 = U(D^*L)$, hence in the mapping (8) it must have come from some full matrix over D^*M, and this shows (7) to be an honest homomorphism, and it completes the proof.

Applying the result with $F = K < Z >, G = K < x, y >$, we obtain the first assertion of

Corollary 2.4. *The embedding $\lambda : D_K < Z > \to D_K < x, y >$ induced by the homomorhism of Th. 2.2 is honest and 1-inert.*

Now 1-inertia follows essentially as in Th. 2.2. If $c \in D_K < Z >$ satisfies $c\lambda = ab$, write $a = a_0 + a_1 + \ldots + a_r$, $b = b_0 + b_1 + \ldots + b_s$ where a_i, b_i are the terms of degree i in x, when these elements are expressed in terms of $x, z_i\lambda (= z_i\beta)$. Then $c = a_0 b_0 + \ldots + a_r b_s$ and we have a contradiction, unless $r = s = 0$; but this leads to a factorization of c in $D_K < Z >$, and it proves λ to be 1-inert.

3. An Example

As an example of a full matrix over $D_K < Z >$ which is also used elsewhere (cf. [6]) we consider the following $n \times n$ matrix $C = (c_{ij})$ suggested by G.M. Bergmann (for use in [6]):

$$c_{ij} = z_{n+j} d z_i, \quad \text{where} \quad d \in D, \ d \neq 0. \tag{1}$$

Our object is to show that this matrix C is full. Let us define, for any $m \times n$ matrix , its *inner rank* or simply *rank* rkA, as the least number r such that A can be written in the form

$$A = PQ, \quad \text{where } P \ \ m \times r \ \text{ and } \ Q \ \text{ is } \ r \times n.$$

We also recall from [5], p.253 the law of nullity: If $UV = 0$, where U is $m \times r$ and V is $r \times n$, then $rk\ U + rk\ V \leq r$.

We shall use induction on n to prove that the matrix C given by (1) is full over $R = D_K < Z >$. If this is not so, then its inner rank r is less than n and we have

$$C = PQ, \quad \text{where } P \ \text{ is } \ n \times r \ \text{ and } \ Q \ \text{ is } \ r \times n. \tag{2}$$

Let us write $Q = (Q_1, Q')$, where Q_1 is the first column of Q and similarly put $C = (C_1, C')$, so that

$$C_1 = PQ_1, \tag{3}$$

$$C' = PQ'. \tag{4}$$

If we omit the first row and column from C, the resulting matrix is full by the induction hypothesis; hence C' has inner rank n-1 and it follows from (4) that $r = n - 1$. Now consider the homomorphism of $D_K < Z >$ obtained by mapping $z_{n+1} \mapsto 0$ and leaving the other variables unchanged. Denoting images under this homomorphism by a bar, we have, by (1), (3), (4)

$$\bar{C}_1 = \bar{P}\bar{Q}_1 = 0, \qquad \bar{C}' = \bar{P}\bar{Q}' = C'.$$

Since C' has inner rank n-1, it follows that $\bar{P}$ has inner rank n-1. By the law of nullity

$$rk\ \bar{P} + rk\ \bar{Q}_1 \leq n - 1,$$

hence $\bar{Q}_1$ has rank 0, i.e. $\bar{Q}_1 = 0$. This means that the elements in the first column of Q lie in the ideal generated by z_{n+1}; similarly the elements of the jth column of Q lie in the ideal generated by z_{n+j} and by a symmetric argument the elements in the ith row of P lie in the ideal generated by z_i. Hence in the product PQ, in any term of degree 2, there is a factor z_i on the left of a factor z_{n+j} but in C these factors are in the opposite order, and so we have a contradiction. This proves C to be full, as claimed. Now we can apply Cor. 2.4 to deduce that under the embedding of $D_F < Z >$ into $D_K < x, y >$, C maps to a full matrix.

References

[1] BERGMANN, G.M.: Commuting elements in free algebras, and related topics in ring theory, Thesis (Harvard University 1967).

[2] COHN, P.M.: The free product of skew fields, *J. Austral. Math. Soc.* **16** (1973), 300-308.

[3] COHN, P.M.: Skew field constructions, *LMS Lecture Notes* **27**, Cambridge University Press, Cambridge 1977.

[4] COHN, P.M.: Algebra Vol. 2, J. Wiley + Sons, Chichester 1977.

[5] COHN, P.M.: Free rings and their relations, Second Edition, *LMS Monographs No.* **19**, Academic Press (London and Orlando 1985).

[6] COHN, P.M.: The specialization lemma for firs, to appear.

Mathematica Pannonica
1/1 (1990), 57 – 71

DREIDIMENSIONALE LIE GRUPPEN UND EBENE KINEMATIK

Manfred **Husty**

Institut für Mathematik und Angewandte Geometrie, Montanuniversität, A-8700 Leoben, Österreich.

Péter **Nagy**

József Attila Tudományegyetem, Bolyai János Intézet, 6720 Szeged, Hongrie.

Received August 1989

AMS Subject Classification: 53 A 17, 22 E 05, 22 E 50

Keywords: Plane cinematics, Lie groups

Abstract: In several papers A. KARGER introduced the notion of moving and fixed directing cones of one parameter motions. By KARGER these cones act in the Lie Algebras belonging to certain Lie Groups. In classical cinematics these Lie Groups act as transformation groups on manifolds (e.g. the plane, the three-space). In this paper we investigate the relation between the concepts of A. KARGER and classical cinematics.

In [8] hat A. KARGER erstmals über die Zusammenhänge zwischen den klassischen ebenen Kinematiken und der Theorie der Lie-Gruppen berichtet. Wir greifen diese Thematik wieder auf und versuchen einige neue Einsichten in die Beziehungen zwischen den, zu den verschiedenen Gruppen gehörenden Lie-Algebren und den ebenen Bewegungsvorgängen zu geben. Hierbei wird besonderes Augenmerk auf

die Gemeinsamkeiten der angesprochenen Liegruppen gelegt, auch weil diese Gruppen in jüngster Zeit vom Standpunkt der Liegruppentheorie (vgl. z.B. MILNOR [10]) einer einheitlichen Behandlung unterzogen wurden.

1. Grundlagen

Schauplatz für die kinematischen Bewegungsvorgänge ist die projektive Ebene P^2, die wir uns jeweils mit den entsprechenden projektiven Metriken im CAYLEY-KLEINschen Sinn ausgestattet denken können (vgl. YAGLOM [16]); P^2 wird dadurch zur *elliptischen Ebene* S^2, zur *hyperbolischen Ebene* H^2, zur *euklidischen Ebene* E^2, zur *pseudoeuklidischen Ebene* $E^{1,1}$ oder auch zur *isotropen Ebene* I^2.
Die zu diesen Metriken gehörenden, zusammenhängenden Komponenten der Isometriegruppen werden in üblicher Weise mit $SO(3)$, $SO(2,1)$, $E(2)$, $E(1,1)$ und $I(2)$[1] bezeichnet. Weiters schreiben wir für die, mit den Tangentialräumen in den Einheitselementen der entsprechenden Isometriegruppen identifizierten Lie Algebren: $\mathfrak{o}(3)$, $\mathfrak{o}(2,1)$, $\mathfrak{e}(2)$, $\mathfrak{e}(1,1)$ bzw. $\mathfrak{i}(2)$. Wird keine Aussage über die metrische Struktur benötigt, so bezeichnen wir die Liegruppe mit G und die zugehörige Liealgebra mit g.

Für die kinematischen Betrachtungen denken wir uns die projektiv metrischen Ebenen in zwei Exemplaren ausgebildet, von denen ein Exemplar als fest anzusehen ist, während das andere gegenüber dem ersten bewegt wird. Wir bezeichnen das feste Exemplar als Rastebene $\bar{P}^2(\bar{S}^2, \bar{H}^2, \ldots)$ und nennen das bewegte Exemplar $P^2(S^2, H^2, \ldots)$ Gangebene. Die Elemente der jeweiligen Liegruppe wirken daher so, daß sie eine Abbildung der Gangebene auf die Rastebene vermitteln. Ein kinematischer Bewegungsvorgang ist nun eine Folge von Abbildungen der Gangebenen auf die Rastebenen und kann daher mit einer Kurve $g(t)$ auf der Lie-Gruppe G identifiziert werden.

Es ist aus der klassischen Kinematik bekannt, daß ein ebener Bewe-

[1] Diese Isometriegruppe ist auch unter dem Namen "Heisenberggruppe" bekannt (vgl. z. B. MILNOR [10])

gungsvorgang durch bogenlängentreues Abrollen von zwei Polkurven erzeugt werden kann. Die Polkurven bestehen dabei aus den jeweiligen Fixpunkten der Momentanbewegungen. Hierbei kann es bei Ausnahmesituationen (Fernpolstellungen in der euklidischen Kinematik (vgl. z.B. B. BLASCHKE [1], S.12f), uneigentliche Fixpunkte in der hyperbolischen Kinematik (vgl. FRANK [2], TÖLKE [15]), isotrope Kinematik (vgl. RÖSCHEL [13], HUSTY [5]) zu gewissen Schwierigkeiten kommen.

Im Gegensatz zu diesem klassischen Konzept hat A. KARGER unter Verwendung der Liegruppentheorie den Bewegungsvorgang mit der Gruppenkurve $g(t)$ identifiziert und den Begriff der Polkegel definiert durch:

$$\begin{aligned} &\text{Bewegter Polkegel:=} && R(t) = g^{-1}\dot{g}; \\ &\text{Fester Polkegel:=} && \bar{R}(t) = \dot{g}g^{-1}. \end{aligned} \tag{1.1}$$

Der Zusammenhang zwischen den beiden Polkegeln wird durch die Abbildung Ad_g geregelt, die entsprechende Vektoren der Polkegel aufeinander abbildet,

$$\bar{R}(t) = g\ R(t)\ g^{-1} = Ad_g\ R(t) \tag{1.2}$$

und damit ein gewisses Analogon zum Rollen der beiden Polkurven in den entsprechenden projektiv metrischen Ebenen bildet.

Ziel dieser Arbeit ist es nun den Zusammenhang zwischen den beiden Konzepten zu durchleuchten.

2. Die metrische Struktur der Lie Algebren $o(3)$, $o(2,1)$, $e(2)$, $e(1,1)$, $i(2)$.

Im folgenden Teil der Arbeit werden Standardkenntnisse über die metrische Struktur der angesprochenen Lie Algebren zusammengestellt und dann wird eine Abbildung zwischen Vektorräumen erklärt.

2.1. $o(3)$ bzw. $o(2,1)$:

Die Killingform κ ist positiv definit (bzw. indefinit). Die Lie Algebra Elemente X können als lineare Operatoren auf einem dreidimensionalen

bzw. pseudoeuklidischen Vektorraum V^3 betrachtet werden. Auf V^3 ist ein Kreuzprodukt definiert durch

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{vmatrix} \varepsilon i & j & k \\ \alpha & \beta & \gamma \\ x & y & z \end{vmatrix} \tag{2.1}$$

$\varepsilon = 1$ (bzw. -1)und i, j, k als Basisvektoren in V^3. Wir definieren eine Abbildung:

Definition 1: $\Phi : X \to \vec{x} \in V^3$

$$\begin{pmatrix} 0 & -\varepsilon\gamma & \varepsilon\beta \\ \gamma & 0 & -\alpha \\ -\beta & \alpha & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}.$$

Diese Abbildung ordnet einem Lie Algebra Element X einen Vektor aus V^3 derart zu, daß gilt:

$$X\vec{y} = \vec{x} \times \vec{y} \quad \forall \vec{y} \in V^3. \tag{2.2}$$

Weiters läßt sich durch eine einfache Rechnung zeigen, daß die Gleichung

$$\vec{x} \times \vec{y} = [X, Y] \tag{2.3}$$

erfüllt ist und zwischen der Killingform $\kappa(X, Y)$ in $o(3)$ bzw. $o(2,1)$ und dem inneren Produkt $<,>$ in V^3 die Beziehung

$$\kappa(X, Y) = -2 < x, y > \tag{2.4}$$

besteht. Unmittelbar ergibt sich dann das leicht zu beweisende

Lemma 1. *(KARGER) Zwischen den Darstellungen von $SO(3)$ auf V^3 und Ad $SO(3)$ auf $o(3)$ ist folgendes kommutative Diagramm gültig:*

$$\begin{tikzcd} V^3 \arrow[r, "g \in SO(3)"] \arrow[d, "\Phi"'] & V^3 \arrow[d, "\Phi"] \\ o(3) \arrow[r, "Ad\ g"] & o(3) \end{tikzcd}$$

Korollar. *Das Lemma ist auch gültig wenn $SO(3)$ durch $SO(2,1)$ und $\mathfrak{o}(3)$ durch $\mathfrak{o}(2,1)$ ersetzt wird.*

2.2. $e(2)$ bzw. $e(1,1)$:

Wir denken uns in P^2 ein homogenes Koordinatensystem eingeführt und erfassen die Punkte durch projektive Koordinaten $(x_0 : x_1 : x_2) \neq = (0 : 0 : 0)$. Dann kann eine Transformation aus $E(2)$ bzw. $E(1,1)$ durch Matrizen der Form

$$(2.5a,b) \qquad \begin{pmatrix} \cos a & \sin a & b \\ -\sin a & \cos a & c \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{bzw.} \qquad \begin{pmatrix} \operatorname{ch} a & \operatorname{sh} a & b \\ \operatorname{sh} a & \operatorname{ch} a & c \\ 0 & 0 & 1 \end{pmatrix}$$

beschrieben werden. Weiters sei durch

(2.6)

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = E_1 \qquad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = E_2 \qquad \begin{pmatrix} 0 & 1 & 0 \\ -\epsilon & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = E_3$$

je eine Basis in den Lie Algebren $e(2)$ $(\epsilon = 1)$ bzw. $e(1,1)$ $(\epsilon = -1)$ bestimmt. Wie man leicht berechnet, gelten für die Basisvektoren E_i die folgenden Beziehungen

$$(2.7) \qquad [E_1, E_2] = 0, \qquad [E_1, E_3] = \epsilon E_2, \qquad [E_2, E_3] = -E_1.$$

Die durch die Killingform $\kappa(X,Y) = Tr\ Ad\ X Ad\ Y$ induzierte invariante quadratische Form hat die Gestalt $\kappa(X,Y) = -2\epsilon x^3 y^3$, wobei x^i die Koordinaten eines Lie Algebra Elements in der angegebenen Basis E_i bedeuten. Damit folgt unmittelbar, daß die Gleichung $\kappa(X,X) = 0$ eine zweidimensionale, kommutative Subalgebra g' mit der Basis E_1, E_2 bestimmt. Es gilt das

Lemma 2. *Auf der durch $\kappa = 0$ bestimmten Subalgebra g' gibt es eine, bis auf einen Faktor eindeutig bestimmte, gegenüber $Ad\ E(2)$ bzw. $Ad\ E(1,1)$ invariante, definite (indefinite) quadratische Form $\lambda(X,Y)$.*

Beweis. Wenn $\lambda(X,Y)$ eine invariante quadratische Form sein soll, so muß gelten

$$(2.8) \qquad \lambda([Z,X],Y) + \lambda(X,[Z,Y]) = 0.$$

Wegen der Kommutativität von g' ist diese Beziehung auf g' automatisch erfüllt, und wir haben diese Gleichung nur für $Z = E_3$ auszuwerten. Wir setzen $X = x^1E_1 + x^2E_2$, $Y = y^1E_1 + y^2E_2$ und erhalten durch Einsetzen in (2.8):

$$(2.9) \qquad -2\lambda(E_1,E_2)x^1y^1 + [\lambda(E_1,E_1) - \epsilon\lambda(E_2,E_2)](x^1y^2 + x^2y^1) + \\ +2\lambda(E_1,E_2)x^2y^2 = 0.$$

Daraus folgt:

$$(2.10) \qquad (E_1,E_1) = \epsilon\lambda(E_2,E_2) := A \qquad \text{bzw.} \qquad \lambda(E_1,E_2) = 0.$$

Insgesamt haben wir damit für die invariante quadratische Form die Gestalt:

$$(2.11) \qquad \lambda(X,Y) = A(x^1y^1 + \epsilon x^2y^2)$$

$\diamond$

Folgerung. Durch diese Konstruktion der invarianten quadratischen Formen erhält der Vektorraum $e(2)$ bzw. $e(1,1)$ eine gegenüber $Ad\ g$ invariante quasielliptische (bzw. indefinit quasielliptische) Struktur (vgl. BLASCHKE [1], S. 177 ff). Diese quasielliptische (indefinit quasielliptische) Struktur kann damit durch Links- und Rechtsschiebung auf die gesamte Lie-Gruppe ausgedehnt werden und bestimmt eine biinvariante, differentialgeometrische Struktur auf $E(2)$ bzw. $E(1,1)$.

Im folgenden Teil sollen nun die strukturellen Beziehungen zwischen den projektiv metrischen Ebenen E^2 bzw. $E^{1,1}$ und den entsprechenden - mit obiger Konstruktion metrischen - Lie Algebren untersucht werden.

Dazu beweisen wir den

Satz 1. *Zwischen V^3 und $e(2)(e(1,1))$ existiert ein eindeutig bestimmter linearer quasielliptischer Isomorphismus Φ, so daß das folgende Diagramm kommutiert:*

$$\begin{array}{ccc} V^3 & \xrightarrow{\quad g \in E(2)\ (E(1,1)) \quad} & V^3 \\ \Big\downarrow \Phi & & \Big\downarrow \Phi \\ e(2), e(1,1) & \xrightarrow{\quad Ad\ g \quad} & e(2), e(1,1) \end{array}$$

Beweis.
A) $e(2)$:
Zur Bestimmung der Abbildung untersuchen wir die Bilder der Basisvektoren $\vec{e}_i$:

$$\begin{aligned} \Phi(\vec{e}_1) &= a_1 E_1 + a_2 E_2 + a_3 E_3 \\ \Phi(\vec{e}_2) &= b_1 E_1 + b_2 E_2 + b_3 E_3 \\ \Phi(\vec{e}_3) &= c_1 E_1 + c_2 E_2 + c_3 E_3. \end{aligned} \tag{2.13}$$

Wegen der Invarianz des isotropen Unterraumes $\{\vec{e}_1, \vec{e}_2\} \to \{E_1, E_2\}$ folgt vorerst $a_3 = b_3 = 0$. Weiters gilt mit einer Matrix A von der Form (2.5,a) für den Basisvektor $\vec{e}_1 = (1,0,0)^t$ (t=Transposition):

$$\Phi(A\vec{e}_1) = (a_1 \cos a - b_1 \sin a)E_1 + (a_2 \cos a - b_2 \sin a)E_2 \tag{2.14}$$

und

$$Ad\ A\ \Phi(\vec{e}_1) = A\ \Phi\ (e_1)\ A^{-1} = (a_1 \cos a + a_2 \sin a)E_1 + \tag{2.15}$$

$$+(a_2 \cos a - a_1 \sin a)E_2.$$

Da die Ausdrücke (2.14) und (2.15) für alle Matrizen A aus $E(2)$ gleich sein müssen, folgt unmittelbar

$$a_2 = -b_1 \qquad \text{und} \qquad b_2 = a_1. \tag{2.16}$$

Eine analoge Rechnung für den Basisvektor $\vec{e}_3 = (0,1,0)^t$ liefert keine neue Bedingungen. Setzt man jedoch $\vec{e}_3 = (0,0,1)^t$ ein, so erhalten wir:

$$b_1 = -c_3, \quad a_1 = b_2 = c_1 = c_2 = 0. \tag{2.17}$$

Weiters folgt aus der Isomorphiebedingung $c_3 = 1$. Damit erhalten wir für die Abbildung der Basisvektoren

$$\begin{aligned} \Phi(\vec{e}_1) &= E_2 \\ \Phi(\vec{e}_2) &= -E_1 \qquad \text{, d.h.} \\ \Phi(\vec{e}_3) &= E_3 \end{aligned} \tag{2.18}$$

ein Vektor $(x, y, z)^t$ aus V^3 wird durch Φ auf ein Lie Algebra Element von der Form

$$\begin{pmatrix} 0 & z & -y \\ -z & 0 & x \\ 0 & 0 & 0 \end{pmatrix} \tag{2.19}$$

abgebildet. Die Umkehrung ist unmittelbar einsichtig.

