Tomus 2. Fasciculus 1.

ACTA CYBERNETICA

FORUM CENTRALE PUBLICATIONUM CYBERNETICARUM HUNGARICUM

Szeged, 1973

Curat: Universitas Szegediensis de Attila József nominata

2. kötet 1. füzet

ACTA CYBERNETICA

A HAZAI KIBERNETIKAI KUTATÁSOK KÖZPONTI PUBLIKÁCIÓS FÓRUMA

Szeged, 1973. március

A szegedi József Attila Tudományegyetem gondozásában

On some aspects of the algebraic description of automaton mappings

By A. ÁDÁM

Introduction

The present paper is devoted to studying the super-finite partitions of finitely generated (non-commutative) free semigroups, i.e. such partitions C for which the relation $C \geqq C^*$ is satisfiable with a right-congruence C^* of finite index. The importance of super-finite partitions arises from the fact that they are in a natural one-to-one correspondence with the automaton mappings realizable by finite automata.

The (sufficiently constructive) description of the super-finite partitions seems to be a difficult task. The present article is intended to make only the first steps to this direction; consequently, the introduction of the concepts and the elucidation of their easily accessible properties take up a remarkable size in the paper.

Chapter I contains a survey of the (more or less known) correspondences between automaton mappings and partitions of semigroups (§§ 1—2); furthermore, after summarizing the previous results on finite right-congruences published in [2] (§ 3), the main purpose of the investigations is exposed (§ 4).[1, 2]

Chapters II, III explain certain suggestions in order to give a description of the super-finite partitions of finitely generated free semigroups, and obtain some results in this direction. These two chapters are independent of each other, the same problem is attacked by two different methods in them. Especially, the results of Chapter II give an answer to the following problem: determine all partitions C of a finitely generated free semigroup $F(X)$ such that C is no right-congruence and, by forming the union of two classes modulo C, a previously given right-congruence C^* is obtained (any other classes mod C remain unchanged). In Chapter III, the critical pairs of a right-congruence of $F(X)$ are characterized.

[1] The results exposed in § 1—2 are given with or without proof; in the latter case, we refer to the paper [1] where related questions are treated.

[2] We note that the basic correspondence, asserted in Proposition 8, was firstly discovered by Nerode [4]; see also [5], [6].

I. The super-finite partitions and their fundamental properties

§ 1

As in [2], we denote by $F(X)$ the free semigroup (non-commutative, with unit element e) generated by the finite set $X=\{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}$.[3] The elements of $F(X)$ are called also *words*, the elements of X are called *generators*, too. The *length* $l(p)$ of a word p is the number of generators whose product equals to p. $\mathfrak{K}_i(p)$, $\mathfrak{V}_i(p)$ are defined by

$$p = \mathfrak{K}_i(p)\cdot\mathfrak{V}_i(p), \quad l(\mathfrak{V}_i(p))=i(\leqq l(p)).$$

Evidently, $\mathfrak{V}_1(\mathfrak{K}_i(p))$ equals to the $(l(p)-i)$-th factor in the "product of generators" representation of p $(0\leqq i<l(p))$.

The *index* ind C of a partition C of $F(X)$ is the number of classes modulo C. We note that $C_1\leqq C_2$ implies ind $C_1\geqq$ind C_2. If the index of C is finite, then we say that C is a *finite partition*. The finite partitions form a sublattice $\mathfrak{L}_1$ of the lattice of all partitions of $F(X)$. A partition C is called a *right-congruence* if the implication

$$p\equiv q \pmod C \Rightarrow px\equiv qx \pmod C$$

is satisfied for every $p(\in F(X))$, $q(\in F(X))$, $x(\in X)$.

Let $\mathfrak{L}_2(\subseteqq\mathfrak{L}_1)$ be the lattice of all finite right-congruences (i.e. all finite partitions being right-congruences) of $F(X)$.

We say that the lattice $\mathfrak{L}$ possesses the *upper finiteness property* (abbreviated: UFP) if to any $C_1(\in\mathfrak{L})$ the relation $C_2>C_1$ is fulfilled only by a finite (possibly zero) number of elements $C_2(\in\mathfrak{L})$. $\mathfrak{L}$ has the *lower infiniteness property* (abbreviated: LIP) if to any $C_1(\in\mathfrak{L})$ there exists a $C_2(\in\mathfrak{L})$ such that $C_2<C_1$. (Consequently, there exists an infinity of C_2's with the desired character.) The lattices $\mathfrak{L}_1$, $\mathfrak{L}_2$ possess clearly both UFP and LIP.

Let C be a finite partition of $F(X)$. We define the partitions $\mathfrak{N}(C)$ and $\mathfrak{M}_1(C)$ by the following rules (see also [1]):

let $p\equiv q$ $(\bmod\ \mathfrak{N}(C))$ be true exactly if $p\equiv q \pmod C$ and $px\equiv qx \pmod C$ for each x (where $p\in F(X)$, $q\in F(X)$, $x\in X$),

let $p\equiv q$ $(\bmod\ \mathfrak{M}_1(C))$ be true exactly if $pr\equiv qr \pmod C$ for each r (where p, q, r are elements of $F(X)$).

Evidently, $\mathfrak{N}(C)\leqq C$ and $\mathfrak{M}_1(C)\leqq C$. We use the shorter notation $\mathfrak{N}^i(C)$ instead of

$$\overset{1}{\mathfrak{N}}(\overset{2}{\mathfrak{N}}(\overset{3}{\mathfrak{N}}(\ldots\overset{i}{\mathfrak{N}}(C)\ldots)))$$

and let $\mathfrak{N}^0(C)$ denote C.

For the (easy) proofs of the following Proposition 1 and Lemma 1, we refer to [1] (see there the assertions (1.2), (1.3), (2.12), (2.13)).

Proposition 1. *For any partition C of $F(X)$, $M_1(C)$ is a right-congruence, moreover, $\mathfrak{M}_1(C)$ is maximal among the right-congruences C^* satisfying $C^*\leqq C$.*

[3] X and $F(X)$ are always considered to be fixed.

Lemma 1. For any $C(\in \mathfrak{L}_1)$ *we have*

$$\text{ind}\, \mathfrak{N}(C) \leqq (\text{ind}\, C)^{n+1}$$

where n denotes the cardinality of X.

Lemma 2. $\mathfrak{N}^i(C) \geqq \mathfrak{M}_1(C)$ *holds for each i.*

Proof. We use induction for i. If $i=0$, then $\mathfrak{N}^0(C)=C\geqq\mathfrak{M}_1(C)$. Suppose $\mathfrak{N}^i(C)\geqq\mathfrak{M}_1(C)$, let $p\equiv q \pmod{\mathfrak{M}_1(C)}$ be true for the words p and q. The right-congruence property of $\mathfrak{M}_1(C)$ implies $px\equiv qx \pmod{\mathfrak{M}_1(C)}$ for each generator x; we get

$$p\equiv q \pmod{\mathfrak{N}^i(C)} \quad \text{and} \quad px\equiv qx \pmod{\mathfrak{N}^i(C)}$$

by the supposition. This means that p, q are congruent modulo $\mathfrak{N}(\mathfrak{N}^i(C))=\mathfrak{N}^{i+1}(C)$.

Proposition 2. *The subsequent three assertions are equivalent for any partition C of F(X):*

(i) *C is a right-congruence,*
(ii) $\mathfrak{N}(C)=C$,
(iii) $\mathfrak{M}_1(C)=C$.

Proof. Our preceding considerations show that $\mathfrak{M}_1(C)\leqq\mathfrak{N}(C)\leqq C$ for each C. Suppose $p\equiv q$ (mod C) where C is a right-congruence. We get $pr\equiv qr$ (mod C) for every word r (by successive application of the right-congruence property), thus $p\equiv q \pmod{\mathfrak{M}_1(C)}$, hence $C\leqq\mathfrak{M}_1(C)$; this implies (ii) and (iii).

Assume that C is not a right-congruence. There exist two words p, q and a generator x such that $p\equiv q$ (mod C) and $px\not\equiv qx$ (mod C). Hence $p\not\equiv q \mod (\mathfrak{N}(C))$, thus $\mathfrak{M}_1(C)\leqq\mathfrak{N}(C)<C$, consequently (ii) and (iii) are not fulfilled.

Proposition 3. *The following three conditions are equivalent for any finite partition C of F(X):*

(1) *There exists a finite right-congruence* C^* *such that* $C^*\leqq C$.
(2) *The right-congruence* $\mathfrak{M}_1(C)$ *is finite.*
(3) *There exists an integer* $i(\geqq 0)$ *such that* $\mathfrak{N}^i(C)=\mathfrak{N}^{i+1}(C)$.

The partitions (belonging to $\mathfrak{L}_1$) that satisfy the conditions posed in Proposition 3 are called *super-finite partitions* of $F(X)$. *This notion is of basic importance in the paper.* The set of super-finite partitions is denoted by $\mathfrak{L}_3$; clearly, $\mathfrak{L}_1\supseteqq\mathfrak{L}_3\supseteqq\mathfrak{L}_2$. We shall see that $\mathfrak{L}_3$ is a lattice, as well (Proposition 4).

Proof of Proposition 3.

(1)⇒(2). If $C^*\leqq C$ and C^* is a right-congruence, then $C^*\leqq\mathfrak{M}_1(C)$ by the minimality stated in Proposition 1, thus the finiteness of C^* implies the finiteness of $\mathfrak{M}_1(C)$.

(2)⇒(3). We prove the assertion indirectly. If (3) does not hold, then

$$C>\mathfrak{N}(C)>\mathfrak{N}^2(C)>\mathfrak{N}^3(C)>\ldots,$$

hence

$$\text{ind}\, \mathfrak{N}^i(C) \geqq i+\text{ind}\, C.$$

On the other hand, Lemma 2 implies

$$\text{ind}\ \mathfrak{M}_1(C) \geqq \text{ind}\ \mathfrak{N}^i(C)$$

for any i; consequently, $\mathfrak{M}_1(C)$ is of infinite index.

(3)⇒(1). If (3) is true, then $\mathfrak{N}^i(C)$ is a right-congruence by Proposition 2. A successive application of Lemma 1 shows that

$$\text{ind}\ \mathfrak{N}^i(C) \leqq (\text{ind}\ C)^{(n+1)^i},$$

therefore $\mathfrak{N}^i(C)$ belongs to $\mathfrak{L}_2$, i.e. $\mathfrak{N}^i(C)$ is a convenient C^* in (1).

Remarks. The equality in (3) implies

$$\mathfrak{M}_1(C) = \mathfrak{N}^i(C) = \mathfrak{N}^{i+1}(C) = \mathfrak{N}^{i+2}(C) = \dots$$

— [1] contains a detailed treatment of the equivalence of (2) and (3).

Proposition 4. *The set $\mathfrak{L}_3$ of super-finite partitions of $F(X)$ is a sublattice of the lattice of all partitions of $F(X)$. The lattice $\mathfrak{L}_3$ possesses both the upper finiteness property and the lower infiniteness property.*

Proof. In order to verify the first assertion, we have to prove that $C_1 \in \mathfrak{L}_3$ and $C_2 \in \mathfrak{L}_3$ imply $C_1 \cap C_2 \in \mathfrak{L}_3$ and $C_1 \cup C_2 \in \mathfrak{L}_3$. There exist two elements C_1^*, C_2^* of $\mathfrak{L}_2$ such that $C_1^* \leqq C_1$ and $C_2^* \leqq C_2$ (by Proposition 3, (1)). $C_1^* \cap C_2^*$ belongs to $\mathfrak{L}_2$ (since $\mathfrak{L}_2$ is a lattice) and the relations

$$C_1^* \cap C_2^* \leqq C_1 \cap C_2 \leqq C_1 \cup C_2$$

are obviously valid. Hence (1) is true for $C_1 \cap C_2$ and $C_1 \cup C_2$, too.

$\mathfrak{L}_3$ has the UFP because $\mathfrak{L}_1$ has; $\mathfrak{L}_3$ has the LIP since $\mathfrak{L}_2$ has.

§ 2

In this §, we treat the natural correspondence between the super-finite partitions of $F(X)$ and the finitely realizable automaton mappings of $F(X)$.

The customary definition of automaton mapping is: an assignment β, defined on $F(X)$, into a free semigroup[4] $F(Y)$ is called an *automaton mapping* (or *sequential function*) if

(1) $l(\beta(p)) = l(p)$ for each $p(\in F(X))$ and

(2) $\mathfrak{R}_1(\beta(p)) = \beta(\mathfrak{R}_1(p))$ for each $p(\in F(X) - \{e\})$.

An automaton mapping β is called to be *proper* if to any $y(\in Y)$ there exists a $p(\in F(X))$ such that $\mathfrak{B}_1(\beta(p)) = y$. The next result shows that the notion of proper automaton mapping is not an essential restriction of the general concept.

Proposition 5. *Let β be an automaton mapping of $F(X)$ into $F(Y)$. Define the set $Y_1(\subseteq Y)$ by the following rule: $y(\in Y)$ belongs to Y_1 exactly if there exists a $p(\in F(X))$ such that y occurs in the representation of $\beta(p)$ as a product of elements of Y (in other words: if there exist p and i such that $y = \mathfrak{B}_1(\mathfrak{R}_i(\beta(p)))$, $0 \leqq i < l(\beta(p))$). Then β is a proper mapping of $F(X)$ into $F(Y_1)$.*

---

[4] X and Y are (not necessarily disjoint) finite sets.

Proof: It is evident that the range of β is included in $F(Y_1)$. Let y be an arbitrary element of Y_1. Then

$$y=\mathfrak{B}_1(\mathfrak{R}_i(\beta(p)))=\mathfrak{B}_1(\beta(\mathfrak{R}_i(p)))$$

where the first equality follows from the definition of Y_1, the second one from property (2) defining the automaton mappings (applied successively i times). Thus β is proper if it is viewed as a mapping into $F(Y_1)$. The proof is completed.

Let β be an automaton mapping. We assign to β a (finite) partition C_β of $F(X)$ in the following way:

$$p\equiv q \pmod{C_\beta} \quad \text{if and only if} \quad \mathfrak{B}_1(\beta(p))=\mathfrak{B}_1(\beta(q)).$$

Now we state an assertion expressing that C_β is common for two mappings (defined on $F(X)$) precisely when they are isomorphic in a certain (natural) sense:

Proposition 6. *Consider two proper automaton mappings β and β' where β maps $F(X)$ into $F(Y)$ and β' maps $F(X)$ into $F(Y')$. The equality $C_\beta=C_{\beta'}$ holds if and only if ($|Y|=|Y'|$ and) there exists a one-to-one correspondence ι between Y and Y' such that*

$$\mathfrak{B}_1(\mathfrak{R}_i(\beta'(p)))=\iota(\mathfrak{B}_1(\mathfrak{R}_i(\beta(p))))$$

for every $p(\in F(X))$ and i $(0\leqq i<l(p))$.

Proof. Suppose $C_\beta=C_{\beta'}$. Let the assignment $\bar\beta$ of the factor set $F(X)/C_\beta$ into Y be defined by $\bar\beta(\bar p)=\mathfrak{B}_1(\beta(p))$ where $\bar p$ is the class (modulo C_β) containing p. $\bar\beta$ is clearly a one-to-one assignment *onto* Y (since β was supposed to be a proper mapping). $\bar\beta'$ can be defined in an analogous manner (with $C_{\beta'}$ instead of C_β). $\bar p$ may denote the class mod $C_{\beta'}$, as well. Introduce the mapping ι by the formula $\iota(y)=\bar\beta'(\bar\beta^{-1}(y))$. Then we have

$$\mathfrak{B}_1(\mathfrak{R}_i(\beta'(p))) = \mathfrak{B}_1(\beta'(\mathfrak{R}_i(p))) = \bar\beta'(\overline{\mathfrak{R}_i(p)}) =$$
$$= \iota(\bar\beta(\overline{\mathfrak{R}_i(p)})) = \iota(\mathfrak{B}_1(\beta(\mathfrak{R}_i(p)))) = \iota(\mathfrak{B}_1(\mathfrak{R}_i(\beta(p)))).$$

Conversely, assume that an assignment ι satisfies the condition and $z=\iota(y)$ (where $y\in Y$, $z\in Y'$). Define the sets $W_y^\beta(\subseteqq F(X))$, $W_z^{\beta'}(\subseteqq F(X))$ by what follows:

$$p\in W_y^\beta \quad \text{if and only if} \quad \mathfrak{B}_1(\beta(p)) = y,$$
$$p\in W_z^{\beta'} \quad \text{if and only if} \quad \mathfrak{B}_1(\beta'(p)) = z.$$

The equivalence

$$\mathfrak{B}_1(\beta(p))=y \;\Leftrightarrow\; (\mathfrak{B}_1(\beta'(p))=)\,\iota(\mathfrak{B}_1(\beta(p)))=\iota(y)(=z)$$

assures $W_y^\beta=W_z^{\beta'}$. This holds for each y and $\iota(y)=z$, consequently $C_\beta=C_{\beta'}$.

Proposition 7. *To any finite partition C of $F(X)$, there exists an automaton mapping β (defined on $F(X)$) such that $C_\beta=C$.*

Proof. Let Y be a set such that $|Y|=\operatorname{ind} C$ and μ be a one-to-one mapping of the factor set $F(X)/C$ onto Y. The mapping β of $F(X)$ into $F(Y)$ defined by

$$\beta(p) = \mu(\overline{\mathfrak{R}_{k-1}(p)})\cdot\mu(\overline{\mathfrak{R}_{k-2}(p)})\cdot\mu(\overline{\mathfrak{R}_{k-3}(p)})\dots\mu(\overline{\mathfrak{R}_1(p)})\cdot\mu(\overline{\mathfrak{R}_0(p)})$$

(where $k=l(p)$) satisfies the requirements. The proof is completed.

The last statement of this § elucidates the close connection between super-finite partitions and finitely realizable automaton mappings. By virtue of this connection, a (sufficiently constructive) description of the super-finite partitions would also mean a description of the mappings in question. For the definitions of finite automaton (of Moore or Mealy type), the mapping realized (in another terminology: induced) by an automaton, moreover for the proof of the following assertion, we refer to [1] (especially, assertions (4.12) and (5.11)) where these questions are discussed in details.

Proposition 8. *The subsequent three conditions are equivalent for an automaton mapping β (defined on $F(X)$):*

(i) *C_β is a super-finite partition of $F(X)$.*
(ii) *There exists a finite Moore automaton realizing β.*
(iii) *There exists a finite Mealy automaton realizing β.*

§ 3

In this § we give a short survey of the matter of the previous paper [2] where a recursion procedure is introduced by which any finite right-congruence of $F(X)$ is obtained precisely once.

We say that the relation $\alpha(p, q)$ is true ($\uparrow$) for the words p, q if there exists a number i $(1 \leqq i \leqq l(q))$ such that $p = K_i(q)$. For any $H(\subseteqq F(X))$, we denote by $\gamma(H)$ the set of words p satisfying $\alpha(p, h) = \uparrow$ with a suitable $h(\in H)$.

A finite subset H of $F(X)$ is called an *independent complete set* (abbreviated: IC-set) if $h_1 = \mathfrak{K}_i(h_2)$ implies $h_1 = h_2$ (and, consequently, $i = 0$) for any two elements h_1, h_2 of H and to almost all words $p(\in F(X))$ there exists an $h(\in H)$ satisfying $h = \mathfrak{K}_i(p)$ with an appropriate $i(\geqq 0)$. If H is an IC-set, then H and $\gamma(H)$ are disjoint.

Let a full ordering $\prec$ be fixed in the set X of generators. We extend this relation to $F(X)$ followingly: $p \prec q$ if either $|p| = |q|$ and p precedes q lexicographically or $|p| < |q|$.

In § 3 of [2], a construction of (all) the IC-sets is given.

Let H be an IC-set, let us fix an arbitrary assignment φ of H into $\gamma(H)$. We define the mapping τ_H^φ of $F(X)$ onto $\gamma(H)$ by the subsequent recursion:
if $p \in \gamma(H)$, then $\tau_H^\varphi(p) = p$,
if $p \in H$, then $\tau_H^\varphi(p) = \varphi(p)$,
if the word p does not belong to $\gamma(H) \cup H$, then

$$\tau_H^\varphi(p) = \tau_H^\varphi(\tau_H^\varphi(\mathfrak{K}_1(p))\,\mathfrak{V}_1(p)).$$

Proposition 9. (The first statement of Proposition 4 and Proposition 6 in [2]). *$\tau_H^\varphi(\mathfrak{K}_1(p))\mathfrak{V}_1(p)$ belongs to $\gamma(H) \cup H$ for any $p(\in F(X))$. The domain of τ_H^φ is the whole semigroup $F(X)$. The range of τ_H^φ is precisely $\gamma(H)$. τ_H^φ is idempotent.*

We define the partition C_H^φ of $F(X)$ such that $p \equiv q \pmod{C_H^\varphi}$ exactly if $\tau_H^\varphi(p) = \tau_H^\varphi(q)$. The mapping φ is called *normal* if $\varphi(p) \prec p$ for any word p.

The main result of [2] is:

Proposition 10. (Theorems 2, 3 in [2].) *Any partition C_H^φ is a finite right-congruence of $F(X)$. If only normal mappings φ are permitted, then each finite right-congruence C^* can be produced in exactly one way in the form C_H^φ.*

In what follows, we shall use also the following facts asserted in [2]:

Proposition 11. (Proposition 7 in [2].) *If p is an arbitrary word and x is an arbitrary generator, then*

$$\tau_H^\varphi(px) = \tau_H^\varphi(\tau_H^\varphi(p)x).$$

Proposition 12. (The first sentence follows from Proposition 8 of [2], the second one from the constructions exposed in [2].) *Any class modulo C_H^φ contains exactly one element g which belongs to $\gamma(H)$. If*

$$g \equiv p \pmod{C_H^\varphi} \qquad (g \in \gamma(H)),$$

then $\tau_H^\varphi(p) = g$ and either $g = p$ or $g \prec p$.

§ 4

In consequence of the propositions stated in § 2, the problem of describing all (essentially different) automaton mappings (defined on $F(X)$) is equivalent to *the problem of the description of all super-finite partitions of the semigroup $F(X)$.*

In § 3 we have sketched a description of the finite right-congruences of $F(X)$; any element of $\mathfrak{L}_2$ was produced uniquely. Unfortunately, this method has the disadvantage that the lattice-theoretical structure of $\mathfrak{L}_2$ remains unexplained, i.e. even if we know H, φ, H', φ', there exists no easy way to decide the validity of the relation $C_H^\varphi \leqq C_{H'}^{\varphi'}$.

If we fix a finite right-congruence C^* and we ask for all the super-finite partitions C satisfying $C \geqq C^*$, then these partitions C can be constructed rather easily (the number of the partitions C is finite by the UFP of $\mathfrak{L}_3$). If C^* is varied, then every super-finite partition C is produced; however, the LIP of $\mathfrak{L}_2$ implies that, for each C, there are infinitely many constructions obtaining C (because of the existence of an infinity of finite right-congruences C^* fulfilling $C^* \leqq C$). Consequently, this simple idea does not give a unique representation of the super-finite partitions of $F(X)$.

By a modification of our previous ideas, the following problem arises: the finite right-congruence C^* is varied and, *for any C^*, it is required to produce uniquely the partitions C satisfying $\mathfrak{M}_1(C) = C^*$.* Then each C is obtained exactly once (for the equality $\mathfrak{M}_1(C) = C^*$ is satisfied by precisely one right-congruence C^*). In what follows, the problem exposed now will be called "*basic problem*".

In Chapter II, we shall make some considerations (being far from completeness) concerning the basic problem. In Chapter III some other related questions will be touched upon.

II. On the description of super-finite partitions by using complexity numbers

§ 5

Let an IC-set H of $F(X)$ be given. Denote the set $\gamma(H)$ by G, too. Let φ be a mapping of H into G. The pair (H, φ) determines a mapping τ_H^φ of $F(X)$ onto G and a right-congruence C_H^φ by virtue of § 3. Since H, φ are throughout fixed, we shall write τ for[5] τ_H^φ.

Let $C^{(G)}$ be an arbitrary partition of the set $G(=\gamma(H))$. Let us assign to $C^{(G)}$ two partitions $\omega(C^{(G)})$, $\omega^*(C^{(G)})$ of $F(X)$ in the following manner:

$$p \equiv q \qquad (\text{mod } \omega(C^{(G)}))$$

exactly if

$$\tau(p) \equiv \tau(q) \qquad (\text{mod } C^{(G)});$$

moreover,

$$p \equiv q \qquad (\text{mod } \omega^*(C^{(G)}))$$

precisely if either $p = q$ or

$$p \in G \quad \& \quad q \in G \quad \& \quad p \equiv q \qquad (\text{mod } C^{(G)})$$

$\big($where $p \in F(X)$, $q \in F(X)\big)$.

Proposition 13. *The equality*

$$\omega(C^{(G)}) = \omega^*(C^{(G)}) \cup C_H^\varphi$$

is valid. The restrictions of the partitions $\omega(C^{(G)})$ and $\omega^*(C^{(G)})$ to G coincide with $C^{(G)}$. Moreover, we have

$$\text{ind } C^{(G)} = \text{ind } \omega(C^{(G)}).$$

Proof. Let us recall Proposition 9 and the definitions of ω, ω^*, C_H^φ. The restriction of $\omega^*(C^{(G)})$ to G equals trivially to $C^{(G)}$. The relation

$$p \equiv q \qquad (\text{mod } \omega(C^{(G)}))$$

implies

$$p \equiv \tau(p), \quad q \equiv \tau(q) \qquad (\text{mod } C_H^\varphi)$$

and

$$\tau(p) \equiv \tau(q) \qquad (\text{mod } \omega^*(C^{(G)}));$$

consequently,

$$\omega(C^{(G)}) \leqq \omega^*(C^{(G)}) \cup C_H^\varphi.$$

On the other hand, if

$$p \equiv q \qquad (\text{mod } \omega^*(C^{(G)}) \cup C_H^\varphi),$$

then

$$\tau(p) \equiv \tau(q) \qquad (\text{mod } \omega^*(C^{(G)}) \cup C_H^\varphi),$$

hence

$$\tau(p) \equiv \tau(q) \qquad \begin{cases} (\text{mod } \omega^*(C^{(G)})) \\ (\text{mod } C^{(G)}) \end{cases}$$

[5] However, we do not use the simple notation C instead of C_H^φ.

(since $\tau(p)$, $\tau(q)$ belong to G and the elements of G are pairwise incongruent mod C_H^φ), thus

$$p \equiv q \qquad (\mathrm{mod}\ \omega(C^{(G)})).$$

The above considerations show also the validity of the assertion on the restriction $\omega(C^{(G)})$ to G and the inequality

$$\mathrm{ind}\ C^{(G)} \leqq \mathrm{ind}\ \omega(C^{(G)}),$$

too. Proposition 12 implies that each class modulo $\omega(C^{(G)})(\geqq C_H^\varphi)$ has a non-empty intersection with G, hence

$$\mathrm{ind}\ C^{(G)} = \mathrm{ind}\ \omega(C^{(G)}).$$

Proposition 14. *The assignment $C^{(G)} \to \omega(C^{(G)})$ is a lattice-theoretical isomorphism (where $C^{(G)}$ runs through all the partitions of $G(=\gamma(H))$. The range of this assignment is exactly the set of the partitions C of $F(X)$ fulfilling $C \geqq C_H^\varphi$.*

Proof. Suppose $C^{(G)} \leqq C_1^{(G)}$ and

$$p \equiv q \qquad (\mathrm{mod}\ \omega(C^{(G)})).$$

Then

$$\tau(p) \equiv \tau(q) \qquad (\mathrm{mod}\ C^{(G)}),$$

hence

$$\tau(p) \equiv \tau(q) \qquad (\mathrm{mod}\ C_1^{(G)}),$$

thus

$$p \equiv q \qquad (\mathrm{mod}\ \omega(C_1^{(G)})).$$

We have proved $\omega(C^{(G)}) \leqq \omega(C_1^{(G)})$.

Now assume that the relation $C^{(G)} \leqq C_1^{(G)}$ does not hold. This means that there exists a pair (p, q) (where $p \in G$, $q \in G$) such that p, q are congruent mod $C^{(G)}$ but incongruent mod $C_1^{(G)}$. The assertion on $\omega(C^{(G)})$ in the second sentence of Proposition 13 ensures that p, q are congruent mod $\omega(C^{(G)})$ but not mod $\omega(C_1^{(G)})$, thus $\omega(C^{(G)}) \leqq \omega(C_1^{(G)})$ cannot be true. The first assertion of the proposition is verified.

Let C be a partition of $F(X)$ satisfying $C \geqq C_H^\varphi$. Denote by $C^{(G)}$ the restriction of C to G. We are going to show that $C = \omega(C^{(G)})$. Indeed, the three relations

$$p \equiv q \qquad (\mathrm{mod}\ C)$$

$$\tau(p) \equiv \tau(q) \qquad (\mathrm{mod}\ C^{(G)})$$

$$p \equiv q \qquad (\mathrm{mod}\ \omega(C^{(G)}))$$

are equivalent (by

$$\left.\begin{array}{l} p \equiv \tau(p) \\ q \equiv \tau(q) \end{array}\right\} \qquad (\mathrm{mod}\ C)$$

and the definition of ω). Thus $C = \omega(C^{(G)})$, hence the range of ω includes the set mentioned in the second sentence of the proposition. The converse inclusion follows from the first assertion of Proposition 13.

§ 6

The following idea seems to be a possible method for investigating the basic problem (exposed in § 4):

(1) we assign a complexity number $c(C)$ (being a non-negative integer) to any super-finite partition C of $F(X)$ (characterizing the "distance" of C and $\mathfrak{M}_1(C)$ in some appropriate manner),

(2) for any pair (C^*, m) (where C^* is a finite right-congruence of $F(X)$ and m is a natural number) we denote by $S(C^*, m)$ the set of partitions $C^{(G)}$ fulfilling $\mathfrak{M}_1(\omega(C^{(G)}))=C^*$ and $c(\omega(C^{(G)}))=m$,

(3) for any finite right-congruence C^* of $F(X)$, we give a description of the partitions lying in $S(C^*, 0)$, $S(C^*, 1)$, ..., $S(C^*, m)$ where m is the largest number such that $S(C^*, m)\neq\emptyset$.

Three different concrete choices of the complexity numbers $c(C)$ seem to be applicable:

(I) Let $c(C)$ be the difference

$$\operatorname{ind}\mathfrak{M}_1(C)-\operatorname{ind} C.$$

(II) Let $c(C)$ be the smallest integer j such that $\mathfrak{N}^j(C)=\mathfrak{M}_1(C)$ (cf. Remarks to Proposition 3).

(III)[6] Let $c(C)$ be $\max\min l(r)$ where the maximum is taken for all pairs p, q such that

$$p\not\equiv q \qquad (\operatorname{mod}\mathfrak{M}_1(C)) \quad (p\in F(X),\, q\in F(X))$$

and (for each pair p, q) the minimum is taken for all words r such that

$$pr\not\equiv qr \qquad (\operatorname{mod} C).$$

In what follows, we adopt the first choice, i.e. we define the *complexity number* of C by

$$c(C) = \operatorname{ind}\mathfrak{M}_1(C)-\operatorname{ind} C.$$

The relation $\mathfrak{M}_1(C)\leqq C$ implies immediately the

Proposition 15. *$c(C)=0$ exactly if $\mathfrak{M}_1(C)=C$ (i.e. if C is a right-congruence).*

Now we return to the former point of view that the IC-set H, the normal mapping φ are fixed and $\tau=\tau_H^\varphi$, $C^*=C_H^\varphi$, $G=\gamma(H)$ are defined by means of H, φ. The following paragraph is devoted to get a certain representation of the partitions $C^{(G)}$ of G satisfying

$$\mathfrak{M}_1(\omega(C^{(G)}))=C^* \quad\text{and}\quad c(\omega(C^{(G)}))=1;$$

this task is the same as that of characterizing the set $S(C^*; 1)$.

Next we state some simple facts. The first of them is obviously valid:

Proposition 16. *Let $C^{(G)}$ be a partition of G such that $\mathfrak{M}_1(\omega(C^{(G)}))=C^*$. Then $c(\omega(C^{(G)}))=0$ exactly if $C^{(G)}$ is the smallest partition of G (i.e. if every class modulo $C^{(G)}$ has only one element).*

[6] This third definition is justified only if the maximum always exists (i.e. if the set consisting of the numbers $\min l(r)$ is bounded). I do not know whether or not this existence is valid for every super-finite partition C.

