Elem. Math. 60 (2005) 25 – 32
0013-6018/05/010025-8

Elemente der Mathematik

An interesting application of algebra to genetics

Helmut Länger

Helmut Länger studied mathematics at the Vienna University of Technology where he received his Ph.D. in 1976. Since 1984 he holds the position of an associate professor at the Institute of Discrete Mathematics and Geometry of the mentioned university. His main research interests are algebra, foundations of axiomatic quantum mechanics and discrete mathematics.

1 Introduction

From the beginnings algebraic methods were used for investigating genetic principles and structures. In particular, this is the case with so-called factor-union phenotype systems introduced by Cotterman ([1]). In these systems a set of properties can be assigned to each gene in such a way that phenotypes are determined by unions of these sets. These properties which can be considered to correspond to imaginary or actual physical factors may help in explaining and understanding the evolution and structure of phenotype systems.

In the literature there exist several algorithms for deciding if a given phenotype system possesses a so-called factor-union representation and for constructing such a representation (cf. e.g. [7], [3] and [4]). (In [5] some results of [3] are generalized.) We mainly follow the method published in [4]. However, the presentation given here explains in more detail the algebraic background and so is giving more insight into the mutual relations between algebra and genetics. Thus, the reader may better understand the main algebraic ideas and methods forming the background for the provided algorithm solving a problem of gene-

Merkmalsausprägungen (sogenannte Phänotypen) bei Individuen werden im einfachsten Fall durch ein Genpaar (einen sogenannten Genotyp), das sich an einem bestimmten Genort befindet, bestimmt. Es ist bekannt, dass verschiedene Genotypen dieselbe Merkmalsausprägung hervorrufen können. Vielfach ist es möglich, dieses Phänomen dadurch zu erklären, dass man jedem Gen gewisse Faktoren zuordnet. Im vorliegenden Artikel geht es um die Frage, wie man erkennen kann, ob eine solche Zuordnung überhaupt existiert, bzw. wie man eine solche finden kann. Dabei gelingt es dem Autor zu zeigen, dass sich sehr allgemeine Konzepte aus dem Gebiet der Algebra bei der Lösung des genannten Problems als nützlich erweisen.

tics. Moreover, it is shown that some of the basic algebraic ideas used in this paper follow from results in universal algebra that can be formulated in a very general way.

We start by explaining some fundamental notions of genetics and then giving an illustrative example.

The fundamental idea of mathematical population genetics is the fact that certain properties of individuals depend on a couple of so-called "genes" which are located at a certain "locus". This couple of genes is called a "genotype". Different genotypes may cause the same property, meaning, they may belong to the same "phenotype". It is natural to assume that certain "factors" assigned to each single gene are responsible for the occurrence of this phenomenon. The following example will illustrate this in more detail:

Example 1.1. The human A_1A_2BO-blood group system is based on the four genes A_1, A_2, B and O. The blood groups (phenotypes) A_1, A_2, A_1B, A_2B, B and O correspond to the following genotypes:

Phenotype system of blood groups

blood group	corresponding genotypes
A_1	A_1A_1, A_1A_2, A_1O
A_2	A_2A_2, A_2O
A_1B	A_1B
A_2B	A_2B
B	BB, BO
O	OO

Now the question arises if this correspondence between blood groups and genotypes can be explained by assigning to each gene x a set $f(x)$ of certain "factors" in such a way that two genotypes yz and uv correspond to the same blood group if and only if $f(y) \cup f(z) = f(u) \cup f(v)$. If we assign to the genes A_1, A_2, B and O some of the factors 1, 2, 3 and 4 according to the following table:

gene	assigned factors
A_1	1, 2, 4
A_2	2, 4
B	3, 4
O	4

then this is the case since to the genotypes A_1A_1, A_1A_2, A_1O, A_2A_2, A_2O, A_1B, A_2B, BB, BO and OO there are then assigned factors according to the table on the top of the next page.

Now the following problems arise:

Problem 1 Decide if a given phenotype system possesses a factor-union representation.

Problem 2 Construct such a representation if it exists.