B) $e(1,1)$
Analoge Rechnungen wie im Beweisteil A) ergeben für die Gruppe $E(1,1)$ als Abbildungsgleichungen der Basisvektoren

$$\begin{aligned} \Phi(\vec{e}_1) &= -E_2 \\ \Phi(\vec{e}_2) &= -E_1 \\ \Phi(\vec{e}_3) &= E_3 \end{aligned} \tag{2.20}$$

und ein Vektor aus V^3 wird auf ein Lie Algebra Element der Form

$$\begin{pmatrix} 0 & z & -y \\ z & 0 & -x \\ 0 & 0 & 0 \end{pmatrix} \tag{2.21}$$

abgebildet.

$\diamond$

2.3. $i(2)$:
In der durch homogene Koordinaten $(x_0 : x_1 : x_2)$ koordinatisierten Ebene können isotrope Isometrien durch Matrizen

$$\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix} \tag{2.22}$$

der Form beschrieben werden. Wir zeichnen in der Lie Algebra $i(2)$ durch

(2.23)

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = E_1 \qquad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = E_2 \qquad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = E_3$$

eine Basis aus und erhalten damit folgende Kommutationsrelationen:

$$[E_1, E_2] = [E_2, E_3] = 0, \qquad [E_1, E_3] = -E_2. \tag{2.24}$$

Wie eine kurze Rechnung zeigt, verschwindet die Killingform auf ganz $i(2)$, und es ist daher notwendig nach invarianten quadratischen Formen zu suchen. Wir beweisen folgendes

Lemma 3. *Jede quadratische Form, die in der Basis E_1, E_2, E_3 die Darstellung $\lambda(X,Y) = \lambda(E_1,E_1)x^1y^1 + \lambda(E_1,E_3)(x^1y^3 + x^3y^1) + \lambda(E_3, E_3)x^3y^3$ besitzt, ist gegenüber Ad $I(2)$ invariant.*

Beweis. Für die Invarianz einer quadratischen Form ist es notwendig und hinreichend, daß (2.8) erfüllt ist. Da diese Relation für E_2 wegen (2.25) automatisch erfüllt ist, können wir setzen $Z = z^1E_1 + z^3E_3$, $X = x^1E_1 + x^2E_2 + x^3E_3$ sowie $Y = y^1E_1 + y^2E_2 + y^3E_3$ und wir erhalten durch Einsetzen in (2.8):

$$\begin{aligned} \lambda(E_1,E_2)[(x^1z^3 - z^1x^3)y^1 + x^1(z^3y^1 - z^1y^3)] + \lambda(E_2,E_2)\cdot \\ \cdot[y^2(x^1z^3 - x^3z^1) + x^2(y^1z^3 - z^1y^3)] + \lambda(E_2,E_3)\cdot \\ \cdot[y^2(x^1z^3 - z^1x^3) + x^2(y^1z^3 - z^1y^3)] = 0. \end{aligned} \tag{2.25}$$

Diese Identität ist für alle x^i, y^i, z^i nur dann erfüllt, wenn $\lambda(E_1,E_2) = \lambda(E_2,E_2) = \lambda(E_2,E_3) = 0$ gilt. Alle anderen $\lambda(E_i,E_j)$ $i,j = 1,3$ können beliebig gewählt werden und damit ist die Behauptung gezeigt.

◊

Auf $g' = [g,g] = \{tE_2\}$ verschwindet die quadratische Form $\lambda(X,Y)$. Wir müssen daher eine invariante quadratische Ersatzform bestimmen. Es gilt:

Lemma 4. *Die Form $\mu(X,Y) = hx^2y^2$ ($h \in \mathbb{R}$) auf g' ist invariant gegen Ad g.*

Beweis. Da E_2 im Zentrum der Lie Algebra liegt, ist die Relation (2.8) automatisch erfüllt und weil die Form nur auf g' definiert ist, folgt die Behauptung unmittelbar.

◊

Satz 2. *Die - bis auf eine Konstante - eindeutig bestimmte, lineare Abbildung Φ zwischen dem durch die projektiv metrische Ebene I^2 bestimmten Vektorraum V^3 und der Lie Algebra $i(2)$ mit dem kommutierenden Diagramm*

$$\begin{array}{ccc} V^3 & \xrightarrow{\quad g \in I(2) \quad} & V^3 \\ \Big\downarrow \Phi & & \Big\downarrow \Phi \\ i(2) & \xrightarrow{\quad Ad\ g \quad} & i(2) \end{array}$$

ist gegeben durch:

$$\Phi : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -y & x & 0. \end{pmatrix}$$

Beweis. Wir nehmen an, daß Φ durch die Bilder der Basisvektoren angegeben ist und bestimmen die Koeffizienten der Abbildungsgleichungen

$$\begin{aligned} \Phi(\vec{e}_1) &= a_1 E_1 + a_2 E_2 + a_3 E_3 \\ \Phi(\vec{e}_2) &= b_1 E_1 + b_2 E_2 + b_3 E_3 \\ \Phi(\vec{e}_3) &= c_1 E_1 + c_2 E_2 + c_3 E_3. \end{aligned} \tag{2.26}$$

Nun gilt für den Basisvektor $\vec{e}_1 = (1,0,0)^t$ mit einer Matrix A aus $I(2)$

$$\Phi(A\vec{e}_1) = (a_1 + ab_1 + bc_3)E_3. \tag{2.27}$$

Andererseits erhalten wir

$$Ad\ A\ \Phi(\vec{e}_1) = a_1 E_1 + (ca_1 + a_2 - aa_3)E_2 + a_3 E_3, \tag{2.28}$$

woraus die Bedingungen

$$a_1 = b_1 = b_3 = c_1 = c_2 = c_3 = 0 \quad \text{und} \quad b_2 = -a_3 \tag{2.29}$$

fließen. Die weiteren Basisvektoren liefern keine neuen Bedingungen, so daß wir für die Abbildungsgleichungen von Φ finden

$$\begin{aligned} \Phi e_1 &= a_3 E_3 \\ \Phi e_2 &= -a_3 E_2 \\ \Phi e_3 &= 0. \end{aligned} \tag{2.30}$$

Umgekehrt ist unmittelbar einsichtig, daß die so gegebene Abbildung das Kommutativitätsdiagramm erfüllt.

$\diamond$

Bemerkung:

1. Wie man sofort sieht, ist der Kern der adjungierten Darstellung der isotropen Bewegungsgruppe $I(2)$ die Untergruppe der isotropen Schiebungen (vgl. SACHS [14]), denn $Ad\ I(2)$ hat in der Basis E_1, E_2, E_3 die Darstellung

$$\begin{pmatrix} 1 & 0 & 0 \\ -c & 1 & -a \\ 0 & 0 & 1. \end{pmatrix} \tag{2.31}$$

Als Folge davon müssen alle isotropen Vektoren aus I^2 bei Φ auf den Nullvektor abgebildet werden.

2. Der Bildraum von Φ besteht aus den Linearkombinationen $\lambda E_2 + \mu E_3$ von E_2 und E_3. Für $\mu = 0$ bestimmt das Lie Algebra Element eine einparametrige Schar isotroper Schiebungen

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda t & 0 & 1 \end{pmatrix}, \tag{2.32}$$

bzw. für $\mu \neq 0$ eine einparametrige isotrope Scherungsgruppe

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda t & \mu t & 1 \end{pmatrix}. \tag{2.33}$$

3. Urbilder der Lie Algebra Elemente bei Φ:

3.1. $i(2)$:
Durch $g/\mathbb{R}$ wird aus dem Vektorraum der Lie Algebra eine projektive Ebene $\widetilde{P}^2$. Wir können damit die Lie Algebra Elemente als Punkte von $\widetilde{P}^2$ ansprechen und die Abbildung Φ vermittelt eine singuläre Kollineation zwischen $\widetilde{P}^2$ und der projektiv metrischen Ebene I^2. Die Bildpunkte dieser Kollineation liegen auf einer Geraden $\widetilde{f}$ wobei der, auf dieser Geraden gelegene, zum Vektor E_2 gehörige, Punkt $\widetilde{F}$ ausgezeichnet ist. Man kann sich daher in natürlicher Weise diese projektive Ebene $\widetilde{P}^2$ als isotrope Ebene metrisiert denken. Man hat dafür den Punkt $\widetilde{F}$ und die Gerade $\widetilde{f}$ als Absolutgebilde dieser isotropen Ebene zu wählen.

Wir werden nun die Urbilder der Bildpunkte näher untersuchen und erhalten den

Satz 3. *Die Urbilder eines Punktes $\widetilde{P} \in \widetilde{f}$ sind für $\widetilde{P} = \widetilde{F}$ die Ferngerade f in I^2 und für $\widetilde{P} \neq \widetilde{F}$ eine isotrope Gerade i; dabei ist $\Phi^{-1}(\widetilde{F}) = f$ die Fixpunktmenge der einparametrigen isotropen Schiebungsgruppe* $\exp(t\vec{f})$, *wo $\widetilde{F} = \mathbb{R}\vec{f}$ ist, und $\Phi^{-1}(\widetilde{P}) = i(\widetilde{P} \neq \widetilde{F})$ ist die Fixpunktmenge der einparametrigen Scherungsgruppe* $\exp(t\vec{p})$ *mit $\widetilde{P} = \mathbb{R}\vec{p}$.*

Beweis. Ein Punkt der Geraden $\widetilde{f}$ kann durch einen Vektor $\vec{p} = \lambda E_2 + \mu E_3$ dargestellt werden. Sein Urbild bei Φ^{-1} ist für $\mu = 0$ die projektive Gerade f bzw. für $\mu \neq 0$ die projektive Gerade $(\mu : -\lambda : t)$, wobei t einen laufenden Parameter bezeichnet. Andererseits ist $epx(t\vec{p})$ gegeben durch (2.33) und es erweist sich als Fixpunktmenge dieser einparametrigen Untergruppen bei $\mu = 0$ die Ferngerade f und bei $\mu = 0$ die projektive Gerade $(\mu : -\lambda : t)$.

$\diamond$

Bemerkung. Für alle anderen Punkte $\widetilde{P}$ der projektiven Ebene $\widetilde{P}^2(\widetilde{P} \notin f, \vec{p} \neq \lambda E_2)$ ist die zugehörige Fixpunktmenge in I^2 der bei jeder isotropen Bewegung fest bleibende absolute Punkt der projektiv metrischen Ebene I^2. Wenn $\vec{p} = \lambda E_2$ gilt, dann ist die zugehörige Fixpunktmenge die Ferngerade $f \in I^2$ (die zugehörigen einparametrigen Untergruppen sind nichtisotrope Schiebungen).

3.2. $o(3)$ bzw. $o(2,1)$

Es hat sich gezeigt, daß die Abbildung Φ im isotropen Fall sehr nützlich war, um die Beziehungen zwischen den Vektoren der Lie Algebra und den Fixpunktskonfigurationen der projektiv metrischen Ebene I^2 zu durchleuchten. Im Falle der klassischen dreidimensionalen Drehungsgruppen sind die Verhältnisse wesentlich einfacher: Wenn wir wieder $g/\mathbb{R}$ mit $\widetilde{P}^2$ bezeichnen, so folgt unmittelbar aus Lemma 1 und dem nachfolgenden Korollar, daß $\widetilde{P}^2$ dieselbe metrische Struktur hat wie die jeweilige projektiv metrische Ausgangsebene. Die Abbildung Φ ist eine Kollineation, welche die jeweilige metrische Struktur erhält.

A) Im Falle der elliptischen Bewegungen $SO(3)$ kann ein Punkt $\widetilde{P} \in \widetilde{P}^2$ über Φ mit dem Fixpunkt der einparametrigen Drehungsgruppe $\exp(t\vec{p})$ identifiziert werden.

B) Im Falle der hyperbolischen Bewegungen $SO(2,1)$ ordnet Φ einem Vektor $\vec{x}$ aus V^3 das Lie Algebra Element X nach Definition 1 zu. Wir bestimmen nun die zu $\exp[t\Phi(\vec{x})]$ gehörige Fixpunktkonfiguration. Die Fixpunkte der einparametrigen Untergruppen werden durch die Eigenvektoren von $\Phi(\vec{x})$ bestimmt. Man berechnet leicht als Eigenwerte von $\Phi(\vec{x})$: $\lambda_1 = 0$, $\lambda_{2/3} = \pm\sqrt{||\vec{x}||^2}$. Der zu $\lambda = 0$ gehörende Eigenvektor stimmt genau mit dem Ausgangsvektor überein. Um die weiteren Eigenvektoren zu untersuchen, können wir o.B.d.A. $||\vec{x}||^2 = 1$ wählen und $\vec{x}$ als den Basisvektor $(0:1:0)^t$ betrachten. Dann erhalten wir die Eigenwerte $\lambda_{2/3} = \pm 1$ und die zugehörigen Eigenvektoren $(1:0:\pm 1)$ sind isotrop. Wir können daher sagen: Φ^{-1} ordnet einem Punkt $\widetilde{P}$ aus $\widetilde{P}^2$ den nicht auf dem absoluten Kegelschnitt gelegenen Fixpunkt von $\exp(t\vec{x})$ zu. Wenn dieser Fixpunkt nicht existiert, dann gibt es einen eindeutig bestimmten Fixpunkt auf dem absoluten Kegelschnitt. Dieser Punkt wird dann durch Φ^{-1} dem Punkt aus $\widetilde{P}^2$ zugeordnet.

3.3. $e(2)$ bzw. $e(1,1)$

A) Ein analoger Prozeß liefert für die euklidischen Bewegungen: Es sei $(x,y,z)^t$ ein Punkt aus E^2. Das durch Φ zugeordnete Lie Algebra Element $\vec{p}$ hat nun einen reellen Eigenwert $\lambda = 0$. Ist $z \neq 0$, dann ist der entsprechende Eigenvektor $(x,y,z)^t$ und der zu diesem Vektor gehörende Punkt ist der einzige eigentliche Fixpunkt der einparametrigen Drehungsgruppe $\exp(t\vec{p})$. Ist jedoch $z = 0$, dann ist $\exp(t\vec{p})$ eine

einparametrige Schiebungsgruppe in Richtung $(-y : x : 0)$ und der Fixpunkt $(x : y : 0)$ ist, als der zur Schiebrichtung orthogonale Fixpunkt ausgezeichnet.

B) Im pseudoeuklidischen Fall sind die isotropen Fernpunkte bei jeder Bewegung fix und ein analoger Prozeß wie im euklidischen Fall liefert: Φ^{-1} ordnet einem Punkt $\widetilde{P}$ aus $\widetilde{P}^2$ den nicht auf der absoluten Geraden gelegenen Fixpunkt von $\exp(t\vec{p})$ zu. Wenn dieser Punkt nicht existiert, dann ist $\exp(t\vec{p})$ eine einparametrige Schiebungsgruppe in Richtung $(-y : -x : 0)$ und $(x : y : 0)$ ist der zur Schiebrichtung orthogonale Fernpunkt.

Literaturverzeichnis

[1] BLASCHKE, W.: Ebene Kinematik, Oldenburg München, 1956.

[2] FRANK, H.: Zur ebenen hyperbolischen Kinematik, *Elem. d. Math.* **26** (1971), 121-131.

[3] GIERING, O.: Vorlesungen über höhere Geometrie, Vieweg-Verlag, Braunschweig Wiesbaden, 1982.

[4] GARNIER, R.: Cours de Cinématique, Tome III, Paris, 1951.

[5] HUSTY, M.: Zur Kinematik winkeltreuer Ähnlichkeiten der isotropen Ebene, Habilschrift Leoben (1988), 1 – 94.

[6] HUSTY, M. und NAGY, P.: Über die Blaschkesche kovariante Ableitung und die kinematische Abbildung, Note di Mathematica, Lecce (1990), (im Druck)..

[7] JACOBSON, N.: Lie Algebras, Wiley & Sons, New York-London, 1962.

[8] KARGER, A.: Lieovy Groupy a Kinematická Geometrie v Roviné, *Časopis pro pěstováni mat.* **93** (1968), 186-200.

[9] KARGER, A. und NOVAK, J.: Space Kinematics and Lie Groups, Gordon and Breach, New York-London-Paris-Montreux-Tokio, 1985.

[10] MILNOR, J.: Curvatures of left invariant metrics on Lie groups, *Advances in Math.* **21** (1975), 293-329.

[11] MÜLLER, H.R.: Kinematik, *Göschen Bd.* **584/584a**, Berlin, 1963.

[12] RÖSCHEL, O.: Zur Kinematik der isotropen Ebene I, *Journ. of Geom.* **21** (1983), 146-156.

[13] RÖSCHEL, O.: Zur Kinematik der isotropen Ebene II, *Journ. of Geom.* **24** (1985), 112-122.

[14] SACHS, H.: Ebene isotrope Geometrie, Vieweg-Verlag, Braunschweig Wiesbaden, 1987.

[15] TÖLKE, J.: Kinematik der Hyperbolischen Ebene I-III, *J. reine angew. Math.* **265** (1974), 145-153; **267** (1974), 143-150; **273** (1975), 99-108.

[16] ЯГЛОМ, И.М.: Проективные мероопределения на плоскости и комплексные числа, Труды семинара по векторному и тензорному анализу при МГУ **7** (1949), 274 – 318.

Mathematica Pannonica
1/1 (1990), 73 – 94

SEMIRINGS WITHOUT ZERO DIVISORS

U. **Hebisch**

Institut für Mathematik, Technische Universität, D-3392 Clausthal-Zellerfeld, Erzstraße 1, BRD

H.J. **Weinert**

Institut für Mathematik, Technische Universität, D-3392 Clausthal-Zellerfeld, Erzstraße 1, BRD

Received August 1989

AMS Subject Classification: 16 A 78, 16 A 56

Keywords: Semirings without zero divisors, multiplicatively cancellative semirings, semirings with zero sums, semiring extensions

Abstract: The main subject of this paper are semirings with an absorbing zero o which are zero divisor free (ZDF), but which have zero sums. We show that each such semiring S contains a greatest subring $R \supset \{o\}$ (in the usual meaning, even if $(S,+)$ is not commutative) which has no α-fiers and of course no zero divisors. Conversely, each ring R of this kind occurs as the greatest subring of some semiring S as above, where S itself is not a ring. In this situation, various structural results on S, R and $U = S \setminus R \neq \emptyset$ are proved, e.g. that each $s \neq o$ in S has infinite additive order. We also deal with semirings which are multiplicatively left or right cancellative or even both (briefly MLC, MRC and MC). For a semiring S with zero, each of these assumptions implies that S is ZDF, but not conversely. We show that each semiring with zero sums is MLC iff it is MRC and thus MC. Moreover, such a semiring S has an absorbing zero, $(S,+)$ is commutative and cancellative, and S is embeddable into a ring which is also MC. Finally, we prove by examples that all our results on proper ZDF semirings S with an absorbing

zero and zero sums are fairly complete. In particular, each ring R satisfying the necessary conditions above can be embedded into such a semiring S which is MC as well as into one which is merely ZDF.