Proposition 17. *Let $C^{(G)}$ be a partition of G such that $\mathfrak{M}_1(\omega(C^{(G)}))=C^*$. Then*

$$c(\omega(C^{(G)}))=1 \tag{6.1}$$

if and only if

$$\text{ind } C^{(G)} = |G|-1. \tag{6.2}$$

Proof. By the definition of the complexity number, (6.1) can be written in the form

$$\text{ind } \omega(C^{(G)}) = \text{ind } \mathfrak{M}_1(\omega(C^{(G)}))-1,$$

this equality is equivalent to (6.2) because

$$\text{ind } C_H^{\varphi}=|G|$$

is implied by Proposition 12.

Proposition 18. *Let $C^{(G)}$ be a partition of index $|G|-1$. The equality*

$$\mathfrak{M}_1(\omega(C^{(G)}))=C^*$$

holds if and only if $\omega(C^{(G)})$ is not a right-congruence.

Proof. We note that $\omega(C^{(G)})\geqq C^*$ and

$$\text{ind } \omega(C^{(G)}) = \text{ind } C^{(G)} = |G|-1 = \text{ind } C^*-1$$

imply $\omega(C^{(G)})>C^*$.

If $\omega(C^{(G)})$ is a right-congruence, then

$$\mathfrak{M}_1(\omega(C^{(G)}))=\omega(C^{(G)})>C^*.$$

If $\omega(C^{(G)})$ is not a right-congruence, then

$$\mathfrak{M}_1(\omega(C^{(G)}))<\omega(C^{(G)})$$

implies

$$\text{ind } \mathfrak{M}_1(\omega(C^{(G)})) > \text{ind } \omega(C^{(G)})(= \text{ind } C^*-1),$$

hence

$$\text{ind } \mathfrak{M}_1(\omega(C^{(G)}))\geqq \text{ind } C^*;$$

on the other hand, Proposition 1 guarantees

$$\mathfrak{M}_1(\omega(C^{(G)}))\geqq C^*.$$

The last two formulae ensure

$$\mathfrak{M}_1(\omega(C^{(G)}))= C^*.$$

§ 7

Let g_1, g_2 be two different elements of G such that $g_1 \prec g_2$. We denote by $C^{(G)}_{g_1,g_2}$ the partition of G in which $\{g_1, g_2\}$ is one of the classes and any other class has one element. In the form $C^{(G)}_{g_1,g_2}$ all the partitions (of G) of index $|G|-1$ (and only these) can be obtained.

THEOREM 1. *The partition* $C=\omega(C^{(G)}_{g_1,g_2})$ *is a right-congruence of* $F(X)$ *if and only if*[7] $g_2X\subseteq H$ *and each* $x(\in X)$ *satisfies one of the following four assertions:*[8]

(i) $g_1x\in G$ & $\varphi(g_2x)=g_1x$.

(ii) $g_1x=g_2$ & $\varphi(g_2x)=g_1$.

(iii) $g_1x\in H$ & $\varphi(g_1x)=\varphi(g_2x)$.

(iv) $g_1x\in H$ & $\{\varphi(g_1x), \varphi(g_2x)\}=\{g_1, g_2\}$.

Proof

Necessity. Suppose that C is a right-congruence. We have $g_ix\in G\cup H$ by $g_i\in G$ and $G=\gamma(H)$ for each $x(\in X)$ (where i may be 1 or 2).

$$g_1\equiv g_2 \qquad (\text{mod } C^{(G)}_{g_1,g_2})$$

implies

$$g_1\equiv g_2 \qquad (\text{mod } C),$$

hence

$$g_1x\equiv g_2x \qquad (\text{mod } C),$$

thus

$$\tau(g_1x)\equiv\tau(g_2x) \qquad (\text{mod } C^{(G)}_{g_1,g_2}).$$

Case 1: $g_1x\in G$ and

$$g_1x\equiv g_2x \qquad (\text{mod } C^*).$$

Then $g_1x\prec g_2x$ (by Proposition 12) and $g_2x(G\cup H)$ cannot belong to G, i.e. $g_2x\in H$. Moreover,

$$g_1x=\tau(g_1x)=\tau(g_2x)=\varphi(g_2x),$$

this means that (i) is satisfied.

Case 2: $g_1x\in G$ and

$$g_1x\not\equiv g_2x \qquad (\text{mod } C^*).$$

In this case

$$g_1x=\tau(g_1x)\neq\tau(g_2x),$$

hence

$$g_1x=g_2 \quad \text{and} \quad \tau(g_2x)=g_1$$

(because $g_1x(=\tau(g_1x))$ and $\tau(g_2x)$ are different elements of G but congruent mod C), consequently

$$g_2x\in H \quad \text{and} \quad \varphi(g_2x)=\tau(g_2x)(=g_1)$$

(by $g_1\prec g_2\prec g_2x$). (ii) is fulfilled.

Case 3: $g_1x\in H$ and

$$g_1x\equiv g_2x \qquad (\text{mod } C^*).$$

Similarly to Case 1, we can deduce $g_2x\in H$ and

$$\varphi(g_1x)=\tau(g_1x)=\tau(g_2x)=\varphi(g_2x),$$

thus (iii) is valid.

Case 4: $g_1x\in H$ and

$$g_1x\not\equiv g_2x \qquad (\text{mod } C^*).$$

---

[7] Usually, g_2X denotes the set of words g_2x where x runs through the elements of X.

[8] The assertions (i), (ii), (iii), (iv) exclude each other.

In analogy with Case 2,

$$\{\tau(g_1x), \tau(g_2x)\}=\{g_1, g_2\}.$$

If $\tau(g_1x)=g_1$ and $\tau(g_2x)=g_2$, then $g_2\neq g_2x$ implies $g_2x\in H$; in case of validity of $\tau(g_1x)=g_2$ & $\tau(g_2x)=g_1$, H must contain g_2x likely to Case 2. In both subcases, $\tau(g_2x)$ equals to $\varphi(g_2x)$, hence (iv) is true.

Sufficiency. Assume $g_2X\subseteqq H$, let p, q be two words being congruent mod C and x be a generator. Suppose that one of (i), (ii), (iii), (iv) is valid for x.

Case 1:

$$p\equiv q \qquad (\text{mod } C^*).$$

Then

$$px\equiv qx \qquad (\text{mod } C^*),$$

hence

$$px\equiv qx \qquad (\text{mod } C).$$

Case 2:

$$p\not\equiv q \qquad (\text{mod } C^*).$$

Then we have

$$\tau(p)=g_1 \quad \text{and} \quad \tau(q)=g_2$$

(possibly after interchanging p and q), thus

$$\tau(px)=\tau(g_1x)\equiv\tau(g_2x)=\tau(qx) \qquad (\text{mod } C^{(G)}_{g_1, g_2})$$

(because the equalities follow from Proposition 11, the congruence is implied by each of (i), (ii), (iii), (iv)), consequently

$$px\equiv qx \qquad (\text{mod } C).$$

The proof of Theorem 1 is finished.

Now we are going to describe a procedure for obtaining the elements C of $S(C^*, 1)$ such that any partition C is produced (not uniquely in general, but) at most $|X|$ times.

Denote by $\mathfrak{R}(p)$ the set of elements q of $F(X)$ fulfilling $q\prec p$ (where $p\in F(X)$).

Construction I. The construction is described in the subsequent rules.

Rule 1. Consider the generators $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ (the superscripts are thought to be fixed), let us choose an arbitrary $x^{(i)}\,(\in X)$.

Rule 2. Denote by G_i the set of elements[9] $g\,(\in G)$ satisfying $gx^{(i)}\in H$.

Rule 3. If $g_2\in G-G_i$, then define the set $\mathfrak{G}_i(g_2)$ by

$$\mathfrak{G}_i(g_2) = G\cap\mathfrak{R}(g_2).$$

Rule 4a. If $g_2\in G_i$, $\mathfrak{B}_2(g_2)=x^{(i)}x^{(i)}$ and $\mathfrak{K}_1(g_2)=\varphi(g_2x^{(i)})$, then define the set $\mathfrak{G}_i'(g_2)$ by

$$\mathfrak{G}_i'(g_2) = \big(G\cap\mathfrak{R}(g_2)\big)-\big(G_i\cup\{\mathfrak{K}_2(g_2), \mathfrak{K}_1(g_2)\}\big).$$

[9] The number of elements of G_i is, in general, small in comparation to $|G|$. This fact has the consequence (advantageous when Construction I is performed practically) that, in what follows, the more complicated Rules 4a, ..., 4d (and more rather Rules 5a, 5b) are executed remarkably fewer times, than the simpler Rule 3.

Rule 4b. If $g_2 \in G_i$, $\mathfrak{B}_1(\varphi(g_2 x^{(i)})) = x^{(i)}$ and $g_2 \neq \varphi(g_2 x^{(i)}) x^{(i)}$, then define $\mathfrak{G}_i'(g_2)$ by

$$\mathfrak{G}_i'(g_2) = (G \cap \mathfrak{R}(g_2)) - (G_i \cup \{\mathfrak{K}_1(\varphi(g_2 x^{(i)}))\}).$$

Rule 4c. If $g_2 \in G_i$, $\mathfrak{B}_2(g_2) = x^{(j)} x^{(i)}$ (where $x^{(j)}$ is a generator, different from $x^{(i)}$) and $\mathfrak{K}_1(g_2) = \varphi(g_2 x^{(i)})$, then define $\mathfrak{G}_i'(g_2)$ by

$$\mathfrak{G}_i'(g_2) = (G \cap \mathfrak{R}(g_2)) - (G_i \cup \{\mathfrak{K}_1(g_2)\}).$$

Rule 4d. If $g_2 \in G_i$, $\mathfrak{B}_1(\varphi(g_2 x^{(i)})) \neq x^{(i)}$ and $g_2 \neq \varphi(g_2 x^{(i)}) x^{(i)}$, then define $\mathfrak{G}_i'(g_2)$ by

$$\mathfrak{G}_i'(g_2) = (G \cap \mathfrak{R}(g_2)) - G_i.$$

Rule 5a. If $g_2 \in G_i$ and $\varphi(g_2 x^{(i)}) = g_2$, then define the set $\mathfrak{G}_i''(g_2)$ as the set of the elements $g_1 (\in G_i \cap \mathfrak{R}(g_2))$ satisfying

$$\varphi(g_1 x^{(i)}) \notin \{g_1, g_2\}.$$

Rule 5b. If $g_2 \in G_i$ and $\varphi(g_2 x^{(i)}) \neq g_2$, then define $\mathfrak{G}_i''(g_2)$ as the set of elements $g_1 (\in G_i \cap \mathfrak{R}(g_2))$ fulfilling at least one of the formulae

$$\varphi(g_1 x^{(i)}) \notin \{g_2, \varphi(g_2 x^{(i)})\}, \qquad \varphi(g_2 x^{(i)}) \notin \{g_1, \varphi(g_1 x^{(i)})\}.$$

Rule 6. If $g_2 \in G_i$, then define the set $\mathfrak{G}_i(g_2)$ by [10]

$$\mathfrak{G}_i(g_2) = \mathfrak{G}_i'(g_1) \cup \mathfrak{G}_i''(g_2).$$

Rule 7. Let us form the set Γ_i of pairs (g_1, g_2) in the following manner: (g_1, g_2) belongs to Γ_i exactly if $g_2 \in G$ and $g_1 \in \mathfrak{G}_i(g_2)$.

Construction I is completed.

THEOREM 2. *The partition $C = \omega(C^{(G)}_{g_1, g_2})$ of $F(X)$ is no right-congruence if and only if the pair (g_1, g_2) is contained in*

$$\Gamma_1 \cup \Gamma_2 \cup \ldots \cup \Gamma_n$$

where $n = |X|$ and any Γ_i (where i can be $1, 2, \ldots, n$) is produced by Construction I.

Proof

Necessity. Suppose that C is no right-congruence. We verify $g_1 \in \mathfrak{G}_i(g_2)$ (with a suitable i) according to several possible cases.

Case 1: there exists a generator $x^{(i)}$ such that $g_2 x^{(i)} \in G$. Then $g_2 \in G - G_i$ (by Rule 2), consequently, $g_1 \in \mathfrak{G}_i(g_2)$ (by Rule 3).

Case 2: $g_2 x \in H$ holds for every $x (\in X)$. Then there exists an $x^{(i)} (\in X)$ which does not satisfy the assertions (i), (ii), (iii), (iv) occurring in Theorem 1.

Case 2a: $g_1 x^{(i)} \in G$ (thus $g_1 \in G - G_i$). If the premissa of Rule 4a are satisfied, then $g_1 \neq \mathfrak{K}_1(g_2)$ (by the falsity of (ii)) and $g_1 \neq \mathfrak{K}_1(\varphi(g_2 x^{(i)})) (= \mathfrak{K}_2(g_2))$ (by the falsity of (i)), hence $g_1 \in \mathfrak{G}_i'(g_2)$. If the premissa of Rule 4b are true, then we get $g_1 \neq \mathfrak{K}_1(\varphi(g_2 x^{(i)}))$ in a similar way, consequently $g_1 \in \mathfrak{G}_i'(g_2)$. If the premissa of Rule 4c hold, then $g_1 \neq \mathfrak{K}_1(g_2)$ (since the contrary would imply (ii)), hence $g_1 \in \mathfrak{G}_i'(g_2)$. In the case of the validity of the premissa of Rule 4d, the membership $g_1 \in \mathfrak{G}_i'(g_2)$ is obvious.

---

[10] The sets $\mathfrak{G}_i'(g_2)$ and $\mathfrak{G}_i''(g_2)$, defined in the previous rules, are obviously disjoint.

Case 2b: $g_1 x^{(i)} \in H$ (thus $g_1 \in G_i$). If the premissa of Rule 5a are fulfilled, then $\varphi(g_1 x^{(i)})$ differs from both g_1 and g_2 (by the falsity of (iii) and (iv)), hence $g_1 \in \mathfrak{G}_i''(g_2)$. If the premissa of Rule 5b hold and $g_1 = \varphi(g_2 x^{(i)})$, then the same inference is valid. If the premissa of Rule 5b are true and $g_1 \neq \varphi(g_2 x^{(i)})$, then the inequality $\varphi(g_1 x^{(i)}) \neq \neq \varphi(g_2 x^{(i)})$ (implied by the falsity of (iii)) guarantees $g_1 \in \mathfrak{G}_i''(g_2)$.

We have obtained $g_1 \in \mathfrak{G}_i(g_2)$ in every case, this membership is equivalent to $(g_1, g_2) \in \Gamma_i$ (by Rules 6, 7).

Sufficiency. Assume $(g_1, g_2) \in \Gamma_i$ for some i, hence $g_1 \in \mathfrak{G}_i(g_2)$ by Rule 7. We are going to show that either $g_2 x^{(i)} \in G$ or each of the assertions (i), (ii), (iii), (iv) is false for g_1, g_2 and the generator $x^{(i)}$.

Case 1: $g_2 \in G - G_i$. Then clearly $g_2 x^{(i)} \in G$.

Case 2: $g_2 \in G_i$. Now $g_2 x^{(i)} \in H$ and the set $\mathfrak{G}_i(g_2)$ (containing g_1) was defined by Rule 6. Thus g_1 belongs either to $\mathfrak{G}_i'(g_2)$ or to $\mathfrak{G}_i''(g_2)$.

Case 2a: $g_1 \in \mathfrak{G}_i'(g_2)$. We can distinguish four situations according as the premissa of Rule 4a or 4b or 4c or 4d are satisfied. In every situation, it is trivial that (iii), (iv) are false (because of $g_1 x^{(i)} \in G$) and it is easy to check that also (i), (ii) do not hold.

Case 2b: $g_1 \in \mathfrak{G}_i''(g_2)$. Then (i), (ii) cannot hold (since $g_1 x^{(i)} \in H$) and, whether the premissa of Rule 5a or the premissa of Rule 5b are valid, we can simply show that (iii), (iv) are false, too.

Theorem 2 and Propositions 17, 18 imply at once

COROLLARY. *Let (g_1, g_2) run through the elements of*

$$\Gamma_1 \cup \Gamma_2 \cup \ldots \cup \Gamma_n.$$

Then each partition $C^{(G)}_{g_1, g_2}$ belongs to $S(C^*, 1)$; conversely, any element of $S(C^*, 1)$ is obtained thus at least once, at most $n = |X|$ times.

(The multiplicity of an element of $S(C^*, 1)$ is here understood from a constructive point of view; i. e. our last assertion corresponds to the facts that Construction I produces the elements of any Γ_i uniquely and, of course, the same pair (g_1, g_2) occurs in $\leqq n$ conponents of the union $\Gamma_1 \cup \ldots \cup \Gamma_n$.)

III. A characterization of the critical pairs of finite right-congruences

§ 8

First we expose three problems concerning the finite partitions of $F(X)$.

(I) Let C_H^φ and $C_{H'}^{\varphi'}$ be two right-congruences of $F(X)$. Let a necessary and sufficient condition of the relation $C_H^\varphi \leqq C_{H'}^{\varphi'}$ be given such that the condition concerns to the pairs (H, φ) and (H', φ').

(II) Let $C^* = C_H^\varphi$ be a right-congruence of $F(X)$. Describe the right-congruences $C^{**} (> C^*)$ satisfying the assertion: if $C^{**} \geqq C' \geqq C^*$ for a right-congruence C', then either $C' = C^{**}$ or $C' = C^*$.

(III) Let $C^* = C_H^\varphi$ be a right-congruence of $F(X)$. Describe the partitions $C (\geqq C^*)$ fulfilling the statement: if $C \geqq C' \geqq C^*$ for a right-congruence C', then $C' = C^*$.

It seems that the solution of (I) would be a remarkable aid for solving (II), furthermore, an analogous relationship exists between the problems (II) and (III). It is clear that (III) is another formulation of the basic problem posed at the end of § 4.

In the remaining part of the paper, we shall make some considerations concerning the problem (I).

Let C be a right-congruence. We say that the unordered pair (p, q) of words is a *critical pair* of C if

$$p \equiv q \pmod{C}$$

and one of the following four assertions hold:

$$p = e,$$

$$q = e,$$

$$\mathfrak{B}_1(p) \neq \mathfrak{B}_1(q),$$

$$\mathfrak{K}_1(p) \not\equiv \mathfrak{K}_1(q) \pmod{C}.$$

The correspondence between a right-congruence and the set Ω of its critical pairs was studied in Chapter II of [3].

Among others, the subsequent result was proved:

Lemma 3. The congruence

$$p \equiv q \pmod{C}$$

is true if and only if there exists a critical pair (p', q') of C and a word r such that $p = p'r$, $q = q'r$.

Proposition 19. *Consider two right-congruences $C = C_H^\varphi$ and $C' = C_{H'}^{\varphi'}$ of $F(X)$. Let Ω, Ω' be the sets of critical pairs of C, C', respectively. The following four statements are equivalent:*

(A) $C \leqq C'$.

(B) *For each* $h(\in H)$

$$h \equiv \varphi(h) \pmod{C'}.$$

(C) *For each* $p(\in F(X))$

$$p \equiv \tau_H^\varphi(p) \pmod{C'}.$$

(D) *For any $(p, q) \in \Omega$ there exist three words p', q', t such that*

$$p = p't, \quad q = q't, \qquad (p', q') \in \Omega'.$$

Proof

(A)$\Rightarrow$(B). h and $\varphi(h)$ are congruent mod C, hence also mod C'.

(B)$\Rightarrow$(C). We shall use induction. The unit element e satisfies (C) obviously (by $\tau_H^\varphi(e) = e$). Assume that (C) holds for $p(\in F(X))$, we show that (C) is true for px, too, instead of p (where $x \in X$).

Case 1: $\tau_H^\varphi(p)x \in G(=\gamma(H))$. Then

$$\tau_H^\varphi(px) = \tau_H^\varphi(p)x \equiv px \pmod{C'}$$

(by Proposition 11 and the right-congruence property of C').

Case 2: $\tau_H^\varphi(p)x \in H$. Then we get

$$\tau_H^\varphi(px) = \varphi(\tau_H^\varphi(p)x) \equiv \tau_H^\varphi(p)x \equiv px \qquad (\mathrm{mod}\, C')$$

by a similar way (using also Proposition 12).

The first statement of Proposition 9 shows that there exists no further possibility.

(C)⇒(A). If

$$p \equiv q \qquad (\mathrm{mod}\, C),$$

then

$$p \equiv \tau_H^\varphi(p) = \tau_H^\varphi(q) \equiv q \qquad (\mathrm{mod}\, C')$$

(by Proposition 12 and the connection of C and τ_H^φ).

(A)⇒(D). If $(p, q) \in \Omega$, then p and q are congruent mod C, thus also mod C'. Lemma 3 assures the validity of (D).

(D)⇒(A). Let p, q be congruent mod C. There exists a critical pair (p_1, q_1) of C such that $p = p_1 r$ and $q = q_1 r$ (by Lemma 3). (D) guarantees $p_1 = p' t$, $q_1 = q' t$ with suitable $(p', q') \in \Omega'$ and $t (\in F(X))$. Consequently, $p = p' tr$, $q = q' tr$, hence $p \equiv q \pmod{C'}$.

In what follows, we shall characterize the critical pairs of a right-congruence represented in the form C_H^φ. First (recalling the first sentence of Proposition 9) we introduce a notation: let $\bar{H}$ be the set of elements $p (\in F(X) - \{e\})$ satisfying

$$\tau_H^\varphi(\mathfrak{K}_1(p))\mathfrak{B}_1(p) \in H.$$

§ 9 will contain certain preparations to the proof of Theorem 3, exposed in § 10. In the remaining part of the paper, we write τ instead of τ_H^φ and C instead of C_H^φ.

§ 9

Lemma 4. $H \subseteqq H'$ and $\bar{H} \cap \gamma(H) = \emptyset$.

Proof. If $p \in (\gamma(H) - \{e\}) \cup H$, then $\mathfrak{K}_1(p) \in \gamma(H)$, hence

$$\tau(\mathfrak{K}_1(p))\mathfrak{B}_1(p) = \mathfrak{K}_1(p)\mathfrak{B}_1(p) = p.$$

This implies $p \in \bar{H}$ or $p \notin \bar{H}$ according to $p \in H$ or $p \in \gamma(H)$, respectively.

Lemma 5. Let p, q be elements of $F(X)$. If $\tau(p)q \in \gamma(H)$, then $\tau(pq) = \tau(p)q$. If $\tau(p)q \in H$, then $\tau(pq) = \varphi(\tau(p)q)$.

Proof. We verify the first statement by induction with respect to the length of q. The assertion is trivial for e (as q). Suppose that it is true for the words of length k, assume $l(q) = k+1$. Denote $\mathfrak{K}_k(q)$ and $\mathfrak{B}_k(q)$ by x and q', respectively (thus $q = xq'$, $x \in X$, $l(q') = k$). We note that the supposition $\tau(p)q \in \gamma(H)$ implies $\tau(p)x \in \gamma(H)$, therefore $\tau(\tau(p)x) = \tau(p)x$. We have

$$\tau(pq) = \tau(pxq') = \tau(px)q' = \tau(\tau(p)x)q' = \tau(p)xq' = \tau(p)q$$

where the second equality is implied by the induction hypothesis and the third one follows from Proposition 11. The first statement is proved.

Suppose $\tau(p)q \in H$ (thus $q \neq e$), write q in the form[11] $q = q'x$ $(x \in X)$. Then $\tau(p)q'$ belongs to $\gamma(H)$ and the inference

$$\tau(pq) = \tau(pq'x) = \tau(\tau(pq')x) = \tau(\tau(p)q'x) = \tau(\tau(p)q) = \varphi(\tau(p)q)$$

is valid (the second equality is again a consequence of Proposition 11).

We are now able to assert a result which yields, supposing $p \in \bar{H}$ particularly, a recursive characterization of $\bar{H}$:

Proposition 20. *Assume* $p \in F(X)$, $q \in F(X) - \{e\}$, $\tau(p)q \in \gamma(H) \cup H$. *If* $\tau(p)q \in \gamma(H)$, *then* $pq \notin \bar{H}$. *If* $\tau(p)q \in H$, *then* $pq \in \bar{H}$.

Proof. In both cases, the condition posed on $\tau(p)q$ implies $\tau(p)\mathfrak{K}_1(q) \in \gamma(H)$, hence

$$\tau(\mathfrak{K}_1(pq))\mathfrak{V}_1(pq) = \tau(p\mathfrak{K}_1(q))\mathfrak{V}_1(q) = \tau(p)\mathfrak{K}_1(q)\mathfrak{V}_1(q) = \tau(p)q$$

(using Lemma 5). The definition of $\bar{H}$ completes the proof.

Lemma 6. If $p \in F(X) - (\{e\} \cup \bar{H})$, *then* $\tau(p) \neq e$, $\mathfrak{K}_1(\tau(p)) = \tau(\mathfrak{K}_1(p))$ *and* $\mathfrak{V}_1(\tau(p)) = \mathfrak{V}_1(p)$.

Proof. $\tau(\mathfrak{K}_1(p))\mathfrak{V}_1(p) \in \gamma(H)$ implies

$$\tau(p) = \tau(\mathfrak{K}_1(p)\mathfrak{V}_1(p)) = \tau(\mathfrak{K}_1(p))\mathfrak{V}_1(p) \ (\neq e)$$

(by Lemma 5); the equalities to be proved follow by applying the operators $\mathfrak{K}_1$, $\mathfrak{V}_1$ for the left-hand and right-hand sides of this equality.

Lemma 7. Let p, q *be elements of* $F(X) - (\{e\} \cup \bar{H})$. *If* $p \equiv q \pmod{C}$, *then* $\mathfrak{K}_1(p) \equiv \mathfrak{K}_1(q) \pmod{C}$ *and* $\mathfrak{V}_1(p) = \mathfrak{V}_1(q)$.

Proof. The supposition implies $\tau(p) = \tau(q)$. Thus

$$\tau(\mathfrak{K}_1(p)) = \mathfrak{K}_1(\tau(p)) = \mathfrak{K}_1(\tau(q)) = \tau(\mathfrak{K}_1(q))$$

and

$$\mathfrak{V}_1(p) = \mathfrak{V}_1(\tau(p)) = \mathfrak{V}_1(\tau(q)) = \mathfrak{V}_1(q)$$

are true by Lemma 6.

Lemma 8. If $p \in \bar{H}$, *then either* $\tau(p) = e$ *or*

$$\mathfrak{K}_1(p) \not\equiv \mathfrak{K}_1(\tau(p)) \pmod{C}$$

or

$$\mathfrak{V}_1(p) \neq \mathfrak{V}_1(\tau(p)).$$

Proof. Suppose that each of the three alternatives, stated in the lemma, is false for $p (\in \bar{H})$, we are going to get a contradiction. The supposition

$$\mathfrak{K}_1(p) \equiv \mathfrak{K}_1(\tau(p)) \pmod{C}$$

implies

$$\tau(\mathfrak{K}_1(p)) = \tau(\mathfrak{K}_1(\tau(p))) = \mathfrak{K}_1(\tau(p))$$

(since $\tau(p)$, $\mathfrak{K}_1(\tau(p))$ belong to $\gamma(H)$).

[11] This notation differs from the previous meaning of q', x.

Denote by i the minimal positive number fulfilling $\mathfrak{K}_i(p)\in\overline{H}\cup\{e\}$. Use the notation $p_1=\mathfrak{K}_i(p)$, $p_2=\mathfrak{B}_i(p)$. We have $\tau(p_1)p_2\in H$ (since $\tau(p_1)p_2\in\gamma(H)$ would imply $p\notin\overline{H}$ and $\tau(p_1)p_2\in F(X)-(\gamma(H)\cup H)$ would lead to a contradiction to the minimality of i by Proposition 20), thus $\varphi(\tau(p_1)p_2)$ is defined (and belongs to $\gamma(H)$). We have the equalities

$$(\gamma(H)\ni)\mathfrak{K}_1(\tau(p_1)p_2)=\tau(p_1)\mathfrak{K}_1(p_2)=\tau(p_1\mathfrak{K}_1(p_2))=$$

$$=\tau(\mathfrak{K}_1(p_1p_2))=\mathfrak{K}_1(\tau(p_1p_2))=\mathfrak{K}_1(\varphi(\tau(p_1)p_2))$$

(the second and last ones follow from Lemma 5). On the other hand,

$$\mathfrak{B}_1(\tau(p_1)p_2)=\mathfrak{B}_1(p_2)=\mathfrak{B}_1(p)=\mathfrak{B}_1(\tau(p))=\mathfrak{B}_1(\tau(p_1p_2))=\mathfrak{B}_1(\varphi(\tau(p_1)p_2)).$$

Hence we get

$$\tau(p_1)p_2=\varphi(\tau(p_1)p_2),$$

this is a contradiction to the disjointness of H and $\gamma(H)$.

§ 10

Let p, q be two elements of $F(X)$ ($p=q$ is permitted). We shall obtain a necessary and sufficient condition for the pair (p, q) in order to be a critical pair. Evidently, $\tau(p)=\tau(q)$ is a necessary condition; however, it is not sufficient.

Denote by i the least positive integer satisfying $\mathfrak{K}_i(p)\in\{e\}\cup\overline{H}$; analogously, by j the least positive integer fulfilling $\mathfrak{K}_j(q)\in\{e\}\cup\overline{H}$.

THEOREM 3. *The pair* (p, q) *is a critical pair of the right congruence* C_H^φ *if and only if one of the subsequent conditions* (i), (ii), (iii) *is satisfied (possibly after interchanging p and q):*

(i) $e=p=\tau(q)$,

(ii) $p\in F(X)-\overline{H}$, $q\in\overline{H}$ *and* $\tau(p)=\tau(q)$,

(iii) $p\in\overline{H}$, $q\in\overline{H}$, $\tau(p)=\tau(q)$ *and either*

$$\mathfrak{K}_i(p)\not\equiv\mathfrak{K}_j(q)\pmod{C_H^\varphi}$$

or

$$\mathfrak{B}_i(p)\neq\mathfrak{B}_j(q).$$

Proof. As we have formulated the theorem, (i) and (ii) do not exclude each other. A non-overlapping system of conditions (equivalent to the system consisting of (i), (ii), (iii)) can be got by replacing (i) by the following condition (i′):

(i′) $q\in F(X)-\overline{H}$ and $e=p=\tau(q)$.

In the verification of the theorem we shall distinguish three cases:

(I) p and q are contained in $F(X)-\overline{H}$.

(II) $p\in F(X)-\overline{H}$ and $q\in\overline{H}$.

(III) p and q belong to $\overline{H}$.

We shall show that, in the cases (I), (II), (III), the criterion for the inclusion $(p, q)\in\Omega$ is (i′), (ii), (iii), respectively.

Case I. Suppose $(p, q) \in \Omega$. If $p \neq e$ and $q \neq e$, then Lemma 4 leads to a contradiction; if $p=e$, then $\tau(q)=\tau(p)=e$, hence (i′) is satisfied. — (i′) implies $(p, q) \in \Omega$ evidently.

Case II. $(p, q) \in \Omega$ implies (ii) trivially. — Conversely, assume that (ii) is valid. Then clearly

$$p \equiv q \qquad (\mathrm{mod}\ C),$$

we are going to prove that either $p=e$ or

$$\mathfrak{K}_1(p) \not\equiv \mathfrak{K}_1(q) \qquad (\mathrm{mod}\ C)$$

or

$$\mathfrak{B}_1(p) \neq \mathfrak{B}_1(q).$$

Suppose $p \neq e$. Then

$$e \neq \tau(p) = \tau(q)$$

by Lemma 6, moreover, one of the inferences

$$\tau(\mathfrak{K}_1(p)) = \mathfrak{K}_1(\tau(p)) = \mathfrak{K}_1(\tau(q)) \not\equiv \tau(\mathfrak{K}_1(q)) \qquad (\mathrm{mod}\ C),$$

$$\mathfrak{B}_1(p) = \mathfrak{B}_1(\tau(p)) = \mathfrak{B}_1(\tau(q)) \neq \mathfrak{B}_1(q)$$

is true (the equalities follow from Lemma 6; either the incongruence or the inequality is implied by Lemma 8). Hence $(p, q) \in \Omega$ in any possible case.