Problem 3 Is the representation (if it exists) unique up to some identification?

Problem 4 If a representation exists, can one find a minimal one (with a minimum number of factors)?

genotype	assigned factors
A_1A_1	1, 2, 4
A_1A_2	1, 2, 4
A_1O	1, 2, 4
A_2A_2	2, 4
A_2O	2, 4
A_1B	1, 2, 3, 4
A_2B	2, 3, 4
BB	3, 4
BO	3, 4
OO	4

E.g., the representation given in Example 1.1 is not minimal (as indicated at the end of the paper).

The aim of this paper is to present the algorithm published in [4] for solving the first two of these problems and to explain the corresponding algebraic background in a clear manner in more detail.

2 Formulation of the problem in mathematical terms

Let G be a fixed finite non-empty set of genes and G_2 denote the set of all one- or two-element subsets of G. G_2 may be considered as the set of all genotypes where each genotype xy is identified with the set $\{x, y\}$. A phenotype system α is nothing else than an equivalence relation on G_2, so may be considered as a subset of $G_2 \times G_2$. By a factor-union representation of α we understand a mapping f assigning to each element of G a certain set such that

$$\Big\{(A, B) \in G_2^2 \;\Big|\; \bigcup_{x \in A} f(x) = \bigcup_{x \in B} f(x)\Big\} = \alpha.$$

α is called a factor-union system if it possesses a factor-union representation. Now the first two of the above questions can be formulated as follows: Is a given phenotype system a factor-union system? If it is a factor-union system, how could one construct a corresponding factor-union representation?

3 Algebraic background

The basic algebraic structure used in the following is that of a semilattice. A semilattice is a commutative idempotent semigroup. There is a natural bijective correspondence between semilattices $(S, \vee)$ and posets $(S, \leq)$ every two elements of which have a supremum. (Here and in the following the term "poset" is used as an abbreviation of the term "partially ordered set".) The correspondence is given by

$$x \leq y \text{ if and only if } x \vee y = y \qquad \text{resp.} \qquad x \vee y := \sup(x, y).$$

If A is an arbitrary set and B denotes the set of all finite non-empty subsets of A then $(B, \cup)$ is a so-called free semilattice with free generating set A where the elements of A are identified with their corresponding singletons. This means that every mapping f from A to the base set S of some semilattice $(S, \vee)$ can be uniquely extended to a homomorphism g from $(B, \cup)$ to $(S, \vee)$, namely via $g(x) := \bigvee_{z \in x} f(z)$ for all $x \in B$. If A coincides with the finite non-empty set G then $B = 2^G \setminus \{\emptyset\}$. From the fact that $(2^G \setminus \{\emptyset\}, \cup)$ is a free semilattice with free generating set G and from the definition of a factor-union representation of a phenotype system one obtains

Remark 3.1. The factor-union systems are exactly the restrictions of the kernels of the homomorphisms from $(2^G \setminus \{\emptyset\}, \cup)$ to semilattices of the form $(2^F, \cup)$ (with an arbitrary set F) to G_2 since they arise by assigning to each element of G a certain subset of F and by extending this mapping f from G to 2^F to a mapping $\bar{f}$ from G_2 to 2^F by defining $\bar{f}(\{x, y\}) := f(x) \cup f(y)$ for all $x, y \in G$. Hence $\bar{f}$ may also be considered as the restriction of the unique extension of f to a homomorphism from $(2^G, \cup)$ to $(2^F, \cup)$ to G_2.

In order to see that these kernels are exactly the congruences on $(2^G \setminus \{\emptyset\}, \cup)$ we need a representation theorem for semilattices. But first we consider a more general situation.