1. Introduction

A (2,2)-algebra $S = (S, +, \cdot)$ is called a *semiring* iff $S(, +)$ and $(S, \cdot)$ are arbitrary semigroups, which are connected by ring-like distributivity, i.e. $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ hold for all $a, b, c \in S$. This rather general concept has been investigated in several papers (e.g. [3], [5], [10], [14], [16], [18], [20], [21]), whereas all semirings occuring in various applications in the last two decades, in particular in different branches of Theoretical Computer Science (cf. e.g. [1], [2], [7], [12], [13], [15]), have commutative addition. Moreover, they mostly have a zero, which is then always assumed to be absorbing (cf. Section 2). Our main purpose is of course to add some knowledge on semirings of the latter kind, but we do not assume that $(S, +)$ is commutative for all results which are in fact independend of this assumption. We also say explicitly if a zero is assumed to be absorbing. We further call a semiring S *non-trivial* iff it contains at least two elements, and *proper* iff S is not a ring.

Let S be a semiring with a zero o. Then S is called *zero divisor free* (briefly ZDF) iff $ab = o$ implies $a = o$ or $b = o$ for all $a, b \in S$. Now either $a + b = o$ implies $a = b = o$ for all $a, b \in S$, or there is at least one pair $(r, s) \in S^* \times S^*$ for $S^* = S \setminus \{o\}$ satisfying $r + s = o$, called a *zero sum* of S. In the first case, for a non-trivial ZDF semiring, $(S^*, +, \cdot)$ is a subsemiring of $(S, +, \cdot)$, and by Lemma 2.1 any semiring T occurs as such a subsemiring S^* of a ZDF semiring S. So the assumption that a semiring S is ZDF provides in this case no more results than those which concern any semiring S^*.

Therefore our interest is with the second case, and the subject of this paper are ZDF semirings which have zero sums, in particular those where the zero is absorbing. We prepare these investigation by some

general concepts and statements on semirings in Section 2.

Then we show that a ZDF semiring S with an absorbing zero o has zero sums iff it contains a nontrivial subring R' with o as zero, in the usual meaning that $(R', +)$ is commutative even if $(S, +)$ is not. Moreover, the greatest subring R of this kind consists of all elements $r, s, \ldots$ of S which occur in zero sums of S (cf. Thm. 3.3). Provided that S itself is not a ring, various structural results on S, R and $U = S \setminus R$ and their interrelation are obtained in Section 3. We only mention here that each element $s \neq o$ of S has infinite additive order, that R is a ring which has no α-fier for any $\alpha \in \mathbb{N}$ as defined in [6] (cf. Section 2 and Thm. 3.7), and that (under a rather general supplementary assumption) S is an Everett-Rédei semiring extension of R as introduced in [14] (cf. Suppl. 3.6).

In the following section we sharpen some of the above results, dealing with semirings which are multiplicatively left or right cancellative or even both (briefly MLC, MRC and MC, cf. Section 2). For a semiring S with a zero, each of these assumptions implies that S is ZDF, but not conversely, and there are even semirings which are e.g. MLC and do not contain any multiplicatively right cancellable element. So it is surprising that each semiring S with zero sums is MLC iff it is MRC and hence MC, which also yields that the zero of S is absorbing and that $(S, +)$ is cancellative and commutative (cf. Thm. 4.1). We further obtain that such a MC semiring S with zero sums is embeddable into a ring and the smallest ring $D(S)$ of this kind is also MC (cf. Thm. 4.4).

In Section 5 we recall that a well-known construction to embed rings in those with an identity transfers similarly to semirings (cf. Prop. 5.1). Using this and results of Section 3, we show that each ZDF semiring S with commutative addition and an absorbing zero is embeddable into a semiring of the same kind which has an identity (cf. Thm. 5.3 and Remark 5.5). Moreover, the constructions mentioned above are also basic for Thm. 5.6 and Constr. 6.1, which leads to Thm. 6.2. The purpose of these theorems will be explained in the following.

Concerning the completeness of our structural statements on proper ZDF or MC semirings S with an absorbing zero and zero sums, the greatest subring R of such a semiring S is of course also ZDF and, as

mentioned above, has no α-fiers. Conversely, each ring R of this kind occurs as the greatest subring of such a semiring S. In fact, we give two general constructions for those embeddings. By the first (Thm. 5.6) we obtain, for each R, a semiring S which is even MC and has an identity, by the second (cf. Thm. 6.2 and Suppl. 6.3) we get semirings S which are ZDF, but neither MLC nor MRC. Moreover, these considerations and some other examples (Expl. 4.2 and 6.4, Remark 4.3) disprove various further conjectures on S, R and U and their interrelation, which have been suggested to us in the context of our investigations.

2. Preliminaries on Semirings

Let $S = (S, +, \cdot)$ be a semiring as defined in Section 1. If there exists a neutral element o of $(S, +)[e$ of $(S, \cdot)]$, it is called the *zero* [the *identity*] of the semiring S. An element $t \in S$ is said to be ***absorbing*** iff $at = ta = t$ holds for all $a \in S$. It is well-known that the zero o of a semiring S need not be absorbing and may even coincide with the identity of S (cf. e.g. [20]). Conditions ensuring that the zero o of a semiring S is absorbing are that $(S, +)$ has no further idempotents or that $(S, +)$ is left or right cancellative. If $(S, +)$ has the last two properties, we call S *additively cancellative* (briefly AC).

Since we consider semirings as (2,2)-algebras $(S, +, \cdot)$, concepts as ***subsemirings, homomorphisms*** etc. are clear and refer merely to the two binary operations, also for semirings which have a zero or an identity. For subsets A, B of S, we define

$$A + B = \{a + b | a \in A,\ b \in B\} \quad \text{and} \quad AB = \{ab | a \in A,\ b \in B\}.$$

In particular, $A \neq \emptyset$ is called an *ideal* of S if $A + A \subseteq A$, $SA \subseteq A$ and $AS \subseteq A$ are satisfied (cf. e.g. [3]).

A semiring S is called *multiplicatively left cancellative* (MLC) iff all $a \in S$ or, for a semiring S with a zero o, all $a \neq o$ of S are left cancellable in $(S, \cdot)$. In the second case this implies (cf. [23]): either the zero o is also left cancellable in $(S, \cdot)$, or o is (from both sides) absorbing. Hence a non-trivial semiring S with an absorbing zero is MLC iff $(S^*, \cdot)$ is a left cancellative subsemigroup of $(S, \cdot)$.

The dual concept and statements for a *multiplicatively right cancellative* (MRC) semiring S are clear, and S is called *multiplicatively cancellative* (MC) iff it is MLC and MRC. Now assume that S has a zero. Then, as already mentioned in Section 1, each of these properties implies that S is ZDF, but not conversely (cf. [5]), contrasting the situation with rings for which all four properties are equivalent.

It is well-known that an absorbing zero can be adjoined to each semiring:

Lemma 2.1. *Let $T = (T, +, \cdot)$ be any semiring and o an element not contained in T. Extend the operations on T to those on $S = T \cup \{o\}$ by*

$$a + o = o + a = a \quad \textit{and} \quad a \cdot o = o \cdot a = o \quad \textit{for all} \quad a \in S.$$

Then $(S, +, \cdot)$ is a semiring with o as absorbing zero, which is ZDF, without zero sums, and contains $T = S^*$ as a subsemiring. Moreover: S has commutative addition or multiplication iff this holds for T; S is AC iff T is AC and has no zero; S is MLC iff T is MLC and has no absorbing zero or consists only of one element.

The semiring $(\mathbb{N}, +, \cdot)$ of positive integers is in a natural way a (left and right) operator domain for each additively commutative semiring $(S, +, \cdot)$ according to

$$\nu s = s\nu = \sum_{i=1}^{\nu} s \quad \text{for all} \quad \nu \in \mathbb{N},\ s \in S. \tag{2.1}$$

The obvious rules $\nu(s + r) = \nu s + \nu r$, $(\nu + \mu)s = \nu s + \mu s$, $(\nu\mu)s = \nu(\mu s)$ and $1s = s$ show that $(S, +)$ is a unitary left (and right) $\mathbb{N}$-semimodule, and one also has $\nu(sr) = (\nu s)r = s(\nu r)$. If S has an absorbing zero o, the semiring $(\mathbb{N}_0, +, \cdot)$ of non-negative integers is also such an operator domain if one extends (2.1) by $0s = o$. Moreover, the ring of integers $(\mathbb{Z}, +, \cdot)$ operates a corresponding way on each ring.

Generalizing a concept introduced for rings in [6], an element a of a semiring S is called an *α-fier* of S for some $\alpha \in \mathbb{N}$ iff

$$as = sa = \alpha s \quad \text{holds for all} \quad s \in S. \tag{2.2}$$

The original purpose of this concept was to describe the epimorphisms φ which occur in Remark 5.2 in the case of rings (cf. also [19]). In the semiring case the situation is similar, but more complicated. Here we need α-fiers in the context of Thm. 3.7.

Let S be a semiring with a zero o. An element $r \in S$ is called *additively invertible* (in S) iff $r + (-r) = (-r) + r = o$ holds for some $-r \in S$, which is then uniquely determined by r. Clearly, all elements of this kind form a subgroup $(R, +)$ of $(S, +)$ with o as neutral element.

Lemma 2.2. *Let S be a semiring with zero o and R the set of all additively invertible elements of S. Then R is an ideal of S iff o is absorbing. In this case, $(R, +, \cdot)$ is a subsemiring and additively a group, but the latter need not be commutative.*

Proof. In the trivial case $R = \{o\}$, clearly $SR = \{o\} = RS$ holds iff o is absorbing. For $R \supset \{o\}$, let R be an ideal of S and $a \in S$. Then $o + o = o$ implies $ao + ao = ao$, so that ao is an idempotent in the group $(R, +)$. This yields $ao = o$, and $oa = o$ follows in the same way. Conversely, let o be absorbing, $a \in S$ and $r \in R$. Then

$$r + (-r) = (-r) + r = o \text{ yields } ar + a(-r) = a(-r) + ar = o,$$

which proves $a(-r) = -(ar)$ and hence $ar \in R$. So we have $SR \subseteq R$ and correspondingly $RS \subseteq R$. For the last statement we note that there are various semirings $(R, +, \cdot)$ such that $(R, +)$ is a non-commutative group, also called additively not commutative rings (cf. [9] and [22], the latter also for more references). However, semirings $(R, +, \cdot)$ of this kind are never ZDF.

As a constrast to the situation in Thm. 3.3, we show that a semiring S may contain subrings $R_1, R_2, \ldots$ which have different zeroes $o_1, o_2, \ldots$, even if S is a ZDF semiring with an absorbing zero:

Example 2.3. Consider a distributive lattice $(L, \cup, \cap)$ as a semiring $(L, +, \cdot)$ and let $T = \{(r, l) | r \in R, l \in L\}$ be the semiring obtained as the direct product of a ring R with L. Then, for each $l_i \in L, T$ contains $R_i = \{(r, l_i) | r \in R\}$ as a subring isomorphic to R, and all corresponding zeroes (o, l_i) are distinct. By Lemma 2.1 one obtains from T a semiring S as claimed above.

3. ZDF Semirings with an Absorbing Zero

In this section we investigate the structure of semirings as indicated by the title. According to the introduction we have to assume that such a semiring has zero sums, since otherwise nothing can be said beyond Lemma 2.1. The following statement will be used several times.

Lemma 3.1. *Let S be any semiring and $a, b, s, r \in S$. If*

(3.1) $\quad$ *as is left and br is right cancellable in $(S, +)$,*

then $ar + bs = bs + ar$ holds.

Proof. Applying the distributive laws to $(a + b)(s + r)$ in both orders of succesion, we obtain

$$as + ar + bs + br = as + bs + ar + br,$$

which yields our statement by the assumptions on as and br.

Lemma 3.2. *Let S be a ZDF semiring with an absorbing zero o. Then*

(3.2) $\quad r + r' = o \;\; \textit{implies} \;\; r' + r = o \;\; \textit{for all} \;\; r, r' \in S.$

Proof. Since (3.2) is trivial for $r = o$, we assume $r \neq o$ and apply Lemma 3.1 for $a = b = r$ and $s = r'$. From $rr + rr' = o$ it follows that $as = rr'$ is left and $br = rr$ is right cancellable in $(S, +)$, which yields $rr + rr' = rr' + rr$. Now $r + r' = o$ implies $rr + rr' = o$. So we get $r(r' + r) = o$ for $r \neq o$, hence $r' + r = o$ as S is ZDF.

Theorem 3.3. *Let S be a ZDF semiring with an absorbing zero o. Then S has zero sums iff S contains a non-trivial subring with o as its zero. If this is the case,*

(3.3) $\quad R = \{r \in S | r + r' = o \;\; \textit{or} \;\; r' + r = o \;\; \textit{for some} \;\; r' \in S\}$

is the greatest subring of S, and even an ideal of S.

Now suppose additionally that S is a proper semiring with zero sums. Then $\{o\} \subset R \subset S$ holds for R as above and $U = S \setminus R$ satisfies

(3.4) $\quad U + S \subseteq U, \;\; S + U \subseteq U \;\; \textit{and hence} \;\; U + U \subseteq U,$

and each element $s \neq o$ of S has infinite additive order. Thus R and the subsemigroup $(U,+)$ of $(S,+)$ are infinite. Moreover, for any $s,t \in S$ and $\alpha \in \mathbb{N}$,

$$\alpha s + ts = o \quad or \quad \alpha s + st = o \quad imply \quad s = o. \tag{3.5}$$

Proof. If S contains any subring $R' \supset \{o\}$, clearly S has zero sums. Conversely, the latter implies $R \supset \{o\}$ for the set R defined by (3.3). Applying Lemma 3.2, we obtain that R consists of all additively invertible elements of S. Hence, by Lemma 2.2, R is a subsemiring and an ideal of S, and $(R,+)$ is a group. To show that $(R,+)$ is commutative, we consider the commutator $p+q+(-p)+(-q)$ for any $p,q \in R$. Again by Lemma 3.2, we have $qr + (-p)r = (-p)r + qr$ for some $r \neq o$ of R, since qr and $(-p)r$ are in R and hence cancellable in $(S,+)$. So we obtain $(p+q-p-q)r = pr + qr - pr - qr = pr - pr + qr - qr = o$ and thus $p+q-p-q = o$ since S is ZDF. So $(R,+)$ is commutative, hence $(R,+,\cdot)$ a ring and obviously the greatest subring of S which contains o. In fact, the latter restriction is superfluous since each subring of S contains o. This is clear if $R = S$ holds, which was not excluded so far, and will follow as a by-product from the following considerations for $R \neq S$.

Now we assume $U = S \setminus R \neq \emptyset$. Then $u + s \in U$ holds for all $u \in U$ and $s \in S$. Otherwise, $u + s = r \in R$ would yield $u + s + (-r) = o$ and thus the contradiction $u \in R$ by (3.3). The other statement of (3.4), $S + U \subseteq U$, follows in the same way. Next we show that each $s \neq o$ of S has infinite additive order (which also yields that any subring of S must have o as its zero). By way of contradiction, assume at first $\nu r = r + \cdots + r = o$ for some $r \neq o$ of R and some $\nu \in \mathbb{N}$. Then $(\nu r)u = r(\nu u) = o$ holds for any $u \in U$. Since S is ZDF, we get $\nu u = u + \cdots + u = o$ and thus the contradiction $u \in R$ by (3.3). Now assume that an element $u \in U$ has finite additive order, which only means that the set $\{\mu u | \mu \in \mathbb{N}\}$ is finite (and not necessarily $\nu u = o$ for some $\nu \in \mathbb{N}$). Then $\{(\mu u)r = \mu(ur) | \mu \in \mathbb{N}\}$ is also finite for any $r \neq o$ of R. Thus $ur \neq o$ of R would have finite additive order, which was already disproved.

For (3.5), assume by way of contradication that e.g. $\alpha s + ts = o$ holds for some $\alpha \in \mathbb{N}$ and $s \neq o$. This yields $u(\alpha s) + uts = o$ for each $u \in U$,

hence $(\alpha u + ut)s = o$ and $\alpha u + ut = o$ as S is ZDF. But the latter implies $u \in R$ by (3.3), a contradiction.

Corollary 3.4. *Let S be a proper finite ZDF semiring with an absorbing zero. Then S has no zero sums.*

Supplement 3.5. *Let S be a proper ZDF semiring with an absorbing zero o and zero sums, R its greatest subring and $U = S \setminus R$.*

a) *Assume $sa = sb$ or $as = bs$ for some $s \neq o$ and $a \neq b$ of S. Then $ra = rb$ and $ar = br$ hold for each $r \in R$, and a and b are in U.*

b) *For all $r \neq o$ of R we have $rR \cap rU = \emptyset$ and hence $rR \subset R$, and correspondingly for Rr and Ur.*

c) *Assume $s + a = s + b$ or $a + s = b + s$ for some s and $a \neq b$ of S. Then s, a and b are in U and $ra = rb$ and $ar = br$ hold for each $r \in R$.*

d) *One has $U + R = R + U = U$ where $u_1 + r = u_2 + r$ implies $u_1 = u_2$ and $u + r_1 = u + r_2$ implies $r_1 = r_2$, and correspondingly for $R + U$. Moreover, there is a subset $W \subset U$ such that each $u \in U$ has a unique presentation $w + r$ for some $w \in W$ and $r \in R$.*

Proof. a) From $sa = sb$ we obtain $sar = sbr$ or $s(ar + b(-r)) = o$ for each $r \in R$, which yields $ar = br$ as S is ZDF. The latter implies $ra = rb$ for each $r \in R$ in same way. Since a ZDF ring is also MC, at most one of a and b can be in R. We may assume $a = q \in R$ and $b \in U$. But then $rq = rb$ for some $r \neq o$ would imply $r(b - q) = o$, hence the contradiction $b = q$.

b) We have just proved that $rq = rb$ for $r \neq o$ of R and $q \in R$, $b \in U$ is impossible, which yields $rR \cap rU = \emptyset$. Note that $rR \subset R$ is also a consequence of (3.5).

c) From $s+a = s+b$ and $a \neq b$ it follows that $s \in U$ since each element of R is clearly cancellable in $(S, +)$. For the same reason, $rs + ra = rs + rb$ implies $ra = rb$ for each $r \in R$ by $rs \in R$. The rest follows from a).

d) The first part is a consequence of (3.4), $o \in R$ and c). Next we state that

$$a + p = b + q \quad \text{for some} \quad p, q \in R, \tag{3.6}$$

i.e. $a = b + r$ for some $r \in R$, defines clearly an equivalence $a \sim b$ on S for which R is one equivalence class. Each set W of representatives for all other classes obviously satisfies the last statement.

It is well known that, for each ideal R of a semiring S, (3.6) defines a congruence κ on $(S, +, \cdot)$ provided that $(S, +)$ is commutative (cf. [3], but observe [4]). The latter can be replaced by $u + R = R + u$ for all $u \in U = S \setminus R$. The converse question to construct all semirings S which contain R as an ideal such that the congruence class semiring S/κ (mostly denoted by S/R as in the ring case) is isomorphic to a given semiring has been settled by Rédei in [14]. The restriction to AC semirings in [14] is unessential. So we can state:

Supplement 3.6. *For S, R and U as in Suppl. 3.5, assume $u + R = R + u$ for all $u \in U$ and define $a\kappa b$ by (3.6). Then S is an Everett-Rédei semiring extension of the ring R by the congruence class semiring $S/\kappa = S/R$, a semiring with an absorbing zero, but without zero sums, which is ZDF iff U is a subsemiring of S.*

Although those extensions are hard to handle in general, we have used the theory given in [14] as a guide-line to obtain some of our examples, which, except Expl. 6.4, are all special cases of extensions according to Suppl. 3.6.