Case III. Assume that (iii) is not fulfilled, we want to show $(p, q) \notin \Omega$. It suffices to study the possibility when $\tau(p)=\tau(q)$. Since (iii) is supposed to be false, we have

$$\mathfrak{B}_i(p) = \mathfrak{B}_j(q) \qquad (\text{thus } i=j)$$

and

$$\mathfrak{K}_i(p) \equiv \mathfrak{K}_j(q) \qquad (\mathrm{mod}\ C).$$

Hence

$$\mathfrak{B}_1(p) = \mathfrak{B}_1(\mathfrak{B}_i(p)) = \mathfrak{B}_1(\mathfrak{B}_j(q)) = \mathfrak{B}_1(q)$$

and

$$\mathfrak{K}_1(p) = \mathfrak{K}_i(p)\mathfrak{K}_1(\mathfrak{B}_i(p)) \equiv \mathfrak{K}_j(q)\mathfrak{K}_1(\mathfrak{B}_j(q)) = \mathfrak{K}_1(q) \qquad (\mathrm{mod}\ C),$$

consequently $(p, q) \in \Omega$. — Suppose $(p, q) \notin \Omega$, our aim is to prove that (iii) is false. This follows trivially unless $\tau(p)=\tau(q)$, $i=j$. Assume that these equalities are true. The suppositions imply

$$\mathfrak{K}_1(p) \equiv \mathfrak{K}_1(q) \qquad (\mathrm{mod}\ C)$$

and

$$\mathfrak{B}_1(p) = \mathfrak{B}_1(q).$$

Apply Lemma 7 for the elements $\mathfrak{K}_h(p)$ and $\mathfrak{K}_h(q)$ (instead of p and q, resp.) where h can be $1, 2, 3, \ldots, i-1$. We get, on the one hand,

$$\mathfrak{K}_i(p) \equiv \mathfrak{K}_i(q) = \mathfrak{K}_j(q) \qquad (\mathrm{mod}\ C),$$

on the other hand,

$$\mathfrak{B}_1(\mathfrak{K}_h(p)) = \mathfrak{B}_1(\mathfrak{K}_h(q)) \qquad (2 \leqq h < i),$$

hence

$$\mathfrak{B}_i(p) = \mathfrak{B}_i(q) = \mathfrak{B}_j(q).$$

О некоторых аспектах алгебраического описания автоматных отображений

Пусть $F(X)$ — свободная полугруппа (с единицей), порождённая конечным множеством X. Изучаются разбиения C полугруппы $F(X)$ так, что отношение $C \geqq C^*$ удовлетворяемо некоторой правой конгруэнтностью C^*, имеющей конечное число классов. Такие разбиения C называются супер-конечными. В §§ 1—2 излагаются некоторые основные (по существу, известные) свойства супер-конечных разбиений, включая их связь с конечно представимыми автоматными отображениями. Кроме других предложений приводится (без доказательства) теорема *Nerode*-а: разбиение соответствует конечно представимому автоматному отображению тогда и только тогда, если оно является супер-конечным.

§ 3 даёт разюме предыдущей статьи [2] автора. В § 4 формулируется следующая проблема: пусть для произвольной правой конгруэнтности C^* с конечным числом классов описаны одинственным образом такие разбиения C, что C^* является наибольшей из всех правых конгруэнтностей меньше чем C. В § 6 вводится число сложности $c(C)$ супер-конечного разбиения C как ind C^* — ind C (где ind C^* — число классов по модулю этой наибольшей правой конгруентности C^*). § 7 содержит метод описывающий супер-конечные разбиения, выполняющие $c(C)=1$, это описание, вообще говоря, не однозначно, но многозначность не превосходит числа элементов множества X.

Пара (p, q) элементов полугруппы $F(X)$ называется критической для правой конгруэнтности C, если $p \equiv q \pmod C$ и хотя бы одно из четырёх условий выполняется: $p=e$, $q=e$, $\mathfrak{V}_1(p) \neq \mathfrak{V}_1(q)$, $\mathfrak{K}_1(p) \not\equiv \mathfrak{K}_1(q) \pmod C$, где e — единица полугруппы, и $\mathfrak{K}_1(p)$, $\mathfrak{V}_1(p)$ определяются отношениями $p=\mathfrak{K}_1(p)\,\mathfrak{V}_1(q)$, $\mathfrak{K}_1(p) \in F(X)$, $\mathfrak{V}_1(p) \in X$. В § 10 устанавливаются критические пары произвольной правой конгруэнтности C с конечным числом классов, предполагая, что C дано методом работы [2].

MATHEMATICAL INSTITUTE
OF THE HUNGARIAN ACADEMY OF SCIENCES
1053 BUDAPEST, HUNGARY
V., REÁLTANODA U. 13–15.

References

[1] ÁDÁM, A., Automata-leképezések, félcsoportok, automaták (Automaton mappings, semigroups, automata), *Mat. Lapok*, v. 19, 1968, pp. 327—343.[12]

[2] ÁDÁM, A., A description of the finite right-congruences of finitely generated free semigroups, *Periodica Math. Hung.*, v. 1, 1971, pp. 135—144.

[3] ÁDÁM, A. & G. POLLÁK, On the congruences of finitely generated free semigroups, *Acta Sci. Math. (Szeged)*, v. 31, 1970, pp. 259—278.

[4] NERODE, A., Linear automaton transformations, *Proc. Amer. Math. Soc.*, v. 9, 1958, pp. 541—544.

[5] RABIN, M. O. & D. SCOTT, Finite automata and their decision problems, *IBM J. Res. Develop.* v. 3, 1959, pp. 114—125.[13]

[6] Любич, Ю. И., О свойствах периодичности событий, представимых в конечных автоматах, *Укр. Мат. Журнал*, v. 16, 1964, pp. 396—402.

(Received January 11, 1972)

[12] Hungarian, with Russian and English summaries. The summary is almost completely contained in *Zentralblatt f. Math.*, v. 179, 1970, p. 23.

[13] Russian translation: *Кибернетический Сборник*, v. 4, 1962, pp. 58—91.

Quasisequentielle Funktionen

Von G. Wechsung

Die Arbeit stellt einen Beitrag zur Theorie der Wortfunktionen dar. Unter einer Wortfunktion über dem Alphabet X verstehen wir eine eindeutige Abbildung der Wortmenge X^* in sich. Eine Wortfunktion f heißt sequentiell [2], wenn sie die Anfangswortrelation invariant läßt, d. h. wenn gilt

$$\forall u \forall v (u, v \in X^* \land u \subseteqq v \to f(u) \subseteqq f(v)).$$

Häufig benützen wir die Sprechweise „sequentielle Funktion" für „sequentielle Wortfunktion".

Ausgangspunkt der vorliegenden Arbeit ist eine Beschreibung der Halbgruppe der längentreuen sequentiellen Wortfunktionen über einem Alphabet X durch den projektiven Limes einer projektiven Familie endlicher Halbgruppen ([3]). Eine naheliegende Verallgemeinerung dieses Zugangs liefert abzählbar viele verschiedene Klassen längentreuer Wortfunktionen, die im allgemeinen nicht mehr retrospektiv sind. Dabei wollen wir eine Wortfunktion f retrospektiv nennen, wenn gilt: Ist $f(x_1 \dots x_n) = y_1 \dots y_n$, so ist y_i höchstens von $x_1, \dots, x_i$ abhängig (für jedes $i \leqq n$). Die neuen Funktionen erscheinen als Morphismen einer bestimmten Funktorkategorie und sollen wegen starker Analogien zu den sequentiellen Funktionen, die im 2. Abschnitt ausgeführt werden, quasisequentiell genannt werden. Unter diesen spielen die Klassen der sogenannten t-sequentiellen Funktionen eine Sonderrolle, weil sie bezüglich der Substitution Halbgruppen bilden, die zur Halbgruppe der sequentiellen Funktionen isomorph sind. Hinsichtlich des Berechnungsverhaltens unterscheiden sich die quasisequentiellen Funktionen jedoch von den sequentiellen. Dies zeigt insbesondere der Satz 6, der den Grad der Nichtretrospektivität der quasisequentiellen Funktionen genau beschreibt. Eine ausführliche Untersuchung der Berechnung quasisequentieller Funktionen und der dazu nötigen Automatentypen soll in einer weiteren Arbeit erfolgen. Der 3. Abschnitt ist den gegenseitigen Beziehungen verschiedener Klassen quasisequentieller Funktionen gewidmet. Die von der Menge aller quasisequentiellen Funktionen erzeugte Halbgruppe Q der sogenannten quasisequentiellen Funktionen im weiteren Sinne erweist sich als freies Produkt der Halbgruppe der sequentiellen Funktionen und einer Gruppe P von gewissen Wortpermutationen.

Bezeichnungen

Nz= Menge der natürlichen Zahlen (einschließlich der 0); X ist ein endliches Alphabet; $|w|$ ist die Länge des Wortes w; $\subseteqq$ ist die Anfangswortrelation $(u \subseteqq v = =_{\text{Def}} \exists w(uw = v))$; e ist das leere Wort;

$$X^n = \{w : w = x_1 \dots x_n \wedge x_1, \dots, x_n \in X\}; \quad X^0 = \{e\};$$

$$X^* = \bigcup_{n=0}^{\infty} X^n;$$

id ist die identische Abbildung von Nz auf sich bzw. von X^* auf sich (Verwechslungen kommen nicht vor);

$$f \circ g(x) = g(f(x))$$

$\mathfrak{F} = \{t\colon t$ allgemein rekursive Funktion von Nz in $Nz \wedge \forall n(n \in Nz \to 1 \leqq t(n) \leqq n)\}$.

1. Quasisequentielle Funktionen

Wir verwenden den Begriff des projektiven (inversen) Limes einer projektiven Familie (vgl. [1]).

Def. 1. $H = [\{H_\lambda : \lambda \in \Lambda\}, \{\alpha_\lambda^\mu : \lambda, \mu \in \Lambda \wedge \mu \geqq \lambda\}]$ heißt projektive Mengenfamilie $=_{\text{Def}}$

a) Λ ist bezüglich $\leqq$ von unten gerichtete Menge.
b) Jedes H_λ ist eine Menge.
c) α_λ^μ ist eine eindeutige Abbildung von H_μ in H_λ für $\mu \geqq \lambda$, und es gilt: α_μ^μ ist die identische Abbildung von H_μ und

$$\alpha_\lambda^\mu \circ \alpha_\nu^\lambda = \alpha_\nu^\mu.$$

Unter einem Faden versteht man eine Familie $\{x_\lambda : \lambda \in \Lambda\}$ mit $x_\lambda \in H_\lambda$ und $\alpha_\lambda^\mu(x_\mu) = x_\lambda$ für $\mu \geqq \lambda$. Die Menge aller Fäden heißt projektiver Limes von H und wird mit $\varprojlim H$ bezeichnet. Sind alle H_λ Halbgruppen und alle α_λ^μ Homomorphismen, so spricht man von einer projektiven Halbgruppenfamilie.

Im Falle einer projektiven Halbgruppenfamilie wird der projektive Limes ebenfalls eine Halbgruppe, wenn das Produkt zweier Fäden komponentenweise (durch die jeweilige Halbgruppenmultiplikation) definiert wird.

Um die quasisequentiellen Funktionen definieren zu können, stellen wir einige Hilfsbegriffe bereit.

Def. 2. Für $1 \leqq k \leqq n+1$ definieren wir die Abbildung $(\Pi_k)_n^{n+1}$ von X^{n+1} auf X^n durch

$$(\Pi_k)_n^{n+1}(x_1 \dots x_{n+1}) =_{\text{Def}} x_1 \dots x_{k-1} x_{k+1} \dots x_{n+1}.$$

Aus Bequemlichkeitsgründen lassen wir manchmal die Indizes n und $n+1$ weg.

Def. 3

1. Es sei $t \in \mathfrak{F}$. Mit F_t bezeichnen wir die projektive Mengenfamilie $[\{X^n : n \in Nz\}, \{(\alpha_t)_n^m : m, n \in Nz \wedge m \geqq n\}]$ mit $(\alpha_t)_n^m =_{\text{Def}} (\Pi_{t(m)})_{m-1}^m \circ \dots \circ (\Pi_{t(n+1)})_n^{n+1}$, $(\alpha_t)_m^m = \text{id}$.

2. $\mathscr{F}$ sei die Kategorie mit der Objektmenge $\{F_t : t \in \mathfrak{T}\}$, deren Morphismen folgendermaßen definiert sind: $(f_0, f_1, \ldots) \in \mathrm{Hom}\,(F_t, F_{t'}) =_{\mathrm{Def}}$

a) Für jedes $n \in Nz$ ist f_n eine eindeutige Abbildung von X^n in sich.

b) Das Diagramm

$$\begin{array}{lcccc} F_t: & \ldots \leftarrow X^n & \xleftarrow{(\Pi_{t\,(n+1)})_n^{n+1}} & X^{n+1} \leftarrow \ldots \\ & \big\downarrow f_n & & \big\downarrow f_{n+1} \\ F_{t'}: & \ldots \leftarrow X^n & \xleftarrow{(\Pi_{t'\,(n+1)})_n^{n+1}} & X^{n+1} \leftarrow \ldots \end{array}$$

ist kommutativ.

Bemerkung. $\mathscr{F}$ kann als Funktorkategorie einer geeigneten Kategorie in die Kategorie der Mengen aufgefaßt werden.

Die Morphismen von $\mathscr{F}$ bilden den Untersuchungsgegenstand der vorliegenden Arbeit. Ist $(f_0, f_1, \ldots) \in \mathrm{Hom}\,(F_t, F_{t'})$, so ist $f =_{\mathrm{Def}} \bigcup_{i=0}^{\infty} f_i$ eine eindeutige längentreue Abbildung von X^* in X^*, aus der die Folge $(f_0, f_1, \ldots)$ durch $f_i = p_i(f) =_{\mathrm{Def}} f \cap \cap (X^i \times X^i)$ zurückgewonnen werden kann. Diese Wortfunktionen, die mit den Elementen von Hom $(F_t, F_{t'})$ identifiziert werden können, sollen $\langle t, t' \rangle$-sequentiell genannt werden. Weiter legen wir folgende Sprechweisen fest.

Def. 4. $S_{tt'} =_{\mathrm{Def}}$ Klasse aller $\langle t, t' \rangle$-sequentiellen Wortfunktionen. Ist $f \in S_{tt'}$, so nennen wir $\langle t, t' \rangle$ den Typ von f. $S_t =_{\mathrm{Def}} S_{tt}$ ist die Klasse der t-sequentiellen (anstatt $\langle t, t \rangle$-sequentiellen) Funktionen. f heißt quasisequentiell$=_{\mathrm{Df}} \exists t \exists t' (t, t' \in \mathfrak{T} \wedge \wedge f \in S_{tt'})$.

Leicht ergibt sich der

Satz 1. Für jedes $t \in \mathfrak{T}$ ist $[S_t, \circ]$ eine Halbgruppe. $[S_{id}, \circ]$ ist die Halbgruppe der längentreuen sequentiellen Funktionen über X.

Beweis. Die erste Behauptung gilt, weil Hom (F_t, F_t) eine Halbgruppe ist. Die zweite ist aus [3] bekannt. Wir bemerken, daß $[S_t, \circ]$ auch für nichtrekursive t eine Halbgruppe ist. Da wir jedoch nur berechenbare Wortfunktionen betrachten, bleiben solche t unberücksichtigt.

Wir erwähnen nun einen Satz, der entscheidenden Aufschluß über die Struktur der Elemente von $S_{tt'}$ gibt.

Setzt man

$$S_{tt',n} =_{\mathrm{Def}} \{h : \exists f (f \in S_{tt'} \wedge p_n(f) = h)\} \qquad (\subseteq \mathrm{Hom}(X^n, X^n))$$

und

$(\Pi_{tt'})_n^{n+k}(f_{n+k}) =_{\mathrm{Def}} f_n$, falls ein $f \in S_{tt'}$ existiert mit $f_{n+k} = p_{n+k}(f)$ und $f_n = p_n(f)$,

so gilt

Satz 2

1. $\mathfrak{S}_{tt'} =_{\mathrm{Def}} [\{S_{tt',n} : n \in Nz\}, \{(\Pi_{tt'})_n^{n+k} : n, k \in Nz\}]$ ist eine projektive Familie. Im Falle $t = t'$ handelt es sich um eine projektive Familie (endlicher) Halbgruppen.

2. $S_{tt'} = \varprojlim \mathfrak{S}_{tt'}$ (mit $\varprojlim \mathfrak{S}_{tt'}$ bezeichnen wir den projektiven Limes von $\mathfrak{S}_{tt'}$).

Der Beweis wird wie in [3] geführt.

Da offenbar (vgl. das obige Diagramm) $(\Pi_{tt'})_{n-1}^{n}([u, v])=[\Pi_{t(n)}(u), \Pi_{t'(n)}(v)]$ ist, kann Satz 2 folgendermaßen anschaulich interpretiert werden: Streicht man im Argumentwort $x_1 \dots x_n$ der quasisequentiellen Funktion $f(x_1 \dots x_n)$ den Buchstaben $x_{t(n)}$, so entsteht der zugehörige Funktionswert aus $f(x_1 \dots x_n)$ durch Streichen des $t'(n)$-ten Buchstaben. Damit haben wir bereits eine deutliche Vorstellung von der grundsätzlichen Beschaffenheit der quasisequentiellen Funktionen.

Die Bedeutung von Satz 2 besteht darin, daß das Studium von $S_{tt'}$ auf die Untersuchung der endlichen Mengen $S_{tt',n}$ zurückgeführt wird. In [3] sind für den Fall $t=t'=\mathrm{id}$ algebraische Eigenschaften der endlichen Halbgruppen $S_{id,n}$ auf den projektiven Limes, d.h. auf die Halbgruppe der sequentiellen Funktionen übertragen worden. Nach Satz 4 der vorliegenden Arbeit gelten jene Überlegungen auch für beliebige S_t.

2. Eigenschaften der quasisequentiellen Funktionen

Die quasisequentiellen Funktionen können durch eine Reihe weiterer Eigenschaften charakterisiert werden, die gegenüber der einführenden Definition den Vorzug größerer Anschaulichkeit haben.

Wir setzen $\Pi_t =_{\mathrm{Def}} \bigcup_{n=1}^{\infty} (\Pi_{t(n)})_{n-1}^{n}$. Damit können wir die Definition der quasisequentiellen Funktionen in nützlicher Weise umformulieren:

Satz 3. $f \in S_{tt'}$ genau dann, wenn gilt $\Pi_t \circ f = f \circ \Pi_{t'}$.

Beweis. Die genannte Bedingung ist offenbar gleichbedeutend damit, daß die zu f gehörige Folge $(f_0, f_1, \dots)$ zu $\mathrm{Hom}\,(F_t, F_{t'})$ gehört. Das ist aber äquivalent zu $f \in S_{tt'}$.

Als Folgerung aus diesem Satz registrieren wir das

Lemma 1. Ist $f \in S_{tt'}$ und $g \in S_{t't''}$, so ist $f \circ g \in S_{tt''}$.

Beweis. Aus $\Pi_t \circ f = f \circ \Pi_{t'}$ und $\Pi_{t'} \circ g = g \circ \Pi_{t''}$ folgt $\Pi_t \circ f \circ g = f \circ \Pi_{t'} \circ g = f \circ g \circ \Pi_{t''}$. Lemma 1 ergibt sich auch unmittelbar aus

$$\mathrm{Hom}\,(F_t, F_{t'}) \circ \mathrm{Hom}\,(F_{t'}, F_{t''}) = \mathrm{Hom}\,(F_t, F_{t''}).$$

Für das weitere Studium der quasisequentiellen Funktionen sind sogenannte Wortpermutationen von Bedeutung.

Def. 5. Eine Wortfunktion s heißt Wortpermutation $=_{\mathrm{Def}}$ Für jedes n gibt es eine Permutation σ_n von $\{1, \dots, n\}$ mit $\forall n \forall x_1 \dots \forall x_n \big(n \in Nz \wedge x_1, \dots, x_n \in X \rightarrow s(x_1 \dots x_n) = x_{\sigma_n(1)} \dots x_{\sigma_n(n)}\big)$. s heißt die zu $\{\sigma_n : n \in Nz\}$ gehörige Wortpermutation.

Lemma 2. In jedem $S_{tt'}$ gibt es genau eine Wortpermutation $s_{tt'}$.

Beweis

1. Eindeutigkeit. Wir zeigen: Sind s und $\bar{s}$ Wortpermutationen in $S_{tt'}$, so gilt $s=\bar{s}$. Dazu beweisen wir $s_n=\bar{s}_n$ $\big(s_n=p_n(s), \bar{s}_n=p_n(\bar{s})\big)$ durch Induktion über n.

a) $s_1=\mathrm{id}_1=\bar{s}_1$.

b) Sei $s_n=\bar{s}_n$. Damit gilt nach Satz 3 für $|w| = n+1$

$$\Pi_{t'(n+1)}\big(s_{n+1}(w)\big) = s_n\big(\Pi_{t(n+1)}(w)\big) = \bar{s}_n\big(\Pi_{t(n+1)}(w)\big) = \Pi_{t'(n+1)}\big(\bar{s}_{n+1}(w)\big). \quad (1)$$

Es sei $s_{n+1}(w)=x_{\sigma_{n+1}(1)}\dots x_{\sigma_{n+1}(n+1)}$ und $\bar{s}_{n+1}(w)=x_{\bar{\sigma}_{n+1}(1)}\dots x_{\bar{\sigma}_{n+1}(n+1)}$. Wegen (1) folgt $\sigma_{n+1}(i)=\bar{\sigma}_{n+1}(i)$ für $i\neq t'(n+1)$. Da σ_{n+1} und $\bar{\sigma}_{n+1}$ Permutationen von $\{1,\dots,n+1\}$ sind, folgt hieraus auch $\sigma_{n+1}(t'(n+1))=\bar{\sigma}_{n+1}(t'(n+1))$, d. h. $s_{n+1}==\bar{s}_{n+1}$.

2. Existenz. Wir setzen

Def. 6. $\sigma_1(1)=_{\text{Def}}1$.

$$\sigma_{n+1}(i)=_{\text{Def}}\begin{cases}\sigma_n(i), & \text{falls } i<t'(n+1)\wedge\sigma_n(i)<t(n+1)\\ 1+\sigma_n(i), & \text{falls } i<t'(n+1)\wedge\sigma_n(i)\geqq t(n+1)\\ \sigma_n(i-1), & \text{falls } i>t'(n+1)\wedge\sigma_n(i-1)<t(n+1)\\ 1+\sigma_n(i-1), & \text{falls } i>t'(n+1)\wedge\sigma_n(i-1)\geqq t(n+1)\\ t(n+1), & \text{falls } i=t'(n+1).\end{cases}$$

Die zugehörige Wortpermutation werde mit $s_{tt'}$ bezeichnet.

Mit Hilfe von Satz 3 zeigen wir, daß $s_{tt'}$ zu $S_{tt'}$ gehört. Denn einerseits gilt nach Definition von σ_{n+1}

$$\Pi_{t'(n+1)}(s_{tt'}(x_1\dots x_{n+1}))=x_{\sigma_{n+1}(1)}\dots x_{\sigma_{n+1}(t'(n+1)-1)}x_{\sigma_{n+1}(t'(n+1)+1)}\dots x_{\sigma_{n+1}(n+1)}$$

$$=x_{\sigma_n(1)+\varepsilon_1}\dots x_{\sigma_n(t'(n+1)-1)+\varepsilon_{t'(n+1)-1}}x_{\sigma_n(t'(n+1))+\varepsilon_{t'(n+1)}}\dots x_{\sigma_n(n)+\varepsilon_n}$$

mit

$$\varepsilon_i=\begin{cases}1 & \text{falls } \sigma_n(i)\geqq t(n+1)\\ 0 & \text{sonst.}\end{cases}$$

Andererseits ist

$$s_{tt'}(\Pi_{t(n+1)}(x_1\dots x_{n+1}))=s_{tt'}(x_1\dots x_{t(n+1)-1}x_{t(n+1)+1}\dots x_{n+1}).$$

Hieraus folgt mit der vorübergehenden Umbenennung

$$y_1=_{\text{Def}}x_1,\ \dots,\ y_{t(n+1)-1}=_{\text{Def}}x_{t(n+1)-1},\ y_{t(n+1)}=_{\text{Def}}x_{t(n+1)+1},\ \dots$$

$$\dots,\ y_n=_{\text{Def}}x_{n+1}=y_{\sigma_n(1)}\dots y_{\sigma_n(n)},$$

woraus sich durch Rückbenennung der y_i unter Benutzung der eben eingeführten ε_i ergibt

$$s_{tt'}(\Pi_{t(n+1)}(x_1\dots x_{n+1}))=x_{\sigma_n(1)+\varepsilon_1}\dots x_{\sigma_n(n)+\varepsilon_n}.$$

Damit haben wir

$$\Pi_{t'(n+1)}(s_{tt'}(x_1\dots x_{n+1}))=s_{tt'}(\Pi_{t(n+1)}(x_1\dots x_{n+1})),$$

also $\Pi_t\circ s_{tt'}=s_{tt'}\circ\Pi_{t'}$ und nach Satz 3 $s_{tt'}\in S_{tt'}$.

Als einfache Folgerung aus der Definition 6 notieren wir

Lemma 3. $s_{tt'}=\mathrm{id}$ genau dann, wenn $t=t'$ ist.

Hieraus ergibt sich

Lemma 4. Es ist $s_{tt'}^{-1}=s_{t't}$.

Beweis. Nach Lemma 1 gehört die Wortpermutation $s_{tt'} \circ s_{t't}$ zu S_t. Demnach ist nach Lemma 3 $s_{tt'} \circ s_{t't} = \mathrm{id}$. Ebenso erhält man $s_{t't} \circ s_{tt'} = \mathrm{id}$, woraus die Behauptung folgt.

Wir sind nun in der Lage, eine zweite Charakterisierung der Funktionen aus $S_{tt'}$ zu beweisen.

Satz 4. $f \in S_{tt'}$ genau dann, wenn es eine sequentielle Funktion φ gibt mit $f = s_{tid} \circ \varphi \circ s_{idt'}$.

Beweis

1. Es sei φ sequentiell. Dann gilt für φ die Beziehung $\Pi_{id} \circ \varphi = \varphi \circ \Pi_{id}$ (nach Satz 1 (zweiter Teil) und Satz 3). Hieraus folgt

$$\Pi_t \circ f = \Pi_t \circ s_{tid} \circ \varphi \circ s_{idt'} = s_{tid} \circ \Pi_{id} \circ \varphi \circ s_{idt'} = s_{tid} \circ \varphi \circ \Pi_{id} \circ s_{idt'} =$$

$$= s_{tid} \circ \varphi \circ s_{idt'} \circ \Pi_{t'} = f \circ \Pi_{t'}.$$

Daher ist $f \in S_{tt'}$ nach Satz 3.

2. Es sei $f \in S_{tt'}$. Wir setzen $\varphi = s_{idt} \circ f \circ s_{t'id}$ und erhalten

$$\Pi_{id} \circ \varphi = \Pi_{id} \circ s_{idt} \circ f \circ s_{t'id} = s_{idt} \circ \Pi_t \circ f \circ s_{t'id} = s_{idt} \circ f \circ \Pi_{t'} \circ s_{t'id} =$$

$$= s_{idt} \circ f \circ s_{t'id} \circ \Pi_{id} = \varphi \circ \Pi_{id}.$$

Daher ist φ sequentiell. Nach Lemma 4 erhalten wir $f = s_{tid} \circ \varphi \circ s_{idt'}$.

Aus Satz 4 ergibt sich erneut, daß jedes S_t eine Halbgruppe bezüglich $\circ$ ist. Darüber hinaus haben wir die

Folgerung. Für beliebiges $t \in \mathfrak{F}$ ist die Halbgruppe $[S_t, \circ]$ isomorph zur Halbgruppe $[S_{id}, \circ]$ der sequentiellen Funktionen.

Denn die Abbildung $\theta(\varphi) =_{\mathrm{Df}} s_{tid} \circ \varphi \circ s_{idt}$ ist offenbar eine eineindeutige, bezüglich $\circ$ relationentreue Abbildung von S_{id} auf S_t.

Damit übertragen sich alle Halbgruppeneigenschaften der sequentiellen Funktionen auf die Halbgruppen S_t.

Um weitere Eigenschaften der quasisequentiellen Funktionen angeben zu können, definieren wir für jedes t eine zweistellige Relation $\subseteqq_t$ auf X^*.

Def. 7. $w_1 \subseteqq_t w_2 =_{\mathrm{Def}} s_{tid}(w_1) \subseteqq s_{tid}(w_2)$.

Lemma 5. $|v| = |u| + 1 \wedge u \subseteqq_t v \leftrightarrow \Pi_t(v) = u$.

Beweis. $|v| = |u| + 1 \wedge u \subseteqq_t v \leftrightarrow |v| = |u| + 1 \wedge s_{tid}(u) \subseteqq s_{tid}(v) \leftrightarrow s_{tid}(u) = \Pi_{id}(s_{tid}(v)) = s_{tid}(\Pi_t(v)) \leftrightarrow u = \Pi_t(v)$.

Damit sind wir in der Lage, die übliche Beschreibung der sequentiellen Funktionen als Homomorphismen der Anfangswortrelation auf die quasisequentiellen Funktionen zu übertragen.

Satz 5. $f \in S_{tt'} \leftrightarrow \forall u \forall v (u \subseteqq_t v \rightarrow f(u) \subseteqq_{t'} f(v))$.

Dieser Satz stellt eigentlich nur eine andere Interpretation der Aussage von Satz 3 dar.

Beweis

1. Es sei $f \in S_{tt'}$. Dann existiert nach Satz 4 eine sequentielle Funktion φ mit

$$s_{tid} \circ \varphi = f \circ s_{t'id}. \qquad (2)$$

Sei nun $u \sqsubseteq_t v$. Nach Definition 7 und wegen der Sequentialität von φ ergibt sich $\varphi(s_{tid}(u)) \sqsubseteq \varphi(s_{tid}(v))$. Wegen (2) erhalten wir

$$s_{t'id}(f(u)) \sqsubseteq s_{t'id}(f(v)),$$

woraus sich nach Definition 7

$$f(u) \sqsubseteq_{t'} f(v)$$

ergibt.

2. f erfülle die Bedingung von Satz 5. Da nach Lemma 5 für jedes $u \in X^*$ die Beziehung $\Pi_t(u) \sqsubseteq_t u$ gilt, folgt insbesondere $f(\Pi_t(u)) \sqsubseteq_{t'} f(u)$. Wegen $|f(u)| = |f(\Pi_t(u))| + 1$ folgt wiederum nach Lemma 5 $f(\Pi_t(u)) = \Pi_{t'}(f(u))$. Also ist $\Pi_t \circ f = f \circ \Pi_{t'}$, und nach Satz 3 ist $f \in S_{tt'}$.

Wird φ die Aussage dieses Satzes als Definition verwendet, so können durch Verzicht auf die Längentreue auch nichtlängentreue quasisequentielle Funktionen definiert werden.

Im folgenden Satz geben wir eine explizite Beschreibung der quasisequentiellen Funktionen an. Es seien s_{tid} und $s_{idt'}$ die gemäß Definition 5 zu $\{\sigma_n : n \in Nz\}$ und $\{\sigma'_n : n \in Nz\}$ gehörigen Wortpermutationen. $s_{tt'}$ gehört offenbar zu $\{\tau_n : n \in Nz\}$ mit $\tau_n = \sigma'_n \circ \sigma_n$.

Def. 8. Es sei $w = x_1 \dots x_n$. Dann setzen wir

$$w_{nj}^{tt'} =_{\mathrm{Def}} x_{\sigma_n(1)} \dots x_{\sigma_n(\sigma'_n(j)-1)}.$$

Wenn keine Verwechslungen vorkommen können, schreiben wir kürzer $w_{nj} = w_{nj}^{tt'}$.