By an algebra we mean a set together with a (possibly infinite) family of finitary operations on it. The corresponding family of the varieties of the operations is called the type of the algebra. A variety is an equationally definable class of algebras of the same type, i.e. the class of all algebras of a fixed type which satisfy a fixed set of laws. For every class $\mathcal{K}$ of algebras of the same type $\mathbf{H}(\mathcal{K})$, $\mathbf{I}(\mathcal{K})$ and $\mathbf{S}(\mathcal{K})$ denote the class of all homomorphic images, isomorphic images and subalgebras of members of $\mathcal{K}$, respectively. By the kernel of a mapping f with domain M we mean the equivalence relation $\{(x, y) \in M^2 \mid f(x) = f(y)\}$ on M. Now we can state the following

Lemma 3.1. *If $\mathcal{K}_1, \mathcal{K}_2$ are classes of algebras of the same type, $\mathbf{H}(\mathcal{K}_1) \subseteq \mathbf{I}(\mathbf{S}(\mathcal{K}_2))$ and $\mathcal{A} \in \mathcal{K}_1$ then the congruences on $\mathcal{A}$ are exactly the kernels of the homomorphisms from $\mathcal{A}$ to members of $\mathcal{K}_2$.*

Proof. Let Θ be a congruence on $\mathcal{A}$. Then $\mathcal{A}/\Theta \in \mathbf{H}(\{\mathcal{A}\}) \subseteq \mathbf{H}(\mathcal{K}_1) \subseteq \mathbf{I}(\mathbf{S}(\mathcal{K}_2))$. Hence there exists some $\mathcal{B} \in \mathcal{K}_2$ and some $\mathcal{C} \in \mathbf{S}(\{\mathcal{B}\})$ with $\mathcal{C} \cong \mathcal{A}/\Theta$. Let f denote the canonical homomorphism from $\mathcal{A}$ to $\mathcal{A}/\Theta$ and g an isomorphism from $\mathcal{A}/\Theta$ to $\mathcal{C}$. Then g can be regarded as a homomorphism from $\mathcal{A}/\Theta$ to $\mathcal{B}$. Since g is injective, $g \circ f$ has the same kernel as f and hence Θ is also the kernel of the homomorphism $g \circ f$ from $\mathcal{A}$ to the member $\mathcal{B}$ of $\mathcal{K}_2$. □

As a consequence we obtain

Corollary 3.1. *If $\mathcal{V}$ is a variety, $\mathcal{K}$ a subclass of $\mathcal{V}$ such that every member of $\mathcal{V}$ can be embedded into some member of $\mathcal{K}$ and $\mathcal{A} \in \mathcal{V}$ then $\mathbf{H}(\mathcal{V}) = \mathcal{V} \subseteq \mathbf{I}(\mathbf{S}(\mathcal{K}))$ and hence the congruences on $\mathcal{A}$ are exactly the kernels of the homomorphisms from $\mathcal{A}$ to members of $\mathcal{K}$.* □

Now we state the above mentioned representation theorem (cf. e.g. [6]; for the case of distributive lattices see [2]).

Theorem 3.1. (Representation theorem for semilattices) *Every semilattice* $(S, \vee)$ *can be embedded into* $(2^S, \cup)$.

Proof. If f denotes the mapping from S to 2^S defined by $f(x) := \{y \in S \mid y \not\geq x\}$ for all $x \in S$ then since $x = \bigwedge(S \setminus f(x))$ for all $x \in S$, f is injective and since for any three elements a, b, c of S, $c \geq a \vee b$ is equivalent to ($c \geq a$ and $c \geq b$), f is a homomorphism from $(S, \vee)$ to $(2^S, \cup)$. □

Combining our results we obtain

Proposition 3.1. *The kernels of the homomorphisms from* $(2^G \setminus \{\emptyset\}, \cup)$ *to semilattices of the form* $(2^F, \cup)$ (*with an arbitrary set* F) *are exactly the congruences on* $(2^G \setminus \{\emptyset\}, \cup)$.