The main purpose of these examples given in Section 5 and 6 is to prove that our statements on S, R and U in Thm. 3.3 and Suppl. 3.5 are fairly complete concerning the general situation (but cf. Thm. 4.1). In particular, we shall see that the subsemigroup $(U, +)$ need not be a subsemiring of S (cf. Remark 4.3), that only the elements of R have to commute in $(S, +)$ (cf. Thm. 6.2 and Expl. 6.4), and that all violations of cancellativity left over by a), c) and d) of our supplement really may occur (cf. Thm. 6.2). Also our statements on R are complete according to the following:

Theorem 3.7. *Let R' be a non-trivial ring with zero o. Then R' is a subring of a proper ZDF semiring S such that o is the absorbing zero of S iff R' is ZDF and satisfies the condition*

C) $\alpha s + ts = o$ *implies* $s = o$ *for all* $s, t \in R'$ *and* $\alpha \in \mathbb{N}$.

The condition C) *can also be formulated with $\alpha s + st = o$ and is equi-*

valent to the fact that R' contains no α-fier for any $\alpha \in \mathbb{N}$. It implies that each $r \neq o$ of R' has infinite additive order.

Moreover, for each ring R' of this kind, S can be chosen in such a way that R' is its greatest subring.

Proof. If R' is contained in a semiring S as assumed above, S has zero sums. Hence the greatest subring R of S is clearly ZDF and satisfies C) and thus the same holds for $R' \subseteq R$.

The converse statement including the last one will be shown in two versions, namely in Thm. 5.6 (where S is even MC) and in Thm. 6.2 (where S is merely ZDF).

Concerning the remarks on C), we consider any ZDF ring R' and assume $\alpha s + ts = o$ for any $s \neq o$. This yields $\alpha r + rt = o$ for each $r \in R$, in particular $\alpha s + st = o$, and in turn $\alpha r + tr = o$. Hence $(-t)$ is an α-fier of R' as defined by (2.2), which conversely implies $\alpha s + ts = o$ even for each $s \in R'$. The last remark is clear, since an element $s \neq o$ of R' of finite additive order satisfies $\alpha s = o$ for some $\alpha \in \mathbb{N}$, which contradicts C).

4. Multiplicative Cancellativity

Let S be a proper ZDF semiring with an absorbing zero o and zero sums. Then, according to Suppl. 3.5 a), left and right cancellativity in $(S, \cdot)$ are closely connected. In particular, it is near by hand to ask whether there are semirings S as above which are not MLC or MRC. We have claimed that without proof in [12], Section 6, and we will show this by the following concrete Expl. 4.2 which can be checked directly (regardless that our constructions in Section 6 will provide lots of those examples as already indicated in the proof of Thm. 3.7). Before that we sharpen the situation by the following result:

Theorem 4.1. *Let S be semiring which has zero sums. Then S is MLC iff S is MRC, hence in turn iff S is MC.*

If this is the case, the zero o of S is absorbing and S is AC and additively

commutative.

Proof. Since all statements are true if S is a ring, we consider a proper semiring. We show at first that MLC implies MRC and the statement on o. So let S be MLC. Then, by a result of [23] cited in Section 2, its zero o is either also multiplicativeley left cancellable in $(S,\cdot)$ or absorbing. Assuming the former, we get from $o(a+a)=(o+o)a=oa$ that $a+a=a$ holds for each $a\in S$. But then $r+r'=o$ for any $r,r'\in S$ implies

$$r=r+o=r+r+r'=r+r'=o$$

and hence $r'=o$. This contradicts that S is assumed to have zero sums. So the zero o of S is absorbing and, since MLC yields ZDF, we can apply our results of Section 3. By way of contradiction, assume that S is not MRC. Then there are elements $s\neq o$ and $a\neq b$ of S satisfying $as=bs$. But this yields $ra=rb$ for all $r\in R\neq\{o\}$ by Suppl. 3.5 a), contradicting that S is MLC. Clearly, MRC implies MLC in the same way.

Now we assume that S is MC and that $s+a=s+b$ or $a+s=b+s$ hold for some $s\in S$ and $a\neq b$ of S. Then we obtain, by Suppl. 3.5 c), $ra=rb$ for all $r\in R\neq\{o\}$. This contradicts that S is MC and proves S to be AC. Hence Lemma 3.1 implies $ac+bc=bc+ac$ for all $a,b,c\in S$ which yields $a+b=b+a$ since S is MC.

Example 4.2. Let $x^0,x^1,x^2,\ldots$ be the elements of the free monoid $(X,\cdot)$ generated by x with x^0 as its identity, and $(H,\cdot)$ the semigroup obtained from $(X,\cdot)$ by adjoining a new identity $e\notin X$. Let

$$D=\{\sum_{i=0}^{n}\gamma_i x^i+\gamma e|\gamma_i,\gamma\in\mathbb{Z}\}$$

be the semigroup ring of $(H,\cdot)$ over the ring $\mathbb{Z}$ of integers. (In other words, D is obtained from the polynomial ring $\mathbb{Z}[x]$ by adjoining a new identity e, or D is the Dorroh-ring $Do(\mathbb{Z},\mathbb{Z}[x])$ in the sence of Section 5). Clearly, D is commutative. Now

$$S=\{\sum_{i=0}^{n}\gamma_i x^i+\gamma e|\gamma_0,\gamma\in\mathbb{N}_0,\ \gamma_i\in\mathbb{Z}\text{ for }i\geq 1\}$$

is a subsemiring of D with $o \in D$ as its absorbing zero. Further, S has zero sums and its greatest subring R consists of all polynomials of $\mathbb{Z}[x]$ satisfying $\gamma_0 = 0$. Moreover, $(1x)(1x^0) = (1x)(1e)$ shows that S is not MC. So it remains to prove that S is ZDF, which is easily checked in a straightforward way.

Remark 4.3. In Expl. 4.2, $U = S \setminus R$ consists of all elements of S satisfying $\gamma_0 + \gamma \neq 0$, and U is a subsemiring of S. But we can change the definition of S e.g. by $\gamma_1, \gamma_0, \gamma \in \mathbb{N}_0$, but also by $\gamma_1, \gamma_0 \in \mathbb{N}_0$ and $\gamma = 0$ (and, clearly, $\gamma_i \in \mathbb{Z}$ for $i \geq 2$). In both cases S remains a ZDF semiring with zero sums, where the greatest subring R consists now of all polynomials of $\mathbb{Z}[x]$ satisfying $\gamma_0 = \gamma_1 = 0$. Hence $U = S \setminus R$ contains in both cases the element $1x$, and $(1x)(1x) = 1x^2 \in R$ shows that U is not a subsemiring of S. Note that S is MC in this second variation of Expl. 4.2, but not in the first one.

Together with statements of Section 3, we obtain a further result from Thm. 4.1. Recall for this purpose that a semiring S is embeddable into a ring iff S is AC and additively commutative (where the former yields that the zero of S is absorbing, if there is one). If this is the case, there exists, unique up to isomorphisms, a smallest ring which contains S as a subsemiring. This ring is called the *difference ring* of S and denoted by $D(S)$, since it consists of all differences $a - b$ for $a, b \in S$, subject to elementary rules.

Now let S be a semiring such that $D(S)$ exists. If S is ZDF but not MRC or even MLC but not MRC, then clearly the properties ZDF or MLC of S do not transfer to $D(S)$. (Otherwise, since ZDF, MLC and MRC are equivalent for the ring $D(S)$, such a transfer would yield that S is also MRC. Cf. also Expl. 4.2 in this context.) But even if S has all these properties, i.e. if S is MC, its difference ring $D(S)$ need not be MC. E.g., consider the congruence class ring $T = \mathbb{Z}[x]/(x^2)$ of the polynomial ring $\mathbb{Z}[x]$ and let S consist of all classes which can be represented by some $\gamma_0 + \gamma_1 x$ for $\gamma_0 > 0$ and $\gamma_1 \geq 0$. Then one checks that S is a MC subsemiring of T, whereas $D(S) = T$ is clearly not MC (cf. [21], p. 221).

Theorem 4.4. *Let S be a semiring with zero sums which is MLC (or MRC). Then S is embeddable into a ring, and the smallest ring $D(S)$*

containing S is MC.

Proof. If S itself is a ring, there is nothing to prove. If S is a proper semiring, we apply Thm. 4.1. Hence S is AC and additively commutative, and so a subsemiring of its difference ring $D(S)$. It remains to show that the latter is ZDF and thus MC (which, according to the above counter-example, depends on some further assumption on S, in our case the existence of zero sums). By way of contradiction, assume $(a-b)(c-d)=o$ for some $a-b \neq o$ and $c-d \neq o$ of $D(S)$. Note that S satisfies all conditions such that it has a greatest subring $R \neq \{o\}$ according to Thm. 3.3. So we obtain $(ra-rb)(cr-dr)=o$ for some $r \neq o$ of R, where ra, rb, cr and dr are in R and hence also $ra-rb$ and $cr-dr$. Since R is MC we get that e.g. $ra-rb=o$ holds, which yields $a=b$ in S, hence the contradiction $a-b=o$.

5. Embedding into Semirings with an Identity

Considerations according to the title will also lead to all our constructions of ZDF semirings with zero sums. For this purpose we need explicitly the well-known result due to Dorroh (cf. [8]) that each ring R can be embedded into a ring with identity in the following way. One defines operations on the set $D=\mathbb{Z}\times R$ by

$$(\nu,s)+(\mu,t)=(\nu+\mu,s+t) \quad \text{and} \tag{5.1}$$

$$(\nu,s)\cdot(\mu,t)=(\nu\cdot\mu,\nu t+\mu s+s\cdot t), \tag{5.2}$$

where νt and μs are defined according to (2.1). Then $(D,+,\cdot)$ is a ring with $(1,o)=e$ as identity. By an obvious isomorphism, one can identify $(0,s)$ with s for each $s\in R$ so that R becomes a subring of D, which also yields the unique presentation

$$(\nu,s)=\nu(1,o)+(0,s)=\nu e+s \quad \text{for the elemtents of } D. \tag{5.3}$$

We call this ring $D=\mathbb{Z}e+R$ the *Dorroh-ring* of R and denote it by $Do(\mathbb{Z},R)$. It is universal in the sense that each ring $\mathbb{Z}e'+R$ generated by R and an identity e' is an R-epimorphic image of $Do(\mathbb{Z},R)$ (cf. [6],[19], and the corresponding Remark 5.2 for semirings).

It is also known that a semiring with non-commutative addition need not be embeddable into one with an identity (cf. [10], but observe [11]), whereas the above statements can be transfered to additively commutative semirings (cf. e.g. [16]):

Proposition 5.1. *Let S be a semiring with an absorbing zero o and commutative addition. Then the above construction applied to $D = \mathbb{N}_o \times S$ yields a proper additively commutative semiring $(D, +, \cdot)$ with $(0, o)$ as absorbing zero and $(1, o) = e$ as identity. Obviously, the subsemiring $\{(0, s) | s \in S\}$ of D is isomorphic to S and can be replaced by the latter, which yields $D = \mathbb{N}_0 e + S$ according to (5.3).*

We call this semiring the *Dorroh-semiring* of S and denote it by $Do(\mathbb{N}_0, S)$. Observe also that $Do(\mathbb{N}_0, S)$ is AC iff S is AC.

(Clearly, Prop. 5.1 applies also to an arbitrary additively commutative semiring T via Lemma 2.1).

Remark 5.2. The semiring $D = Do(\mathbb{N}_0, S)$ is universal in the following sense. Let S be a subsemiring of any additively commutative semiring $\overline{T}$ with an identity, say e'. Then

$$T = \mathbb{N}_0 e' + S = \{\nu e' + s | \nu \in \mathbb{N}_0,\ s \in S\}$$

is a subsemiring of $\overline{T}$ with $o \in S$ as absorbing zero and e' as identity, and there is an epimorphism

$$\varphi : (D, +, \cdot) \to (T, +, \cdot) \text{ given by } \nu e + s \to \nu e' + s.$$

Since φ leaves each $s \in S$ fixed, we call it an *S-epimorphism.* One checks that a typical example of such an epimorphism φ satisfying $\varphi(\alpha e') = \varphi(a)$ for a fixed α-fier a of S according to (2.2) is obtained from the congruence on D defined by

$$(\gamma - \sigma\alpha,\ \sigma a + c) \equiv (\gamma - \tau\alpha,\ \tau a + c)$$

for any $(\gamma, c) \in D$ and any $\sigma, \tau \in \mathbb{N}_0$ satisfying $\gamma \geq \tau\alpha$ and $\gamma \geq \sigma\alpha$.

Now we obtain a rather general result on ZDF semirings:

Theorem 5.3. *Each proper ZDF semiring S with commutative addition and an absorbing zero o can be embedded into a semiring of the*

same kind which has an identity. In particular, the Dorroh-semiring $Do(\mathbb{N}_0, S)$ of S is such a semiring.

Proof. We only have to show that $Do(\mathbb{N}_0, S)$ is ZDF. By way of contradiction, assume $(\nu, s)(\mu, t) = (0, o)$ for some $(\nu, s) \neq (0, o) \neq (\mu, t)$ of $Do(\mathbb{N}_0, S)$. To obtain $\nu\mu = 0$ in (5.2), we assume at first $\nu = 0$, which yields $s \neq o$ and $\mu s + st = o$. Clearly, $\mu = 0$ and hence $t \neq o$ contradicts that S is ZDF. But $\mu \neq 0$ implies that the proper semiring S has zero sums.

So we can apply Thm. 3.3, where (3.5) states that $\mu s + st = o$ for $\mu \in \mathbb{N}$ yields $s = o$, again a contradiction. The case $\mu = 0$ follows in the same way via $\nu t + st = o$.

Remark 5.4. Due to [16], the first part of Thm. 5.3 remains true if one replaces ZDF by MC. However, $D = Do(\mathbb{N}_0, S)$ itself need not be MC if S is. In the contrary, there is a unique S-epimorphic image $T = D/\kappa$ of D which is MC (cf. Remark 5.2), where the corresponding congruence κ on D is given by

$$(\nu, s)\kappa(\nu', s') \text{ iff } \nu t + ts = \nu' t + ts' \text{ for some } t \neq o \text{ of } S.$$

Remark 5.5. The first part of Thm. 5.3 remains also true if $S = R$ is a ZDF ring and hence also MC. But again the semiring $Do(\mathbb{N}_0, R)$ as well as the ring $Do(\mathbb{Z}, R)$ need not be ZDF. Due to [17], there is a unique R-epimorphic image $\mathbb{Z}e' + R \cong Do(\mathbb{Z}, R)/\mathbf{a}$ of the Dorroh-ring by a suitable ideal $\mathbf{a}$, the smallest ZDF ring containing R and an identity (cf. also [19]). Clearly, the R-epimorphism of $Do(\mathbb{Z}, R)$ induces one for its subsemiring $Do(\mathbb{N}_0, R)$.

A special case of the last remark provides, as announced in Section 3, our first construction of a ZDF semiring which contains a given ring R (satisfying the necessary conditions) as its greatest subring:

Theorem 5.6. *Let R be a non-trivial ZDF ring which satisfies the condition C) of Thm. 3.7. Then the Dorroh-semiring $Do(\mathbb{N}_0, R)$ of R is a proper ZDF semiring S with zero sums containing R as its greatest subring. In fact, S is even MC and contains an identity.*

Proof. Clearly, R is the greatest subring of $Do(\mathbb{N}_0, R)$, and it remains to show that $Do(\mathbb{N}_0, R)$ is MC, due to the conditions on R. For the

latter, we prove that the Dorroh-ring $Do(\mathbb{Z}, R)$ is ZDF and hence MC. By way of contradiction, assume $(\nu, s)(\mu, t) = (0, o)$ for some $(\nu, s) \neq$ $\neq (0, o) \neq (\mu, t)$ of $Do(\mathbb{Z}, R)$. There is no loss of generality in assuming that ν and μ are in $\mathbb{N}_0$. So we can use the proof of Thm. 5.3 and obtain for $\nu = 0$ clearly $s \neq o$, but also $\mu s + st = o$ for some $\mu \in \mathbb{N}$, which contradicts C). For $\mu = 0$ and $t \neq o$ we get $\nu t + st = o$ for some $\nu \in \mathbb{N}$, which is also excluded by C) and completes our proof.

Note that each MC semiring S which contains R as a subring such that both have the same absorbing zero has to be AC and additively commutative by Thm. 4.1. So the fact that $S = Do(\mathbb{N}_0, R)$ has these properties corresponds to this situation.

6. Further Constructions of ZDF Semirings

Our next point is to show that each ring R satisfying the necessary conditions of Thm. 3.7 is the greatest subring of a proper ZDF semiring which is not MLC (and hence not MRC by Thm. 4.1). For this purpose we generalize the construction of the Dorroh-semiring $Do(\mathbb{N}_0, S)$ given in Prop. 5.1 as follows.

Construction 6.1. Let W_0 be any semiring with an absorbing zero o and ψ a homomorphism of $(W_0, +, \cdot)$ into $(\mathbb{N}_0, +, \cdot)$. To simplify our notation, we write $\psi(v) = |v|$ for each $v \in W_0$. Note that $|o| + |o| = |o|$ yields $|o| = 0$. Let S be a semiring with commutative addition and an absorbing zero, also denoted by o. Then we define operations on the set $D = W_0 \times S$ in replacing (5.1) and (5.2) by

$$(v, s) + (w, t) = (v + w, s + t) \quad \text{and} \tag{6.1}$$

$$(v, s) \cdot (w, t) = (v \cdot w, |v|t + |w|s + s \cdot t), \tag{6.2}$$

where $|v|t$ and $|w|s$ are defined by the natural operation (2.1) of $\mathbb{N}_0$ on S. It is straightforward to check that $(D, +, \cdot)$ is a semiring with (o, o) as absorbing zero. Moreover, by obvious isormorphisms, we can identify (o, s) with s for each $s \in S$ and (v, o) with v for each $v \in W_0$. Then W_0 and S become subsemirings of D, and we have the unique

presentation

$$(6.3) \qquad (v,s) = (v,o) + (o,s) = v + s \quad \text{for the elements of } D.$$

We denote this semiring by $Do(W_0, \psi, S)$.

Theorem 6.2. *Let R be a non-trivial ZDF ring which satisfies the condition* C) *of Thm. 3.7. Let W_0 be a non-trivial semiring with a zero o and $\psi : W_0 \to \mathbb{N}_0$ a homomorhism satisfying*

$$(6.4) \qquad \psi(v) = |v| = 0 \iff v = o \quad for\ all \quad v \in W_0,$$

which yields that the zero o of W_0 is absorbing and that W_0 is ZDF and has no zero sums.

Then the semiring $S = Do(W_0, \psi, R)$ constructed above with $(o,o) = o$ as absorbing zero is ZDF and has zero sums, and $R = \{(o,r) | r \in R\}$ is its greatest subring. Clearly, S is additively commutative or AC iff W_0 has the same property, and S has an identity, namely (e,o), iff W_0 has an identity e. However, S is neither MLC nor MRC iff there are elements $w \neq w'$ of W_0 satisfying $|w| = |w'|$.