Satz 6. f gehört genau dann zu $S_{tt'}$, wenn es eine eindeutige Abbildung α von X^* in X^X gibt, die jedem $p \in X^*$ ein $\alpha_p \in X^X$ zuordnet, so daß für alle $n \in Nz$ und alle $w = x_1 \dots x_n \in X^n$ gilt

$$f(w) = \alpha_{w_{n1}}(x_{\tau_n(1)}) \dots \alpha_{w_{nn}}(x_{\tau_n(n)}).$$

Beweis

1. Es sei $f \in S_{tt'}$. Nach Satz 4 gibt es ein sequentielles φ mit $\varphi = s_{tid} \circ \varphi \circ s_{idt'}$. Das bedeutet

$$f(w) = s_{idt'}(\varphi(s_{tid}(w))) = s_{idt'}(\varphi(x_{\sigma_n(1)} \dots x_{\sigma_n(n)})) =$$
$$= s_{idt'}(\varphi(x_{\sigma_n(1)}) \varphi_{x_{\sigma_n(1)}}(x_{\sigma_n(2)}) \dots \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(n-1)}}(x_{\sigma_n(n)})).$$

Zur Erleichterung der Rechnung führen wir vorübergehend die Abkürzung $y_i =_{\mathrm{Def}} \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(i-1)}}(x_{\sigma_n(i)})$ ein. Damit ergibt sich

$$f(w) = s_{idt'}(y_1 \dots y_n) = y_{\sigma'_n(1)} \dots y_{\sigma'_n(n)}$$

$$f(w) = \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(\sigma'_n(1)-1)}}(x_{\sigma_n(\sigma'_n(1))}) \dots \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(\sigma'_n(n)-1)}}(x_{\sigma_n(\sigma'_n(n))}) =$$
$$= \varphi_{w_{n1}}(x_{\tau_n(1)}) \dots \varphi_{w_{nn}}(x_{\tau_n(n)}) \qquad (3)$$

mit den in Definition 8 eingeführten w_{nj}. Die im Satz genannte Abbildung α ordnet jedem p den zugehörigen Zustand φ_p zu.

2. f habe eine Darstellung der in Satz 6 angegebenen Form. Damit verfügen wir über eine sequentielle Abbildung α, die durch $\alpha(x_1 \dots x_n) = \alpha_e(x_1)\alpha_{x_1}(x_2) \dots \alpha_{x_1 \dots x_{n-1}}(x_n)$ definiert ist. Die Umkehrung der im ersten Teil des Beweises durchgeführten Rechnung mit α an Stelle von φ liefert

$$f(x_1 \dots x_n) = \alpha_{w_{n1}}(x_{\tau_n(1)}) \dots \alpha_{w_{nn}}(x_{\tau_n(n)}) = s_{idt'}(\alpha(s_{tid}(x_1 \dots x_n))).$$

Daher gilt $f = s_{tid} \circ \alpha \circ s_{idt'}$, und nach Satz 4 ist $f \in S_{tt'}$.

Zur Erläuterung von Satz 6 geben wir ein Beispiel einer $\langle t, t' \rangle$-sequentiellen Funktion an. Die Permutationen σ_n, σ'_n, τ_n, durch die wie oben die Wortpermutationen s_{tid}, $s_{idt'}$ und $s_{tt'}$ definiert sind, seien durch Tabellen gegeben, von denen wir die ersten 6 Zeilen angeben wollen:

σ	1	2	3	4	5	6
1	1					
2	2	1				
3	3	1	2			
4	3	1	2	4		
5	4	1	3	5	2	
6	5	2	4	6	3	1

σ'	1	2	3	4	5	6
1	1					
2	1	2				
3	3	1	2			
4	3	1	2	4		
5	3	1	5	2	4	
6	3	1	5	2	4	6

τ	1	2	3	4	5	6
1	1					
2	2	1				
3	2	3	1			
4	2	3	1	4		
5	3	4	2	1	5	
6	4	5	3	2	6	1

In der i-ten Zeile der σ-Tabelle stehen von links nach rechts $\sigma_i(1) \dots \sigma_i(i)$. Ebenso sind die anderen Tabellen zu lesen. φ sei eine sequentielle Funktion, und $f = s_{tid} \circ \varphi \circ s_{idt'}$. Dann ist zum Beispiel

$$f(x_1 x_2 \dots x_6) = \varphi_{x_5 x_2}(x_4)\varphi(x_5)\varphi_{x_5 x_2 x_4 x_6}(x_3) \cdot \varphi_{x_5}(x_2)\varphi_{x_5 x_2 x_4}(x_6)\varphi_{x_5 x_2 x_4 x_6 x_3}(x_1).$$

Die Verwandtschaft zur Sequentialität wird deutlicher bei Betrachtung der Klassen S_{tid}, $S_{idt'}$, S_t (hierbei sind alle σ'_n bzw. σ_n bzw. τ_n die identischen Permutationen).

a) Ist $f \in S_{tid}$, so gilt mit passendem φ

$$f(x_1 \dots x_n) = \varphi(x_{\sigma_n(1)})\varphi_{x_{\sigma_n(1)}}(x_{\sigma_n(2)}) \dots \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(n-1)}}(x_{\sigma_n(n)}).$$

b) Ist $f \in S_{idt}$, so gilt mit passendem φ

$$f(x_1 \dots x_n) = \varphi_{x_1 \dots x_{\sigma'_n(1)-1}}(x_{\sigma'_n(1)}) \dots \varphi_{x_1 \dots x_{\sigma'_n(n)-1}}(x_{\sigma'_n(n)}).$$

c) Ist $f \in S_t$, so gilt mit passendem φ

$$f(x_1 \dots x_n) = \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(\sigma'_n(1)-1)}}(x_1) \dots \varphi_{x_{\sigma_n(1)} \dots x_{\sigma_n(\sigma'_n(n)-1)}}(x_n).$$

3. Beziehungen der Klassen $S_{tt'}$ untereinander

Zunächst beweisen wir den einfachen

Satz 7

1) Jedes $f \in S_{tt'}$ mit allgemein rekursiver zugehöriger sequentieller Funktion φ (vgl. Satz 4) ist allgemein rekursiv.

2) Nicht jede längentreue allgemein rekursive Funktion von X^* in X^* ist quasisequentiell.

Beweis

1. Da mit t und t' auch s_{tid} und $s_{idt'}$ allgemein rekursiv sind, folgt die Behauptung aus Satz 4.

2. Die durch

$$f(x_1 \dots x_{2n}) = x_{2n-1}x_{2n}x_{2n-2} \dots x_2x_1x_3 \dots x_{2n-3}$$

$$f(x_1 \dots x_{2n+1}) = x_{2n+1}x_{2n}x_{2n-2} \dots x_2x_1x_3 \dots x_{2n-3}x_{2n-1}$$

festgelegte Funktion f ist offenbar allgemein rekursiv, aber sie gehört zu keiner Klasse $S_{tt'}$. Denn man prüft leicht folgendes nach: Streicht man einen beliebigen Buchstaben aus dem Wort $x_1x_2x_3x_4x_5$ und wendet man darauf f an, so kann das entsprechende Bildwort nicht aus $f(x_1x_2x_3x_4x_5)$ durch Streichen eines passenden Buchstabens gewonnen werden.

Satz 8. Ist $t_1 \neq t_3$ oder $t_2 \neq t_4$, so gibt es in $S_{t_1t_2}$ Funktionen, die nicht in $S_{t_3t_4}$ liegen.

Beweis

1. Fall. Es sei $t_1 \neq t_3$. Dann gibt es ein n mit $i =_{\text{Def}} t_1(n) \neq t_3(n) =_{\text{Def}} j$. Ferner sei $t_4(n) = k$. s_{t_1id} sei durch $\{\sigma_n^1 : n \in Nz\}$ gegeben. Nach Definition 6 gilt $\sigma_n^1(n) = 1$. Aus (3) (im Beweis von Satz 6) folgt hiermit, daß der j-te Buchstabe des Arguments $x_1 \dots x_n$ von f an wenigstens zwei Stellen in das Bildwort eingeht. Daher ist die Angabe einer Funktion $f \in S_{t_1t_2}$ so möglich, daß

$$f(x_1 \dots x_{j-1}x_jx_{j+1} \dots x_n) = y_1 \dots y_l \dots y_n$$

und für $x_j' \neq x_j$

$$f(x_1 \dots x_{j-1}x_j'x_{j+1} \dots x_n) = y_1' \dots y_l' \dots y_n'$$

gilt, wobei $y_l \neq y_l'$ wenigstens für ein $l \neq k$ vorkommt. Dieses f gehört nicht zu $S_{t_3t_4}$, weil sonst einerseits

$$f(x_1 \dots x_{j-1}x_{j+1} \dots x_n) = \Pi_k(y_1 \dots y_l \dots y_n)$$

andererseits aber

$$f(x_1 \dots x_{j-1}x_{j+1} \dots x_n) = \Pi_k(y_1' \dots y_l' \dots y_n')$$

wäre, was wegen $\Pi_k(y_1 \dots y_l \dots y_n) \neq \Pi_k(y_1' \dots y_l' \dots y_n')$ unmöglich ist.

2. Fall. Es sei $t_2 \neq t_4$. Dann gibt es ein n mit $m =_{\text{Def}} t_2(n) \neq t_4(n) =_{\text{Def}} k$. Wir Betrachten nur den Fall $m < k \wedge i = t_1(n) \leqq j = t_3(n)$. Die anderen Fälle lassen sich analog behandeln. Es sei $w = x_1 \dots x_i \dots x_j \dots x_n$ mit $x_i = x_{i+1} = \dots = x_j$, so daß $\Pi_i(w) = \Pi_j(w)$. Die Funktion $f \in S_{t_1t_2}$ wird so gewählt, daß

$$f(w) = y_1 \dots y_m y_{m+1} \dots y_n \quad \text{mit} \quad y_m \neq y_{m+1}$$

gilt. Dies läßt sich erreichen, denn nach Satz 6 ist

$$y_m = \alpha_{w_{nm}}(\tau_n(m)),\ y_{m+1} = \alpha_{w_{n,m+1}}(\tau_n(m+1))$$

und

$$w_{nm} \neq w_{n,m+1}.$$

Wäre $f \in S_{t_3t_4}$, so müßte $\Pi_{t_3} \circ f = f \circ \Pi_{t_4}$ gelten, d.h.

$$f(\Pi_{t_3}(w)) = y_1 \dots y_m y_{m+1} \dots y_{k-1}y_{k+1} \dots y_n.$$

Wegen $f \in S_{t_1 t_2}$ haben wir

$$f(\Pi_{t_1}(w)) = \Pi_{t_2}(f(w)) = y_1 \dots y_{m-1} y_m \dots y_n,$$

und aus

$$\Pi_{t_3}(w) = \Pi_j(w) = \Pi_i(w) = \Pi_{t_1}(w)$$

folgte $f(\Pi_{t_3}(w)) = f(\Pi_{t_1}(w))$, was wegen $y_m \neq y_{m+1}$ unmöglich ist.

Folgerung 1. Jede Klasse $S_{tt'}$ mit $t \neq \mathrm{id}$ oder $t' \neq \mathrm{id}$ enthält nichtsequentielle Funktionen.

Beweis. Man wende Satz 8 für $t_3 = t_4 = \mathrm{id}$ an.

Folgerung 2. Für $t_1 \neq t_3$ oder $t_2 \neq t_4$ ist niemals $S_{t_1 t_2} \subseteq S_{t_3 t_4}$.
Aus Satz 6 ergibt sich leicht der

Satz 9. Jede Klasse S_t enthält alle Homomorphismen, (d.h. alle sequentiellen Funktionen vom Gewicht 1).

Beweis. Jeder Homomorphismus kann gemäß Satz 6 bei beliebigem t mit Hilfe einer konstanten Abbildung α realisiert werden.

Wir untersuchen jetzt $S_{t_1} \cap S_{t_2}$ für $t_1 \neq t_2$. Es sei $f \in S_{t_1} \cap S_{t_2}$. Nach Satz 6 sind dann zwei verschiedene Darstellungen für f möglich:

$$f(x_1 \dots x_n) = \alpha_{w^1_{n1}}(x_1) \dots \alpha_{w^1_{nn}}(x_n),$$

$$f(x_1 \dots x_n) = \bar{\alpha}_{w^2_{n1}}(x_1) \dots \bar{\alpha}_{w^2_{nn}}(x_n).$$

Dabei sind die w^1_{ni} bzw. w^2_{ni} die nach Definition 8 zu t_1 bzw. t_2 gehörigen Wörter.

Hieraus ergibt sich als notwendige Bedingung für $f \in S_{t_1} \cap S_{t_2}$: $\alpha_{w^1_{ni}} = \bar{\alpha}_{w^2_{ni}}$ für $i = 1, \dots, n$. Um eine notwendige und hinreichende Bedingung zu erhalten, definieren wir

$$p \sim_{t_1 t_2} q =_{\mathrm{Def}} \exists v \exists w \exists n \exists m \exists j \exists k (p = w^1_{nj} \wedge q = v^1_{mk} \wedge w^2_{nj} = v^2_{mk}).$$

$\approx_{t_1 t_2}$ sei die transitive Hülle von $\sim_{t_1 t_2}$.

Ist f gemäß Satz 6 durch α dargestellt, so setzen wir

$$p \equiv_\alpha q =_{\mathrm{Def}} \alpha_p = \alpha_q.$$

Dann gilt offenbar der

Satz 10. $f \in S_{t_1}$ gehört genau dann zu S_{t_2}, wenn gilt $\forall p \forall q (p \approx_{t_1 t_2} q \rightarrow p \equiv_\alpha q)$, d.h., wenn $\equiv_\alpha$ eine Vergröberung der transitiven Hülle von $\sim_{t_1 t_2}$ ist.

Man prüft auf Grund dieses Satzes leicht nach, daß beispielweise $S_1 \cap S_{id} =$ = Menge aller Homomorphismen ist. Dabei ist 1 die konstante Funktion 1. Dasselbe ergibt sich für $S_t \cap S_{id}$ mit

$$t(n) =_{\mathrm{Def}} \begin{cases} n-1 & \text{für } n > 1, \\ 1 & n = 1. \end{cases}$$

Bisher ist die Menge der quasisequentiellen Funktionen als Kategorie betrachtet worden. Wir stellen nun die Frage nach der durch diese Menge erzeugten Halbgruppe $[Q, \circ]$. Die Elemente von Q nennen wir quasisequentielle Funktionen im weiteren Sinne. Q enthält nämlich mehr als die quasisequentiellen Funktionen. Um dies einzusehen, betrachten wir noch einmal die Funktion f aus dem zweiten Teil des Beweises von Satz 7. f ist das Produkt zweier quasisequentieller Funktionen g und h, die folgendermaßen definiert sind

$$g(x_1 \dots x_{2n}) = x_{2n} \dots x_4 x_2 x_1 x_3 \dots x_{2n-1}$$

$$g(x_1 \dots x_{2n+1}) = x_{2n} \dots x_4 x_2 x_1 x_3 \dots x_{2n-1} x_{2n+1}$$

und

$$h(x_1 \dots x_n) = x_n x_1 \dots x_{n-1}.$$

$$g \in S_{idt_1} \text{ mit } t_1(n) = \begin{cases} 1, & \text{falls} \quad n \equiv 0(2), \\ n, & \text{falls} \quad n \equiv 1(2). \end{cases}$$

Ist t_2 eine beliebige Funktion aus $\mathfrak{F}$ mit $\forall n(n > 1 \to t_2(n) < n)$ und $t_3 = t_2 + 1$, so ist $h \in S_{t_2 t_3}$. Demnach ist $f = g \circ h$ quasisequentiell im weiteren Sinne. f ist jedoch nicht quasisequentiell, wie aus dem Beweis von Satz 7 bekannt ist.

Es sei P die Gruppe der Wortpermutationen, die von $\{s_{tid} : t \in \mathfrak{F}\}$ erzeugt wird. Bezeichnet $*$ die Bildung des freien Produkts, so gilt trivialerweise der

Satz 11. $Q = S_{id} * P$.

Damit ist das Studium der quasisequentiellen Funktionen im weiteren Sinne im wesentlichen auf die Untersuchung der Wortpermutationen zurückgeführt. Diese lassen sich indessen nicht durch einfache Eigenschaften charakterisieren.

Квазипоследовательностные функции

Обобщается определение последовательностных функции как проективный предел проективного семейства конечных полугрупп. Возникают таким образом новые классы так называемых квазипоследовательностных функции разных типов. Исследуются свойства этих функции и их соотношения к последовательностным функциям.

SEKTION MATHEMATIK
FRIEDRICH SCHILLER UNIVERSITÄT
69 JENA, DDR
UNIVERSITÄTS HOCHHAUS

Literatur

[1] Bourbaki, N., *Théorie des ensembles*, Chapitre III, Ensembles ordonnés, cardinaux, nombres entiers, Hermann, Paris, 1956.
[2] Raney, G. N., Sequential functions, *J. Assoc. Comput. Mach.* v. 5, 1958, pp. 177—180.
[3] Wechsung, G., Die Gruppe der längentreuen eindeutigen sequentiellen Funktionen über festem Alphabet, *Elektron. Datenverarbeit. Kybernetik*, v. 8, 1972, pp. 335—352.

(Eingegangen am 11. April 1972)

Generalized context-free grammars

By J. GRUSKA

1. *Introduction*. Generalized context-free grammars can be thought of as context-free grammars all rules of which are of the form $A \to \alpha$ where α is a regular expression. Generalized context-free grammars and their representation by a set of finite-state diagrams are a convenient tool to describe context-free languages. In this paper a classification of context-free languages according to the minimal number of non-terminals of generalized context-free grammars is studied and the corresponding decision problems are investigated.

2. *Definitions*. By a generalized context-free grammar we mean a quadruple $G=\langle V, \Sigma, P, \sigma \rangle$ where V, Σ and σ have the same meaning as for context-free grammars (see [2]) and P is a set (maybe infinite) of context-free rules such that for any nonterminal $A \in V-\Sigma$, the set $\{w; A \to w \in P\} \subset V^*$ is regular. The relations $\Rightarrow$ and $\overset{*}{\Rightarrow}$ for a generalized context-free grammar are defined in the same way as for context-free grammars.

It is obvious that a language L is context-free if and ony if $L=L(G)$ for a generalized context-free grammar G.

3. *Representations*. A generalized context-free grammar $G=\langle V, \Sigma, P, \sigma \rangle$ can be represented by a finite set of rules $A \to \alpha$, one for each nonterminal in $V-\Sigma$, where α is a regular expression over V. This in turn means that a generalized context-free grammar can be represented by a finite set of transition diagrams, one for each nonterminal of G, each of which represents a finite-state automaton which is capable of recursively calling other finite state automata [1], or G can be represented by a finite set of the so-called flag diagrams, one for each nonterminal of G [4].

4. *Problems*. As suggested by Kalmár [4], for a context-free language L let $N(L)$ be the minimum of the number of non-terminals of generalized context-free grammars generating L. Since $N(L)$ is also the minimum of transition diagrams for L, $N(L)$ may be thought of as a measure of non-finite state character of L.

5. *Results*. It will be shown now that for any integer n there is a context-free language L_n such that $N(L_n)=n$ and that there is no effective way to calculate $N(L)$.

Theorem 1. For any integer n there is a context-free language $L_n \subset \{a, b\}^*$ such that $N(L_n)=n$.

Proof. The case $n=1$ is trivial. Let now $n>1$ and let L_n be the language generated by the context-free grammar

$$\sigma \to a\sigma b, \qquad \sigma \to aba^2A_2bab$$
$$A_i \to a^iA_ib, \quad A_i \to ba^{i+1}A_{i+1}ba \qquad 2 \leqq i \leqq n-1$$
$$A_n \to a^nA_nb, \quad A_n \to b\sigma a, \quad A_n \to b^2a^2.$$

Let G be a generalized context-free grammar generating L_n and such that no generalized context-free grammar for L_n has fewer nonterminals. It means that from any nonterminal of G an infinite set of terminal words can be derived. All words of L_n posses a very regular structure. It holds

(1) If $x \in L_n$, then $x=ub^2a^2v$, u (v) is uniquelly determined by v (by u) and neither u nor v contains b^2a^2 as a subword.

From (1) it follows

(2) All rules of G are of the form $A \to uBv$ or $A \to ub^2a^2v$ where $u, v \in \Sigma^*$ and $B \in V-\Sigma$.

(3) If $A \to uBv$, $A \to u'Bv$ or $A \to ub^2a^2v$ and $A \to u'b^2a^2v$ are rules of G, then $u=u'$. If $A \to uBv$, $A \to uBv'$ or $A \to ub^2a^2v$, $A \to ub^2a^2v'$ are rules of G, then $v=v'$.

Since for any nonterminal A of G, the set $\{w; A \to w \in P\}$ is regular, it follows easily from (1) to (3) that the set P must be finite and therefore G is a "normal" context-free grammar. It was shown in [3], that the language L_n can not be generated by a context-free grammar having less than n nonterminals and therefore $N(L_n) \geqq n$. Since $N(L_n) \leqq n$ is obviously true we get the theorem.

Theorem 2. Let $n \geqq 1$ be an integer. It is undecidable for an arbitrary context-free grammar G whether or not $N(L(G))=n$.

Proof. Let us first consider the case $n=1$. Let x and y be arbitrary m-tuples of non-empty words over the alphabet $\{a, b\}$. Let $L(x)$, $L(x, y)$ and L_s be the languages defined by

$$L(x) = \{ba^{i_1}ba^{i_2} \dots ba^{i_k}cx_{i_k} \dots x_{i_2}x_{i_1};\ 1 \leqq i_j \leqq m\}$$
$$L(x,y) = L(x)cL^R(y)$$
$$L_s = \{w_1cw_2cw_2^Rcw_1^R;\ w_1w_2 \in \{a,b\}^*\}$$

where, for a word w, w^R is the reverse of w and for a language L, $L^R = \{w^R; w \in L\}$.

By [2], given x and y, one can effectively construct a context-free grammar $G_{x,y}$ generating the language

$$L_{x,y} = \{a, b, c\}^* - L(x, y) \cap L_s.$$

If $L(x, y) \wedge L_s = \emptyset$, then obviously $N(L_{x,y})=1$. Let us now consider the case $L(x, y) \wedge L_s \neq \emptyset$ and let us assume that again $N(L_{x,y})=1$. Then there is a generalized context-free grammar $G=\langle V, \Sigma, P, \sigma \rangle$ with only one nonterminal σ which generates the language $L_{x,y}$.

Since $L(x, y) \wedge L_s \neq \emptyset$, there are indices $i_1, \dots, i_k$ such that if we denote

$$I = ba^{i_1} \dots ba^{i_k}, \quad X = x_{i_k} \dots x_{i_1}, \quad j = I^R, \quad Y = X^R$$

then $I^rcX^rcY^rcJ^r \in L_{x,y}$ for no integer $r \geqq 1$.

Since the set $R = \{\alpha; \sigma \to \alpha \in P\}$ is regular, there must exists an integer N such that if $i>N$, then $z_i = I^icX^{i+1}cY^{i+1}cJ^{i+1} \notin R$ and, moreover, if $u_i\sigma v_i \in R$, $u_iv_i \neq \varepsilon$,

$u_i \in \{a, b, c\}^*$, $u_i \sigma v_i \overset{*}{\Rightarrow} z_i$, then u_i does not contain the symbol c. Hence there exists a word $\bar{u}_i c \bar{v}_i \in L(G)$ such that $\bar{u}_i \in \{a, b\}^*$ and $u_i \bar{u}_i c \bar{v}_i v_i = z_i$. But then the word $\bar{u}_i I c \bar{v}_i$ is also in $L_{x,y}$ and therefore $L(G)$ generates the word $u_i \bar{u}_i I c \bar{v}_i v_i = I^{i+1} c X^{i+1} c Y^{i+1} c J^{i+1} \notin$ $\notin L_{x,y}$ what is a contradiction. Thus $N(L_{x,y}) = 1$ if and only if $L(x, y) \wedge L_s = \emptyset$. Since it is undecidable for arbitrary x and y whether or not $L(x, y) \wedge L_s = \emptyset$ [2], we get the theorem for the case $n = 1$.

For $n > 1$ we proceed as follows. By Theorem 2, for $n > 2$ there is a context-free language $L_{n-2} \subset \{d, e\}^*$ such that $N(L_{n-2}) = n - 2$. For $n = 2$ let us consider the language $L_{x,y,2} = \{a, b, c\}^* - L(x, y) \wedge L_s \cup \{f\}$ and for $n > 2$ let $L_{x,y,n} = L_{x,y} \cup \{f\} \cup L_{n-2}$ where f, d, e are new symbols. It is easy to verify that $N(L_{x,y,n}) = n$ if and only if $L(x, y) \wedge L_s = \emptyset$ and now the theorem for the case $n > 1$ follows in the same way as for $n = 1$.

Corollary. There is no effective way to construct for an arbitrary context-free grammar G a generalized context-free grammar with fewest states and generating the language $L(G)$.

It follows from this corollary that there is no effective way to determine for an arbitrary context-free grammar G the minimum of transition diagrams for the language $L(G)$. Can we, however, at least to minimize effectively the overall number of states of transition diagrams for $L(G)$? It was shown implicitly in the course of the proof of Theorem 2 that the answer is again in negative.

Обовщенные контекстно-свободные грамматики

Обобщенные контекстно-свободные грамматики — это грамматики имеющие правила вида $A \to \alpha$, где A вспомогательный символ и α регулярное выражение над основыми и вспомогательными символами. В работе установлена классификация контекстно-свободных языков в зависимости от минимального числа вспомогательных символов обощенных контекстно-свободных грамматик, которые порождают данный контекстно-свободный язык. Доказана алгоритмическая неразрешимость основных проблем связанных с этой классификацией, как напр. проблема построить минимальную грамматику для данного языка.

References

[1] Conway, M. E., Design of a separable transition-diagrams compiler, *Comm. ACM.* v. 6, 1963, pp. 396—408.

[2] Ginsburg, S., *The mathematical theory of context-free languages*, McGraw-Hill, New York, 1966.

[3] Gruska, J., Some classifications of context-free languages, *Information and Control*, v. 14, 1969, pp. 152—173.

[4] Kalmár, L., An intuitive representation of context-free languages, COLING, *The proceedings of the International Conference on Computational Linguistics*, Sånga—Säby, 1969.

(Received April 18, 1972)

Eine verbandstheoretische Klassifikation der längentreuen Wortfunktionen

Von G. Wechsung

Feststehende Bezeichnungen: Nz = Menge der natürlichen Zahlen; $f \circ g(x) = g(f(x))$; X^* = Menge aller Wörter über dem Alphabet X einschließlich des leeren Wortes e; $|w|$ = Länge des Wortes w; $\mathfrak{P}(M) = \{U : U \subseteq M\}$ — die Potenzmenge von M.

1. Einleitung

Die vorliegende Arbeit ist dem Studium algebraischer Eigenschaften von Klassen längentreuer Wortfunktionen über festem Alphabet X gewidmet. Dabei sollen vorerst keine Berechenbarkeitsforderungen gestellt werden. Die Einschränkung auf Längentreue, die bei allen betrachteten Funktionen auch ohne ausdrückliche Erwähnung stets gelten soll, ermöglicht eine deutlichere Beschreibung der Abhängigkeit der Buchstaben der Bildwörter von den Buchstaben der Originalwörter. Die Unterscheidung verschiedener Arten eines solchen Abhängigkeitsverhaltens führt zur Einteilung der Wortfunktionen in Klassen von Funktionen gleichen Typs. Dies wird präzisiert durch die

Definition 1

1. Eine eindeutige Abbildung aus Nz^2 in $\mathfrak{P}(Nz)$ heißt Typ $=_{\text{Def}}$

a) $\forall n \forall m (m \leqq n \rightarrow T(n, m) \subseteq \{1, \ldots, n\})$,

b) $T(n, m)$ ist nicht definiert für $n < m$.

An Stelle von T(n,m) wollen wir künftig auch T_n(m) schreiben.

2. Ist T Typ, $w = x_1 \ldots x_n \in X^*$ und $T_n(m) = \{i_1, \ldots, i_s\}$ mit $i_1 < i_2 < \cdots < i_s$, so setzen wir

$$w_m^T =_{\text{Def}} x_{i_1} \ldots x_{i_s}.$$

Werden nicht mehrere T gleichzeitig betrachtet, so schreiben wir einfacher w_m.

3. Die Wortfunktion f heißt vom Typ T (wir schreiben dafür kurz $f \tau T$) $=_{\text{Def}}$ Für jedes $n \in Nz$ existieren eindeutige Abbildungen $f_{n1}, \ldots, f_{nn}$ von X^* in X, so daß gilt

$$\forall w \forall n (w \in X^* \wedge |w| = n \rightarrow f(w) = f_{n1}(w_1^T) \ldots f_{nn}(w_n^T)).$$

4. $W =_{\text{Def}}$ Menge aller (längentreuen) Wortfunktionen über X, $A =_{\text{Def}}$ Menge aller Typen, $F_T =_{\text{Def}} \{f : f \in W \wedge f \tau T\}$ für $T \in A$, $\mathfrak{K} =_{\text{Def}} \{F_T : T \in A\}$.

$T_n(m)$ interpretieren wir als die Menge derjenigen Stellen des Originalworts der Länge n, von denen der m-te Buchstabe des Bildworts abhängt.

Für $T_n(m)=\{1, \ldots, n\}$ entstehen genau die sequentiellen Funktionen. Die nichtretrospektiven Wortfunktionen f sind diejenigen, zu denen ein n und ein m existieren mit $m<n$ und $(T_f)_n(m) \cap \{m+1, \ldots, n\} \neq \emptyset$, wobei T_f in Satz 1b erklärt wird. Bei diesen Funktionen hängt nicht für jedes Wort der m-te Buchstabe des Bildworts nur von den ersten m Buchstaben des Originalworts ab. Ein einfaches Beispiel dafür ist die Funktion

$$f(x_1 \ldots x_n) = x_2 \ldots x_n x_1,$$

für die ein Typ T durch

$$T_n(m) = \begin{cases} \{m, m+1\} & \text{für} \quad m<n, \\ \{1, m\} & \text{für} \quad m=n, \end{cases}$$

gegeben ist. Dabei ist

$$f_{nm}(xy) = \begin{cases} y & \text{für} \quad m<n, \\ x & \text{für} \quad m=n, \quad x, y \in X. \end{cases}$$

Es zeigt sich, daß die Menge A der Typen in natürlicher Weise zu einem vollständigen atomaren Booleschen (und damit totaldistributiven) Verband gemacht werden kann, der isomorph zu einem Verband mit der Trägermenge $\mathfrak{K}$ und damit zu einem Teilbund des Potenzmengenverbandes von W ist. Dieser Teilbund ist insofern ausgezeichnet, als er aus allen und nur den Mengen besteht, die bezüglich eines Hüllenoperators abgeschlossen sind, der von der zur Relation τ gehörigen Galoisverbindung abgeleitet ist (Abschnitt 2). Der 3. Abschnitt führt zu der Feststellung, daß F_T dann und nur dann eine Halbgruppe bezüglich der Substitution bildet, wenn jedes T_n ein abgeschlossener Operator ist (vgl. Definition 4). Insbesondere gehören dazu diejenigen sogenannten topologischen Typen, bei denen jedes T_n ein topologischer Hüllenoperator ist.

2. Der Verband der Typen

Jede längentreue Wortfunktion f kann in der in Definition 1 angegebenen Weise dargestellt werden. Dazu setze man $T_n(m)=\{1, \ldots, n\}$ für jedes n und jedes $m \leqq n$ und $f_{nm}(x_1 \ldots x_n) = m$-ter Buchstabe von $f(x_1 \ldots x_n)$. W kann daher als F_T mit dem eben angegebenen Typ T angesehen werden.

Definition 2. Für $T, S \in A$ definieren wir $T \cap S$, $T \cup S$, $\overline{T}$ durch

$$(T \cap S)_n(m) = T_n(m) \cap S_n(m),$$

$$(T \cup S)_n(m) = T_n(m) \cup S_n(m),$$

$$(\overline{T})_n(m) = \{1, \ldots, n\} \setminus T_n(m).$$

Unmittelbar klar ist damit die

Folgerung. $[A, \cap, \cup, ^-]$ ist ein vollständiger atomarer Boolescher Verband. Die zugehörige Halbordnung ist durch

$$T \leqq S =_{\text{Def}} \forall n \forall m (m \leqq n \rightarrow T_n(m) \subseteq S_n(m))$$

gegeben. Die Atome von A sind diejenigen T^{ijk} mit

$$T_n^{ijk}(m) = \begin{cases} \{k\} & \text{für} \quad n = i \wedge j = m, \\ \emptyset & \text{sonst.} \end{cases}$$

Daß der Typ einer gegebenen Wortfunktion nicht eindeutig bestimmt ist, zeigt der triviale

Satz 1a. Ist $f \in F_T$ und ist $T \leqq S$, so ist auch $f \in F_S$.
Dieser Satz kann jedoch ergänzt werden durch

Satz 1b. Zu jeder Wortfunktion $f \in W$ gibt es einen eindeutig bestimmten kleinsten Typ T_f in A.