Proof. This follows from Theorem 3.1 and Corollary 3.1 by specializing $\mathcal{V}$ to the variety of semilattices, $\mathcal{K}$ to the class of all algebras of the form $(2^F, \cup)$ (with an arbitrary set F) and $\mathcal{A}$ to the algebra $(2^G \setminus \{\emptyset\}, \cup)$. □

Combining Remark 3.1 with Proposition 3.1 yields (cf. [4])

Corollary 3.2. *The factor-union systems are exactly the restrictions of the congruences on* $(2^G \setminus \{\emptyset\}, \cup)$ *to* G_2. □

This result can be sharpened as follows (cf. [4]):

Proposition 3.2. *A phenotype system* α *is a factor-union system if and only if it is the restriction of the congruence on* $(2^G \setminus \{\emptyset\}, \cup)$ *generated by* α *to* G_2.

Proof. If α is the restriction of a congruence Φ on $(2^G \setminus \{\emptyset\}, \cup)$ to G_2 and Θ denotes the congruence on $(2^G \setminus \{\emptyset\}, \cup)$ generated by α then $\Theta \subseteq \Phi$ and hence

$$\alpha \subseteq \Theta \cap G_2^2 \subseteq \Phi \cap G_2^2 = \alpha$$

which shows $\alpha = \Theta \cap G_2^2$. The assertion of the lemma now follows from Corollary 3.2. □

How can one construct the congruence on $(2^G \setminus \{\emptyset\}, \cup)$ generated by a given phenotype system? Since an equivalence relation Θ on the base set S of a semilattice $(S, \vee)$ is a congruence on $(S, \vee)$ if and only if $(x, y) \in \Theta$ and $z \in S$ imply $(x \vee z, y \vee z) \in \Theta$, the following result is easy to verify (cf. [4]):

Lemma 3.2. *If* α *is a phenotype system then the congruence on* $(2^G \setminus \{\emptyset\}, \cup)$ *generated by* α *is the transitive closure of* $\{(x \cup z, y \cup z) \mid (x, y) \in \alpha, z \subseteq G\}$. □

Now we can present a method for constructing a factor-union representation of a factor-union system.

Theorem 3.2. (Construction of a factor-union representation) *If* α *is a factor-union system and* Θ *denotes the congruence on* $(2^G \setminus \{\emptyset\}, \cup)$ *generated by* α *then the mapping* f *from* G *to* $(2^{(2^G \setminus \{\emptyset\})/\Theta}, \cup)$ *defined by* $f(x) := \{y \in (2^G \setminus \{\emptyset\})/\Theta \mid y \not\geq [\{x\}]\Theta\}$ *for all* $x \in G$ *is a factor-union representation of* α.

Proof. Since Θ is the kernel of the canonical homomorphism g from $(2^G \setminus \{\emptyset\}, \cup)$ to $((2^G \setminus \{\emptyset\})/\Theta, \cup)$ and the mapping h from $(2^G \setminus \{\emptyset\})/\Theta$ to $2^{(2^G \setminus \{\emptyset\})/\Theta}$ defined by $h(x) := \{y \in (2^G \setminus \{\emptyset\})/\Theta \mid y \not\geq x\}$ for all $x \in (2^G \setminus \{\emptyset\})/\Theta$ is an embedding of $((2^G \setminus \{\emptyset\})/\Theta, \cup)$ into $(2^{(2^G \setminus \{\emptyset\})/\Theta}, \cup)$ according to the proof of Theorem 3.1, $h \circ g$ is a homomorphism from $(2^G \setminus \{\emptyset\}, \cup)$ to $(2^{(2^G \setminus \{\emptyset\})/\Theta}, \cup)$ with kernel Θ which together with $\Theta \cap G_2^2 = \alpha$ (which holds according to Proposition 3.2) shows that the mapping f from G to $2^{(2^G \setminus \{\emptyset\})/\Theta}$ defined by $f(x) := \{y \in (2^G \setminus \{\emptyset\})/\Theta \mid y \not\geq [\{x\}]\Theta\}$ for all $x \in G$ is a factor-union representation of α. □

Remark 3.2. If a given phenotype system α with n genes has a factor-union representation then 2^n factors are sufficient. Hence the problem formulated in the beginning could be solved in a finite number of steps by taking a fixed 2^n-element set F of factors and checking all $(2^{2^n})^n = 2^{n2^n}$ mappings from G to 2^F if they are factor-union representations of α or not. In [7] it was proved that even α factors suffice.