Proof. All statements on W_0 claimed as consequences of (6.4) are checked straightforwardly. In particular, W_0 has no zero sums. Hence, by (6.1), all zero sums of S are of the form $(o,r) + (o,-r) = (o,o)$ for some $r \in R$, hence R is the greatest subring of S. Next we show that S is ZDF and assume, by way of contradiction, $(v,s)(w,t) = (o,o)$ for some $(v,s) \neq (o,o) \neq (w,t)$. From $vw = o$ by (6.2) and since W_0 is ZDF, we get $v = o$ or $w = o$, and it is enough to consider the first case. Then we get $s \neq o$ and, by (6.2), $|w|s + st = o$, which yields $|w| = 0$ due to the assumed condition C) for R. But the latter implies $w = o$ by (6.4), and $st = o$ yields $t = o$ since R is ZDF. Thus we have the contradiction $(w,t) = (o,o)$. Finally, if ψ is injective, $Do(W_0, \psi, R)$ is R-isomorphic to a subsemiring of $Do(\mathbb{N}_0, R)$ and hence MC by Thm. 5.6. Otherwise, there are $w \neq w'$ in W_0 satisfying $|w| = |w'|$, which yields $(o,s)(w,t) = (o,s)(w',t)$ for all $s \neq o$ of R. So S is not MLC and hence not MRC by Thm. 4.1.

We remark without proof that all elements $(v,s) \in S$ with $v \neq o$ are multiplicatively left as well as right cancellable in S.

Note that each element of the semiring $S = Do(W_0, \psi, R) = W_0 + R$ has the unique presentation $(v, r) = v + r$ for $v \in W_0$ and $r \in R$, according to (6.3). In particular, $W = W_0 \setminus \{o\}$ is a subset of $U = S \setminus R$ such that $u = w + r$ is the unique presentation of the elements of U as described in Suppl. 3.5 d). So, in order to obtain by Thm. 6.2 embeddings of R into ZDF and not MC semirings S, as we have announced in Section 3, it remains to show:

Supplement 6.3. *There are semirings W_0 with a zero o which have non-injective homomorphisms $\psi : W_0 \to \mathbb{N}_0$ satisfying (6.4). In particular, there are semirings W_0 of this kind which are not AC or not additively commutative.*

Proof. Let A be any non-trivial semiring and W the direct product of $\mathbb{N}$ and A, which means that the set $W = \{(\nu, a) | \nu \in \mathbb{N}, a \in A\}$ is endowed with componentwise addition and multiplication. Let W_0 be obtained from W by adjoining an absorbing zero $o \notin W$ according to Lemma 2.1. Then $\psi((\nu, a)) = \nu$ for all $(\nu, a) \in W$ and $\psi(o) = 0$ define obviously a homomorphism $\psi : W_0 \to \mathbb{N}_0$ as claimed above. Moreover, W_0 is AC or additively commutative iff A has the same property. There are clearly semirings which even violate both properties. E.g., let A be any set of at least two elements and define $a + b = a$ and $a \cdot b = c$ for all $a, b \in A$ and any fixed element $c \in A$.

As noted above, for each semiring $S = Do(W_0, \psi, R)$ obtained in this way by Thm. 6.2 and Suppl. 6.3, W is a subsemiring of $U = S \setminus R$. Hence we also see by the last statements that $(U, +)$ need neither be cancellative nor commutative (observe again Thm. 4.1 in this context). So we have settled all our announcements given before Thm. 3.7 concerning the structure of proper ZDF semirings S satisfying $S \supset R \supset \{o\}$ as considered in Thm. 3.3 and Suppl. 3.5, except an example such that $u + r \neq r + u$ holds for some $u \in U$ and $r \in R$. Although this could also be done in a rather general way, we restrict ourselves to present a concrete case.

Example 6.4. On $W_0 = 2\mathbb{N}_0 \times 2\mathbb{N}_0$, where $2\mathbb{N}_0$ denotes the set of even non-negative integers, define operations by

$$(\nu_1, \nu_2) + (\mu_1, \mu_2) = (\nu_1 + \mu_1, \nu_2 + \mu_2) \quad \text{and} \tag{6.5}$$

$$(\nu_1, \nu_2) \cdot (\mu_1, \mu_2) = (0, (\nu_1 + \nu_2)(\mu_1 + \mu_2)). \tag{6.6}$$

One easily checks that $(W_0, +, \cdot)$ is a semiring with $(0,0) = o$ as absorbing zero, and that $\psi(\nu_1, \nu_2) = \nu_1 + \nu_2$ defines a homomorphism $\psi : W_0 \to \mathbb{N}_0$ which is not injective and satisfies (6.4). Note that there is an automorphism χ of W_0 given by $\chi(\nu_1, \nu_2) = (\nu_2, \nu_1)$ which satisfies

$$\psi(\chi(v)) = \psi(v) \quad \text{for all} \quad v = (\nu_1, \nu_2) \in W_0. \tag{6.7}$$

Let R be the subring of $\mathbb{Z}[x]$ given by

$$R = \{f(x) = \sum_{i=1}^{n} \gamma_1 x^i | \gamma_i \in \mathbb{Z}\}.$$

Then $Do(W_0, \psi, R) = \{(v, f(x)) | v \in W_0, f(x) \in R\}$ provides a ZDF semiring $(S, +, \cdot)$ according to Thm. 6.2 which contains R as its greatest subring, whose addition is of course commutative. The semiring we want to construct will be $(S, \oplus, \cdot)$, obtained from $(S, +, \cdot)$ in defining a new addition by

$$(v, f(x)) \oplus (w, g(x)) = \begin{cases} (v + w, f(x) + g(x)) & \text{if } \gamma_1 \text{ is even} \\ (v + \chi(w), f(x) + g(x)) & \text{if } \gamma_1 \text{ is odd,} \end{cases}$$

where $\gamma_1 \in \mathbb{Z}$ denotes the coefficient of x^1 in $f(x)$. This clearly yields $f(x) \oplus w = \chi(w) \oplus f(x) \neq w \oplus f(x)$ for all $w \neq o$ of $W \subseteq U = S \setminus R$ and all $f(x) \in R$ for which γ_1 is odd.

To show that $(S, \oplus, \cdot)$ is again a semiring, one has to check that $(S, \oplus)$ is a semigroup and, since $(S, \cdot)$ is commutative, one of the two distributive laws. This can be done in a straightforward manner. But we note that the associativity of $(S, \oplus)$ is known, since the latter is a special semidirect product of the semigroups $(W_0, +)$ and $(R, +)$. Moreover, the distributivity depends essentially on (6.7) and on the restriction of W_0 to pairs of even integers, the latter since then in (6.2),

$$(v, f(x))(w, g(x)) = (v \cdot w, |v| g(x) + |w| f(x) + f(x) \cdot g(x)),$$

the crucial coefficient of x^1 of the polynomial of the right hand side is always even.

Clearly, $(S, \oplus, \cdot)$ is a ZDF semiring like $(S, +, \cdot)$, and $(R, \oplus, \cdot) = (R, +, \cdot)$ is again the greatest subring of $(S, \oplus, \cdot)$. But $w \in U \setminus S$ and $f(x) \in R$ do not always commute as noted above.

References

[1] AHO, A.V., HOPCROFT, J.E. and ULLMANN, J.D.: *The Design and Analysis of Computer Algorithms.* Addison-Wesley, 1974.

[2] BERSTEL, J. and REUTENAUER, C.: *Rational Series and Their Languages.* Springer, 1988.

[3] BOURNE, S.: The Jacobson radical of a semiring. *Proc. Nat. Acad. Sci. USA*, **37**: 163 – 170, 1951.

[4] BOURNE, S.: On the homomorphism theorem for semirings. *Proc. Nat. Acad. Sci. USA*, **38**: 118 – 119, 1952.

[5] BOURNE, S.: On multiplicative idempotents of a potent semiring. *Proc. Nat. Acad. Sci. USA*, **42**: 632 – 638, 1956.

[6] BROWN, B. and McCOY, N.H.: Rings with unit element which contain a given ring. *Duke Mathematical Journal*, **13**: 9 – 20, 1946.

[7] CARRÉ, B.: *Graphs and Networks.* Clarendon Press, 1979.

[8] DORROH, I.L.: Concerning adjunctions to algebra. *Bull. Amer. Math. Soc.*, **38**: 85 – 88, 1932.

[9] FURTWÄNGLER, P. and TAUSSKY, O.: Über Schiefringe. *Sitzungsberichte Akad. Wissensch. Wien*, **145**: 525, 1936.

[10] GRIEPENTROG, R.D. and WEINERT, H.J.: Embedding semirings in semirings with identy. *Coll Math. Soc. J. Bolyai*, **20.** *Algebraic Theory of Semigroups*, North-Holland, 225 – 245, 1979.

[11] GRIEPENTROG, R.D. and WEINERT, H.J.: Correction and remarks to our paper "Embedding semirings in semirings with identy". *Coll. Math. Soc. J. Bolyai*, **39** *Semigroups*, North-Holland, 491 – 493, 1985.

[12] HEBISCH, U. and WEINERT, H.J.: Generalized semigroup semirings which are zero divisor free or multiplictively left cancellative. *Theoretical Computer Science*, to appear.

[13] KUICH, W. and SALOMAA, A.: *Semirings, Automata, Languages.* Springer, 1986.

[14] RÉDEI, L.: Die Verallgemeinerung der Schreierschen Erweiterungstheorie. *Acta Sci. Math.*, **13**: 252 – 273, 1952.

[15] ROTE, G.: A systolic array algorithm for the algebraic path problem (shortest paths; matrix inversion). *Computing*, **34:** 191 – 219, 1985.

[16] STEINFELD, O.: Über Semiringe mit multiplikativer Kürzungsregel. *Acta Sci. Math.*, **24**: 190 – 195. 1963.

[17] SZENDREI, J.: On the extension of rings without divisors of zero. *Acta Sci. Math.*, **13**: 231 – 234, 1949/50.

[18] VANDIVER, H.S.: Note on a simple type of algebra in which the cancellation law of addition does not hold. *Bull. Am. Math. Soc.*, **40**: 920, 1934.

[19] WEINERT, H.J.: Über die Einbettung von Ringen in Oberringe mit Einselement. *Acta Sci. Math. Szeged.*, **22**: 91 – 105, 1961.

[20] WEINERT, H.J.: Über Halbringe und Halbkörper I. *Acta Math. Acad. Sci. Hung.*, **13**: 365 – 378, 1962.

[21] WEINERT, H.J.: Über Halbringe und Halbkörper II. *Acta Math. Acad. Sci. Hung.*, **14**: 209 – 227, 1963.

[22] WEINERT, H.J.: Ringe mit nichtkommutativer Addition I. *Jber. Deutsch. Math.-Verein*, **77**: 10 – 27, 1975.

[23] WEINERT, H.J.: Multiplicative cancellativity of semirings and semigroups. *Acta Math. Acad. Sci. Hung.* **35**: 335 – 338, 1980.

Mathematica Pannonica
1/1 (1990), 95 – 109

ON THE DUAL SPACE OF AN MS-ALGEBRA

T.S. **Blyth***

Mathematical Institute, University of St. Andrews, Scotland.

J.C. **Varlet***

Institut de Mathématique, Université de Liège, Belgique.

Received June 1988

AMS Subject Classification: 06 A 10, 06 D 30

Keywords: Ockham algebra, MS-algebra, Priestley space, topological duality

Abstract: We provide a characterisation of all subvarieties of the variety **MS** of MS-algebras via their dual spaces. It consists of universal sentences in disjunctive normal form which involve only one variable. We apply this result to the construction of distributive lattices on which there can be defined (up to isomorphism) a unique MS-algebra which belongs to a preassigned class.

In 1983 we introduced the notion of an MS-algebra as a common abstraction of a de Morgan algebra and a Stone algebra [3]. Precisely, an MS-algebra is a bounded distributive lattice L endowed with a unary operation $a \to a^\circ$ such that

$$(\forall a \in L) a \leq a^{\circ\circ};$$

* NATO Research Grant 0532/85 is gratefully acknowledged.

$$(\forall a, b \in L)(a \wedge b)^\circ = a^\circ \vee b^\circ;$$
$$1^\circ = 0.$$

Clearly, an MS-algebra is a distributive Ockham algebra ([2], [6] and [9]).

The class **MS** of MS-algebras is equational and all its subclasses were described in [4] by identities that involve at most two variables. We keep the numbering which was adopted in [4, page 159].

R. Beazer [1] and ourselves [5] showed the role that duality theory can play in the study of **MS**. Throughout we assume familiarity with H.A. Priestley's topological duality for bounded distributive lattices as it is presented in [8]. We only recall the facts we need.

A *Priestly space* X is a compact totally order disconnected space, the property of *total order disconnectedness* being defined as follows:
(TOD) given $x \not\leq y$ in X, there exists a clopen order ideal $V \subseteq X$ such that $x \notin V$ and $y \in V$.

The lattice of clopen order ideals of X is denoted by $\mathcal{O}(X)$ and is isomorphic to L_X, the dual algebra of X. In any Priestley space X, for each $x \in X$ there is $y \leq x$ such that y is minimal with respect to the partial order. The set of all minimal points of X is denoted by $\min X$.

Since MS-algebras are bounded distributive lattices, they are dually equivalent to some suitable category of Priestley spaces. In fact, an MS-*space* X is a Priestley space endowed with a continuous order reversing map $g : X \to X$ which satisfies

$$(\forall x \in X) x \geq g^2(x).$$

In [5] we observed that the latter condition implies

$$(\forall x \in X) g^3(x) = g(x)$$

and that to determine such a mapping g it suffices to find a closed subspace X_1 of X which possesses a dual order isomorphism h, then a decreasing order preserving retraction $f : X \to X_1$, and to take $g = h \circ f$. Clearly, $g^2(x) = x$ if and only if $x \in X_1$. It follows that $X_1 \supseteq \min X$.

The unary operation $^\circ$ on L_X is defined by

$$\begin{aligned} I^\circ &= X \setminus g^{-1}(I) \qquad (I \in \mathcal{O}(X)) \\ &= \{x \in X : g(x) \notin I). \end{aligned}$$

Consequently,

$$I^{\circ\circ} = \{x \in X : g^2(x) \in I\} \supseteq I.$$

We use the symbols $\gtrless$ and $\|$ to indicate that two elements are comparable and incomparable respectively. The signs $\subseteq$ and $\subset$ are employed for inclusion and strict inclusion respectively. The expression "L properly belongs to **X**" means that **X** is the least subvariety of **MS** to which L belongs.

After establishing the main result of this paper, that is, the characterisation of all the subvarieties of **MS** via their dual spaces, we formulate some direct consequences which highlight the crucial role played by duality theory. The characterisation theorem is then used to solve the following problem: given a subvariety **X** of **MS**, how to construct a distributive lattice on which there can be defined a unique **MS**-algebra which belongs to **X**. We show that this is possible except for the class **S** of Stone algebras.

Theorem 1. *Let $(L;^\circ)$ be an* MS-*algebra and $(X_L;g)$ its dual space. Then $(L;^\circ)$ satisfies the identity on the left if and only if $(X_L;g)$ satisfies the corresponding formula on the right:*

(2)	$a \vee a^\circ = 1$	(II)	$x = g(x)$
(2_d)	$a \wedge a^\circ = 0$	(II_d)	$g(x) = g^2(x)$
(3)	$a = a^{\circ\circ}$	(III)	$x = g^2(x)$
(4)	$a \wedge a^\circ = a^{\circ\circ} \wedge a^\circ$	(IV)	$x = g^2(x)$ or $x > g(x)$
(4_d)	$a \vee a^\circ = a^{\circ\circ} \vee a^\circ$	(IV_d)	$x = g^2(x)$ or $x < g(x)$
(5)	$(a \wedge a^\circ) \vee b \vee b^\circ = b \vee b^\circ$	(V)	$x \gtrless g(x)$
(6)	$(a \wedge a^\circ) \vee b^{\circ\circ} \vee b^\circ = b^{\circ\circ} \vee b^\circ$	(VI)	$g(x) \gtrless g^2(x)$
(7)	$(a \wedge a^\circ) \vee b \vee b^\circ = (a^{\circ\circ} \wedge a^\circ) \vee b \vee b^\circ$	(VII)	$x = g^2(x)$ or $x \lessgtr g(x)$
(8)	$a \vee b^\circ \vee b^{\circ\circ} = a^{\circ\circ} \vee b^\circ \vee b^{\circ\circ}$	(VIII)	$x = g^2(x)$ or $g^2(x) \leq g(x)$
(9)	$(a \wedge a^\circ) \vee b^\circ \vee b^{\circ\circ} = (a^{\circ\circ} \wedge a^\circ) \vee b^\circ \vee b^{\circ\circ}$	(IX)	$x = g^2(x)$ or $g(x) \gtrless g^2(x)$.

Proof. <u>(2) $\Leftrightarrow$ (II).</u>

$$\begin{aligned}
(2) \quad &\Leftrightarrow I \cup I^\circ = X && (\forall I \in \mathcal{O}(X)) \\
&\Leftrightarrow x \in I \text{ or } g(x) \notin I && (\forall I \in \mathcal{O}(X)) \\
&\Leftrightarrow x \notin I \text{ implies } g(x) \notin I && (\forall I \in \mathcal{O}(X)).
\end{aligned}$$

Let (2) be satisfied. Then if $x \neq g(x)$, by (TOD) there is $V \in \mathcal{O}(X)$ which separates the elements x and $g(x)$, which constradicts the last equivalence. The converse is straightforward. $\diamondsuit$

Observe that (II) implies that X is an antichain. In fact, if $y \geq x$ then

$y = g(y) \leq g(x) = x$, hence $x = y$.

$(2_d) \Leftrightarrow (\mathrm{II}_d)$.

$$\begin{aligned}
(2_d) \quad & \Leftrightarrow I \cap I^\circ = \emptyset && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow \{x \in X : x \in I \text{ and } g(x) \notin I\} = \emptyset && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow x \in I \text{ implies } g(x) \in I && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow g(x) \leq x && (\forall x \in X) \text{ by (TOD)} \\
& \Leftrightarrow g(x) = g^2(x) && (\forall x \in X).
\end{aligned}$$

Indeed, if $g(x) \leq x$ identically, then $g^2(x) \leq g(x)$ but also $g^2(x) \geq g(x)$ since g is order reversing. It follows that $g(x) = g^2(x)$. The other direction is clear since $g^2(x) \leq x$ always. $\diamond$

The condition (II_d) is equivalent to

(II'_d) every connected component A of X contains exactly one element a of $\min X$ and $g(A) = \{a\}$.

Clearly (II'_d) implies (II_d). Conversely, suppose that (II_d) holds (i.e. $g = g^2$). Since g is order reversing and g^2 is order preserving, it follows from $x \gtrless y$ that $g(x) = g(y)$. Hence, since A is connected, $g(A)$ is a singleton, necessarily a minimal element, and (II'_d) is verified. $\diamond$

Note also that (II'_d) has as direct consequence the well-known fact that in a Stone algebra every prime ideal contains exactly one minimal prime ideal.

$(3) \Leftrightarrow (\mathrm{III})$.

$$\begin{aligned}
(3) \quad & \Leftrightarrow I = I^{\circ\circ} && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow I^{\circ\circ} \subseteq I && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow g^2(x) \in I \text{ implies } x \in I && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow g^2(x) = x && (\forall x \in X).
\end{aligned}$$

The last equivalence is justified as follows: if $x > g^2(x)$, then by (TOD) there is $V \in \mathcal{O}(X)$ such that $g^2(x) \in V$ and $x \notin V$, a contradiction. $\diamond$

$(4) \Leftrightarrow (\mathrm{IV})$.

$$\begin{aligned}
(4) \quad & \Leftrightarrow I \cap I^\circ = I^{\circ\circ} \cap I^\circ && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow I^{\circ\circ} \cap I^\circ \subseteq I && (\forall I \in \mathcal{O}(X)) \\
& \Leftrightarrow (g^2(x) \in I \text{ and } g(x) \notin I) \text{ implies } x \in I && (\forall I \in \mathcal{O}(X)).
\end{aligned}$$

Let (4) be satisfied. If $x > g^2(x)$ and $g(x) \not\leq x$, then by (TOD) there exist $V, W \in \mathcal{O}(X)$ such that $x \in V$, $g(x) \notin V$, $g^2(x) \in W$, $x \notin W$.