Beweis. T_f ist das Infimum aller Typen von f.
Satz 1a kann wesentlich verschärft werden zu dem

Satz 2. $\mathfrak{K}$ ist bezüglich der mengentheoretischen Inklusion $\subseteqq$ ein zu $[A, \leqq]$ isomorpher Verband.

Beweis

1. Die Zuordnung $T \to F_T$ ist eineindeutig. Es sei nämlich $T \neq S$. Dann gibt es ein n und ein m mit $T_n(m) \neq S_n(m)$. O. B. d. A. existiert daher ein j mit $j \in S_n(m)$, aber $j \notin T_n(m)$. Damit gibt es eine Funktion $f \in F_S$ und $x_1, \ldots, x_n, y_1, \ldots, y_n \in X$ mit der Eigenschaft $f(x_1 \ldots x_j \ldots x_n) = y_1 \ldots y_m \ldots y_n$, und y_m hängt echt von x_j ab, d.h. y_m ändert sich, wenn sich x_j bei sonst gleichbleibenden x_i $(i \neq j)$ ändert. Keine Funktion aus F_T dagegen darf wegen $j \notin T_n(m)$ dieses Verhalten aufweisen. Also ist $F_T \neq F_S$.
2. Die Aussage $T \leqq S \to F_T \subseteqq F_S$ ist genau Satz 1a.
3. Wir zeigen noch: $F_T \subseteqq F_S \to T \leqq S$. Es sei $F_T \subseteqq F_S$. Für jedes n und $m \leqq n$ gibt es ein $f \in F_T$, so daß, wenn $f(x_1 \ldots x_n) = y_1 \ldots y_n$ gilt, y_m von allen x_i mit $i \in T_n(m)$ wirklich abhängt. Da f nach Voraussetzung zu F_S gehört, muß $S_n(m) \supseteqq T_n(m)$ sein. Also gilt $T \geqq S$, womit die behauptete Isomorphie bewiesen ist (vgl. [2]).

Bemerkungen

1. Den Verband $[A, \cap, \cup, ^{-}]$ nennen wir den Typenverband. Wegen Satz 2 wollen wir diese Bezeichnung auf $[\mathfrak{K}, \subseteqq]$ übertragen und die Klassen F_T gelegentlich als Typen ansprechen.
2. Das kleinste Element von $\mathfrak{K}$ ist der Typ der Konstanten, das größte ist die Klasse W (siehe den Beginn dieses Abschnitts).
3. Man überzeugt sich leicht davon, daß die Infimumbildung in $\mathfrak{K}$ die gewöhnliche Durchschnittsbildung ist. Damit ergibt sich für $\mathfrak{M} \subseteqq \mathfrak{K}$ ([2])

$$\sup \mathfrak{M} = \cap \{F : F \in \mathfrak{K} \wedge \forall C(C \in \mathfrak{M} \to C \subseteqq F)\},$$

insbesondere

$$\sup (F_T, F_S) = \cap \{F_{T'} : F_{T'} \supseteqq F_T \cup F_S\}.$$

$\mathfrak{K}$ ist ein Teilbund des Potenzmengenverbandes von W, dessen Bedeutung noch deutlicher wird, wenn man die zu der Relation τ gehörige Galoisverbindung und die zugehörigen Hüllenoperatoren betrachtet.

Definition 3. Für $B \subseteqq A$ und $U \subseteqq W$ setzen wir

$$\gamma(B) =_{\text{Def}} \{f : f \in W \wedge \forall T(T \in B \rightarrow f \tau T)\},$$
$$\delta(U) =_{\text{Def}} \{T : T \in A \wedge \forall f(f \in U \rightarrow f \tau T)\},$$
$$\Gamma =_{\text{Def}} \delta \circ \gamma, \quad \Delta =_{\text{Def}} \gamma \circ \delta.$$

$[\mathfrak{P}(A), \mathfrak{P}(W), \gamma, \delta]$ ist eine Galoisverbindung, Γ und Δ sind Hüllenoperatoren, die bezüglich Γ und Δ abgeschlossenen Mengen bilden je einen vollständigen Verband, und beide Verbände sind zueinander dual isomorph (vgl. [1], [2]). Wir wollen diese Verbände genauer beschreiben. Dazu stellen wir zunächst fest

$$\gamma(B) = F_{\inf B}, \delta(U) = \{T : T \geqq \bigcup_{f \in U} T_f\},$$

wobei T_f wie in Satz 1b verstanden wird. Damit ergibt sich

$$\Gamma(U) = F_{\bigcup_{f \in U} T_f}, \quad \Delta(B) = \{T : T \geqq \inf B\}.$$

Weil $\bigcup_{f \in U} T_f$ alle Elemente von A annehmen kann, sind die Γ-abgeschlossenen Mengen genau die oben eingeführten Typen F_T, und der Verband $[\mathfrak{K}, \subseteqq]$ erscheint somit als Verband der Γ-abgeschlossenen Mengen für den von der Galoisverbindung erzeugten Hüllenoperator Γ. Ordnet man jeder Δ-abgeschlossenen Menge ihr Infimum zu, so erhält man einen dualen Isomorphismus vom Verband der Δ-abgeschlossenen Teilmengen von A auf den Verband $[A, \leqq]$.

3. Halbgruppen

Es erhebt sich die Frage, wann mit f und g stets auch $f \circ g$ vom Typ T ist. Um die Antwort besser formulieren zu können, fassen wir T_n als Operator auf $\{1, \ldots, n\}$ auf

Definition 4. $T_n(\emptyset) = \emptyset$, $T_n(\{i_1, \ldots, i_k\}) = T_n(i_1) \cup \ldots \cup T_n(i_k)$.
T_n heißt einbettend $=_{\text{Def}}$, $\forall M(M \subseteqq \{1, \ldots, n\} \rightarrow M \subseteqq T_n(M))$,
T_n heißt abgeschlossen $=_{\text{Def}}$, $\forall M(M \subseteqq \{1, \ldots, n\} \rightarrow T_n(T_n(M)) \subseteqq T_n(M))$.
Dann kann der folgende Satz ausgesprochen werden.

Satz 3. Die Klasse F_T ist genau dann eine Halbgruppe bezüglich $\circ$, wenn für jedes n der Operator T_n abgeschlossen ist.

Beweis

1. Es seien $f, g \in F_T$, und T erfülle die Bedingung des Satzes. Dann ist

$$g(f(x_1 \ldots x_n)) = g(f_{n1}(w_1) \ldots f_{nn}(w_n)) = y_1 \ldots y_n.$$

Hierbei hängt y_m höchstens von den $f_{nj}(w_j)$ mit $j \in T_n(m)$ ab. Diese $f_{nj}(w_j)$ hängen höchstens von den x_k mit $k \in T_n(j)$ ab. Insgesamt hängt y_m daher höchstens von den x_k mit

$$k \in \bigcup_{j \in T_n(m)} T_n(j) = T_n(T_n(m)) = T_n(m)$$

ab. Damit ist $f \circ g \in F_T$.

2. Es sei $[F_T, \circ]$ Halbgruppe und $T_n(T_n(m)) \not\subseteqq T_n(m)$. Dann gibt es ein $j \in T_n(T_n(m))$ mit $j \notin T_n(m)$. Dazu gibt es ein $i \in T_n(m)$ mit $j \in T_n(i)$. Werden $T_n(i)$

(bzw. $T_n(m)$) der Größe nach aufgezählt, so seien j (bzw. i) das s-te (bzw. r-te) Element. Wir wählen jetzt zwei Funktionen $f, g \in F_T$ mit den Eigenschaften

$$f_{ni}(w_i) = s\text{-ter Buchstabe von } w_i = x_j,$$

$$y_m = g_{nm}((f(x_1 \dots x_n))_n) = r\text{-ter Buchstabe von } (f(x_1 \dots x_n))_m = f_{ni}(w_i) = x_j.$$

(Die Bezeichnungen stimmen dabei mit denen von Punkt 1 dieses Beweises überein.) Also hängt y_m echt von x_j ab. Das ist aber nicht möglich, weil nach Voraussetzung $f \circ g \in F_T$ ist und daher $j \in T_n(m)$ sein müßte, was der obigen Annahme widerspricht. Infolgedessen ist für jedes n und $m \leqq n$ $T_n(T_n(m)) \subseteq T_n(m)$, womit die zu beweisende Eigenschaft von T_n aus Definition 4 folgt.

Definition 5. T (bzw. F_T) heißt abgeschlossen (topologisch) $=_{\text{Def}}$ Für jedes n ist der in Definition 4 erklärte Operator T_n ein abgeschlossener Operator (topologischer Hüllenoperator).

Satz 4. Die Menge $\mathfrak{H}$ aller Halbgruppen F_T bildet einen vollständigen Teilbund (keinen Teilverband) von $[\mathfrak{K}, \subseteq]$. Die Menge H aller abgeschlossenen Typen bildet einen vollständigen Teilbund (keinen Teilverband) von $[A, \leqq]$.

Beweis. Ist $\mathfrak{M}$ eine Teilmenge von $\mathfrak{H}$, so gehört wegen der Vollständigkeit von $[\mathfrak{K}, \subseteq] F =_{\text{Def}} \bigcap \mathfrak{M}$ zu $\mathfrak{K}$. Da F als Durchschnitt von Halbgruppen auch eine Halbgruppe ist, gehört F zu $\mathfrak{H}$. Daraus ergibt sich, daß die abgeschlossenen Klassen einen vollständigen Teilbund von $\mathfrak{K}$ bilden. Es handelt sich dabei nicht um einen Teilverband von $\mathfrak{K}$, weil mit zwei abgeschlossenen Typen ihre Vereinigung nicht notwendig abgeschlossen ist, was wir durch folgendes Beispiel zeigen. Wir setzen

$$T_3(1) = \{1\}, \quad T_3(2) = \{2, 3\}, \quad T_3(3) = \{2, 3\},$$

$$S_3(1) = \{1\}, \quad S_3(2) = \{2\}, \quad S_3(3) = \{1, 3\},$$

und wählen T und S ansonsten so, daß auch für $n \neq 3$ die Operatoren T_n und S_n abgeschlossen sind. $T \cup S$ ist nicht abgeschlossen, weil

$$(T \cup S)_3((T \cup S)_3(2)) = \{1, 2, 3\} \nsubseteq (T \cup S)_3(2) = \{2, 3\}.$$

Bemerkung. Die beiden Aussagen von Satz 4 sind wegen Satz 2 völlig gleichwertig. Dieser Satz rechtfertigt auch die zur Vereinfachung des Beweises angewendete Methode, einen Teil der Behauptung (vollständiger Teilbund zu sein) in $\mathfrak{K}$ zu zeigen und den Rest (kein Teilverband zu sein) in A durchzuführen. Ähnlich verfahren wir beim Beweis des nächsten Satzes.

Satz 5. Die Menge $\mathfrak{T}$ der topologischen Typen F_T bildet einen vollständigen Teilverband von $\mathfrak{H}$. Die Menge $\text{Top} \subseteq A$ aller topologischen Typen bildet einen vollständigen Teilverband von H.

Beweis

1. Für beliebige $M \subseteq \text{Top}$ ist $\bigcap M \in \text{Top}$. Denn nach Satz 4 ist $\bigcap M$ abgeschlossen, und daß $\bigcap M$ wieder einbettend ist, ist offenkundig. Daher ist $\bigcap M \in \text{Top}$ und Top ein vollständiger Verband.

2. Wir zeigen noch. Ist $M \subseteqq \text{Top}$, so ist $\inf_H M = \inf_{\text{Top}} M$ und $\sup_H M = \sup_{\text{Top}} M$. Trivialerweise ist $\inf_H M = \inf_{\text{Top}} M = \bigcap M$. Dagegen ist

$$\sup_H M = \bigcap \{T: T \in H \wedge \forall S (S \in M \rightarrow S \leqq T)\},$$

$$\sup_{\text{Top}} M = \bigcap \{T: T \in \text{Top} \wedge \forall S (S \in M \rightarrow S \leqq T)\}.$$

Wegen $\text{Top} \subseteqq H$ ist immer $\sup_H M \leqq \sup_{\text{Top}} M$. Ist $M \subseteqq \text{Top}$, so ist darüber hinaus jedes Element T, das bei der Durchschnittsbildung von $\sup_H M$ vorkommt, abgeschlossen (wegen $T \in H$) und notwendig einbettend (weil ein nicht einbettendes T kein einbettendes S majorisieren kann), also topologisch und kommt daher auch bei der Durchschnittsbildung von $\sup_{\text{Top}} M$ vor. Daraus folgt $\sup_{\text{Top}} M \leqq \sup_H M$, also $\sup_{\text{Top}} M = \sup_H M$ für jedes $M \subseteqq \text{Top}$.

Folgerung. Zu jeder Klasse F_T gibt es eine kleinste sie umfassende abgeschlossene Klasse $\theta(F_T)$ und eine kleinste sie umfassende topologische Klasse $\theta^*(F_T)$. Dasselbe gilt für die Typen aus A.

Beweis. $\theta(F_T)$ ist der Durchschnitt aller F_T umfassenden Halbgruppen F_S, $\theta^*(F_T)$ ist der Durchschnitt aller F_T umfassenden topologischen Klassen F_S.

Man könnte vermuten, daß $\theta(F_T)$ mit der von F_T erzeugten Halbgruppe F_T^* zusammenfällt. Wir zeigen, daß dies i.a. nicht zutrifft. Wir benützen die Bezeichnungen θ und θ^* auch für die Typen $T \in A$. Dann haben wir

Lemma 1. Für jedes $T \in A$ gilt $\theta(F_T) = F_{\theta(T)}$. Die gleiche Beziehung gilt auch für θ^*.

Beweis

1. $\theta(T)$ ist abgeschlossener Typ oberhalb von T. Damit ist $F_{\theta(T)}$ Halbgruppe oberhalb von F_T und demnach $\theta(F_T) \subseteqq F_{\theta(T)}$.

2. Wäre $F_S =_{\text{Def}} \theta(F_T) \subset F_{\theta(T)}$, so wäre $T \leqq S < \theta(T)$. Da S abgeschlossen ist, wäre $\theta(T)$ nicht der kleinste abgeschlossene Typ oberhalb von T.

Ab jetzt beschränken wir uns auf einbettende T. Für diese ist offenbar $\theta(T) = \theta^*(T)$. Für einbettende T gilt stets

$$T_n(m) \subseteqq T_n^2(m) \subseteqq T_n^3(m) \subseteqq \ldots \subseteqq \{1, \ldots, n\}.$$

Demnach gibt es für jedes n und $m \leqq n$ ein k_{nm} mit

$$T_n^{k_{nm}}(m) = T_n^{k_{nm}+1}(m).$$

Mit diesen Bezeichnungen gilt

Lemma 2. Für einbettende Typen T ist $(\theta(T))_n(m) = T_n^{k_{nm}}(m)$.

Beweis. R sei abgeschlossener Typ mit $T \leqq R$. Das bedeutet nach Definition 2

$$T_n(m) \subseteqq R_n(m)$$

für alle n und $m \leqq n$. Hieraus folgt unter Beachtung von Definition 4

$$T_n^2(m) \subseteqq T_n(R_n(m)) \subseteqq R_n^2(m) \subseteqq R_n(m)$$

und durch vollständige Induktion

$$T_n^k(m) \subseteqq R_n(m)$$

für jedes k. Jeder abgeschlossene Typ oberhalb von T umfaßt daher den in der Aussage des Lemmas angegebenen Typ. Da dieser offenbar selbst abgeschlossen ist, ist das Lemma bewiesen.

Definition 6. Der einbettende Typ T heißt von endlicher Ordnung $=_{\text{Def}} \exists k \forall n (T_n^k = T_n^{k+1})$.

Satz 6. Ist der einbettende Typ T nicht von endlicher Ordnung, so ist $F_T^* \neq \theta(F_T)$.

Beweis. Für jedes $f \in F_T^*$ ist $T_f < \theta(T)$. Ist nämlich $f = f_1 \circ \ldots \circ f_r$, so ist jedenfalls $(T_f)_n(m) \subseteqq T_n^r(m)$ für alle n und $m \leqq n$ (vgl. Beweis von Satz 3). Nach Voraussetzung existiert eine solche unendliche Folge $k_{n_1} < k_{n_2} < \ldots$, daß $T_{n_i}^{k_{n_i}} \neq (\theta(T))_{n_i}$ ist. Man wähle nun i so groß, daß $r \leqq k_{n_i}$ ist. Dann gilt für passendes m

$$(T_f)_{n_i}(m) \subseteqq T_{n_i}^r(m) \subseteqq T_{n_i}^{k_{n_i}}(m) \neq (\theta(T))_{n_i}(m).$$

Wäre $F_T^* = \theta(F_T) = F_{\theta(T)}$, so müßte jedes f mit $T_f = \theta(T)$ in F_T^* liegen, was dem eben bewiesenen widerspricht.

Folgerung. Die Klassen F_T mit abgeschlossenem T sind nicht die einzigen Unterhalbgruppen von W.

Beweis. Wir brauchen nur zu zeigen, daß es Typen gibt, die keine endliche Ordnung haben. Als Beispiel wählen wir

$$T_n(m) = \begin{cases} \{m-1, m\} & \text{für } 1 < m \leqq n, \\ \{1\} & \text{für } m = 1. \end{cases}$$

F_T ist die Klasse der 1-stabilen sequentiellen Wortfunktionen, während $\theta(F_T)$ die Klasse der sequentiellen Funktionen ist. Da T offenbar nicht von endlicher Ordnung ist, gehört F_T^* nicht zu $\mathfrak{K}$, weil sonst $F_T^* = \theta(F_T)$ gelten müßte, was nach Satz 6 unmöglich ist. (Diesem Sachverhalt entspricht die bekannte Tatsache, daß nicht jede sequentielle Funktion als Produkt 1-stabiler sequentieller Funktionen darstellbar ist.)

Die Umkehrung von Satz 6 ist i.a. nicht richtig.

Описание словарьных функции, которые сохраняют длину, при помощи понятий из теории структур

Пусть T однозначное отображение из множества всех пар натуральных чисел в множество всех конечных множеств натуральных чисел со свойствами

(1) $\forall n \forall m (m \leqq n \rightarrow T(n, m) \subseteqq \{1, \ldots, n\})$,

(2) $T(n, m)$ не определенно для $m > n$.

Словарьная сохраняющая длину функция f называется функцией типа T, когда выполняется следующее условие: Когда $f(x_1 \ldots x_n) = y_1 \ldots y_n$, y_m зависит только от букв множества

$$\{x_i : i \in T(n,m)\}.$$

Множество всех типов естественным образом образует польную, атомарную булевую структуру, которая изоморфна структуре всех F_T (F_T является множеством всех функций типа T). Рассматрываются замкнутые и топологические типи и исследуются их свойства. Например, F_T тогда и только тогда является полугруппой, когда T замкнутый тип.

SEKTION MATHEMATIK
FRIEDRICH SCHILLER UNIVERSITÄT
69 JENA, DDR
UNIVERSITÄTS HOCHHAUS

Literatur

[1] COHN, P. M., *Universal Algebra*, Harper & Row Publishers, New York, Evanston and London, 1965.
[2] SZÁSZ, G., *Einführung in die Verbandstheorie*, Teubner Verlagsgesellschaft, Leipzig, 1962.

(Eingegangen am 2. Februar 1972)

The notion and some applications of generalized initial segment

By Á. Makay

Introduction

In computational practice we are often confronted with situations where we have to handle strings composed of elements that can be put into two disjoint classes according to the way they influence the outcome of the operations to be performed on these strings. E.g. this happens in the case of translating higher-level programming languages. A programme written in such a language is a combination of terminators (operation symbols, parentheses, etc.) and quantities (identifiers and numbers, these are composed of letters and numerals). During translation the terminators control the compiler, while quantities play only a passive role: the translation process is not influenced e.g. by replacing each occurrence of an identifier with occurrences of another, while a similar statement for terminators is obviously false. Other types of formula-handling algorithms serve also as good examples in which such mixed-sequence situations arise.

In the present article we are going to define the notion of generalized initial segment for strings composed of elements of two disjoint sets. This notion simultaneously extends those of initial segment and of subsequence. We also present an algorithm which can decide whether or not any string is a generalized initial segment of another. The applicability in practice of the algorithm is illustrated with two examples. The second one of these deals with the Universal Decimal Classification (UDC), which is generally used in library practice. The ideas outlined in this context have been used in an information-retrieval system here working on the basis of UDC.

Definitions

Let V be a finite set and Σ a subset of V. The elements of V are usually called signs, the elements of Σ letters, and the elements of $V-\Sigma$ terminators.

Let V^* designate the set of all finite sequences which can be formed of elements of V (i.e. the free semigroup over the set V). We mean by the length $|x|$ of a sequence $x \in V^*$ the number of signs (including repeats) in x, while $\|x\|$ denotes the number of terminators (again including repeats) in x. Obviously $0 \leqq \|x\| \leqq |x|$ and $|x|$ is 0 if and only if $x=\varepsilon$, where ε is the empty sequence.

Assume $x, y \in V^*$. The notion of the generalized initial segment is defined by recursion on $\|x\|$. In case $\|x\|=0$, x is called generalized initial segment of y if

$$x=j_1 j_2 \ldots j_n, \quad y=j_1 j_2 \ldots j_n z,$$

where $n \geqq 0$, $z \in V^*$, $j_1, j_2, \ldots, j_n \in \Sigma$ (i.e. x is an initial segment of y and contains no terminators). Assume now $\|x\| = n+1$. We say that x is the generalized initial segment of y if there are sequences $x_1, x_2, y_1, y_2 \in V^*$ such that

$$x=x_1 x_2, \quad y=y_1 x_2 y_2, \quad x_2=t j_1 j_2 \ldots j_m,$$

where x_1 is the generalized initial segment of y_1, and, moreover, $m \geqq 0$, $t \in V-\Sigma$, and $j_1, j_2, \ldots, j_m \in \Sigma$.

In other words, x is a generalized initial segment of y if and only if the following holds: y contains any such sequence as is either an initial segment of x containing no terminators or is a subsequence of x such that its first element and only this is a terminator, moreover, choosing any set of such nonoverlapping sequences, these occur in y in the same order as in x.

The generalized initial segments of a sequence y consisting only of elements of Σ are simply the initial segments of y in the usual sense. Thus if $\Sigma = V$, the generalized initial segments are also initial segments. However if Σ is the empty set, then any subsequence of y is a generalized initial segment as well.

We will define an algorithm which decides for any $x, y \in V^*$ whether x is a generalized initial segment of y. By selecting Σ in a suitable way, the same algorithm can decide whether x is an initial segment or a subsequence of y.

We define the algorithm as an ALGOL-60 [1] Boolean function. The signs of the formal parameter-strings x and y are denoted by $x_1, x_2, \ldots, x_{|x|}$ and by $y_1, y_2, \ldots \ldots, y_{|y|}$, respectively. For lucidity's sake we use the nonstandard notations $|x|$, $\|x\|$, $|y|$, $\notin$ as well. The outcome of the procedure is the value *true* if x is a generalized initial segment of y, otherwise it is *false*.

```
boolean procedure GEN IN SEG (x, y); string x, y;
  begin integer array F, G [1:‖x‖]; integer i, j, k;
    i:= j:= k:= 1;
    C: if j>|x| then GEN IN SEG:= true else
       if i>|y| then GEN IN SEG:= false else
        if x_j = y_i then
   begin if x_j ∉ Σ then begin F[k]:= i; G[k]:= j; k:= k+1 end;
     i:= i+1; j:= j+1; goto C
   end else if x_j ∉ Σ then begin i:= i+1; goto C end else
   begin k:= k-1; if k≠0 then
   begin i:= F[k]+1; j:= G[k]; goto C end else GEN IN SEG:= false
   end
  end GEN IN SEG;
```

Applications

The notion of the generalized initial segment and the algorithm defined above is quite widespread. Here we disregard the special cases of the initial segment and subsequence, and give two applications which are useful in information retrieval systems.

Example 1. Let us suppose that in our information system abstracts written in a natural language serve as descriptions of the content of documents. The request for retrieval is written in words or expressions consisting of several words. In the request the single words are given as root-words and in the abstracts in their inflected forms. Disregarding rootchanges, when comparing a request consisting of only one word and an abstract our task is to ascertain of the word in the request whether it agrees with the begginning of any of the words of the abstract. If the request is an expression we have to ascertain of several words whether all these words agree with the beginnings of some words in the abstract, that have the order same as given in the request. To sum up, if hyphens and spaces between words are regarded as terminators, it is to be decided whether the word or expression in the request is a generalized initial segment of any of the sentences of the abstract.

Example 2. The most widespread system of content classification of the library practice is the Universal Decimal Classification (UDC) [2]. The tremendous amount of time, money and spirit devoted to the system makes it imperative that these results and forms should be used by up-to-date automatic information systems. Below a formal definition is given for the following statement: the notion denoted by UDC number y belongs to the category denoted by UDC number x.

When setting up the UDC system mainly manual methods had been in mind, their application in automatic systems was not considered. Therefore certain transformations for computer information systems are required. This can be done by decomposing UDC-numbers into parts (general subject, facets separately) [3]. Another way (which we follow here) is the mechanical transformation of the UDC-numbers. In particular we omit redundant signs („right-hand",)) and, moreover, instead of sings consisting of several characters we use only a single character (−0,.0,.00,(0,(=). The schematic description [1] gives all these transformations. However, the categories of the general connective (+), the inclusive connective (/) and the relative connective(:) are left undefined: instead of them in the information system various UDC-numbers can be used or there are several other ways of excluding them [2].

⟨UDC-number⟩ ::= ⟨general subject⟩ ⟨subordinate facets⟩ ⟨facets⟩

⟨general subject⟩ ::= ⟨empty⟩|⟨decimal number⟩|⟨synthetic connectiv⟩

⟨synthetic connectiv⟩ ::= ⟨general subject⟩′⟨decimal number⟩

⟨subordinate facets⟩ ::= ⟨empty⟩|⟨subordinate facet⟩|⟨subordinate facets⟩ ⟨subordinate facet⟩

⟨subordinate facet⟩ ::= ⟨special auxiliary⟩|⟨point of view⟩

⟨special auxiliary⟩ ::= −⟨decimal number⟩|.0⟨non-0 decimal number⟩

⟨point of view⟩ ::=.00⟨decimal number⟩

⟨facets⟩ ::= ⟨empty⟩|⟨facet⟩|⟨facets⟩ ⟨facet⟩

⟨facet⟩ ::= ⟨language⟩|⟨form of work⟩|⟨place⟩|⟨race⟩|⟨time⟩

⟨language⟩ ::== ⟨non-0 decimal number⟩

⟨form of work⟩ ::= (0 ⟨decimal number⟩

⟨place⟩ ::=(⟨non-0 decimal number⟩

⟨race⟩ ::= (=⟨decimal number⟩

⟨time⟩ ::="⟨decimal number⟩

⟨decimal number⟩ ::=⟨digit⟩|⟨decimal number⟩ ⟨digit⟩

⟨non-0 decimal number⟩ ::= ⟨non-0 digit⟩|⟨non-0 decimal number⟩ ⟨digit⟩

⟨digit⟩ ::= 0|⟨non-0 digit⟩

⟨non-0 digit⟩ ::= 1|2|3|4|5|6|7|8|9

⟨empty⟩ ::=

If the UDC-numbers are decomposed in parts in an information system (general subject, facets), the requests are also to be built of these units. Therefore, if we want to formulate the simultaneous existence of two elements, we are in need of the logical operation of "conjunction", which is a notion unknown in library practice.

The form of formulating requests is brought closer to the UDC system when complete UDC-numbers are used as units. In view of the above syntactic rules it is easy to see that if the numeral digits are regarded as letters and all the other signs as terminators, the notion denoted by an UDC number y belongs to the category denoted by an UDC number x if and only if x is a generalized initial segment of y.

Понятие обобщенного начального сегмента и некоторые его применения

В вычислительной технике довольно частый случай, когда занимаемся такими последовательностями, которые с точки зрения обработки, построены из элементов двух друг от друга хорошо различиных множеств. Такое же положение тогда, когда, например, изготавливаем транслятор языка для программировании более высокой степени, в этом случае числа и буквы служат для обозначения «величин», а последовательность скобок и знаков операций содержит структуру програмы.

Статья, с помощью понятия «обобщенного начального сегмента», одновременно и обобщает понятие начального сегмента и частичной последовательности таких «смещенных» последовательностей, а также публикуется алгоритм решение того, что последовательность является ли обобщенным начальным сегментом другой последовательности.

Вторая из двух сообщенных возможностей применения обращает внимание на свойство алгоритма, который может быть применен в системах обратного отыскания информации в системе Универсальной Десятичной Классификации (УДК).

RESEARCH GROUP ON MATHEMATICAL LOGIC
AND THEORY OF AUTOMATA OF THE HUNGARIAN
ACADEMY OF SCIENCES,
SZEGED, HUNGARY

References

[1] NAUR, P., et al., Revised Report on Algorithmic Language ALGOL 60, *Comm.* ACM., v. 6, 1963, pp. 1—17.

[2] *Universal Decimal Classification*, Auxiliary tables (in Hungarian), Közgazdasági és Jogi Könyvkiadó, Budapest, 1969.

[3] FREEMAN, R. R. & P. ATHERTON, *File organization and search strategy using the universal decimal classification in mechanized reference retrieval systems*, American Institute of Physics, Report No. AIP/UDC-5, 1967.

(Received March 11, 1972)

Combined application of drawing and steepest descent in generating initial estimates for subsequent optimization

By K. VARGA and P. FEJES

Introduction

The determination of the unrestricted local minimum of a function of several variables by the "direct search" methods consists in the sequential examination of function values belonging to randomly selected vectors as the independent variables. A comparison of each trial solution with the "best" one up to that time helps locate the approximate value of the minimum. Even the more sophisticated gradient methods [1, 2, 3] cannot dispense with a similar method for generating initial estimates for subsequent iterative optimization. In certain cases (e.g. in the Gauss—Newton method) a favourably chosen initial value is a prerequisite of convergence, whereas in others it allows the gradient method to be used for finding the *absolute* minimum of a function in a bounded region. Finally, the number of iterations can be reduced considerably if a good initial estimate of the minimum is available.

We have found that the effectiveness of this method can be substantially improved if the initial estimate is chosen not as the vector corresponding to the lowest function value but, starting with the drawn vector, a step is performed according to the principles of the steepest descent and the function values are compared *in these modified points.*

Both methods are described from the viewpoint of probability, and some problems of application encountered in practice are discussed.

Statistical basis of the direct search method

For the sake of simplicity let us consider the case of minimization. Let the function to be optimized be the scalar-vector function

$$f(p_1, p_2, \ldots, p_n) \equiv f(p)$$

assumed to be continuous and single-valued in the bounded n-dimensional rectangle $T \subset E_n$. Let us suppose further that function $f(p)$ exhibits a minimum in an *inner* point $p_{\min}$ of the region.

In order to find the point $p_{\min} \in T$ that makes the function to attain its minimum, let us select some optimizing method which unambiguously determines a T_{conv} (eventually multiply interrelated) parameter interval characterized in such a way that any point of this interval could serve as initial value for the method to yield p_{min}, in short: *the method is convergent.* By using the usual set theory notation

$$T_{conv} \equiv \{p;\text{ from which the gradient method converges to } p_{min}\}.$$

If the initial value is selected from the $(T-T_{conv})$ region, the method "gets stuck" in local minima or some other special points, i.e. it is *divergent.* In other words, this region contains those points of T where such unfavourable properties of an "ill-conditioned" function may be experienced as e.g. saddle points or, if the Hessian matrix is used, points where the Hessian is singular or not positive definite.

If the problem is solved by successive iteration, a preliminary knowledge of some initial value $p_1^* \in T$ is an important condition. Should no other information be available, drawing is to be used for selecting an appropriate p_1^*. Let us suppose that a process is available for drawing n numbers at random with equal probability to represent the components of p_1^*. We shall now examine the probability of drawing a "good" point that lies within T_{conv}.