Remark 3.3. The number of factors used in the factor-union representation described in Theorem 3.2 can be reduced by using an improved version of the representation theorem for semilattices. As a sharpening of the result in Theorem 3.1 it can be proved that every semilattice $(S, \vee)$ can be embedded into the power sets over a subset of S. In order to see this let us define meet-irreducible elements of a poset.

An element of a poset is called meet-irreducible if it is not the meet of two other elements. A poset is said to satisfy the ascending chain condition if every ascending chain is finite.

Now we prove the following lemma:

Lemma 3.3. *In every poset $(P, \leq)$ satisfying the ascending chain condition every element a is the meet of finitely many meet-irreducible elements.*

Proof. Let M denote the set of all meet-irreducible elements of $(P, \leq)$. If $a \in M$ we are done. Otherwise there exist $b, c \in P \setminus \{a\}$ with $a = b \wedge c$. If $b, c \in M$ we are done. If $b \notin M$ then there exist $d, e \in P \setminus \{b\}$ with $b = d \wedge e$. Then $a = d \wedge e \wedge c$. Since $(P, \leq)$ satisfies the ascending chain condition, the described procedure has to terminate after a finite number of steps thus finally arriving at finitely many elements of M the meet of which is a. □

A direct consequence of Lemma 3.3 is

Corollary 3.3. *In every poset satisfying the ascending chain condition every element is the meet of its meet-irreducible upper bounds.* □

Now we are ready to prove (cf. e.g. [6]; for the case of distributive lattices see [2])

Theorem 3.3. (Improved version of the Representation theorem for semilattices) *Every semilattice $(S, \vee)$ satisfying the ascending chain condition can be embedded into $(2^M, \cup)$ where M denotes the set of all meet-irreducible elements of $(S, \leq)$.*

Proof. If f denotes the mapping from S to 2^M defined by $f(x) := \{y \in M \mid y \not\geq x\}$ for all $x \in S$ then, since $x = \bigwedge(M \setminus f(x))$ for all $x \in S$ according to Corollary 3.3, f is an injective homomorphism from $(S, \vee)$ to $(2^M, \cup)$ which follows in an analogous way as in the proof of Theorem 3.1. □

The improved version of our theorem describing the construction of a factor-union representation can now be formulated as follows (cf. [4]):

Theorem 3.4. (Construction of a smaller factor-union representation) *If α is a factor-union system, Θ denotes the congruence on $(2^G \setminus \{\emptyset\}, \cup)$ generated by α and M denotes the set of all meet-irreducible elements of $((2^G \setminus \{\emptyset\})/\Theta, \cup)$ then the mapping f from G to 2^M defined by $f(x) := \{y \in M \mid y \not\geq [\{x\}]\Theta\}$ for all $x \in G$ is a factor-union representation of α.* □

4 The algorithm

Now we can present an algorithm for solving the first two of the problems stated at the beginning.

Algorithm for checking if a given phenotype system α is a factor-union system and for constructing a corresponding factor-union representation (cf. [4])

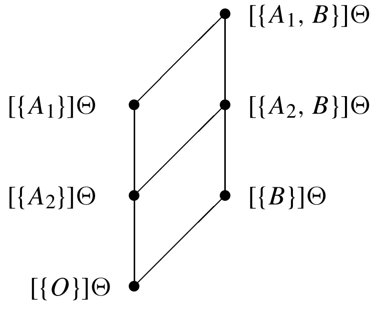
Construct the congruence Θ on $(2^G \setminus \{\emptyset\}, \cup)$ generated by α by forming the transitive closure of $\{(x \cup z, y \cup z) \mid (x, y) \in \alpha, z \subseteq G\}$ (Lemma 3.2). If $\Theta \cap G_2^2 \neq \alpha$ then α is not a factor-union system (Proposition 3.2). Otherwise construct the Hasse diagram of $((2^G \setminus \{\emptyset\})/\Theta, \leq)$. Let M denote the set of all meet-irreducible elements of $((2^G \setminus \{\emptyset\})/\Theta, \leq)$. Then the mapping f from G to 2^M defined by $f(x) := \{y \in M \mid y \not\geq [\{x\}]\Theta\}$ for all $x \in G\}$ is a factor-union representation of α (Theorem 3.4).