We thus have $g^2(x) \in V \cap W$ and $g(x) \notin V \cap W$ whereas $x \notin V \cap W$, contradicting (4). It follows that $x > g(x)$.

Conversely, let (IV) be satisfied. If $g^2(x) \in I$ and $g(x) \notin I$ for some $I \in \mathcal{O}(X)$, then $g(x) \not\leq g^2(x)$, hence $g(x) \not\leq x$, and, by (IV), $x = g^2(x)$, $x \in I$ and (4) holds. $\diamond$

<u>$(4_d) \Leftrightarrow (\mathrm{IV}_d)$.</u>

$$\begin{aligned}
(4_d) \;&\Leftrightarrow I^\circ \cup I^{\circ\circ} \subseteq I^\circ \cup I && (\forall I \in \mathcal{O}(X))\\
&\Leftrightarrow I^{\circ\circ} \subseteq I^\circ \cup I && (\forall I \in \mathcal{O}(X))\\
&\Leftrightarrow g^2(x) \in I \text{ implies } (x \in I \text{ or } g(x) \notin I) && (\forall I \in \mathcal{O}(X))\\
&\Leftrightarrow (x \notin I \text{ and } g(x) \in I) \text{ implies } g^2(x) \notin I && (\forall I \in \mathcal{O}(X)).
\end{aligned}$$

Let (4_d) be satisfied and $x > g^2(x)$. If $x = g(x)$ then $x = g^2(x)$, contradicition. If $x \not\leq g(x)$, then there is $V \in \mathcal{O}(X)$ such that $g(x) \in V$ and $x \notin V$. Since by assumption $x \not\leq g^2(x)$, there is $W \in \mathcal{O}(X)$ such that $g^2(x) \in W$ and $x \notin W$. Thus we have $x \notin V \cup W$, $g(x) \in V \cup W$ and nevertheless $g^2(x) \in V \cup W$.

Conversely, let (IV_d) be satisfied. If $x = g^2(x)$ then (4_d) is trivially satisfied. If $g(x) > x$, then every order ideal which contains $g(x)$ contains x as well and (4_d) holds. $\diamond$

<u>$(5) \Leftrightarrow (\mathrm{V})$.</u>

$$\begin{aligned}
(5) \;&\Leftrightarrow I \cap I^\circ \subseteq J \cup J^\circ && (\forall I, J \in \mathcal{O}(X))\\
&\Leftrightarrow (x \in I \text{ and } g(x) \notin I) \text{ implies}\\
&\qquad (x \in J \text{ or } g(x) \notin J) && (\forall I, J \in \mathcal{O}(X)).
\end{aligned}$$

Let (5) be satisfied. If $g(x)||x$, then there are $V, W \in \mathcal{O}(X)$ such that $x \in V$, $g(x) \notin V$, $g(x) \in W$ and $x \notin W$, contradiciting the preceding implication.

Now suppose that (V) is satisfied. The case $g(x) \leq x$ is straightforward. Now if $g(x) > x$, any decreasing subset which does not contain x does not contain $g(x)$ either. $\diamond$

<u>$(6) \Leftrightarrow (\mathrm{VI})$.</u>

$$\begin{aligned}
(6) \;&\Leftrightarrow I \cap I^\circ \subseteq J^\circ \cup J^{\circ\circ} && (\forall I, J \in \mathcal{O}(X))\\
&\Leftrightarrow I^{\circ\circ} \cap I^\circ \subseteq J^\circ \cup J^{\circ\circ} && (\forall I, J \in \mathcal{O}(X))\\
&\Leftrightarrow (g^2(x) \in I \text{ and } g(x) \notin I) \text{ implies}\\
&\qquad (g^2(x) \in J \text{ or } g(x) \notin J) && (\forall I, J \in \mathcal{O}(X)).
\end{aligned}$$

The proof is similar to the preceding one, just changing x into $g^2(x)$. $\diamond$

(7) $\Leftrightarrow$ (VII).

$$\begin{aligned}
(7) \quad &\Leftrightarrow (I^{\circ\circ} \cap I^{\circ}) \cup J \cup J^{\circ} \subseteq (I \cap I^{\circ}) \cup J \cup J^{\circ} && (\forall I, J \in \mathcal{O}(X)) \\
&\Leftrightarrow I^{\circ\circ} \cap I^{\circ} \subseteq I \cup J \cup J^{\circ} && (\forall I, J \in \mathcal{O}(X)) \\
&\Leftrightarrow (g^2(x) \in I \text{ and } g(x) \notin I) \text{ implies} \\
&\quad\ (x \in I \text{ or } x \in J \text{ or } g(x) \notin J) && (\forall I, J \in \mathcal{O}(X)) \\
&\Leftrightarrow x \text{ satisfies (IV) or (V)} && (\forall x \in X) \\
&\Leftrightarrow x = g^2(x) \text{ or } x \gtrless g(x). \quad \diamond
\end{aligned}$$

(8) $\Leftrightarrow$ (VIII).

$$\begin{aligned}
(8) \quad &\Leftrightarrow I^{\circ\circ} \cup J^{\circ} \cup J^{\circ\circ} \subseteq I \cup J^{\circ} \cup J^{\circ\circ} && (\forall I, J \in \mathcal{O}(X)) \\
&\Leftrightarrow I^{\circ\circ} \subseteq I \cup J^{\circ} \cup J^{\circ\circ} && (\forall I, J \in \mathcal{O}(X)) \\
&\Leftrightarrow g^2(x) \in I \text{ implies} \\
&\quad\ (x \in I \text{ or } g^2(x) \in J \text{ or } g(x) \notin J) && (\forall I, J \in \mathcal{O}(X)) \\
&\Leftrightarrow (g^2(x) \in I \text{ and } x \notin I) \text{ implies} \\
&\quad\ (g^2(x)s \in J \text{ or } g(x) \notin J) && (\forall I, J \in \mathcal{O}(X))
\end{aligned}$$

Let (8) be satisfied and $x > g^2(x)$. If $g^2(x) \not\leq g(x)$, then there is $V \in \mathcal{O}(X)$ such that $g(x) \in V$ and $g^2(x) \notin V$, which contradicts (8).

Conversely, let (VIII) be satisfied. Since every decreasing subset which does not contain $g^2(x)$ does not contain $g(x)$ either, (8) is satisfied. $\diamond$

(9) $\Leftrightarrow$ (IX).

The proof goes along the same lines as in (7) $\Leftrightarrow$ (VII).

$$\begin{aligned}
(9) \quad &\Leftrightarrow x \text{ satisfies (IV) or (VI)} && (\forall x \in X) \\
&\Leftrightarrow x = g^2(x) \text{ or } g(x) \gtrless g^2(x). \quad \diamond
\end{aligned}$$

Corollary 1. *All the subvarieties of* **MS** *can be characterised via their spaces by the disjunction of at most three universal sentences which involve only one variable.*

The results are recorded in the subvariety lattice represented on the page 102. Note that g° means id_X.

Proof. Theorem 1 yields the characterisation of the subvarieties which are defined by a unique identity. The other non-trivial subvarieties are characterised by the conjunction of two or three conditions. An easy computation provides the corresponding conditions on the dual space in

disjunctive form. For instance, $L \in \mathbf{S} \vee \mathbf{K}$ if and only if $(X_L; g)$ satisfies (IV),(V) and (VIII). The conjunction of (IV) and (VIII) is equivalent to the disjunction of $x = g^2(x)$ and $x > g(x) \geq g^2(x)$, which in turn is equivalent to the disjunction of $x = g^2(x)$ and $g(x) = g^2(x)$. Finally, the conjunction of (IV), (V) and (VIII) is equivalent to the disjunction of $x = g^2(x) \gtrless g(x)$ and $g(x) = g^2(x)$. $\diamondsuit$

Corollary 2.- *If X has at least two connected components A, B and $g(A) \subseteq B$, then L does not satisfy* (6). *If moreover $g(A) \subset B$, then L properly belongs to* $\mathbf{M}_1$.

Proof. The first part is obvious. As for the second part, observe that there is $x \in B \setminus g(A)$ such that $x \neq g^2(x)$ and $g(x) || g^2(x)$, hence L does not satisfy (9). $\diamondsuit$

Corollary 3. *If $X_1 \subset X$ and X_1 is convex, then L does not satisfy* (4_d).

Proof. If L satisifies (4_d) but not (3), then there is $x \in X \setminus X_1$ such that $g^2(x) < x < g(x)$ and X_1 is not convex. $\diamondsuit$

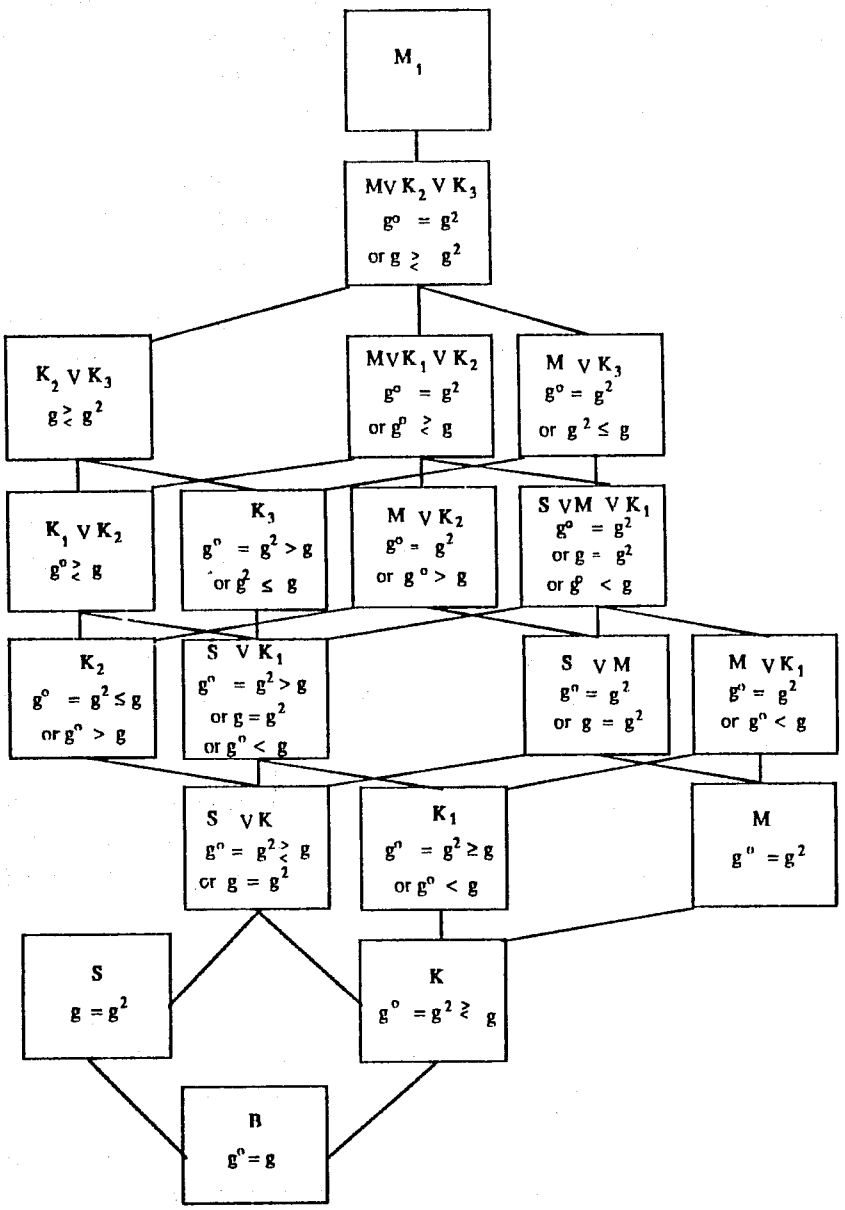
Corollary 4. *If X_1 is open and convex, then the dual of X_1 is a principal ideal of L the generator of which is the least element a of L such that $a^\circ = 0$.*

Proof. Since X_1 is always closed, it is clopen; since it contains $\min X$ and is assumed to be convex, it is decreasing. Its dual is a principal ideal $a^\downarrow$ of L. By its very definition $X_1^\circ = \emptyset$. Let $Y \in \mathcal{O}(X)$. Clearly $Y \supseteq X_1$ if and only if $Y^\circ = \emptyset$. So the least element Y of $\mathcal{O}(X)$ such that $Y^\circ = \emptyset$ is X_1. $\diamondsuit$

We already noticed that (III) is equivalent to $X = X_1$, and that (II) implies $X = X_1$. Can the verification of some of the other axioms be restricted to X_1? The answer is affirmative as shown by

Corollary 5. *The axioms* (II_d) *and* (VI) *are satisfied if (and only if) they are so on X_1. If X_1 is totally ordered, then L satisfies* (6).

Proof. Just observe that $X_1 = \{g(x) : x \in X\} = \{g^2(x) : x \in X\}$. $\diamondsuit$


M_1
$M \vee K_2 \vee K_3$
$g^o = g^2$
or $g \gtrless g^2$
$K_2 \vee K_3$
$g \gtrless g^2$
$M \vee K_1 \vee K_2$
$g^o = g^2$
or $g^o \gtrless g$
$M \vee K_3$
$g^o = g^2$
or $g^2 \leq g$
$K_1 \vee K_2$
$g^o \gtrless g$
K_3
$g^o = g^2 > g$
or $g^2 \leq g$
$M \vee K_2$
$g^o = g^2$
or $g^o > g$
$S \vee M \vee K_1$
$g^o = g^2$
or $g = g^2$
or $g^o < g$
K_2
$g^o = g^2 \leq g$
or $g^o > g$
$S \vee K_1$
$g^o = g^2 > g$
or $g = g^2$
or $g^o < g$
$S \vee M$
$g^o = g^2$
or $g = g^2$
$M \vee K_1$
$g^o = g^2$
or $g^o < g$
$S \vee K$
$g^o = g^2 \gtrless g$
or $g = g^2$
K_1
$g^o = g^2 \geq g$
or $g^o < g$
M
$g^o = g^2$
S
$g = g^2$
K
$g^o = g^2 \gtrless g$
B
$g^o = g$

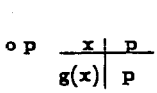
There is a significant difference between Stone algebras and de Morgan algebras: when a bounded distributive lattice can be made into a Stone algebra, this can be done in only one way (in other words, the lattice structure determines the unary operation $^\circ$); on the contrary, many bounded distributive lattices admit various definitions of the unary operation $^\circ$ which satisfy the axioms of a de Morgan algebra (more generally, of an MS-algebra of a given class other than **S** or **B**). This provokes the question as to whether we can find, for a given subvariety **X** of MS-algebras, a distributive lattice L on which there can be defined to within isomorphism a *unique* MS-algebra structure such that $(L,^\circ) \in \mathbf{X}$. A subvariety **X** for which this is the case will be called *saturated.*

Theorem 2. *All subvarieties of MS-algebras, other than* **S**, *are saturated.*

Proof. We first show that the subvariety **S** of Stone algebras is not satureted. Suppose that L is a distributive lattice on which there can be defined an MS-algebra structure such that $(L,^\circ)$ belongs properly to **S**. Let $(X_L; g)$ be the corresponding MS-space. We have $X_L = (A_i)_{i\in I}$ where the A_i are the connected components of X_L. By (II'$_d$) every A_i has a least element a_i. Moreover, not every A_i consists of a singleton, for otherwise $(L,^\circ) \in \mathbf{B}$. For a given A_i that is not a singleton, say A_{i*}, choose an element $x_{i*} \neq a_{i*}$ and take $X_1 = \{a_i : i \in I\} \cup \{x_{i*}\}$. Since $a_{i*} = g(x_{i*}) \neq g^2(x_{i*}) = x_{i*}$,we obtain an MS-algebra that does not belong to **S**.

Now, for each non-trivial subvariety **X**, other than **S**, we give an example of a distributive lattice L that can be made into an MS-algebra in only one way with $(L,^\circ) \in \mathbf{X}$. This we achieve by considering in each case an appropriate MS-space. In all examples the space X is connected and, except for the class **B** of Boolean algebras, $|\min X| \geq 2$ since otherwise X_1 could be chosen in various ways. The black circles correspond to the elements of X which do not belong to X_1, and to the meet-irreducible elements of L_X other than 1.