Let $\mu(X)$ be the Lebesgue measure of some region $X \subseteq T \subset E_n$; then the probability of finding a point within T_{conv} is given by

$$P(p_1^* \in T_{conv}) = \mu(T_{conv})/\mu(T) \equiv \varrho_{conv}.$$

The probability of obtaining a good initial value can be increased by making several independent drawings one after another. The probability that at least one vector from among the set: $p_1^*, \ldots, p_m^*$ chosen independently lies within T_{conv} can be written according to the binomial distribution as

$$P_1(m) \equiv P(\text{for at least one } i\text{: } p_i^* \in T_{conv};\quad i=1, 2, \ldots, m) =$$

$$= 1-\binom{m}{0}\varrho_{conv}^0(1-\varrho_{conv})^m = 1-(1-\varrho_{conv})^m. \tag{1}$$

The number of drawings to be made if we want to find at least one point in T_{conv} with a predetermined probability P is

$$m_1'(P) = \log(1-P)/\log(1-\varrho_{conv}) \tag{2}$$

whence the number of drawings is given by

$$m_1(P) = \text{entier}\,[m_1'(P)]+1$$

$m_1'(P)$ being not an integer number.

Now, provided that ϱ_{conv} is known, in case of some certainty we can perform in principle as many drawings as are necessary for determining at least one suitable initial value. The only problem is: which one of the $m_1(P)$ vectors belongs to T_{conv}?

The generation of initial parameters by drawing for minimization is an evident yet not widely applied possibility just because it is difficult to give a reliable answer to the above question. In practice there is no other possibility of selection from the vectors drawn than to compare the pertaining function values. This method, however, is not reliable because the vector resulting in the lowest function value often does not belong to T_{conv}. Therefore a comparison of this kind, as a method of distinction, might be misleading.

In order to evalute this error characteristic of the direct search methods, let us consider now the probability that the vector with the lowest function value in a set of m vectors is an element of T_{conv}.

Let A_k be the event that k of the selected m points are elements of T_{conv} and introduce the following notation

$$P_a(m, k) \equiv P(\min_{\substack{i \\ p_i^* \in T_{\text{conv}}}} f(p_i^*) < \min_{\substack{j \\ p_j^* \notin T_{\text{conv}}}} f(p_j^*) | A_k).$$

Further on, let p^* be the vector for which

$$f(p^*) \equiv \min_{i=1,2,\ldots,m} f(p_i^*).$$

Now, considering that upon drawing, the probability of finding k points in T_{conv} is

$$P_b(m, k) \equiv P(A_k) \equiv \binom{m}{k} \varrho_{\text{conv}}^k (1-\varrho_{\text{conv}})^{m-k}$$

the probability we are looking for can be expressed as

$$P^*(m) \equiv P(p^* \in T_{\text{conv}}) = \sum_{k=1}^{m} P_a(m, k) \cdot P_b(m, k) =$$

$$= \sum_{k=1}^{m} P(\min_{i=1,2,\ldots,k} f(p_i^*) < \min_{j=k+1,\ldots,m} f(p_j^*)) \binom{m}{k} \varrho_{\text{conv}}^k (1-\varrho_{\text{conv}})^{m-k}. \tag{3}$$

This equation is, however, unsuitable for practical calculations. But even so, it reflects the uncertainty involved in the comparison of function values. A simple consequence of the equation is for example the inequality

$$P^*(m) \leqq P_1(m), \tag{4}$$

which follows from

$$P^*(m) \leqq \sum_{k=1}^{m} \binom{m}{k} \varrho_{\text{conv}}^k (1-\varrho_{\text{conv}})^{m-k} = 1-(1-\varrho_{\text{conv}})^m = P_1(m).$$

This inequality implies that a comparison of the function values provides absolute certainty for the selection of a point in the convergence region (provided that such a point is contained in the set) only under ideal conditions (i.e. if the function to be optimized has appropriate characteristics), whereas in other cases the comparative technique may impair the efficiency of the search by drawing. This means that more than $m_1(P)$ drawings should be carried out in order to be able to single out a point which is an element of T_{conv} with probability P. Although no certain distinction is possible between the points belonging to T_{conv} and others (e.g. those lying in the vicinity of a local minimum), generally there always exists a region $T_{\min} \subset T$, $\mu(T_{\min}) \neq 0$, where the selection on the basis of function values leads to correct results. In other words, there exists a region where no mistake arises, and this is nothing else but the largest neighbourhood around $p_{\min}$ for all the points of which $f(p') < f(p'')$ is true, unless $p' \in T_{\min}$ and $p'' \in (T-T_{\min})$. The exact definition of $T_{\min}$ is

$$T_{\min} \equiv \{p; f(p) < \min_{p \in (T-T_{\text{conv}})} f(p)\}. \tag{5}$$

It follows from the definition that $T_{\min} \subseteq T_{\text{conv}}$.

The probability of finding a point in T_{min} among m drawn vectors by comparing the function values can be calculated by the binomial distribution; for the selection does not lead to error in this region

$$P_2(m) \equiv P(p^* \in T_{min}) = 1-(1-\varrho_{min})^m, \tag{6}$$

using for ϱ_{min} the simple relation

$$\varrho_{min} \equiv \mu(T_{min})/\mu(T).$$

Finally, considering Eq. (5), it can be proved that

$$P_2(m) \leqq P^*(m). \tag{7}$$

The probabilities $P_1(m)$ and $P_2(m)$ are easy to calculate for any value of m and through definitions (4) and (7), they determine the lower and upper bounds of the probability function $P^*(m)$

$$1-(1-\varrho_{min})^m \leqq P^*(m) \leqq 1-(1-\varrho_{conv})^m. \tag{8}$$

The inverse $m(P)$ of the function $P^*(m)$ shows how many drawings are to be carried out to get one point lying in the convergence region in case of a predetermined certainty P. Although the inverse function cannot be calculated, an estimate of $m(P)$ can be made using Eq. (8)

$$m_1(P) \leqq m(P) \leqq m_2(P),$$

where

$$m_2(P) = \text{entier}\,[m_2'(P)]+1$$

and

$$m_2'(P) = \log(1-P)/\log(1-\varrho_{min}), \tag{9}$$

respectively which follows from (6) by analogy to Eq. (2).

The upper and lower bounds defined in this way are, unfortunately, far from each other because in reality $\mu(T_{min})$ is several orders of magnitude smaller than $\mu(T_{conv})$, consequently

$$\varrho_{min} \ll \varrho_{conv}.$$

On the other hand, from Eqs. (2) and (9)

$$m_2'(P) \cdot \log(1-\varrho_{min}) = m_1'(P) \cdot \log(1-\varrho_{conv}).$$

Now considering that the function $x \log(1-x)$ is negative and monotoneously decreasing in the interval (0, 1), we get

$$\frac{m_2'(P)}{m_1'(P)} = \frac{\log(1-\varrho_{conv})}{\log(1-\varrho_{min})} \geqq \frac{\varrho_{conv}}{\varrho_{min}},$$

therefore

$$m_2(P) \approx m_2'(P) \gg m_1'(P) \approx m_1(P).$$

The modified method. Search from random sets modified by a step of steepest descent

From the above considerations we conclude that a comparison of the function values belonging to a number of $m_2(P)$ vectors leads to $p_1^* \in T_{min}$. The smaller set $m_1(P)$ also contains at least one vector belonging to T_{conv}, but we cannot find it owing to *the lack of a perfect method for selection.*

Since the comparison of function values as a principle of selection cannot be replaced with anything else, the error can be reduced only by generating a point within T_{min}. Such a point could be made available only if an $m_2(P)$ number of drawings had been performed. This task becomes especially hard when practically nothing is known about ϱ_{min}; e.g. in the case when, upon increasing the number of drawings from 100 to 1000, still no $p^*_{min} \in T_{min}$ can be expected with certainty. We can get out of this apparent deadlock by not generating the parameter vector in T_{min} *by simple drawing.*

Let us carry out one drawing, then modify this point by moving along the direction-grad $f(p)|_{p_i^*}$ until a minimum of the function at p_1^{**} on this line is found. The number of drawings and searches be altogether $m_1(P)$. It is a practical experience, which can be proved for a number of functions also theoretically, that if $p_i^* \in T_{conv}$, then $p_1^{**} \in T_{min}$ is also valid. Therefore, by applying this strategy, an erroneous decision is practically out of the question.

This statement has been verified in many practical applications. In parameter estimations and also in the case of the test functions to be shown later it has been found that the probability $P^{**}(m) \equiv P(p^{**} \in T_{conv})$ by far exceeded $P^*(m)$. The vector p^{**} is that for which

$$f(p^{**}) = \min_{i=1,2,\ldots,m} f(p_i^{**}).$$

As a consequence, if the random vectors are modified by a search for the minimum along the gradient direction, it is sufficient to make only $m_1(P)$ drawings corresponding to the lower bound in Eq. (9).

Examples

We have succesfully used the modified method of generating initial values for optimization in the determination of rate constants in reaction kinetics. The function to be optimized was the sum of squares function

$$f(p) = \sum_{i=1}^{t} (y_i' - y_i(p))^2$$

where

$$y' \equiv (y_1', y_2', \ldots, y_t')$$

stands for the experimental data and

$$y(p) \equiv (y_1(p), y_2(p), \ldots, y_t(p))$$

for the response function. The values of the response function $y(p)$ in kinetic work can be obtained only after laborious calculations involving expansion, numerical integration, etc. For optimization of the sum of squares functions, we have used the Fletcher—Powell [4] method and a procedure we developed by modifying the Newton—Gauss type of iteration [5]. The modified drawing, when combined with one of the known gradient methods, is well suited according to our experiences for generating initial estimates in practical optimization procedures [5, 6]. Nevertheless, this paper should be confined to a lesser job, i.e. to illustrate the application of the method on two test functions.

In Table I, the results obtained with two *test functions* exhibiting several local minima in the parameter-space are shown. The absolute minimum for both functions lies at $p_{min}=(0, 0)$, where $f(p_{min})=0$. The functions themselves, and the parameter and the convergence regions, expected on the basis of the analytical properties of the functions, are specified in the first part of the table. In the second part of the table are listed the lower and upper bounds evaluated from probability functions (1) and (6), as well as the relative frequencies for both standard and modified drawings obtained as a result of several hundred computer runs. The values of $\varrho_{min,I}=0.01$ and $\varrho_{min,II}=0.13$ used for the computation of the lower bounds have been estimated from the analytical properties of functions I and II, respectively. The upper bounds have been calculated not from the trivial measure of the convergence region defined in the table but from the relative frequencies found for single drawings, making use of the definition of T_{conv}, i.e. putting $\mu(\varrho_{conv})=P^*(1)=P^{**}(1)$.

I. Table

1. Optimizations	I.	II.
test functions	$f(p) = (25-p_1^2)\sin^2 \Pi p_1+p_2^2$	$f(p) = p_1^2((p_2^2-4)^2+1)+p_2^2$
parameter region	$T = \{p; \|p_1\| \leqq 4.5, \|p_2\| \leqq 1.5\}$	$T = \{p; \|p_1\| \leqq 2.5, \|p_2\| \leqq 2.5\}$
convergence region	$T_{conv} \cong \{p; \|p_1\|<0.5, \|p_2\|<1.5\}$	$T_{conv} \cong \{p; \|p_1\|<1.2, \|p_2\|<2.5\}$

2. Probability of convergence depending on number of drawings

number of drawings	1	3	6	9	12	1	2	4	6	8
lower limit	0,010	0,030	0,059	0,087	0,114	0,130	0,243	0,427	0,566	0,672
simple drawing	0,137	0,35	0,50	0,63	0,68	0,550	0,57	0,74	0,84	0,88
modified drawing	0,137	0,35	0,55	0,71	0,83	0,550	0,78	0,92	0,97	1,00
upper limit	0,137	0,357	0,587	0,735	0,843	0,550	0,798	0,959	0,992	0,998

By comparing the lines of the table, it becomes obvious that the modified drawing, as expected, is more efficient than the simple one. Also, the relative frequencies for modified drawings are in good agreement with the theoretical maxima, thus the method seems to be suitable for the elimination of the error involved in simple selection.

A similar result has been obtained in multiparameter fits. The illustration, however, would be more complicated in this case, owing to the features of the sum of squares functions mentioned above and the excessive computer time needed for setting up a similar table.

Though the modified method practically eliminates the error in the comparison of function values, the problem of generating initial estimates for optimization is far from being solved. As yet, no satisfactory answer has been found to the main question: how many drawings have to be made in a given case (with or without modification of the vectors drawn) in order to obtain good initial values. No $m(P)$ inverse can be given for the theoretical function in Eq. (3), and the limits $m_1(P)$ and $m_2(P)$, which might be of theoretical interest and were applied successfully in this work, cannot be calculated in practical problems. (It is easy to show that the computation of ϱ_{conv} and $\varrho_{\min}$ would be a more complex problem than finding the optimum itself.) The only definite statement which can be made is that the number of drawings needed to assure convergence in the modified search is always of a lower order of magnitude than that needed in the direct search. In practice it proved to be a good strategy to try to find the initial value by modified search from as many drawings as there are parameters involved and to repeat the whole of optimization in case of divergence.

Summary

When one tries to determine the unrestricted local minimum of a function of several variables by an iterative algorithm, it frequently happens that the algorithm is successful only if a sufficiently good estimate of the starting vector can be provided. Authors consider the following process: generate n random vectors, and apply one iteration of the steepest descent method for each of them; select as starting vector for subsequent optimization one that yields the least function value. The paper deals with the probability theory foundation of the modified drawing method, and with the discussion of the experiences of its application. It is proved that this strategy enhances the probability of convergence in practical optimization procedures.

Совместное применение разыгрыша и метода "steepest descent" в задаче определения начальных значений для дальнейшей оптимизации

Если локальный минимум функции от нескольких переменных нужно определить итеративным алгоритмом, тогда операция в большинстве случаев только в том случае удачная, если можно предписать относительно хорошие начальные значения со стороны переменных. Авторы и предлагают следующий метод для определения таких начальных значений:

после генерации n случайных векторов, исходя из них, осуществляем по одной итерации методом "steepest descent" и из полученных векторов тот нужно выбрать для исходного значения, который является меньшим значением функции. Статья занимается новым методом, как теоретическим сформулированием задачи теории вероятностных исчислений, и дает результат практических опытов, которые доказывают, что такая стратегия в действительности увеличивает вероятность конвергенции метода оценки параметров.

INSTITUTE OF RADIOCHEMISTRY
JÓZSEF ATTILA UNIVERSITY
SZEGED, HUNGARY

Literature

[1] RALSTON, A., H. S. WILF, *Mathematical methods for digital computers*, John Wiley and Sons, Inc., 1960.
[2] LINNIK, YU. V., *Methods of least squares and principles of the theory of observations*, Pergamon Press, 1961.
[3] BARD, Y., L. LAPIDUS, *Catalysis Reviews*, v. 2, 1968, p. 67.
[4] FLETCHER, R., M. J. D. POWELL, *Comput. J.* v. 6, 1963, p. 163.
[5] VARGA, K., *Acta Chim. Acad. Sci. Hungar.* v. 73, 1972, p. 411.
[6] FEJES, P., P. KÖNIG, *Acta Chim. Acad. Sci. Hungar.* v. 71, 1972, p. 217.

(Received May 16, 1972)

Some remarks on the paper of K. Varga and P. Fejes

By P. HUNYA

In practical applications of the iterative unconstrained optimization methods (for example in the gradient method discussed in [1]) a difficult problem is to find the initial estimation of the solution. Generally m points are randomly chosen from the region containing the optimum place and that which represents the minimum (maximum) value of function is considered as the initial estimation. Following this strategy the sequence generated by the iterative process converges with a probability generally less then 1. By the modified method proposed in [1] one step of the iteration is performed for all the m points before the selection and this may improve the probability of the convergency. However the verification of this property is heuristic, it is based on a number of experimental calculations with various type of functions. The upper and lower limits of the convergency are also given in [1].

In this paper a generalization of the modified method is described and the probability of the convergency is discussed in detail. (The problem of minimization is exemined, since all considerations are analogous in the case of maximization.)

Let us consider the continuous function of real values $F(x)$ defined on the complete metric space S. Let T be a subset of S with nonzero measure and let us suppose that F takes its minimum on T. Let $y=Mx$ be a mapping of T into T with the property

$$F(Mx) \leqq F(x). \tag{1}$$

There exists obviously a $T_{\text{conv}} \neq \emptyset$ subset of T for the elements x of which the sequence $x=M^0x, M^1x, M^2x, \dots, M^nx, \dots$ is convergent and

$$\lim_{n\to\infty} M^n x = \bar{x}, \tag{2}$$

where

$$F(\bar{x}) = \min_{x\in T} F(x);$$

that is, the iteration process generated by M converges to the solution of the optimization problem. To simlify the considerations we suppose the uniqueness of $\bar{x}$, however, this fact is not essential in the following proofs.

Suppose further that

$$\varrho(T_{\text{conv}}) \neq 0 \tag{3}$$

($\varrho(t)$ means the measure of the set $t \subset S$) and

$$\varrho(T_{\min}) \neq 0 \tag{4}$$

where $T_{\min}$ is a subset of T_{conv} defined by

$$T_{\min} = (x|\, F(x) < \inf_{y \in T - T_{\text{conv}}} F(y)). \tag{5}$$

Now let us consider the following procedure:

a) let the points $x_1, x_2, \ldots, x_m$ be chosen from T independently, with homogeneous distribution on T;

b) then form the sequence

$$M^n x_1, M^n x_2, \ldots, M^n x_m$$

for $n \geqq 0$ and

c) let a point $\bar{x}^n = M^n x_{k*}$ be chosen for which

$$F(M^n x_{k*}) = \min_{1 \leqq i \leqq m} F(M^n x_i).$$

We shall prove that by the previous conditions for $n \to \infty$ the probability $P = P(\bar{x}^n \in T_{\text{conv}})$ converges to the limit $\bar{P}$ while

$$P_n \leqq P \leqq \bar{P} \tag{6}$$

where P_n is a monotonic increasing sequence i. e.

$$P_n \leqq P_{n+1} \tag{7}$$

and

$$\lim_{n \to \infty} P_n = \bar{P}. \tag{7a}$$

In other words: by increasing the number of the iterative steps before selection the probability of the convergency approximates its upper limit with an arbitrary degree of accuracy. (The dependence on m is not considered in this paper.)

For the proof we define the following events:

1) A_n denotes the event that among the points $M^n x_1, \ldots, M^n x_m$ there exists element of T_{conv}

$$A_n = \{\exists i (M^n x_i \in T_{\text{conv}})\}. \qquad (n = 0, 1, 2, \ldots) \tag{8}$$

As a consequence of the definition of T_{conv} we have

$$A_n = A_k (= A), \tag{9}$$

since it is obvious that

$$M^k x \in T_{\text{conv}} \Leftrightarrow M^{k+1} x \in T_{\text{conv}}$$

for $k = 0, 1, 2, \ldots$.

2) B_n denotes that the selected point $\bar{x}^n$ is element of T_{conv}, that is, the iterational process converges:

$$B_n = \{\bar{x}^n \in T_{\text{conv}}\}. \tag{10}$$

3) C_n denotes the event that $\bar{x}^n$ is element of $T_{\min}$

$$C_n = \{\bar{x}^n \in T_{\min}\}. \tag{11}$$

4) And finally D_n denotes the event that at least one of the points $M^n x_1, M^n x_2, \ldots, M^n x_m$ belongs to $T_{\min}$

$$D_n = \{\exists i (M^n x_i \in T_{\min})\}. \tag{12}$$

From the definitions 1)—4) immediately follows that for all n.

$$D_n \Leftrightarrow C_n \Rightarrow B_n \Rightarrow A_n \Leftrightarrow A, \tag{13}$$

consequently, the probabilities of A_n, B_n, C_n, D_n, A satisfy the relations

$$P(D_n) = P(C_n) \leqq P(B_n) \leqq P(A_n) = P(A). \tag{14}$$

For $n=0$ (14) contains as a special case one of the results of [1] for the values

$$P_0 = P(C_0) = 1-(1-\varrho_{\min})^m$$
$$\bar{P} = P(A_0) = 1-(1-\varrho_{\text{conv}})^m.$$

Using the notations $P_n = P(C_n) = P(D_n)$ and $\bar{P} = P(A) = P(A_n)$ (14) proves (6) also.

As a consequence of the condition (1) and the definition 4) we have

$$D_n \Rightarrow D_{n+1} \tag{15}$$

and this implies the inequality (7).

Let us consider now the sequence $D_0, D_1, \ldots, D_n, \ldots$. We shall prove that

$$\sum_{i=1}^{\infty} D_i = A. \tag{16}$$

From the continuity of the function $F(x)$ follows that for the elements $\underline{x}$ of T_{conv} we can find a natural number $n(x)$, such that $n > n(x)$ implies $M^n x \in T_{\min}$ Let T^i denote the subset of T_{conv} for the elements of which $n(x) = i$ ($i = 0, 1, 2, \ldots$) and define the event A^i as follows

$$A^i = \{\exists j (x_j \in T^i\}.$$

By the definition above $A_i \Rightarrow D_i$. On the other hand if there exists x_j ($1 \leqq j \leqq m$) such that $M^i x_j \in T_{\min}$ is true, then $x_j \in T_i$ also holds; that is, D_i implies A^i. Thus we have $A^i = D_i$. It is obvious that $A = \sum_{i=0}^{\infty} A^i$ so we get (16).

Because of (15) and (16) one of the basic limit theorems of the probability theory ([2], § 2.2) can be applied to the sequence $D_0, D_1, \ldots, D_n, \ldots$, therefore

$$\lim_{n\to\infty} P(D_n) = P(A)$$

so using the notations introduced previously we get (7a)

$$\lim_{n\to\infty} \underline{P}_n = \bar{P}.$$

Несколько замечаний к работе К. Варга и Р. Фейеш

В итерационных методах безусловной оптимизации при случайно выбранном начальном значении, получается последовательность сходящаяся к решению задачи, только с вероятностью P (обычно меньшей единицы). Вероятность сходимости может увеличиться, если из m случайно выбранных точек, считаем начальной ту точку, в которой достигается мини-

мум (максимум). Дальнейшее улучшение получается при выпольнении n шагов итерационной процедуры перед выбором исходной точки.

Доказывается, что при довольно общих условиях, нижний предел вероятности P является монотонной функцией от n, и с ростом n, P сходится к своему верхнему пределу, обеспечивая этим увеличение вероятности конвергенции.

LABORATORY OF CYBERNETICS
JÓZSEF ATTILA UNIVERSITY
SZEGED, HUNGARY

References

[1] VARGA, K., P. FEJES, Combined application of drawing and steepest descent in generating initial estimates for subsequent optimization, *Acta Cybernet.*, v. 2, 1973, pp. 53—59.
[2] PRÉKOPA, A., *Valószínűségelmélet*, Budapest, 1962.

(Received June 28, 1972)

Neuron counting in three dimensions; a proposal

By C. ARCELLI and S. LEVIALDI

Abstract

The preparation of histological specimens for neuron cell counting is briefly reviewed. Problems related to the observation of biological sections are discussed and a proposal for neuron counting in three dimensions is made. A parallel algorithm for this purpose is described.

Introduction

For the last century people have investigated the nature and structure of the nervous system through its elementary units, the neurons. For this purpose they have developed different techniques that, with an improving technology, have successively furnished a better insight into the numerous problems existing in the field. We will point out the significant phases of the preparation of biological specimens as well as the difficulties that arise when using the conventional techniques. Our proposal is aimed at: a) the three-dimensional reconstruction of a specimen by means of a digital matrix and b) the use of a special algorithm that operates in three dimensions on all elements of the matrix simultaneously, to count the digitized neurons contained in the previous matrix.

In the preparations of histological specimens various substances are needed in order to obtain the slide that will be subsequently analyzed. One of these substances is the fixative, employed to stain the relevant elements to be observed either by optical or electronic means. For neurons the Nissl method with cresyl violet staining is commonly employed [1]. This method, as well as all others, produces large changes in tissue volume. Some authors rate this change up to about a 70% reduction [2] in the volume of the original sample. The embedding medium must satisfy the following requirements [3]: high resistivity for constant section thickness, discrete rigidity for sections of constant width, adequate elasticity to contain samples of any size. Two materials commonly employed are paraffin and celloidin. Paraffin allows fast embedding and thinner sections, while celloidin is more resistant. A good compromise is reached by using tissuemat and combining the advantages of paraffin with those of celloidin. As for the thickness of the section, the following considerations should be kept in mind. If the section is thick, then fewer sections for a given histological preparation are required, less time and space are needed, and fewer cells, nuclei and nucleoli are split. This increases the probability for accurate counting. On the other hand, if we use thin sections less counting error will be introduced since fewer neurons will overlap. Neurons, as seen through an optical microscope, will be overlapping [4] and only certain parts within the section may be simultaneously in focus. Look at fig. 1a), b) and c). The thickness of the sections varies from 10 to 50 microns, usually about 15—20 microns.

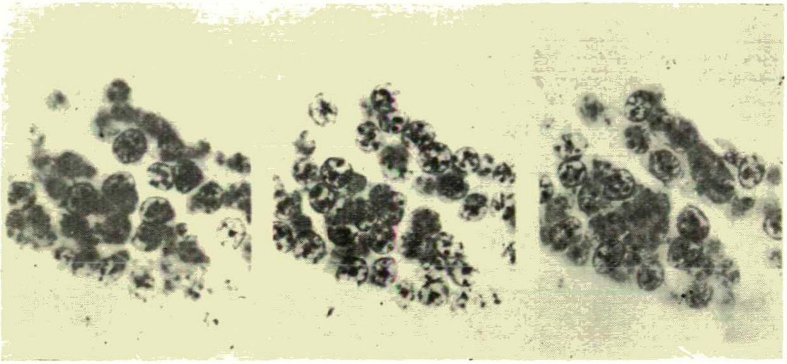

Fig. 1a), b), c)
Successive focal planes of habenular nucleus of frog (Rana esculenta).
Transverse section, *x* 1200

Although there are no absolute rules for the choice of a sectioning plane, one plane is chosen to show a specific structure on the basis of the interpretation of the biological material when cut along three orthogonal directions. If one could have a three-dimensional representation of the specimen then this one could suggest the best sectioning plane.

The problems most commonly met when counting neurons using the previously described specimens are: 1) contour definition of cells (the presence of neurons on different focal planes further complicates the picture); 2) the appearance of the same neurons in adjacent sections. To solve the first problem some authors have tried to outline, as precisely as possible, the contour of the nucleus with a very sharp pencil and Higgins green ink. To compensate for the poor discrimination between different planes some specific correction factors have been introduced, (e.g., Abercrombie, [5] 1946). The difficulty in assessing the exact number of neurons in a given specimen may be described by a figure ranging between 2% and 10% [2] of the total number of counts on the same specimen.

Three-dimensional representation

We have seen that when investigators use a specific technique for counting neurons with an optical microscope, different focal planes are inspected. We might ask ourselves how it could be possible to three-dimensionally reconstruct the sample so as to eliminate picture noise due to overlapping of cells and to the presence of non-focussed components. If this were possible then not only would we obtain a fully focussed representation but also a preservation of size and shape.

Under this assumption we propose a three-dimensional matrix in which biological information from the specimen will be stored. Let us consider a three-dimensional matrix of x, y parallel planes ordered along the z-axis. With each of these planes we may associate one focal plane, such that all histological sections will be stored in the matrix. From an operative standpoint, a focussed component of a picture has a contrast ratio above a specified threshold value. We must remark that before storing the information on a matrix plane, some processing of the picture must be performed in order to extract only those components which are focussed. This can be achieved by using techniques [6, 7] for eliminating spurious noise in digitized images. We then have a matrix which contains all relevant elements present in the specimen in digital form.

We may note that each element of the matrix has a grey scale value: this is due to the fact that the patterns involved are not black and white but rather continually varying in their intensity as a result of the use of staining techniques and of the complex structure of cells.

Three-dimensional matrices can also be considered as arrays for storing tactile sensory information from objects in space [8]. After making contact with the object a special sensor could, in principle, trace it, obtaining a quantized contour for every section along the z-axis. For this specific case we are involved in binary matrices and only information relevant to the surface will be stored. Once the three-dimensional matrix is obtained patterns stored in it can be processed according to the set of rules dictated by the task.

Problems existing in two dimensions regarding connectivity, adjacency, geometrical operations, should be reconsidered for three dimensions.

For this reason, for example, let us compare processing of two-dimensional patterns with three-dimensional ones. The memory occupation will obviously be larger since more data are needed because of the presence of an extra dimension but we must also note that more operations will be required to test certain properties in this space. As an example, when we define two elements $a(i, j, g)$; $b(h, k, m)$ to be d_2-adjacent if

$$d_2\{a(i,j,g); b(h,k,m)\} = \max(|i-h|, |j-k|, |g-m|) = 1$$

then, an isolated element test will require 26 check operations while only 8 are necessary in two dimensions.

If, instead of using sequential algorithms we are interested in performing parallel processing, further memory must be employed as a buffer unit for storage.

Neuron counting

For the problem of neuronal counts we propose an algorithm developed for counting objects in three dimensions [9]. This algorithm operates in parallel, first shrinking all objects and then normalizing them to single isolated elements. This procedure was obtained from the superposition of a two-dimensional algorithm acting along three orthogonal planes. It may operate on d_1- and d_2-connected objects where d_1 connectivity in three dimensions can be defined as follows: a set S of elements is d_1-connected if, any two elements in S having been considered, a path exists joining them through successive elements $(a(i, j, g); b(h, k, m))$ all in S such that

$$d_1\{a(i, j, g); b(h, k, m)\} = |i-h|+|j-k|+|g-m| = 1.$$

The d_2 connectivity definition can be obtained from the d_2 adjacency formula. The algorithm is a parallel one since every element is processed simultaneously with all others and independently. The transformed state b^* at every step depends on the state of elements belonging to a $2\times2\times2$ window. In fig. 2 we can see that one vertex corresponds to element b. The small letters represent the 0, 1 states of each element.

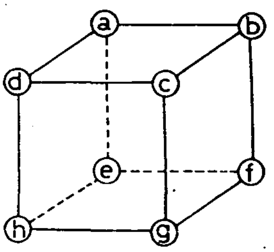

Fig. 2
$2\times2\times2$ window, b^* (transformed state of b) will depend on the state of elements contained in this window

The following relations hold for d_1- and d_2-connected objects respectively, if

$$u(t)=0, \quad t<0 \quad \text{and} \quad u(t)=1, \quad t\geqq 0,$$

$$b^* = u[u(a+b-2)+u(c+b-2)+u(f+b-2)+u(a+d+ +c-3)+u(a+e+f-3)+u(c+g+f-3)],$$

$$b^* = u[u(a+b+g-2)+u(b+c+e-2)+u(a+c+f-2)+ +u(b+h-2)+u(b+f+d-2)].$$

Our algorithm is direction oriented since symmetrical configurations, e.g., eight elements forming a cube, must not be completely erased. For example, if we consider a parallelepiped circumscribing the object, the process of shrinking compels every element to move towards one vertex and, precisely, for the chosen disposition of elements in the formula, towards the top, right, backward vertex (vertex b, fig. 2). After a finite number of steps all objects will be shrunk to single elements (vertex elements), then extracted and counted. For a correct counting to be performed two conditions must be satisfied: no object must disconnect itself during the process of shrinking, and no two or more objects must merge during the same process. The first condition is always satisfied while the second is verified only if the parallelepipeds circumscribing the objects are not adjacent, i.e., the distance between them $d_2\geqq 2$. Thus, the number of steps necessary to process all objects placed in a matrix depends only on the dimensions of the largest parallelepiped circumscribing the object.

Conclusion

A procedure is introduced to represent a biological specimen in digital form which preserves the spatial organization of its components. Every histological section is stored in a three-dimensional array in which every x, y plane corresponds to a single focal plane as seen by optical inspection. In this way, only strictly focussed components are preserved. Our processing takes into account the three-dimensional nature of the chosen description of the world and should not be seen as a set of successive differing processes in two dimensions. From this point of view picture processing methods should be reconsidered with respect to space geometry. As an example, a parallel shrinking algorithm which could perform counting of cells has been proposed.

Подсчёт нейроновых клеток в трёх димензиях

Количественные анализы в нейронной анатомии требуют метода автоматизации из-за необходимости исследования большого количества гистологических материалов. Первые шаги в этом направлении были сконцентрированы на расширение каналов передачи данных от микроскопа к вычислительной машине для того, чтобы хранить и как можно лучше обрабатывать данные, биологическую информацию, для решения задачи.

Коротко даётся обзор специфичных методов окраски и подготовки нейроновых клеток. Освещаются некоторые проблемы исследования человеком морфологической структуры клеток. Даются предложения к методу обработки, исследований биологических образцов, в котором вся информация трёхмерной реальной структуры клеток хранится и обрабатывается в памяти вычислительной машины.