Now we return to our introductory example.

Example 4.1. We have

$$\begin{aligned}
G &= \{A_1, A_2, B, O\},\\
\alpha &= \{\{A_1\}, \{A_1, A_2\}, \{A_1, O\}\}^2 \cup \{\{A_2\}, \{A_2, O\}\}^2 \cup \{\{A_1, B\}\}^2 \cup \{\{A_2, B\}\}^2 \cup\\
&\quad \cup \{\{B\}, \{B, O\}\}^2 \cup \{\{O\}\}^2,\\
\Theta &= \{\{A_1\}, \{A_1, A_2\}, \{A_1, O\}, \{A_1, A_2, O\}\}^2 \cup \{\{A_2\}, \{A_2, O\}\}^2 \cup\\
&\quad \cup \{\{A_1, B\}, \{A_1, A_2, B\}, \{A_1, B, O\}, \{A_1, A_2, B, O\}\}^2 \cup \{\{A_2, B\}, \{A_2, B, O\}\}^2 \cup\\
&\quad \cup \{\{B\}, \{B, O\}\}^2 \cup \{\{O\}\}^2,
\end{aligned}$$

where Θ denotes the congruence on $(2^G \setminus \{\emptyset\}, \cup)$ generated by α. The Hasse diagram of $((2^G \setminus \{\emptyset\})/\Theta, \leq)$ looks as follows:



Hence, the mapping f from G to 2^M (where M denotes the set $\{[\{A_1\}]\Theta, [\{A_1, B\}]\Theta, [\{A_2, B\}]\Theta, [\{B\}]\Theta\}$ of all meet-irreducible elements of $((2^G \setminus \{\emptyset\})/\Theta, \leq))$ defined by

$$\begin{aligned} f(A_1) &:= \{[\{B\}]\Theta, [\{A_2, B\}]\Theta\}, \\ f(A_2) &:= \{[\{B\}]\Theta\}, \\ f(B) &:= \{[\{A_1\}]\Theta\}, \\ f(O) &:= \emptyset \end{aligned}$$

is a factor-union representation of α.

Investigating the computational complexity of the proposed algorithm seems to be very difficult. Forming the transitive closure of the described binary relation may be a long procedure if G is large. If Θ has k classes then $\binom{k}{2}$ comparisons are necessary in order to determine the factor poset $((2^G \setminus \{\emptyset\})/\Theta, \leq)$. In order to determine the meet-irreducible elements one has to consider the possible infimum of any two distinct elements of the factor poset. The number of these pairs is again $\binom{k}{2}$. Software packages for algebraic structures may be used in order to apply the proposed algorithm in an as effective as possible way.

References

[1] Cotterman, C.W.: Factor-union phenotype systems. Computer Applications in Genetics (ed. by N.E. Morton), Univ. of Hawaii Press, 1969, 1–19.

[2] Grätzer, G.: *Lattice Theory*. Freeman, San Francisco 1971.

[3] Karigl, G.: Factor-union representation in phenotype systems. *Contr. General Algebra* 6 (1988), 123–130.

[4] Länger, H.: Factor-union representation of phenotype systems. *Math. Pannon.* 1 (1990), 107–110.

[5] Länger, H.: A lattice-theoretical description of phenotype systems. *Contr. General Algebra* 7 (1991), 247–250.

[6] Markowsky, G.: The representation of posets and lattices by sets. *Algebra Universalis* 11 (1980), 173–192.

[7] Markowsky, G.: Necessary and sufficient conditions for a phenotype system to have a factor-union representation. *Math. Biosci.* 66 (1983), 115–128.

Helmut Länger
Institute of Discrete Mathematics and Geometry
Vienna University of Technology
Wiedner Hauptstraße 8–10
A-1040 Wien, Austria
e-mail: h.laenger@tuwien.ac.at