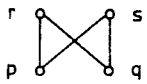
B



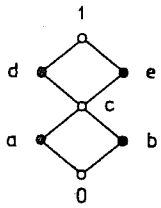
x	p
g(x)	p

α	0	1
α^0	1	0

K

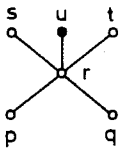


x	p	q	r	s
g(x)	r	s	p	q

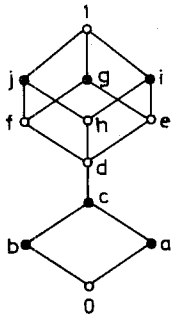


α	0	a	b	c	d	e	1
α^0	1	e	d	c	b	a	0

S ∨ K

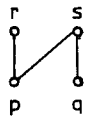


x	p	q	r	s	t	u
q(x)	s	t	r	p	q	r

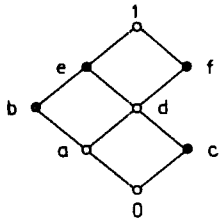


α	0	a	b	c	d	e	f	g	h	i	j	1
α^0	1	j	i	h	c	b	a	0	c	b	a	0

M

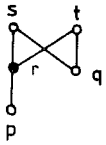


x	p	q	r	s
g(x)	s	r	q	p

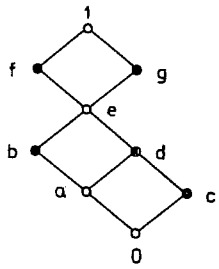


α	0	a	b	c	d	e	f	1
α^0	1	e	b	f	d	a	c	0

K_1

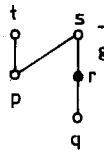


x	p	q	r	s	t
g(x)	s	t	s	p	q

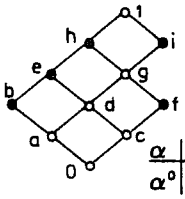


α	0	a	b	c	d	e	f	g	1
α^0	1	g	g	f	e	e	c	b	0

M_1

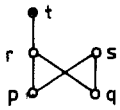


x	p	q	r	s	t
g(x)	s	t	t	p	q

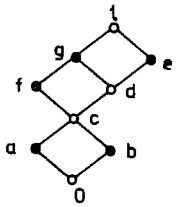


α	0	a	b	c	d	e	f	g	h	i	1
α^0	1	h	b	i	g	a	i	g	a	f	0

K_2

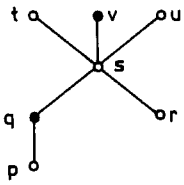


x	p	q	r	s	t
g(x)	r	s	p	q	p

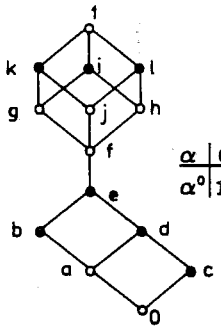


α	0	a	b	c	d	e	f	g	1
α^0	1	f	e	c	b	b	a	0	0

$S \vee K_1$

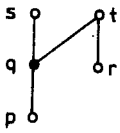


x	p	q	r	s	t	u	v
g(x)	t	t	u	s	p	r	s

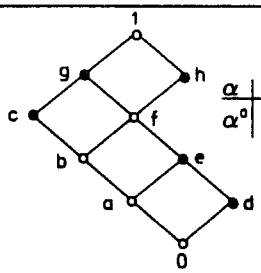


α	0	a	b	c	d	e	f	g	h	i	j	k	l	1
α^0	1	1	1	k	j	j	e	c	b	0	e	c	b	0

$M \vee K_1$

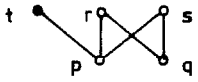


x	p	q	r	s	t
g(x)	t	t	s	r	p

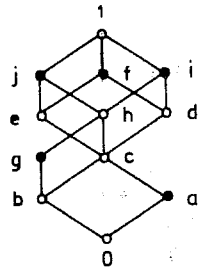


α	0	a	b	c	d	e	f	g	h	1
α^0	1	g	g	c	h	f	f	b	d	0

K_3

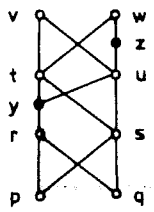


x	p	q	r	s	t
g(x)	r	s	p	q	r

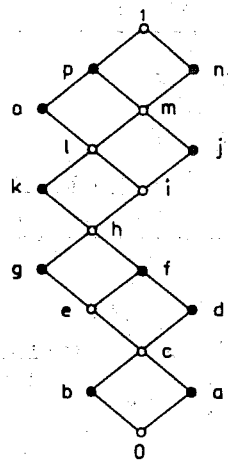


α	0	a	b	c	d	e	f	g	h	i	j	1
α^0	1	j	i	h	g	a	0	i	h	g	a	0

$K_1 \vee K_2$



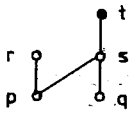
x	p	q	r	s	t	u	v	w	y	z
g(x)	v	w	t	u	r	s	p	q	t	s



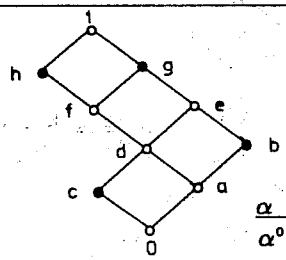
α	0	a	b	c	d	e	f	g	h	i	j
α^0	1	p	n	m	k	j	h	j	h	g	g

α	k	l	m	n	o	p	1
α^0	d	c	c	b	a	a	0

$M \vee K_2$

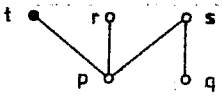


x	p	q	r	s	t
g(x)	s	r	q	p	p

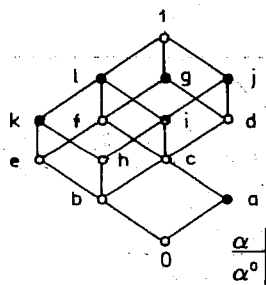


α	0	a	b	c	d	e	f	g	h	1
α^0	1	e	b	h	d	a	c	0	c	0

$M \vee K_3$

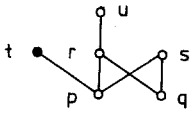


x	p	q	r	s	t
g(x)	s	r	q	p	s

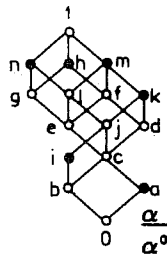


α	0	a	b	c	d	e	f	g	h	i	j	k	l	1
α^0	1	j	l	i	a	k	h	0	l	i	a	k	h	0

$K_2 \vee K_3$

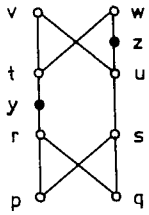


x	p	q	r	s	t	u
g(x)	r	s	p	q	r	p

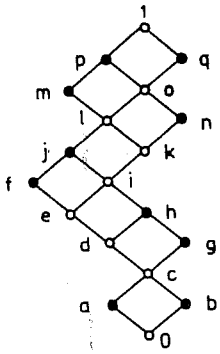


α	0	a	b	c	d	e	f	g	h	i	j	k	l	m	n	1
α^0	1	n	k	j	i	a	0	a	0	k	j	i	a	0	a	0

$M \vee K_1 \vee K_2$

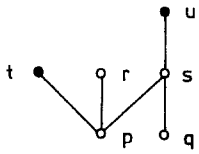


x	p	q	r	s	t	u	v	w	y	z
g(x)	w	v	u	t	s	r	q	p	u	r

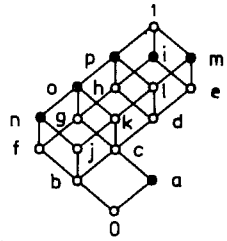


α	0	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	1
α^0	1	p	q	0	j	j	f	n	i	i	e	g	c	a	g	c	a	b	0

$M \vee K_2 \vee K_3$

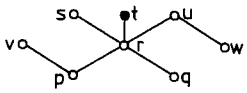


x	p	q	r	s	t	u
g(x)	s	r	q	p	s	p

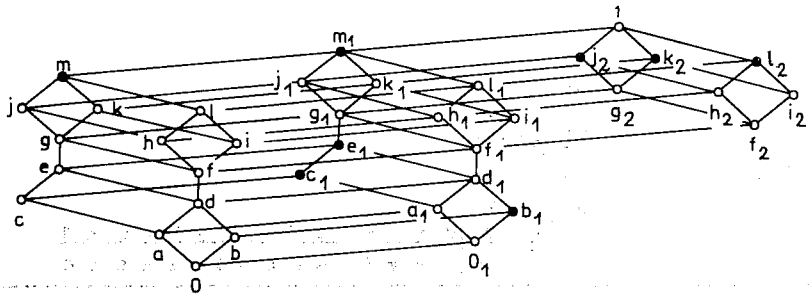


α	0	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	1
α^0	1	i	o	g	a	a	o	g	a	a	n	f	0	0	n	f	0	0

S ∨ M



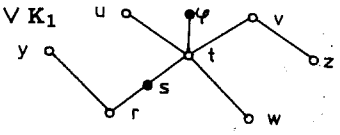
x	p	q	r	s	t	u	v	w
g(x)	u	s	r	q	r	p	w	v



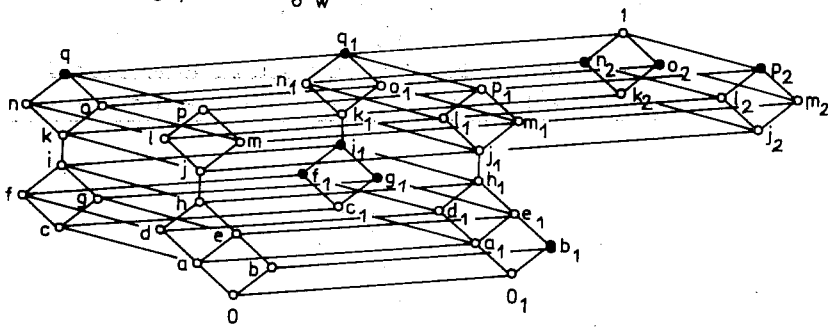
α	0	a	b	c	d	e	f	g	h	i	j	k	l	m	0_1	a_1	b_1	c_1	d_1	e_1	f_1	g_1	h_1	i_1	j_1
α^o	1	m_1	k_2	m	k_1	k	e_1	e	c_1	e_1	c	e	c_1	c	l_2	l_1	i_2	l	i_1	i	d_1	d	a_1	d_1	a

α	k_1	l_1	m_1	f_2	g_2	h_2	i_2	j_2	k_2	l_2	1
α^o	d	a_1	a	b_1	b	0_1	b_1	0	b	0_1	0

S ∨ M ∨ K_1



x	r	s	t	u	v	w	y	z	φ
g(x)	v	v	t	w	r	u	z	y	t



α	0	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	0_1	a_1	b_1	c_1	d_1	e_1	f_1
α^o	1	q_1	o_2	q	q_1	o_1	q	o	o_1	o	i_1	i	f_1	i_1	f	f_1	i	f	p_2	p_1	m_2	p	p_1	m_1	p

α	g_1	h_1	i_1	j_1	k_1	l_1	m_1	n_1	o_1	p_1	q_1	j_2	k_2	l_2	m_2	n_2	o_2	p_2	1
α^o	m	m_1	m	h_1	h	d_1	h_1	d	h	d_1	d	b_1	b	0_1	b_1	0	b	0_1	0

References

[1] BEAZER, R.: Injectives in some small varieties of Ockham algebras, *Glasgow Math. J.* **25** (1984), 183 – 191.

[2] BERMAN, J.: Distributive lattices with an additional unary operation, *Aequationes Math.* **16** (1977), 165 – 171.

[3] BLYTH, T.S. and VARLET, J.C.: On a common abstraction of de Morgan algebras and Stone algebras, *Proc. Roy. Soc. Edinburgh Sect. A* **94** (1983), 301 – 308.

[4] BLYTH, T.S. and VARLET, J.C.: Subvarieties of the class of MS-algebras, *Proc. Roy. Soc. Edinburgh Sect. A* **95** (1983), 157 – 169.

[5] BLYTH, T.S. and VARLET, J.C.: MS-algebras definable on a distributive lattice, *Bull. Soc. Roy. Sci. Liège* **54** (1985), 167 – 182.

[6] GOLDBERG, M.S.: Distributive Ockham algebras: free algebras and injectivity, *Bull. Austral. Math. Soc.* **24** (1981), 161 – 203.

[7] PRIESTLEY, H.A.: The construction of spaces dual to pseudocomplemented distributive lattices, *Quart. J. Math. Oxford* **26** (1975), 215 – 228.

[8] PRIESTLEY, H.A.: Ordered sets and duality for distributive lattices, *Ann. Discrete Math.* **23** (1984), 39 – 60.

[9] URQUHART, A.: Distributive lattices with a dual homomorphic operation, *Studia Logica* **38** (1979), 201 – 209.

Mathematica Pannonica
1/1(1990), 111 - 116

ON AN INTEGRAL INEQUALITY FOR CERTAIN ANALYTIC FUNCTIONS

Petru T. **Mocanu**

Department of Mathematics, Babes-Bolyai University, RO-3400 Cluj, Romania.

Received August 1988

AMS Subject Classification: 30 C 80, 30 C 45

Keywords: Analytic function, inequality, subordination.

Abstract: Let g be an analytic function on the unit disc $U = \{z; |z| < 1\}$, with $g(0) = g'(0) - 1 = 0$ and let $f(z) = \int_0^z [g(t)/t]dt$. It is shown that if g satisfies the inequality $|g'(z) - 1| < 8/(2 + \sqrt{15}) = 1.362\ldots$ for $z \in U$, then $|zf'(z)/f(z) - 1| < 1$, which is equivalent to $Re \int_0^1 [g(uz)/ug(z)]du > 1/2$, for $z \in U$.

1. Introduction

Let A denote the class of functions f, which are analytic on the unit disc $U = \{z; |z| < 1\}$, with $f(0) = 0$ and $f'(0) = 1$. In a recent paper we obtained the following result [3, Corollary 4.2].

If $g \in A$ satisfies $|g'(z) - 1| < 1$, for $z \in U$, then

$$Re \int_0^1 \frac{g(uz)}{ug(z)}\, du > \frac{1}{2}, \quad \text{for } z \in U.$$

If we let

$$f(z) = \int_0^1 \frac{g(uz)}{u}\, du,$$

then this last inequality is equivalent to

$$\left| \frac{zf'(z)}{f(z)} - 1 \right| < 1, \quad \text{for } z \in U.$$

In the present paper we improve the above result, by showing that the same conclusion holds under the less restrictive condition $|g'(z) - 1| < 8/(2+\sqrt{15}) = 1.362\ldots$

2. Preliminaries

If f and g are analytic functions on U, then we say that f is subordinate to g, written $f \prec g$, or $f(z) \prec g(z)$, if g is univalent, $f(0) = g(0)$ and $f(U) \subset g(U)$.

We shall use the following lemmas to prove our results.

Lemma 1 [1,p.192]. *Let h be a convex function on U (i.e. h is univalent and $h(U)$ is a convex domain). If p is analytic in U and satisfies the differential subordination*

$$p(z) + zp'(z) \prec h(z),$$

then

$$p(z) \prec \frac{1}{z} \int_0^z h(t)\, dt.$$

Lemma 2 [2,p.201]. *Let E be a set in the complex plane $\mathbb{C}$ and let q be an analytic and univalent function on U. Suppose that the function $H : \mathbb{C}^2 \times U \to \mathbb{C}$ satisfies*

$$H[q(\zeta), m\zeta q'(\zeta); z] \notin E,$$

whenever $m \geq 1$, $|\zeta| = 1$ and $z \in U$. If p is analytic on U, and satisfies $p(0) = q(0)$ and

$$H[p(z), zp'(z); z] \in E, \quad \text{for } z \in U,$$

then $p \prec q$.

For use in Section 4 we need the following elementary sharp inequalities.

Lemma 3. *If* $z \in \mathbb{C}$ *then* $|\sin z| \leq \operatorname{sh}|z|$; *if* $z \in \mathbb{C}$ *and* $|z| < \pi/2$ *then* $|\tan z| \leq \tan|z|$.

3. Main results

Theorem 1. *If* $f \in A$ *satisfies*

$$|f'(z) + zf''(z) - 1| < M, \quad z \in U, \tag{1}$$

where $M \leq M_0 = 8/(2+\sqrt{15}) = 1.362\ldots$, *then*

$$\left|\frac{zf'(z)}{f(z)} - 1\right| < 1, \quad z \in U. \tag{2}$$

Proof. Since the inequality (1) can be rewritten as

$$f'(z) + zf''(z) \prec 1 + Mz,$$

by using Lemma 1, we deduce $f'(z) \prec 1 + Mz/2$ and

$$\frac{f(z)}{z} \prec 1 + \frac{M\ z}{4}. \tag{3}$$

Let $p(z) = zf'(z)/f(z)$ and $P(z) = f(z)/z$. Since (3) implies $P(z) \neq 0$, the function p is analytic in U and the inequality (1) becomes

$$|P(z)[zp'(z) + p^2(z)] - 1| < M, \quad z \in U. \tag{4}$$

The inequality (2) is equivalent to

$$p(z) \prec 1 + z \tag{5}$$

and in order to show that (5) holds, by Lemma 2, it is sufficient to check the inequality

$$|P(z)[m\zeta + (1+\zeta)^2] - 1| \geq M, \tag{6}$$

for all $m \geq 1$, $|\zeta| = 1$ and $z \in U$.

If we let $\zeta = e^{i\theta}$, then

$$\begin{aligned} L(m,\theta,z) &\equiv |P(z)[m\zeta + (1+\zeta)^2] - 1|^2 = \\ &= |P(z)\zeta(\zeta + \bar{\zeta} + m + 2) - 1|^2 = \\ &= (2\cos\theta + m + 2)\{(2\cos\theta + m + 2)|P(z)|^2 - \\ &\quad - 2Re[e^{i\theta}P(z)]\} + 1. \end{aligned}$$

From (3) we deduce $|P(z) - 1| < M/4$ and $|P(z)| > 1 - M/4$. For $m \geq 1$ we have

$$\begin{aligned} \frac{\partial L}{\partial m} &= (2\cos\theta + m + 2)|P(z)|^2 - Re[e^{i\theta}P(z)] = \\ &= (m+2)|P(z)|^2 + Re\{e^{i\theta}P(z)[2\overline{P(z)} - 1]\} \geq \\ &\geq |P(z)|\{(3|P(z)| - |2P(z) - 1|\} \geq |P(z)|(2 - \frac{5M}{4}) > 0, \end{aligned}$$

which shows that L is an increasing function of m. Hence we deduce

$$\begin{aligned} L(m,\theta,z) \geq L(1,\theta,z) &= (2\cos\theta + 3)[3|P|^2 - 2Re[e^{i\theta}P(\bar{P} - 1)] + 1 \\ &\geq (2\cos\theta + 3)|P|[3|P| - 2|P - 1|] + 1 \geq \\ &\geq \left(1 - \frac{M}{4}\right)\left[3\left(1 - \frac{M}{4}\right) - \frac{M}{2}\right] + 1 \equiv K(M). \end{aligned}$$

Since $0 < M \leq M_0$, where M_0 is the positive root of the equation $K(M) = M^2$, we deduce $L(m,\theta,z) \geq M^2$, which yields (6). Hence the subordination (5) holds and we obtain (2), which completes the proof of Theorem 1.

The following two theorems are integral versions of Theorem 1.

Theorem 2. *If $g \in A$ satisfies $|g'(z) - = 1| < M_0 = 8/(2 + \sqrt{15})$ then*

$$\left|\frac{zf'(z)}{f(z)} - 1\right| < 1, \quad \text{for } z \in U,$$

where

$$f(z) = \int_0^z \frac{g(t)}{t}\, dt = \int_0^1 \frac{g(uz)}{u}\, du.$$

Theorem 3. *If $g \in A$ satisfies $|g'(z) - 1| < M_0 = 8/(2 + \sqrt{15}$ then*

$$Re \int_0^1 \frac{g(uz)}{ug(z)}\, du > \frac{1}{2}, \quad \text{for } z \in U.$$

4. Examples

Example 1. If we let $g(z) = (\sin \lambda z)/\lambda$, where

$$|\lambda| \leq ln[1 + M_0 + \sqrt{M_0(M_0 + 2)}] = 1.504\ldots$$

then, by using Lemma 3, we have

$$|g'(z) - 1| = 2|\sin^2 \frac{\lambda z}{2}| \leq 2\ sh^2 \frac{|\lambda z|}{2} < 2\ sh^2 \frac{|\lambda|}{2} \leq M_0,$$

for $z \in U$ and by Theorem 3 we deduce

$$Re \frac{\mathrm{Si}(z)}{\sin z} > \frac{1}{2}, \quad \text{for} \quad |z| < 1.504\ldots$$

where

$$\mathrm{Si}(z) = \int_0^1 \frac{\sin uz}{u}\, du = \int_0^z \frac{\sin t}{t}\, dt.$$

Example 2. If we let $g(z) = (e^{\lambda z} - 1)/\lambda$, where

$$|\lambda| \leq ln(1 + M_0) = 0.859\ldots$$

then $|g'(z) - 1| \leq M_0$, for $z \in U$ and by Theorem 3 we deduce

$$Re \int_0^1 \frac{e^{uz} - 1}{u(e^z - 1)}\, du > \frac{1}{2}, \quad \text{for} \quad |z| < 0.859\ldots$$

Example 3. If we let $g(z) = [ln(1 + \lambda z)]/\lambda$, where

$$|\lambda| \leq \frac{M_0}{1 + M_0} = 0.576\ldots$$

then $|g'(z) - 1| < M_0$, for $z \in U$ and by Theorem 3 we deduce

$$Re \int_0^1 \frac{\ln(1+uz)}{u\ ln(1+z)}\, du > \frac{1}{2}, \quad \text{for} \quad |z| < 0.576\ldots$$

Example 4. If we let $g(z) = (\tan \lambda z)/\lambda$, where

$$|\lambda| \leq \arctan \sqrt{M_0} = 0.862\ldots$$

then, by Lemma 3, we have

$$|g'(z) - 1| = |\tan^2 \lambda z| \leq \tan^2 |\lambda z| < \tan^2 |\lambda| \leq M_0,$$

for $z \in U$ and by Theorem 3 we deduce

$$Re \int_0^1 \frac{\tan\ uz}{u\ \tan\ z}\, du > \frac{1}{2}, \quad \text{for} \quad |z| < 0.862\ldots$$

References

[1] HALLENBECK, D.J. and RUSCHEWEYH, S.: Subordination by convex functions, *Proc. Amer. Math. Soc.* **52** (1975), 191 – 195.