Далее рассматривается алгоритм для специфичной задачи подсчёта нейроновых клеток, который симультанно и независимо друг от друга оперирует над всеми элементами трёхмерного изображения образца.

LABORATORIO DI CIBERNETICA
DEL CNR
ARCO FELICE, NAPOLI

References

[1] Kemali, M. & V. Braitenberg , *Atlas of the frog's brain*, Springer Verlag, 1969.
[2] Pakkenberg, H., The number of nerve cells in the cerebral cortex of man, *J. Comp. Neurology*, v. 128, 1966, pp. 17—20.
[3] Konigsmark, B. W., Methods for the counting of neurons, *Contemporary research methods in neuroanatomy*, Springer Verlag, 1970, pp. 315—333.
[4] Agduhr, E., A contribution to the technique of determining the number of nerve cells per volume unit of tissue, *Anatomical Record*, v. 80, 1941, pp. 191—202.
[5] Abercrombie, M., Estimation of nuclear population from microtome sections, *Anatomical Record*, v. 94, 1946, pp. 239—247.
[6] Prewitt, J. M. S., Object enhancement and extraction, *Picture processing and psychopictorics*, Academic Press, New York, 1970, pp. 75—149.
[7] Rosenfeld, A., Picture processing by computer, Academic Press, New York, 1969.
[8] Aida S., L. Cordella, N. Ivancevic, A method of visualtactile symbiotic system for stereometric pattern recognition, presented at *the Second International Joint Conference on Artificial Intelligence*, London, 1971.
[9] Arcelli, C. & S. Levialdi, Parallel shrinking in three dimensions, to appear in *Computer Graphics and Image Processing*, v. 1, 1972.

(Received January 26, 1972)

Notes on maximal congruence relations, automata and related topics

By H. ANDRÉKA*, S. HORVÁTH**, I. NÉMETI*

Abstract

The paper starts from the fact that if r_0 is an equivalence relation on a free semigroup **A**, then (uniquely) exists a greatest right compatible refinement of r_0 (see e.g. [3, chapter 9] and [4, 1. §]).

In Part 1, the authors generalize the above question and investigate it in the case when **A** is an arbitrary semigroup. They present a constructive proof for one of the concerning theorems (Theorem 1′) e.g. they show that if r_0 is an equivalence relation on **A**, then the relation

$$r_m \overset{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in r_0 \wedge (\forall a, b)[a, b \in \mathbf{A} \Rightarrow (\langle ax, ay\rangle, \langle xb, yb\rangle, \langle axb, ayb\rangle \in r_0)]\}$$

is the greatest congruent refinement of r_0 in the sense that whenever r_1 is a congruence relation on **A** and $r_1 \subset r_0$, then $r_1 \subset r_m$.

In an interesting way, it turns out that in the definition of r_m, requiring $\langle axb, ayb\rangle \in r_0$ too (in addition to $\langle ax, ay\rangle, \langle xb, yb\rangle \in r_0$), is not superfluous: generally it does not follow from the other two.

The most general theorem of Part 1 is proved by using lattice-theoretical considerations (Theorem 1).

In Part 2, it is proved (Theorems 2 and 2′) that a partial reverse of Theorem 1 is equivalent to **A** having some sort of the special "quasi-trivial" structure (Definition 1).

In part 3, we represent every equivalence class of initially connected Moore automata, the elements of which induce the same automaton mapping $\bar{f}$, by the function f, derived from $\bar{f}$ by putting for every $w \in X^*$ (X is the input alphabet)

$$f(w) \overset{\text{def}}{=} \text{"the last letter of } \bar{f}(w)\text{"}.$$

These functions f we simply call automata. We draw a short parallel between the notion of an automaton f and the classical notion of a Moore automaton. During this the theorems of Part 1 prove to be directly applicable to the automata f, and in this way classical results concerning Moore automata can be deduced (e.g. the Corollary of Statement 1).

As a generalization of the fact that the semigroup of a finite Moore automaton is also finite, we prove (Statement 2) that if r is a right congruence relation of finite

index, on a semigroup **A**, then r can always be refined into a congruence relation of finite index.

In connection with the general investigation of the semigroup of the so-called semigroup-machine $\langle \mathbf{A}, \mathbf{A}, \delta\rangle$, where **A** is an arbitrary semigroup and $(\forall a, b \in \mathbf{A})$ $\delta(a, b) \overset{\text{def}}{=} ab$; we introduce the "congruence relations of right uniformity, left uniformity and uniformity" (Def. 8).

At the end of Part 3, we prove that the possibility of simulating an automaton f by an automaton g, depends essentially on the semigroups of f and g, and is independent of their input alphabets which may be different.

1. Maximal compatible refinements of equivalence relations; generalizations

In this paper by the word *relation* we shall always mean a binary relation r over some nonvoid set **A** i.e.

$$r \subset \mathbf{A} \times \mathbf{A} = \mathbf{A}^2.$$

If we define an associative binary operation "$\circ$" on **A**, we have the semigroup $\langle \mathbf{A}, \circ\rangle$. For the sake of simplicity, we shall refer to **A** *as a semigroup* simply by the same letter **A**, instead of $\langle \mathbf{A}, \circ\rangle$ and instead of $x \circ y$ we shall write xy. If an equivalence relation r on **A** has the property

$$(\forall x, y, u, w)[(\langle x, y\rangle \in r \wedge \langle u, w\rangle \in r) \Rightarrow \langle xu, yw\rangle \in r], \tag{1.1}$$

we call it a *congruence relation* on **A** (as a semigroup). If we regard only "one half" of (1.1), namely

$$(\forall x, y, u)[(\langle x, y\rangle \in r \wedge u \in \mathbf{A}) \Rightarrow \langle xu, yu\rangle \in r] \tag{1.2}$$

or

$$(\forall x, y, u)[(\langle x, y\rangle \in r \wedge u \in \mathbf{A}) \Rightarrow \langle ux, uy\rangle \in r], \tag{1.3}$$

then we call r a *right* or *left congruence relation* respectively. Of course, a congruence relation is at the same time a right congruence relation as well as a left one. Conversely, because of the transitivity of r (as an equivalence relation)

$$(\forall r)[((1.2) \wedge (1.3)) \Rightarrow (1.1)].$$

Hence

$$(\forall r)[((1.2) \wedge (1.3)) \Leftrightarrow (1.1)]$$

i.e. r is a congruence relation iff

$$(\forall x, y, u)[(\langle x, y\rangle \in r \wedge u \in A) \Rightarrow (\langle xu, yu\rangle, \langle ux, uy\rangle \in r)]. \tag{1.4}$$

We shall always use (1.4) instead of (1.1).

The following notations will also prove useful

$$\mathscr{C}\mathbf{A} \overset{\text{def}}{=} \{r \mid r \text{ is an equivalence relation on } \mathbf{A}\},$$

$$\mathscr{R}\mathbf{A} \overset{\text{def}}{=} \{r \mid r \text{ is a reflexive relation on } \mathbf{A}\},$$

$$\mathscr{S}\mathbf{A} \overset{\text{def}}{=} \{r \mid r \text{ is a symmetric relation on } \mathbf{A}\},$$

$$\mathscr{T}\mathbf{A} \overset{\text{def}}{=} \{r \mid r \text{ is a transitive relation on } \mathbf{A}\},$$

$$\mathcal{TS}\mathbf{A} \stackrel{\text{def}}{=} \mathcal{T}\mathbf{A} \cap \mathcal{S}\mathbf{A},$$
$$\mathcal{TR}\mathbf{A} \stackrel{\text{def}}{=} \mathcal{T}\mathbf{A} \cap \mathcal{R}\mathbf{A},$$
$$\mathcal{SR}\mathbf{A} \stackrel{\text{def}}{=} \mathcal{S}\mathbf{A} \cap \mathcal{R}\mathbf{A},$$
$$\mathcal{C}_\Omega \mathbf{A} \stackrel{\text{def}}{=} \{r \mid r \text{ is a congruence relation on } \mathbf{A}\},$$
$$\mathcal{C}_{\Omega R} \mathbf{A} \stackrel{\text{def}}{=} \{r \mid r \text{ is a right congruence relation on } \mathbf{A}\},$$
$$\mathcal{C}_{\Omega L} \mathbf{A} \stackrel{\text{def}}{=} \{r \mid r \text{ is a left congruence relation on } \mathbf{A}\}.$$

Of course, by definition, $\mathcal{C}\mathbf{A} = \mathcal{R}\mathbf{A} \cap \mathcal{S}\mathbf{A} \cap \mathcal{T}\mathbf{A}$ and by the equivalence of (1.1) and (1.4)

$$\mathcal{C}_\Omega \mathbf{A} = \mathcal{C}_{\Omega R} \mathbf{A} \cap \mathcal{C}_{\Omega L} \mathbf{A}.$$

Further notations

$$\pi \mathbf{X} \stackrel{\text{def}}{=} \{\mathbf{Y} \mid \mathbf{Y} \subset \mathbf{X}\}$$

(here and all along the symbol "$\subset$" may stand for "$=$" too),

$$1_{\mathbf{X}} \stackrel{\text{def}}{=} \{\langle z, z\rangle \mid z \in \mathbf{X}\}.$$

If $r \subset \mathbf{A}^2$ and n is a natural number, the n-th power of r we define as

$$r^n \stackrel{\text{def}}{=} \{\langle x, y\rangle \mid (\exists z_0, z_1, \ldots, z_n)[(z_0, z_1, \ldots, z_n \in A) \wedge$$
$$\wedge z_0 = x \wedge z_n = y \wedge (\langle z_0, z_1\rangle, \ldots, \langle z_{n-1}, z_n\rangle \in r)]\}$$

and the *transitive closure* of r is

$$\hat{r} \stackrel{\text{def}}{=} \bigcup_{i=1}^{\infty} r^i. \tag{1.5}$$

As is well known, for any set $\mathbf{X}$, $\pi\mathbf{X}$ forms a complete lattice with respect to the partial ordering $\subset$.

In this case, the meet and join operations are the following

$$(\forall \mathbf{Z} \subset \pi\mathbf{X}) \begin{cases} \underset{z \in \mathbf{Z}}{\frown} z \stackrel{\text{def}}{=} \bigcap_{z \in \mathbf{Z}} z & (1.6) \\ \text{and} \\ \underset{z \in \mathbf{Z}}{\smile} z \stackrel{\text{def}}{=} \bigcup_{z \in \mathbf{Z}} z & (1.7) \end{cases}$$

where $\frown$, $\smile$ denote the lattice-theoretical operations and $\bigcap$, $\bigcup$ are the usual symbols of the set-theoretical intersection and union respectively. We agree (as usual) that

$$\bigcap_{z \in \emptyset} z = \mathbf{X}, \quad \bigcup_{z \in \emptyset} z = \emptyset.$$

E.g. if $\mathbf{X} = \mathbf{A}^2$, $\pi\mathbf{A}^2$ is a complete lattice with meet operation (1.6) and join operation (1.7). However, if we replace $\pi\mathbf{A}^2$ with $\mathcal{T}\mathbf{A}$, we must modify the join operation of (1.7) for $\mathcal{T}\mathbf{A}$ to be a complete lattice (under the partial ordering $\subset$)

$$\underset{r \in \mathbf{Z}}{\smile} r \stackrel{\text{def}}{=} \coprod_{r \in \mathbf{Z}} r \stackrel{\text{def}}{=} \widehat{\bigcup_{r \in \mathbf{Z}} r}. \tag{1.8}$$

The reason why transitive closure (1.5) has entered is just the transitivity of the elements of $\mathcal{T}\mathbf{A}$. It can easily be checked that with the operations $\cap$ and $\amalg$, $\langle \mathcal{T}\mathbf{A}, \subset \rangle$ is indeed a complete lattice.

Using the following notation for any two lattices V and W, $V \to W =$ "W is a complete sublattice of V", the following "directed graph" is valid

$$\begin{array}{ccccccccc} & \nearrow & \mathcal{T}\mathcal{S}\mathbf{A} & \searrow & & \nearrow & \mathcal{C}_{\Omega R}\mathbf{A} & \searrow & \\ \mathcal{T}\mathbf{A} & & & & \mathcal{C}\mathbf{A} & & & & \mathcal{C}_{\Omega}\mathbf{A}. \\ & \searrow & \mathcal{T}\mathcal{R}\mathbf{A} & \nearrow & & \searrow & \mathcal{C}_{\Omega L}\mathbf{A} & \nearrow & \end{array} \tag{1.9}$$

(The relation "$\to$" is itself a partial ordering over the complete sublattices of any complete lattice, as it is reflexive, antisymmetric and transitive.)

The "edges" in (1.9) between $\mathcal{T}\mathbf{A}$ and $\mathcal{C}\mathbf{A}$ may be verified simply by using definitions (1.6) and (1.8), while for those between $\mathcal{C}\mathbf{A}$ and $\mathcal{C}_{\Omega}\mathbf{A}$ we must take into account (1.2), (1.3) and (1.4) also (to show that the meet and join operations always result in an appropriate — belonging to $\mathcal{C}_{\Omega R}\mathbf{A}$ etc. — relation). This is a routine calculation. ($\pi\mathbf{A}^2 \to \mathcal{T}\mathbf{A}$ is not true, because the join operation in $\mathcal{T}\mathbf{A}$ (see (1.8)) differs from that in $\pi\mathbf{A}^2$ (see (1.7))).

The common unit element of all these complete lattices is $\mathbf{A}^2$, while the zero element of $\mathcal{T}\mathcal{R}\mathbf{A}$, $\mathcal{C}\mathbf{A}$, $\mathcal{C}_{\Omega R}\mathbf{A}$, $\mathcal{C}_{\Omega L}\mathbf{A}$ and $\mathcal{C}_{\Omega}\mathbf{A}$ is $1_{\mathbf{A}}$, and that of $\mathcal{T}\mathbf{A}$ and $\mathcal{T}\mathcal{S}\mathbf{A}$ is $\emptyset$. For any two relations r, r_1 for which $r_1 \subset r$, we say that r_1 is less than or equal to r, or r is greater than or equal to r_1, or (equivalently) r_1 is a refinement of r.

Theorem 1. If $\mathbf{A}$ is a semigroup and

(a) $$r_0 \in \mathcal{T}\mathcal{R}\mathbf{A} \quad \text{and} \quad \mathbf{M} \in \{\mathcal{C}\mathbf{A}, \mathcal{C}_{\Omega L}\mathbf{A}, \mathcal{C}_{\Omega R}\mathbf{A}, \mathcal{C}_{\Omega}\mathbf{A}\},$$

or

(b) $$r_0 \in \mathcal{T}\mathbf{A}, \quad \pi r_0 \cap \mathcal{S}\mathbf{A} \neq \emptyset \quad \text{and} \quad \mathbf{M} = \mathcal{T}\mathcal{S}\mathbf{A},$$

then the set $\mathbf{H} \overset{\text{def}}{=} \mathbf{M} \cap \pi r_0$ has a (unique) greatest element r_g

$$(\exists r_g \in \mathbf{H})(\forall r)[r \in \mathbf{H} \Rightarrow r \subset r_g]. \tag{1.10}$$

Proof

(a) By the definition of r_0, $1_{\mathbf{A}} \subset r_0$, so $\mathbf{H} \neq \emptyset$ (the case is not trivial). Being $\mathbf{M}$ a complete lattice and $\mathbf{H} \subset \mathbf{M}$, there is in $\mathbf{M}$ a least upper bound of $\mathbf{H}$ (see (1.8) and (1.9))

$$r_g \overset{\text{def}}{=} \coprod_{r \in \mathbf{H}} r \tag{1.11}$$

for which $r \in \mathbf{H} \Rightarrow r \subset r_g$ of (1.10) holds. So we have only to prove that

$$r_g \in \pi r_0. \tag{1.12}$$

Being r_g the transitive closure of a subset $\left(\bigcup_{r \in \mathbf{H}} r\right)$ of r_0 (see (1.11), (1.8) and (1.5)) and r_0 being transitive,

$$r_g \subset r_0 \tag{1.13}$$

i. e. (1.12) holds.

(b) Again $\mathbf{H} \neq \emptyset$ (the case is not trivial) and by an argument, similar to that of (a), we again have (1.13) i.e. (1.12).

Now we proceed by giving a *constructive proof* for a special case of Theorem 1, part (a).

Theorem 1'. If $\mathbf{A}$ is a semigroup, $r_0 \in \mathscr{C}\mathbf{A}$, $\mathbf{M} \in \{\mathscr{C}_{\Omega R}\mathbf{A}, \mathscr{C}_{\Omega L}\mathbf{A}, \mathscr{C}_{\Omega}\mathbf{A}\}$, and $\mathbf{H} \stackrel{\text{def}}{=} \mathbf{M} \cap \cap \pi r_0$, then (1.10) holds.

Proof. First we deal with the case $\mathbf{M} = \mathscr{C}_{\Omega}\mathbf{A}$ and then point out the obvious differences for the case $\mathbf{M} \in \{\mathscr{C}_{\Omega R}\mathbf{A}, \mathscr{C}_{\Omega L}\mathbf{A}\}$.

Let

$$r_m \stackrel{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in r_0 \wedge (\forall a, b)[a, b \in \mathbf{A} \Rightarrow (\langle ax, ay\rangle, \langle xb, yb\rangle, \langle axb, ayb\rangle \in r_0)]\}. \quad (1.14)$$

Obviously, $r_m \subset r_0$ i.e. $r_m \in \pi r_0$ and it can easily be verified that r_m satisfies condition (1.4), so

$$r_m \in \mathscr{C}_{\Omega}\mathbf{A} \cap \pi r_0 = \mathbf{H}.$$

It remained to prove (1.10) for r_m in place of r_g

$$(\forall r)[r \in \mathbf{H} \Rightarrow r \subset r_m]. \quad (1.15)$$

By the definition of $\mathbf{H}$

$$r \in \mathbf{H} \Leftrightarrow \left\{\begin{matrix} \text{(i)} \quad r \in \mathscr{C}_{\Omega}\mathbf{A} \\ \text{and} \\ \text{(ii)} \quad r \subset r_0 \end{matrix}\right\}. \quad (1.16)$$

From (i) of (1.16) follows (see (1.4)) that

$$(\forall a, b, x, y \in \mathbf{A})(\forall r \in \mathbf{H})[\langle x, y\rangle \in r \Rightarrow (\langle ax, ay\rangle, \langle xb, yb\rangle, \langle axb, ayb\rangle \in r)]. \quad (1.17)$$

From (1.17), (ii) of (1.16), and (1.14) we get that

$$(\forall x, y \in \mathbf{A})(\forall r \in \mathbf{H})[\langle x, y\rangle \in r \Rightarrow \langle x, y\rangle \in r_m],$$

which is equivalent to (1.15).

If e.g. $\mathbf{M} = \mathscr{C}_{\Omega L}\mathbf{A}$, then b, $\langle xb, yb\rangle$ and $\langle axb, ayb\rangle$ above must be deleted etc.

Remark. Condition (1.4) suggests that requiring $\langle axb, ayb\rangle \in r_0$ too in (1.14) is perhaps superfluous, but this is not at all the case

Fact. In definition (1.14), condition $(\forall a, b \in \mathbf{A})[\langle axb, ayb\rangle \in r_0]$ does not follow from

$$(\forall a, b \in \mathbf{A})[\langle ax, ay\rangle, \langle xb, yb\rangle \in r_0].$$

Proof. We construct an example. Let $\mathbf{A} = \{1, 2\}^*$ (the free monoid, generated by the set $\{1, 2\}$) and

$$r_0 \stackrel{\text{def}}{=} \{\langle u\alpha v, w\alpha z\rangle | u, v, w, z \in \{1, 2\} \wedge \alpha \in \{1, 2\}^*\} \cup \{\langle \Lambda, \Lambda\rangle, \langle 1, 1\rangle, \langle 1, 2\rangle, \langle 2, 1\rangle, \langle 2, 2\rangle\},$$

where

$$\Lambda \stackrel{\text{def}}{=} \text{the empty word (of any free monoid).}$$

It can easily be seen that $r_0 \in \mathscr{C}\mathbf{A}$, because $\langle \beta, \gamma\rangle \in r_0$ means that (denoting the length of the words in $\mathbf{A}$ by "lg") $\lg(\beta) = \lg(\alpha)$ and if $\lg(\beta) > 2$, then removing the first and last symbols from β and γ, the remaining word will be the same.

Constructing from this r_0 relations

$$r'_m \overset{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in r_0 \wedge (\forall a \in A)[\langle ax, ay\rangle, \langle xa, ya\rangle \in r_0]\}$$

and r_m — the latter according to (1.14) —, then by an easy calculation we get that $r'_m \supset r_m$, $r'_m \neq r_m$, $r'_m \in \mathscr{C}\mathbf{A}$, $r_m \notin \mathscr{C}_\Omega \mathbf{A}$. Namely, $r_m = 1_\mathbf{A}$ and $r'_m - r_m = \{\langle 1, 2\rangle, \langle 2, 1\rangle\}$.

Remark. If **A** is a monoid (i.e. a semigroup, having a unit element) then (1.14) becomes simpler

$$r_m \overset{\text{def}}{=} \{\langle x, y\rangle | (\forall a, b)[a, b \in \mathbf{A} \Rightarrow \langle axb, ayb\rangle \in r_0]\}. \tag{1.18}$$

2. A characterization of quasi-trivial semigroups

We shall introduce the following

Definition 1. We call the semigroup **A** *right quasi-trivial* iff $|\mathbf{A}| \geqq 3$ and there is a decomposition of $\mathbf{A}: \mathbf{A} = \mathbf{A}_{1R} \cup \mathbf{A}_{2R}$, $\mathbf{A}_{1R} \cap \mathbf{A}_{2R} = \emptyset$, for which there is a function $f_{R\mathbf{A}}: \mathbf{A}_{2R} \to \mathbf{A}_{2R}$ (in case $A_{2R} \neq \emptyset$) and $f_{R\mathbf{A}} \upharpoonright \mathscr{R}(f_{R\mathbf{A}}) = 1_{\mathscr{R}(f_{R\mathbf{A}})}$[1] and

$$(\forall x \in \mathbf{A})(\forall y \in \mathbf{A}) \left[xy = \begin{cases} x, & \text{if } \ y \in \mathbf{A}_{1R} \\ f_{R\mathbf{A}}(y), & \text{if } \ y \in \mathbf{A}_{2R} \end{cases} \right].$$

We analogously interpret the *left quasi-trivial* property. We can refer to both of right and left quasi-triviality by saying simply *quasi-trivial.* We call the semigroup **A** *strongly quasi-trivial,* iff the structure of **A** is one of the following three alternatives

$$\begin{aligned} &\text{(i)} \quad (\forall x, y)[x, y \in \mathbf{A} \Rightarrow xy = x], \\ &\text{(ii)} \quad (\forall x, y)[x, y \in \mathbf{A} \Rightarrow xy = y], \\ &\text{(iii)} \quad (\exists c \in \mathbf{A})(\forall x, y)[x, y \in \mathbf{A} \Rightarrow xy = c]. \end{aligned} \tag{2.1}$$

Obviously, if **A** is strongly quasi-trivial, then it is quasi-trivial also, but the converse is not true. Further, it can easily be checked that the quasi-trivial structure is associative.

As a characterization of quasi-trivial and strongly quasi-trivial semigroups, we prove the following theorem, which is a partial reverse of Theorem 1'.

Theorem 2. If **A** is a semigroup with $|\mathbf{A}| \geqq 3$, $r_0 \subset \mathbf{A}^2$ and $\mathbf{H} \overset{\text{def}}{=} \mathbf{M} \cap \pi r_0$, where

$$\begin{aligned} &\text{(i)} \quad \mathbf{M} = \mathscr{C}_{\Omega R} \mathbf{A}, \\ &\text{(ii)} \quad \mathbf{M} = \mathscr{C}_{\Omega L} \mathbf{A}, \\ &\text{(iii)} \quad \mathbf{M} = \mathscr{C}_{\Omega} \mathbf{A}, \end{aligned}$$

then the needful and sufficient condition of

$$(\forall r_0)[((r_0 \in \mathscr{S}\mathscr{R}\mathbf{A}) \wedge (1.10)) \Rightarrow r_0 \in \mathscr{T}\mathbf{A}] \tag{2.2}$$

[1] As usual, $\mathscr{D}$ and $\mathscr{R}$ stand for "domain" and "range" respectively. The symbol "$\upharpoonright$" is used to denote the restriction of functions.

is that

$$\begin{array}{ll}\text{(i)} & \mathbf{A} \text{ is right quasi-trivial,}\\ \text{(ii)} & \mathbf{A} \text{ is left quasi-trivial,}\\ \text{(iii)} & \mathbf{A} \text{ is strongly quasi-trivial,}\end{array} \tag{2.3}$$

respectively.

Remark. If in (2.2) we change "$\Rightarrow r_0 \in \mathcal{T}\mathbf{A}$" into "$\Rightarrow r_0 \in \mathcal{C}\mathbf{A}$", then (2.2) remains the same.

Proof. First of all, transform (2.2) into an equivalent form

$$(\forall r_0)\big[((r_0 \in \mathcal{S}\mathcal{R}\mathbf{A}) \wedge (r_0 \notin \mathcal{T}\mathbf{A})) \Rightarrow \neg(1.10)\big]. \tag{2.4}$$

If (2.4) is true, then it must hold for every r_0 of the form

$$\begin{gathered} r_0' \overset{\text{def}}{=} 1_{\mathbf{A}} \cup \{\langle a, b\rangle, \langle b, a\rangle, \langle a, c\rangle, \langle c, a\rangle\},\\ a, b, c \in \mathbf{A}, \qquad a \neq b \neq c \neq a. \end{gathered} \tag{2.5}$$

(Evidently, for any such r_0', $r_0' \in \mathcal{S}\mathcal{R}\mathbf{A}$ and $r_0' \notin \mathcal{T}\mathbf{A}$.)

(i) $\mathbf{M} = \mathcal{C}_{\Omega R}\mathbf{A}$. If one of the two equivalence relations

$$r_b \overset{\text{def}}{=} 1_{\mathbf{A}} \cup \{\langle a, b\rangle, \langle b, a\rangle\} (\subset r_0')$$

and

$$r_c \overset{\text{def}}{=} 1_{\mathbf{A}} \cup \{\langle a, c\rangle, \langle c, a\rangle\} (\subset r_0')$$

is not a right congruence relation, then (2.4) does not hold for $r_0 = r_0'$. (Because if e.g. $r_b \notin \mathcal{C}_{\Omega R}\mathbf{A}$ and $r_c \in \mathcal{C}_{\Omega R}\mathbf{A}$, then taking $r_g = r_c$, (1.10) will hold; and if also $r_c \notin \mathcal{C}_{\Omega R}\mathbf{A}$, then $r_g = 1_{\mathbf{A}}$ will satisfy (1.10).) On the other hand, if $r_b, r_c \in \mathcal{C}_{\Omega R}\mathbf{A}$, then $\sup(\mathbf{H})$, by virtue of its belonging to $\mathcal{T}\mathbf{A}$, must contain $\langle b, c\rangle$ (as $\langle b, a\rangle, \langle a, c\rangle \in r_0'$) therefore in this case $\sup(\mathbf{H}) \in \mathbf{H}$, i.e. (2.4) holds. This argument is valid for any r_0' of the type (2.5), so an equivalent transcription of (2.4) is the following

$$(\forall a, b)[(a, b \in A) \Rightarrow (1_{\mathbf{A}} \cup \{\langle a, b\rangle, \langle b, a\rangle\}) \in \mathcal{C}_{\Omega R}\mathbf{A}]. \tag{2.6}$$

Using criterion (1.2), (2.6) is further equivalent to

$$(\forall a, b, x)\big[(a, b, x \in A) \Rightarrow \big(ax = bx \vee (\{ax, bx\} \subset \{a, b\})\big)\big]. \tag{2.7}$$

Now we shall deduce (2.3) (i) from (2.7) (the converse is obvious: if $\mathbf{A}$ is right quasi-trivial, then (2.7) holds). Indeed, define the subset $\mathbf{A}_{1R}$ of $\mathbf{A}$ so

$$\mathbf{A}_{1R} \overset{\text{def}}{=} \{x \mid x \in \mathbf{A} \wedge (\forall y)[y \in \mathbf{A} \Rightarrow yx = y]\} \tag{2.8}$$

(obviously, $\mathbf{A}_{1R}$ may be empty), and let

$$\mathbf{A}_{2R} \overset{\text{def}}{=} \mathbf{A} - \mathbf{A}_{1R}. \tag{2.9}$$

Fix an arbitrary

$$x \in \mathbf{A}_{2R} \quad (\text{if } \mathbf{A}_{2R} \neq \emptyset).$$

By the definition of $\mathbf{A}_{2R}$, there is a $y \in \mathbf{A}$, for which

$$yx \neq y, \quad \text{say} \quad yx = z. \tag{2.10}$$

Then because of $|\mathbf{A}|\geqq 3$, there is a $u\in\mathbf{A}$, $u\neq y$, $u\neq z$. According to (2.7)

$$(z=)yx=ux\vee(\{yx,ux\}\subset\{y,u\}).$$

As $z\neq y$ (see (2.10)) and, by its choosing, $z\neq u$, $yx=z\notin\{y,u\}$, so

$$(\forall u)[(u\neq y\wedge u\neq z)\Rightarrow ux=z]. \tag{2.11}$$

Let us now examine zx. On the basis of (2.7), if $u\neq y$ and $u\neq z$

$$(zx=yx\vee(\{zx,yx\}\subset\{z,y\}))\wedge(zx=ux\vee(\{zx,ux\}\subset\{z,u\}). \tag{2.12}$$

As $yx=ux=z$ — from (2.10) and (2.11) —, (2.12) is not other than

$$zx=z\vee(\{zx,z\}\subset(\{z,y\}\cap\{z,u\})(=\{z\}))$$

i.e.

$$zx=z.$$

Summing up, if $x\in\mathbf{A}_{2R}$, then the value of wx does not depend on w

$$(\exists f_{\mathbf{RA}}:\mathbf{A}_{2R}\to\mathbf{A})(\forall w,x)[(w\in\mathbf{A}\wedge x\in\mathbf{A}_{2R})\Rightarrow wx=f_{\mathbf{RA}}(x)]. \tag{2.13}$$

Taking now into consideration that the structure of $\mathbf{A}$ is associative; if $x\in\mathbf{A}_{2R}$ and $w,s\in\mathbf{A}$, then

$$f_{\mathbf{RA}}(x)=(ws)x=w(sx)=wf_{\mathbf{RA}}(x),\quad wf_{\mathbf{RA}}(x)=f_{\mathbf{RA}}(x),$$

independently of w, i.e. $f_{\mathbf{RA}}(x)\in\mathbf{A}_{2R}$ and

$$f_{\mathbf{RA}}(f_{\mathbf{RA}}(x))=wf_{\mathbf{RA}}(x)=f_{\mathbf{RA}}(x),$$

from which we conclude, that in (2.13)

$$(\mathscr{R}(f_{\mathbf{RA}})\subset\mathbf{A}_{2R})\wedge(f_{\mathbf{RA}}\restriction\mathscr{R}(f_{\mathbf{RA}})=1_{\mathscr{R}(f_{\mathbf{RA}})}) \tag{2.14}$$

i.e. $\mathbf{A}$ is right quasi-trivial.

(ii) $\mathbf{M}=\mathscr{C}_{\Omega L}\mathbf{A}$. The argument is analogous to that of case (i).

(iii) $\mathbf{M}=\mathscr{C}_{\Omega}\mathbf{A}$.The left counterpart of (2.7) being

$$(\forall a,b,x)[a,b,x\in\mathbf{A}\Rightarrow(xa=xb\vee(\{xa,xb\}\subset\{a,b\}))], \tag{2.15}$$

we get in a similar way as in case (i), that in case (iii) — using condition (1.4) among others — (2.2) is equivalent to (2.7)$\wedge$(2.15), i.e.

$$(\forall a,b,x)[a,b,x\in\mathbf{A}\Rightarrow((ax=bx\vee(\{ax,bx\}\subset\{a,b\}))\wedge$$
$$\wedge(xa=xb\vee(\{xa,xb\}\subset\{a,b\})))]. \tag{2.16}$$

There are two distinct (disjoint) subcases of case (iii)

$$(\alpha)\quad(\forall a,b)[(a,b\in\mathbf{A}\wedge a\neq b)\Rightarrow ab\subset\{a,b\}],$$

$$(\beta)\quad(\exists a,b)[a,b\in\mathbf{A}\wedge a\neq b\wedge ab\notin\{a,b\}]\quad(\text{i.e.}\ \neg(\alpha)).$$

(α) Fix a, b for which, say, let

$$ab=a\quad(a,b\in\mathbf{A},\ a\neq b). \tag{2.17}$$

If $c \notin \{a, b\}$, then as now case (α) is valid,

$$cb \in \{c, b\} \tag{2.18}$$

and on the basis of (2.7) and (2.17)

$$cb \in \{a, c\}. \tag{2.19}$$

As $a \neq b$, from (2.18) and (2.19) follows

$$cb = c. \tag{2.20}$$

Further (2.7), (2.17) and (2.20) give that

$$bb \in \{a, b\} \wedge bb \in \{c, b\}$$

i.e. because of $a \neq c$;

$$bb = b.$$

From the above we can conclude that

$$(\forall x)[x \in \mathbf{A} \Rightarrow xb = x]. \tag{2.21}$$

Now let

$$y \neq b \neq x \neq y. \tag{2.22}$$

On the basis of (2.15), (2.21) and (2.22)

$$(xy = xb \vee (\{xy, xb\} \subset \{y, b\})) \wedge (xb = x \notin \{y, b\})$$

i.e.

$$xy = x \quad (x \neq y). \tag{2.23}$$

From (2.23), quite in a similar way as starting from (2.17), we can deduce that (2.21) is true for y in place of b. And finally, as $y(\neq b)$ was arbitrary, we get

$$(\forall u, w)[u, w \in \mathbf{A} \Rightarrow uw = u]. \tag{2.24}$$

If at the beginning in (2.17) we alter $ab = a$ into $ab = b$, then the final result will be

$$(\forall u, w)[u, w \in A \Rightarrow uw = w]. \tag{2.25}$$

(β) We can start with

$$(a, b, c \in A) \wedge (a \neq b \neq c \neq a) \wedge ab = c. \tag{2.26}$$

From (2.15) and (2.26)

$$(aa = ab \vee (\{aa, ab\} \subset \{a, b\})) \wedge (ab = c \notin \{a, b\}),$$

from which

$$aa = c. \tag{2.27}$$

Similarly, by means of (2.7) and (2.26), we get

$$bb = c.$$

To determine ac, using (2.15), (2.26) and (2.27), we can write

$$\big(ac = ab \vee (\{ac, ab\} \subset \{c, b\})\big) \wedge \big(ac = aa \vee (\{ac, aa\} \subset \{c, a\})\big) \wedge$$
$$\wedge ab = c \wedge aa = c \wedge (a \neq b \neq c \neq a)$$

i.e.

$$ac = c. \tag{2.28}$$

Likewise

$$bc = ca = cb = c.$$

For ba, using (2.7), (2.26) and (2.27)

$$\big(ba = aa \vee (\{ba, aa\} \subset \{b, a\})\big) \wedge (aa = c \notin \{b, a\})$$

and from this

$$ba = c. \tag{2.29}$$

At last, starting from (2.15), (2.26) and $cb = ca = c$, in the same way as leading to (2.28), we have

$$cc = c. \tag{2.30}$$

Summing up (2.26) to (2.30)

$$(\forall x, y)[(x, y \in \{a, b, c\}) \Rightarrow xy = c]. \tag{2.31}$$

If $z \in \mathbf{A} - \{a, b, c\}$, then in the same fashion as in (2.29) we have

$$zb = za = c.$$

Analogously to deducing (2.31), we conclude, that

$$(\forall u, w)[(u, w \in \{z, b, c\}) \Rightarrow uw = c] \tag{2.32}$$

and

$$(\forall u, w)[(u, w \in \{z, a, c\}) \Rightarrow uw = c]$$

and similarly, if $w \in \mathbf{A} - \{z, a, b, c\}$, then

$$wz = zw = ww = c. \tag{2.33}$$

Summarizing (2.31), (2.32) and (2.33)

$$(\forall x, y)[x, y \in \mathbf{A} \Rightarrow xy = c]. \tag{2.34}$$

As (2.24), (2.25) and (2.34) correspond to (2.1) (i), (2.1) (ii) and (2.1) (iii) respectively, we are ready.

Remark. In the proof of part (iii) and up to (2.13) in that of part (i) (analogous statement holds true of part (ii)) we did not make use of associativity.

In the following we give a second proof for part (iii), on the basis of (2.13) and its left counterpart, without making use of property (2.14) and the left counterpart of it i.e. again not taking into account associativity.

Second proof for part (iii) *of Theorem 2.* Let

$$\mathbf{A} = \mathbf{A}_{1R} \cup \mathbf{A}_{2R} (\mathbf{A}_{1R} \cap \mathbf{A}_{2R} = \emptyset)$$

the decomposition of A, defined by (2.8) and (2.9) (this decomposition exists — and is unique — for any semigroup A), and let

$$\mathbf{A} = \mathbf{A}_{1L} \cup \mathbf{A}_{2L} (\mathbf{A}_{1L} \cap \mathbf{A}_{2L} = \emptyset)$$

be the left counterpart of the former decomposition.

If one of $\mathbf{A}_{1R}$ and $\mathbf{A}_{1L}$ is A itself, we are ready, evidently having (2.1) (i) or (2.1) (ii) respectively.

If

$$\mathbf{A}_{1R} \neq \emptyset \wedge \mathbf{A}_{1L} \neq \emptyset \tag{2.35}$$

then

$$(\forall x, y)[(x \in \mathbf{A}_{1L} \wedge y \in \mathbf{A}_{1R}) \Rightarrow y = xy = x]$$

i.e.

$$|\mathbf{A}_{1R}| = |\mathbf{A}_{1L}| = 1,$$

$$\{e\} \stackrel{\text{def}}{=} \mathbf{A}_{1R} = \mathbf{A}_{1L}$$

and consequently

$$\mathbf{A}_{2R} = \mathbf{A}_{2L} = \mathbf{A} - \{e\}$$

(e is the — unique — identity element of A).

Furthermore

$$(\forall x, y)[(x \in \mathbf{A}_{2L} \wedge y \in \mathbf{A}_{2R}) \Rightarrow xy = f_{L\mathbf{A}}(x) = f_{R\mathbf{A}}(y)]$$

i.e.

$$f_{R\mathbf{A}} = f_{L\mathbf{A}} = \text{constant}. \tag{2.36}$$

Being $|\mathbf{A}| \geqq 3$, $|A_{2R}| (= |A_{2L}|) \geqq 2$, so there are $x, y \in \mathbf{A}_{2R}$, $x \neq y$, for which on one hand $ex = f_{R\mathbf{A}}(x) = f_{R\mathbf{A}}(y) = ey$, while on the other hand $ex = x \neq y = ey$, which is a contradiction, and therefore (2.35) is impossible. Thus, let e.g.

$$\mathbf{A}_{2R} = \mathbf{A} \wedge \mathbf{A}_{2L} \neq \emptyset$$

(the symmetric counterpart is quite analogous).

From this immediately follows (2.36) with $\mathbf{A}_{2R} = \mathbf{A}_{2L} = \mathbf{A}$ i.e. (2.1) (iii).

To close Part 2 of our paper, we formulate the following

Theorem 2'. If A is a semigroup and $|\mathbf{A}| \geqq 3$, then the following three statements are equivalent

$$\text{(a)} \begin{cases} \text{(i)} & \mathbf{M} = \mathscr{C}_{\Omega R}\mathbf{A}, \\ \text{(ii)} & \mathbf{M} = \mathscr{C}_{\Omega L}\mathbf{A}, \\ \text{(iii)} & \mathbf{M} = \mathscr{C}_{\Omega}\mathbf{A}, \end{cases} r_0 \in \mathbf{A}^2, \mathbf{H} \stackrel{\text{def}}{=} \mathbf{M} \cap \pi r_0 \text{ and } (\forall r_0)[(r_0 \in \mathscr{S}\mathscr{R}\mathbf{A} \wedge (1.10)) \Rightarrow r_0 \in \mathscr{T}\mathbf{A}]$$

$$\text{(b)} \begin{cases} \text{(i)} & \mathbf{A} \text{ is right quasi-trivial,} \\ \text{(ii)} & \mathbf{A} \text{ is left quasi-trivial,} \\ \text{(iii)} & \mathbf{A} \text{ is strongly quasi-trivial,} \end{cases}$$

$$\text{(c)} \begin{cases} \text{(i)} & \mathscr{C}\mathbf{A} = \mathscr{C}_{\Omega R}\mathbf{A}, \\ \text{(ii)} & \mathscr{C}\mathbf{A} = \mathscr{C}_{\Omega L}\mathbf{A}, \\ \text{(iii)} & \mathscr{C}\mathbf{A} = \mathscr{C}_{\Omega}\mathbf{A} \end{cases}$$

(i.e. $(a)(x) \Leftrightarrow (b)(x) \Leftrightarrow (c)(x)$ for $x = $ i, ii, iii).

Proof. It follows from the proof of Theorem 2 ((*a*) ⇔ (*b*) is Theorem 2 itself).

3. Some questions of the semigroups and the simulation of automata

In this part of our paper the focus will be on automata, and we shall take known several widely accepted notions and notations of automata theory.

The set of all *initially connected Moore automata*, having the same input alphabet X and output alphabet Y, can be partitioned into equivalence classes, regarding two automata equivalent iff they induce the same *automaton mapping*

$$\bar{f}: X^* \to (Y^* - \{\Lambda\}) \tag{3.1}$$

with the following property

$$(\forall u \in X^*)(\forall w \in X)(\exists z \in Y)[\bar{f}(uw) = \bar{f}(u)z] \wedge f(\Lambda) \in Y.$$

From this easily follows that

$$(\forall u \in X^*)[\lg(\bar{f}(u)) = \lg(z) + 1].$$

As is known, the functions $\bar{f}$ defined in (3.1) are in one-to-one correspondence with the functions

$$f: X^* \to Y \tag{3.2}$$

(if for all $u \in X^*$, $f(u)$ is the last symbol of $\bar{f}(u)$).

In the following — unless otherwise stated — by the word *automaton* we shall always mean a function f of the type (3.2) and the (not necessarily finite) non-void sets X and Y we shall take given.

As a generalization of right and left compatible partitions of the semigroup **A**, we formulate the following

Definition 2. If $r \in \mathscr{C}_{\Omega R}\mathbf{A}$, the partition p is a *right compatible partition on (the set of classes)* $\mathbf{A}/r$ iff

$$(\forall x)(\forall Z_1, Z_2, W_1, W_2)[(x \in \mathbf{A} \wedge (Z_1, Z_2, W_1, W_2 \in \mathbf{A}/r) \wedge$$
$$\wedge (Z_1\{x\} \subset W_1) \wedge (Z_2\{x\} \subset W_2) \wedge \langle Z_1, Z_2 \rangle \in p) \Rightarrow \langle W_1, W_2 \rangle \in p].^2$$

The meaning of *left compatible partition on a partition* is analogous.

Remark. "Compatible partition on a compatible partition r" is an ordinary compatible partition on the factor semigroup $\mathbf{A}/r$.

Definition 3. Given a set **Z** and $r \in \mathscr{C}\mathbf{Z}$, we call the function

$$\operatorname{nat} r: \mathbf{Z} \to \mathbf{Z}/r \tag{3.3}$$

which has the following property

$$(\forall x)[x \in Z \Rightarrow x \in (\operatorname{nat} r)(x)],$$

the natural mapping belonging to the partition r.

[2] If a binary operation, written as multiplication is defined on a set **S**, and $\mathbf{T}, \mathbf{U} \subset \mathbf{S}$, then $T \cdot U = TU \overset{\text{def}}{=} \{tu \mid t \in \mathbf{T} \wedge u \in \mathbf{U}\}$.

The composition (consecutive application) of two functions f and g we write in the form

$$g \circ f,\ (g \circ f)(x) \stackrel{\text{def}}{=} f(g(x)). \tag{3.4}$$

Definition 4. Given a semigroup **A**, $r \in \mathscr{C}_{\Omega R}\mathbf{A}$ and the set Y, we call the function $k:\mathbf{A}/r \to Y$ *right compatible-free* (in short **RCF**) iff

$$(\forall q, s)[(k = (\text{nat}\, q) \circ s) \Rightarrow q = 1_{\mathbf{A}/r}]$$

where q is a right compatible partition on $\mathbf{A}/r$ and $s:(\mathbf{A}/r)/q \to Y$ (s is uniquely defined by q) (see Definition 2, Definition 3, (3.3) and (3.4)).

The meaning of *left compatible-free* (**LCF**) is analogous. Iff above $r \in \mathscr{C}_\Omega\mathbf{A}$ and $q \in \mathscr{C}_\Omega(\mathbf{A}/r)$, we call the function k *homomorph-free* (in short **HF**).

For any function f, we define the following equivalence relation

$$f^0 \stackrel{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in \mathscr{D}(f) \wedge f(x) = f(y)\}. \tag{3.5}$$

Now we are ready to prove the following

Statement 1. Given a function $f:\mathbf{A} \to Y$ where **A** is a semigroup, the decomposition

$$f = (\text{nat}\, r) \circ k \quad (\text{where } r \in \mathscr{C}\mathbf{A})$$

(exists and) is unique if at least one of the following conditions holds

$$\text{(i)}\quad r \in \mathscr{C}_{\Omega R}\mathbf{A} \quad \text{and} \quad k \quad \text{is} \quad \mathbf{RCF},$$

$$\text{(ii)}\quad r \in \mathscr{C}_{\Omega L}\mathbf{A} \quad \text{and} \quad k \quad \text{is} \quad \mathbf{LCF},$$

$$\text{(iii)}\quad r \in \mathscr{C}_{\Omega}\mathbf{A} \quad \text{and} \quad k \quad \text{is} \quad \mathbf{HF}$$

(see (3.3), (3.4) and Definition 4).

Proof

(i) Let r be the greatest ("roughest") right compatible refinement of f^0 (see (3.5)) which exists (and is unique) on the basis of Theorem 1′. If $f = (\text{nat}\, \bar{r}) \circ k'$ is another decomposition, for which $\bar{r} \neq r$, then according to Theorem 1′, $\bar{r} \subset r$ and there is a right compatible partition $q \neq 1_{\mathbf{A}/r}$ on $\mathbf{A}/r$ (see Definition 2), for which $k' = (\text{nat}\, q) \circ k''$ (for some k'') i.e. k' is not **RCF**.

(ii) Quite analogous to case (i).

(iii) The argument needs only slight and obvious modifications on that of case (i).

Definition 5. We supply r and k (which we have introduced in Statement 1) with subscripts R, L and C according to cases (i), (ii) and (iii) in Statement 1 respectively and write

$$\text{(i)}\quad f = (\text{nat}\, r_{Rf}) \circ k_{Rf},\ \text{nat}\, r_{Rf} \stackrel{\text{def}}{=} R_f,$$

$$\text{(ii)}\quad f = (\text{nat}\, r_{Lf}) \circ k_{Lf},\ \text{nat}\, r_{Lf} \stackrel{\text{def}}{=} L_f,$$

$$\text{(iii)}\quad f = (\text{nat}\, r_{Cf}) \circ k_{Cf},\ \text{nat}\, r_{Cf} \stackrel{\text{def}}{=} C_f.$$

We call R_f, L_f and C_f the *greatest right compatible*, the *greatest left compatible* and the *greatest homomorphic component of (or contained in) f*, respectively, while r_{Cf} we call the *congruence relation of f*.

Remark. As a consequence of Theorem 1′, for any $f(:A \to Y)$

$$r_{Cf} \subset r_{Rf} \subset f^0 \quad \text{and} \quad r_{Cf} \subset r_{Lf} \subset f^0. \tag{3.6}$$

Corollary of Statement 1, part (i). For any equivalence class K of initially connected Moore automata, the elements of which induce the same automaton mapping $\bar{f}$ (see (3.1)) there is a (unique) automaton $\bar{A}$ in K, which is the state-homomorphic image of all members in K.

Proof. It easily follows from (3.1), (3.2) and part (i) of Statement 1) (cf. [3, Chapter 9], [4, 4. §], [6, § 1.11] and [7, § 3.1]).

Definition 6. For an automaton f, the factor-semigroup

$$\mathbf{S}_f \overset{\text{def}}{=} X^*/r_{Cf}$$

we call the *semigroup (characteristic semigroup) of f*.[3]

The usual way of defining the semigroups of automata is found in the following

Definition 7. If $M = \langle Q, X, \delta \rangle$ is an automaton without output (with state-set Q, input alphabet X and next-state function δ), the semigroup of M is

$$\mathbf{S}(M) \overset{\text{def}}{=} X^*/\varrho(M), \tag{3.7}$$

where $\varrho(M)$ is the congruence relation of M and

$$\varrho(M) \overset{\text{def}}{=} \{\langle x, y\rangle | x, y \in X^* \wedge (\forall q)[q \in Q \Rightarrow qx = qy]\}. \tag{3.8}$$

(It can easily be checked using (1.4) that indeed $\varrho(M) \in \mathscr{C}_\Omega X^*$.)

Remarks

(a) On the basis of Theorem 1′ (see (1.14) and the end of the proof of Theorem 1′, and (1.18) in the Remark at the end of Part 1) using the notations of Definition 5

$$\begin{aligned} r_{Rf} &= \{\langle x, y\rangle | x, y \in X^* \wedge (\forall a)[a \in X^* \Rightarrow f(xa) = f(ya)]\}, \\ r_{Lf} &= \{\langle x, y\rangle | x, y \in X^* \wedge (\forall a)[a \in X^* \Rightarrow f(ax) = f(ay)]\}, \\ r_{Cf} &= \{\langle x, y\rangle | x, y \in X^* \wedge (\forall a, b)[a, b \in X^* \wedge f(axb) = f(ayb)]\}. \end{aligned} \tag{3.9}$$

(b) (3.9) is a more explicit formulation of (3.6), and further we can write

$$\begin{aligned} r_{Cf} &= \{\langle x, y\rangle | \langle x, y\rangle \in r_{Rf} \wedge (\forall a)[a \in X^* \Rightarrow \langle ax, ay\rangle \in r_{Rf}]\}, \\ r_{Cf} &= \{\langle x, y\rangle | \langle x, y\rangle \in r_{Lf} \wedge (\forall a)[a \in X^* \Rightarrow \langle xa, ya\rangle \in r_{Lf}]\}. \end{aligned} \tag{3.10}$$

[3] See (3.2) and our agreement following it; and Def. 5.

(c) From the Corollary of Statement 1, Definitions 6 and 7, and equations (3.9) and (3.10), easily follows that (if A corresponds to f)

$$\mathbf{S}(\bar{A}) = \mathbf{S}_f \tag{3.11}$$

and

$$\varrho(\bar{A}) = r_{Cf}. \tag{3.12}$$

(d) If the state-set Q (of $\bar{A}$) is finite, then we can even deduce from equations (3.7) to (3.12) that $\mathbf{S}(\bar{A})$ is finite too. More generally, in the language of semigroups

Statement 2. If $\mathbf{A}$ is a semigroup, $r \in \mathscr{C}_{\Omega R}\mathbf{A}$ and $|\mathbf{A}/r| < \infty$, then there exists an $r' \subset r$ and $r' \in \mathscr{C}_{\Omega}\mathbf{A}$, for which $|\mathbf{A}/r'| < \infty$. (Analogous statement is true of $r \in \mathscr{C}_{\Omega L}\mathbf{A}$.)

Proof. Let (like (3.10))

$$r' \stackrel{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in r \wedge (\forall a)[a \in \mathbf{A} \Rightarrow \langle ax, ay\rangle \in r]\}, \tag{3.13}$$

from which we can see at once using (1.4) that $r' \in \mathscr{C}_{\Omega}\mathbf{A}$ (and evidently $r' \subset r$). To prove the finiteness of $\mathbf{A}/r'$, we rewrite (3.13) in the following way

$$r' \stackrel{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in r \wedge (\forall a, b)[\langle a, b\rangle \in r \Rightarrow \langle ax, by\rangle \in r]\}. \tag{3.14}$$

(3.14) $\Rightarrow$ (3.13) is obvious. (3.14) can be obtained from (3.13) by taking into account that $r \in \mathscr{C}_{\Omega R}\mathbf{A}$, so $\langle a, b\rangle \in r \Rightarrow \langle ay, by\rangle \in r$ and being $r \in \mathscr{T}\mathbf{A}$, $(\langle ax, ay\rangle \in r \wedge \langle ay, by\rangle \in r) \Rightarrow$ $\Rightarrow \langle ax, by\rangle \in r$. Now, with each element $x \in \mathbf{A}$, we can associate a function

$$(\varphi_x : \mathbf{A}/r \to \mathbf{A}/r) \wedge (\forall C)[C \in \mathbf{A}/r \Rightarrow C\{x\} \subset \varphi_x(C)] \tag{3.15}$$

(this was hinted by F. Gécseg). With the functions of (3.15), an equivalent form of (3.14) is

$$r' \stackrel{\text{def}}{=} \{\langle x, y\rangle | \langle x, y\rangle \in r \wedge \varphi_x = \varphi_y\}.$$

By the definition of the functions φ_x (see (3.15))

$$\{\varphi_x | x \in \mathbf{A}\} \subset \{\varphi | \varphi : \mathbf{A}/r \to \mathbf{A}/r\} \stackrel{\text{def}}{=} F,$$

so r' can be obtained from r by splitting each class in $\mathbf{A}/r$ into not more than $|F|$ subclasses and therefore

$$|\mathbf{A}/r'| \leqq |\mathbf{A}/r| \cdot |F|. \tag{3.16}$$

Taking

$$|\mathbf{A}/r| \stackrel{\text{def}}{=} m < \infty,$$

then $|F| = m^m$ and from (3.16) we get

$$|\mathbf{A}/r'| \leqq m \cdot m^m = m^{m+1} < \infty. \tag{3.17}$$

Remarks

(a) (3.17) is also valid for m's of any cardinality, but only $m < \infty$ has practical significance.

(b) Several authors declare that "any semigroup is isomorphic to the semigroup of an automaton" (in the sense of Definition 7), but this is wrong: we must say "any *monoid*" instead of "any semigroup" and so the statement will already be true. This easily follows from (3.7) and (3.8), or more generally from the simple fact: every factor-semigroup of a monoid is again a monoid. The mistake in the "proof" of the former defective assertion, which uses the so-called *semigroup machine*

$$M_{\mathbf{A}} \stackrel{\text{def}}{=} \langle \mathbf{A}, \mathbf{A}, \delta \rangle, \quad (\forall s_1, s_2)[\delta(s_1, s_2) \stackrel{\text{def}}{=} s_1 s_2]$$

(where $\mathbf{A}$ is any semigroup) is that even if $\mathbf{A}$ has no identity element, $\mathbf{A}^*$ does have, when applying (3.8) to $M_{\mathbf{A}}$. We cannot even be sure of

$$\mathrm{S}(M_{\mathbf{A}}) = \mathbf{A}_{\mathbf{I}} \tag{3.18}$$

(for any semigroup $\mathbf{A}$, $\mathbf{A}_{\mathbf{I}} \stackrel{\text{def}}{=} \mathbf{A}$, if $\mathbf{A}$ is a monoid and if not, then $\mathbf{A}_{\mathbf{I}} \stackrel{\text{def}}{=}$ "the monoid which we get by attaching to $\mathbf{A}$ an external unit element"), because if $\mathbf{A}$ is not a monoid, then it can well have *right uniform* elements. The notion of right uniform elements we introduce in the following

Definition 8. In the semigroup $\mathbf{A}$, the elements c and c' are said *right uniform* iff

$$(\forall x)[x \in A \Rightarrow xc = xc']$$

and the relation of right uniformity in the semigroup $\mathbf{A}$ we denote with $u_R(\mathbf{A})$.

The meaning of *left uniformity* is analogous and the notation for the corresponding relation is $u_L(\mathbf{A})$. At last, the relation $u(\mathbf{A}) \stackrel{\text{def}}{=} u_R(\mathbf{A}) \cap u_L(\mathbf{A})$ we call the relation of *uniformity* on $\mathbf{A}$.

Remark. Evidently $u_R(\mathbf{A})$, $u_L(\mathbf{A})$, $u(\mathbf{A}) \in \mathscr{C}_\Omega \mathbf{A}$. As an example, suppose $\mathbf{A}$ is right (left) quasi-trivial (see Def. 1), then $u_R(\mathbf{A}) = \mathbf{A}^2_{1R} \cup f^0_{R\mathbf{A}}$ $\left(u_L(\mathbf{A}) = \mathbf{A}^2_{1L} \cup f^0_{L\mathbf{A}}\right)$ (see (3.5)). In this case $u(\mathbf{A}) \neq 1_{\mathbf{A}}$ iff. $\mathbf{A}_{1R} = \emptyset$ $(\mathbf{A}_{1L} = \emptyset)$ and $f^0_{R\mathbf{A}} = 1_{\mathbf{A}}$ $(f^0_{L\mathbf{A}} = 1_{\mathbf{A}})$, so there exist $\mathbf{A}$'s for which $u_R(\mathbf{A}) \neq u(\mathbf{A})$ $\left(u_L(\mathbf{A}) = u(\mathbf{A})\right)$.

A trivial example for uniform elements is the case when A is strongly quasi-trivial and (2.1) (iii) is valid (Def. 1). A less trivial example is the following: take an arbitrary semigroup $\mathbf{A}_0$ and choose a $c \in \mathbf{A}_0$ and let $c' \notin \mathbf{A}_0$, $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{A}_0 \cup \{c'\}$. If we define the operations in $\mathbf{A}$ so

$$(\forall x, y \in \mathbf{A})[x, y \in \mathbf{A}_0 \Rightarrow (xy(\text{in } \mathbf{A}) = xy(\text{in } \mathbf{A}_0) \wedge c'x(\text{in } \mathbf{A}) =$$

$$= cx(\text{in } \mathbf{A}_0) \wedge xc'(\text{in } \mathbf{A}) = xc(\text{in } \mathbf{A}_0) \wedge c'c'(\text{in } \mathbf{A}) = cc(\text{in } \mathbf{A}_0))],$$

then $c \equiv c' \pmod{u(\mathbf{A})}$. Of course, by this method an unbounded number of uniform elements can be achieved. (If, furthermore, we randomly select some pairs $\langle x, y \rangle$ for which $xy = c$ (in $\mathbf{A}$) and change their result into c', then $\mathbf{A}$ will remain a semigroup and c' will play a more active role).

Now if $\mathbf{A}$ has right uniform elements, then (3.18) will not hold, because when forming $\mathrm{S}(M_{\mathbf{A}})$ according to (3.8) and (3.7), the right uniform elements of $\mathbf{A}$ will "coincide" in $\mathrm{S}(M_{\mathbf{A}})$. This can be expressed in the following

Fact. For any semigroup $\mathbf{A}$, $\mathrm{S}(M_{\mathbf{A}}) \cong (\mathbf{A}/u_R(\mathbf{A}))_{\mathbf{I}}$, and $\mathrm{S}(M_{\mathbf{A}}) \cong \mathbf{A}$ iff $\mathbf{A}$ is a monoid (see Def.'s 7 and 8).

Proof. Easy from Def.'s 7 and 8.

Now, let us come to the question of the simulation of automata by each other.

We say that the automaton f *can simulate* (in short: simulates) automaton f' (both f and f' correspond to (3.2)), iff there are suitable functions h and p, for which

$$f' = h \circ f \circ p, \tag{3.19}$$

where (3.19) we interpret in the sense of (3.4). Here

$$f: X^* \to Y \quad \text{and} \quad f': X_1^* \to Y.$$

A glance at (3.1) and (3.2) convinces us that in (3.19)

$$h: X^* \simeq X_1^*$$

("$\simeq$" and "$\cong$" are the usual symbols for denoting homorphic and isomorphic mappings respectively).

First we prove that the possibility of simulation depends essentially on the semigroups of the automata in question, and is independent of the input alphabet

Theorem 3. Let $f: X_f^* \to Y$ and $g: X_g^* \to Y$ two automata, $i: \mathbf{S}_f \cong \mathbf{S}_g$ and

$$k_{C_f} = i \circ k_{C_g}. \tag{3.20}$$

Then f and g can simulate each other.[4]

Proof. It is enough to prove, that g can simulate f. (In the following proof, the definitions, relations etc. mentioned in footnote 4, will be widely used without further explanation.)

Let

$$h_1: X_f \to X_g^* \tag{3.21}$$

be such that

$$(\forall x \in X_f)[h_1(x) \in (C_f \circ i)(x)]. \tag{3.22}$$

From (3.21) easily follows, that h_1 can be uniquely extended into a homomorphism

$$h: X_f^* \to X_g^*,$$

for which automatically $h(\Lambda) = \Lambda$ (otherwise the reader is likely to know the verification of the existence and uniqueness of h, from the theory of free semigroups).

As as consequence of (3.22), it can easily be seen that

$$(\forall w \in X_f^*)[h(w) \in (C_f \circ i)(w)]. \tag{3.23}$$

(It is usual also to require from h_1, that for every $x \in X_f$, $\lg(h_1(x))$ is the least possible, but this is not necessary for our purposes.)

(3.23) eans that

$$(\forall w \in X_f^*)[(h \circ C_g)(w) = (C_f \circ i)(w)],$$

i.e.

$$h \circ C_g = C_f \circ i.$$

[4] See Def.'s 5, 6, equation (3.19) and convention (3.4).

Multiplying this equation with equation

$$k_{Cg}\circ 1_Y = k_{Cg},$$

we get

$$h\circ(C_g\circ k_{Cg})\circ 1_Y = C_f\circ(i\circ k_{Cg})$$

and taking into account (3.20)

$$h\circ g\circ 1_Y = f,$$

i.e. g can simulate f.

*COMPUTING CENTER OF THE HUNGARIAN MINISTRY OF HEAVY INDUSTRIES, BUDAPEST

**RESEARCH GROUP ON MATHEMATICAL LOGIC AND THE THEORY OF AUTOMATA OF THE HUNGARIAN ACADEMY OF SCIENCES, SZEGED

Замечания о максимальных конгруенциях, автоматах и смежных темах

Статья состоит из трех частей.

В 1-ой части авторы занимаются следующим обобщением: для данной сверх некоторой полугруппы A эквивалентностной реляции однозначно существует уточнение по максимальной конгруенции, доказано, что вместо эквивалентности и для более обобщенных реляций однозначно существуют максимальные уточнения более общего типа, чем конгруенция.

Во 2-ой части показывается, что некоторая возможная инвертность результатов 1-ой части взаимнооднозначно соответствует определенной специальной операционной структуре полугруппы A.

В 3-ей части исследуются вопросы, связанные с полугруппами автоматов Мура и их симуляцией, исходя из эквивалентностных и конгруентных реляций, выходящих из трансформаций авт омата Мура, и используя результаты 1-ой части.

References

[1a] COHN, P. M., *Universal algebra*, Harper and Row, New York, Evanston and London, 1965.
[1b] KON, P., *Un'iversalnaya alghebra* (Russian edition), Mir, Moscow, 1968.
[2] SCHMIDT, E. T., *Kongruenzrelationen Algebraischer Strukturen*, VEB Deutscher Verlag der Wissenschaften, Berlin, GDR, 1969.
[3] HARRISON, M. A., *Introduction to switching and automata theory*, McGraw-Hill, New York, London, Sydney, 1965.
[4] ÁDÁM, A., Automata-leképezések, félcsoportok, automaták (in Hungarian, with English and Russian abstracts), *Matematikai Lapok*, Budapest, v. 19, 1968, pp. 327—343.
[5] ARBIB, M. A. (editor), *Algebraic theory of machines, languages and semigroups* (Collection of Papers), Academic Press, New York, London, 1968.
[6] GÉCSEG, F., I. PEÁK, *Algebraic theory of automata*, Akadémiai Kiadó, Budapest, Hungary, 1972.
[7] ARBIB, M. A., *Theories of abstract automata*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1969.

(Received June 5, 1972)

INDEX—TARTALOM

Felelős szerkesztő és kiadó: Kalmár László
A kézirat a nyomdába érkezett: 1972. december hó
Megjelenés: 1973. március hó
Példányszám: 1000. Terjedelem: 7,75 (A/5) ív
Készült monószedéssel, íves magasnyomással
az Msz 5601-90 és ez Msz 5602-55 szabvány szerint
72-5150—Szegedi Nyomda