[2] MILLER, S. S. and MOCANU, P.T.: Differential subordinations and inequalities in the complex plane, *J. of Differential Equations* **67**, 2 (1987), 199 – 211.

[3] MOCANU, P.T.: Starlikeness conditions for analytic functions, *Rev. Roumaine Math. Pures Appl.*, **33**, 1 – 2 (1988), 117 – 124.

Mathematica Pannonica
1/1 (1990), 117 – 123

SU UN ASSIOMA INTERESSANTE LA TEORIA DEGLI "INTERVAL ORDERS"

Romano **Isler**

Dipartimento di Matematica Applicata "Bruno de Finetti", Università, I-34100 - TRIESTE, Piazzale Europa 1, Italia.

Received September 1988

AMS Subject Classification: 06 A 10, 05 A 15

Keywords: Interval order, representation theorem

Abstract: Given a set X with a strong order, satisfying the usual axioms $a \not\prec a \ \forall a$ and transitivity, if we add axiom: $(a_1 \prec a_2) \wedge (b_1 \prec b_2) \Rightarrow$ $\Rightarrow (a_1 \prec b_2) \vee (b_1 \prec a_2)$ we obtain the so called "interval order" which was deeply studied first by P.C. Fishburn.

If we substitute such axiom with the following:
$(a_1 \prec a_2) \wedge (b_1 \prec b_2) \Rightarrow (a_1 \prec b_1) \vee (b_1 \prec a_1) \vee (a_2 \prec b_2) \vee (b_2 \prec a_2)$
we obtain a different order structure which implies the "interval order". Such order structure is analized and many results are obtained including a very simple proof of a theorem of representability, analogous to a fundamental theorem on interval orders due to Fishburn.

In un mio precedente lavoro avevo introdotto un assioma che indeboliva, in un certo senso nel modo più blando possibile, gli assiomi dell'ordine debole o concordante in un insieme E.

Precisamente, considerati gli assiomi usuali dell'ordine forte (o stretto)

(1) $a \not\prec a \quad \forall a$

(2) $a \prec b \Rightarrow b \not\prec a$

(3) $a \prec b,\ b \prec c \Rightarrow a \prec c$

posto l'assioma della negativa transitività:

(4) $a \not\prec b,\ b \not\prec c \Rightarrow a \not\prec c$

si aveva gli assiomi (2), (4) che individuavono l'ordine debole erano equivalenti agli assiomi (2), (3), (5) dove

(5) $a\ I\ b \Rightarrow (a \prec x \Rightarrow b \prec x) \wedge (x \prec a \Rightarrow x \prec b)$ con

$a\ I\ b \Leftrightarrow (a \prec b) \wedge (b \prec a)$.

L'assioma (5) definiva una relazione concordante e dunque gli assiomi (2), (3), (5) un ordine concordante in quanto, per parlare d'ordine, pretenderemo sempre verificate le (1), (2), (3).

Inoltre, introdotti gli assiomi

(4*) $a \prec a' \Rightarrow (a \prec b) \vee (b \prec a') \quad \forall b \in E$

(4^*_*) $a_1 \prec a_2 \prec a_3 \Rightarrow (b \prec a_2) \vee (a_2 \prec b) \vee (a_1 \prec b \prec a_3) \quad \forall b \in E$

valevano le (4) $\Leftrightarrow$ (4*) $\Leftrightarrow$ (4^*_*).

Assiomi che indeboliscano un ordine debole sono i seguenti

(4') $a_1 \prec a_2 \prec a_3 \Rightarrow (a_1 \prec b) \vee (b \prec a_3)$

(4") $(a_1 \prec a_2) \wedge (b_1 \prec b_2) \Rightarrow (a_1 \prec b_2) \vee (b_1 \prec a_2)$

(4"') $(a_1 \prec a_2) \wedge (b_1 \prec b_2) \Rightarrow (a_1 \prec b_1) \vee (b_1 \prec a_1) \vee (a_2 \prec b_2) \vee (b_2 \prec a_2)$

In una relazione d'ordine si ha (4"') $\Rightarrow$ (4"). Per il resto (4'), (4") e (4"') sono fra loro slegati.

L'assioma (4") definisce quello che, in letteratura, è noto come *"interval order"* mentre gli assiomi (4')$\wedge$(4") il "semiorder". Non essendomi noto uno studio, nè tantomeno l'introduzione, dell'assioma (4"'), mi propongo di studiare la struttura d'ordine che esso induce, con particolare riguardo all'esistenza di funzioni di valutazione.

Poichè (4"') $\Rightarrow$ (4"), i risultati che otteremo comprenderanno anche quelli riguardanti gli interval orders.

Ci proponiamo dunque di studiare la struttura d'ordine nel caso dell'assioma (4"') che indebolisce l'ordine concordante in un certo senso meno dell'assioma (4") (toglie una confrontabilità, ossia una relazione di preferenza, contro le due dell'assioma (4")).

Sia E un insieme con una relazione $\prec$ per la quale valgano gli assiomi (2), (3), (4"'). Indicheremo con $|A|$ la cardinalità dell'insieme A. Con $a \to b$ la $a \prec b$.

Proposizione 1.1. *Sia C_α una catena massimale, $|C_\alpha| > 1$ e $b \notin C_\alpha$, b non isolato. Allora b è confrontabile con C_α ossia $\exists a \in C_\alpha : b \not I\, a$.*

Dim. Sia per esempio $b \to b_1$, $b_1 \notin C_\alpha$. Sia b inconfrontabile con ogni $a \in C_\alpha$. Poichè $|C_\alpha| > 1$ $\exists a_1 \in C_\alpha : (a_1 \to a) \vee (a \to a_1)$. Se $a_1 \to a$, poichè b è inconfrontabile con a_1 e con a, ne viene, per (4"), $a \to b_1$. Se invece $a \to a_1$ ne viene $a_1 \to b_1$ da cui $a \to b_1$. Ma allora $a \to b_1$ $\forall a \in C_\alpha$, ossia C_α non è massimale: impossible.

Proposizione 1.2. *Sia $b \notin C_\alpha$. $\exists \bar{a} \in C_\alpha : bI\bar{a}$*

Dim. Ovvia, stante la massimalità.

Proposizione 1.3. *Sia $b \notin C_\alpha$ non isolato. Se esiste $b_1 : b \to b_1$, l'intervallo di inconfrontabilità con b di C_α è superiormente limitato.*

Dim. Se $b_1 \in C_\alpha$ è ovvio. Sia $b_1 \notin C_\alpha$. Sia $a \in C_\alpha : b\ I\ a$. Se fosse $b\ I\ a_1$ $\forall a_1 \in C_\alpha : a \to a_1$, seguirebbe, per (4"), $a_1 \to b_1$ e quindi C_α non massimale.

Analoga proposizione vale se esiste $b_1 : b_1 \to b$. Allora l'intervallo di inconfrontabilità con b : $I(C_\alpha, b)$ è inferiormente limitato.

Proposizione 1.4. *Sia $b \notin C_\alpha$ non isolato. Se esiste $b_1 \notin C_\alpha : (b_1 \to \to b) \vee (b \to b_1)$, allora $I(C_\alpha, b)$ è un singleton.*

Dim. Sia per esempio $b \to b_1$. Sia $a \in C_\alpha : a\ I\ b$. Sia $a_1 \in C_\alpha : a_1\ I\ b$ e sia per assurdo $a_1 \neq a$; se $a \to a_1$ l'elemento $\bar{a}$ di inconfrontabilità con b_1 deve seguire $a : a \to \bar{a}$; infatti da (4") segue $a_1 \to b_1$ e se fosse $\bar{a} \overset{\to}{=} a$, poichè $a \to a_1 \to b_1$, seguirebbe $\bar{a} \to b_1$. Ma allora, da $a \to \bar{a}$, $b \to b_1$ si giunge ad un assurdo per (4"). Analogmente se $a_1 \to a$.

Corollario 1.5. *Nelle ipotesi precedenti, $I(C_\alpha, b) = I(C_\alpha, b_1)$.*

Dim. Pressochè ovvia.

Corollario 1.6. *Dette C_α e C_β due catene massimali diverse, $C_\alpha - C_\beta$ e $C_\beta - C_\alpha$ sono due intervalli di mutua indifferenza di cui almeno uno un singleton.*

Dim. In base alla 1.4, se C_α e C_β differiscono per più di un elemento, per esempio $|C_\beta - C_\alpha| > 1$, $b \in C_\beta - C_\alpha$, allora $C_\alpha - C_\beta = I(C_\alpha, b) =$

$= \{a\}$ e $C_\beta - C_\alpha = I(C_\beta, a)$. Se $|C_\alpha - C_\beta| = |C_\beta - C_\alpha| = 1$, allora $I(C_\alpha, b) = \{a\}$ e $I(C_\beta, a) = \{b\}$.

Proposizione 1.7. *Sia $b \notin C_\alpha$ non isolato. Se b è massimale, l'insieme $I(C_\alpha, b)$ è una semiretta positiva.*

Dim. Pressochè evidente, atteso che $\exists a \in C_\alpha : a\ I\ b$.

Definizione 1.8. Diremo che *due catene massimali sono equivalenti* e scriveremo $C_\alpha \approx C_\beta$ se, $\forall a \in C_\alpha$, $\forall b \in C_\beta$, $|I(C_\alpha, b)| = |I(C_\beta, a)| = 1$.

Proposizione 1.9. *Se vale* (4"), *allora $\approx$ è una equivalenza.*

Dim. Le proprietà riflessiva e simmetrica sono ovvie. Siano ora $C_\alpha \approx$ $\approx C_\beta$, $C_\beta \approx C_\gamma$ e dimostriamo che $C_\alpha \approx C_\gamma$.

Sia $a \in C_\alpha$. O $a \in C_\beta$ e non c'è nulla da dimostrare oppure $a \notin C_\beta$. Allora $I(C_\beta, a) = \{b\} \neq \{a\}$. In tal caso C_α e C_β differiscono per i due punti a e b e, per 1.6, solo per tali due punti. Ma $|I(C_\gamma, b)| = 1$. Se $I(C_\gamma, b) = \{b\}$ o $= \{a\}$ siamo a posto. Altrimenti $I(C_\gamma, b) = \{c\}$ $(\neq \{a\}, \neq \{b\})$. Dimostriamo che $a\ I\ c$. Se fosse $c \to a$, preso $a_1 \in C_\alpha$: : $a_1 \to a$, sarebbe $a_1 \to b$ e quindi, essendo $b\ I\ c$, $a_1 \to c$. Allora C_α non sarebbe massimale. Analogamente se fosse $a \to c$. Dunque $c \in I(C_\gamma, a)$. Se poi esistesse $c_1 \to c$, $c_1 \in I(C_\gamma, a)$, ne verrebbe $c_1 \to b$; inoltre $\forall b_1 \to b$, $b_1 \in C_\beta$, sarà anche $b_1 \in C_\alpha$, $b_1 \to a$ da cui ne segue $b_1 \to c_1$ $(\in I(C_\gamma, a))$. Allora $b_1 \to c_1 \to b$ e C_β non sarebbe massimale. Analogamente non può esistere c_1 tale che $c \to c_1$, $c_1 \in I(C_\gamma, a)$. Dunque $|I(C_\gamma, a)| = 1$. Analogamente $|I(C_\alpha, c)| = 1$.

Definizione 1.10. Diremo che $C_\alpha \prec C_\beta \Leftrightarrow |C_\beta - C_\alpha| > 1$.

Osservazione. Se $C_\alpha \prec C_\beta$, e se vale (4"), $|C_\alpha - C_\beta| = 1$ e quindi $C_\alpha \prec C_\beta \Rightarrow C_\beta \not\prec C_\alpha$.

Proprietà 1.11. *Per la relazione in* 1.10 *vale la proprietà transitiva:* $C_\alpha \prec C_\beta$, $C_\beta \prec C_\gamma \Rightarrow C_\alpha \prec C_\gamma$.

Dim. Sia $\{a\} = C_\alpha - C_\beta$; $\{b\} = C_\beta - C_\gamma$ e consideriamo $I(C_\beta, a) =$ $= C_\beta - C_\alpha$. Siano $b_1, b_2 \in C_\beta - C_\alpha$; per esempio $b_1 \to b_2$; siano $c_1, c_2 \in$ $\in C_\gamma - C_\beta$; per esempio $c_1 \to c_2$. Allora $b \in I(C_\beta, a)$. Se fosse $b \notin$ $\notin I(C_\beta, a)$, sarebbe ad esempio $b \to a$. Ma allora $b \to b_1$ e, siccome $b\ I\ c_1$, sarebbe $c_1 \to b_1$. Da $b \to a$, $c_1 \to b_1$ seguirebbe, per (4"),

$(b \not\to c_1) \vee (a \not\to b_1)$; impossibile. Ora $c_1 \neq a$. Se fosse $c_1 = a$, da $c_1 \to c_2$, seguirebbe $b \to c_2$, impossibile. Analogamente $c_2 \neq a$. Inoltre $c_1 \notin C_\alpha$, altrimenti $(c_1 \to a) \vee (a \to c_1)$ da cui $(c_1 \to b) \vee (b \to c_1)$, impossibile. Analogamente $c_2 \notin C_\alpha$.

Proposizione 1.12. *La relazione $\approx$ è compatibile con la $\prec$.*

Dim. Sia $C'_\alpha \approx C''_\alpha \prec C_\beta$. Dimostriamo che $C'_\alpha \prec C_\beta$. Siano $b_1, b_2 \in$ $\in C_\beta$, inconfrontabili con $\bar{a} \in C''_\alpha$. Se $\bar{a} \in C'_\alpha$ siamo a posto. Altrimenti $\bar{a} = a'' = C''_\alpha - C'_\alpha$; sia $\{a'\} = C'_\alpha - C''_\alpha$. Potrebbe essere $b_1 \circ b_2 = a'$? Se fosse per esempio $b_2 = a'$, avremmo $b_1 \to a'$. Se $a' = \min C'_\alpha$, avremmo un assurdo (C'_α massimale). Altrimenti $\exists a \to a'$, $a \in C'_\alpha$. Allora $a \to a''$. $\forall a \to a'$, $a \in C'_\alpha$ avremmo $a \to a''$, $b_1 \to a' \Rightarrow a \to b_1$ (non può essere $b_1 \to a$, altrimenti $b_1 \to a''$). Ma allora C'_α non sarebbe massimale. Da ciò la tesi.

Proposizione 1.13. *Dati C_α, C_β è $(C_\alpha \prec C_\beta) \vee (C_\alpha \approx C_\beta) \vee (C_\alpha \succ$ $\succ C_\beta)$ e quindi $(E/\approx, \prec)$ è un ordine lineare.*

Dim. Evidente in base alle precendenti proprietà dimostrate.

Teorema 1.14. *Sia E senza punti isolati e poniamoci in $E/\approx$. Posto $C = \bigcup_\alpha C_\alpha - \bigcup_{\alpha\beta}(C_\alpha - C_\beta)$ con $|C_\alpha - C_\beta| = 1$, C_α famiglia delle catene massimali a 2 a 2 non equivalenti, C è una catena massimale.*

Dim. Siano $x_1, x_2 \in C$; $x_1 \neq x_2$. Allora $\exists \bar{\alpha} : \{x_1\} \cup \{x_2\} \in C_{\bar{\alpha}}$ da cui $(x_1 \to x_2) \vee (x_2 \to x_1)$. Dunque C è una catena. Proviamo che è massimale. Sia C non massimale. Esiste allora una catena massimale $C_{\bar{\alpha}} \supset C$. Sia $y \in C_{\bar{\alpha}} - C$. Allora $\exists \beta_i, \alpha_i : \{y\} = C_{\beta_i} - C_{\alpha_i}$; ma allora y sarebbe inconfrontabile con i punti di $C_{\alpha_i} - C_{\beta_i}$ che sono più di uno. Ma y è confrontabile con ogni punto di C ed in C stanno i punti di ciascuna catena C_α escluso al più uno. Dalla contraddizione si ha la tesi.

Corollario 1.15. *Esiste in $(E/\approx, \prec)$ una catena massimale massima.*

Dim. $|C - C_\alpha| \neq 1$ $\forall \alpha$. Dunque C, nell'ordine $(E/\approx, \prec)$ è le massima catena massimale.

Sia allora C tale catena massimale. Ogni altro punto x di E o è

isolato o è inconfrontabile con un intervallo di C. Se tale intervallo è un singleton $\{a\}$, allora $C \approx (C - \{a\}) \cup \{x\}$. Altrimenti è un intervallo $I(C, x)$. Si ha allora

Proposizione 1.16. $x_1 \neq x_2 \quad \Rightarrow \quad [I(C, x_1) \subset I(C, x_2)] \vee [I(c, x_1) \supset \supset I(C, x_2)]$.

Dim. Se $I(C, x_1)$ o $I(C, x_2) = C$, la proposizione è vera. Siano allora diversi da C e supponiamo la tesi non valga. Esiste allora $a_1 \in I(C, x_1)$, $a_1 \notin I(C, x_2)$ e $a_2 \in I(C, x_2)$, $a_2 \notin I(C, x_1)$. Sia per esempio $a_1 \to \to a_2$. D'altra parte $x_1 \; I \; x_2$ perchè altrimenti C non sarebbe massimale. Allora dev'essere $x_1 \to a_2$ e $a_1 \to x_2$, da cui, per (4"'), dovrebbe seguire $(x_1 \not\to a_1) \vee (x_2 \not\to a_2)$, impossibile.

Corollario 1.17. *Data la catena C di* 3.15, *l'insieme degli intervalli di inconfrontabilità di C con gli $x \notin C$ è linearmente ordinato per inclusione.*

Dim. Ovvia.

Proposizione 1.18 (Theorema di Fishburn). *Se $E/\approx$ è numerabile, esistono due funzioni a valori reali $v(x)$ e $\sigma(x)$, con $\sigma(x) \geq 0$, tali che $x_1 \to x_2 \Leftrightarrow v(x_1) + \sigma(x_1) < v(x_2)$.*

Dim. Sia $\bar{v}(x)$ una qualunque funzione di valutazione su C con $\inf \bar{v}(C) > -\infty$ e $\sup \bar{v}(C) < +\infty$. Posto $v(x) = \bar{v}(x)$ e $\sigma(x) = 0 \; \forall x \in C$ e $v(x) = \inf \bar{v}(I(C, x))$, $\sigma(x) = \sup \bar{v}(I(C, x)) - \inf \bar{v}(I(C, x))$, si ha la tesi.

Nota. Si noti che l'assioma (4"') è più debole, ossia implica, l'assioma (4"). Dunque la dimostrazione del teorema avviene in una situazione assiomaticamente più semplice che nel teorema originale di Fishburn, dove ci si mette nel caso degli interval orders, ossia dell'assioma (4"). Tuttavia ritengo sia possibile usare tecniche analoghe, ancorchè più complesse, per dimostrare il teorema originale e mi riprometto di ritornare sull'argomento.

Bibliografia

[1] FISHBURN, P.C.: Utility theory for decision making, Wiley, New York. Reprinted by Krieger, Huntington, New York, 1979.

[2] FISHBURN, P.C.: Interval orders and interval graphs, Wiley interscience Series in Discret' Mathematics, New York, 1985.

[3] ISLER, R.: Su alcuni assiomi in teoria delle decisioni, *Quad. matem. Dip. mat. appl. "B. de Finetti"*, 4/88, Trieste.

[4] FRENCH, S.: Decision Theory, Ellis Horwood Series in Mathematics and its applications, 1986, Chichester, England.