PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série, tome 38 (52), 1985, pp. 69–82

EMBEDDING SEMIGROUPS IN GROUPS: A GEOMETRICAL APPROACH

Sava Krstić

Abstract. A way to visualize Mal'cev quasi-identities is presented. As a consequence an analogy, expressed in a geometric language, is found between Mal'cev and Lambek quasi-identities. These are known to be of a special form which is called stable here; it is proved that certain geometrically characterized sets of stable quasi-identities axiomatize the class of embeddable semigroups. The results of Mal'cev and Lambek are obtained as corollaries. The method of diagrams, borrowed from group theory, enabled us to give a unified treatment which seems to be conceptually simpler than those previously employed.

1. Introduction. In order to be embeddable in a group a semigroup has to satisfy the cancellation laws $xz = yz \Rightarrow x = y$ and $zx = zy \Rightarrow x = y$. That this is not sufficient was shown by Mal'cev [**9**] who also found the first set of conditions which are both necessary and sufficient [**10**]. Mal'cev's system contains an infinity of formulas generalizing the cancellation laws, each formula being a quasi-identity (*quid*, for short), i.e. of the form "a conjuction of identities implies an identity". In a subsequent paper [**11**] Mal'cev proved that no finite set of quids could serve the same purpose.

After some time another solution in the form of a different infinite set of quids was offered by Lambek [**7**]. A feature of Lambek's proof, which contrasts the linear arguments of Mal'cev, is the usage of polyhedra as geometric means of describing quids.

For obvious reasons all conditions for embeddability are satisfied by any group. In addition to this trivial common property there is a striking similarity between quids comprising Mal'cev's and Lambek's systems. Namely, every quid $\sigma_1 \wedge \cdots \wedge \sigma_n \Rightarrow \sigma_0$ occurring in either of them involves variables $x_1, \ldots, x_p, y_1, \ldots, y_q$ (for some $p, q \geq 1$) so that every $\sigma_m,\ ) \leq m \leq n$, is of the form $x_i y_j = x_k y_l$, $i \neq k, j \neq l$, and every x-variable as well as every y-variable occurs exactly twice in $\sigma_1 \wedge \cdots \wedge \sigma_n \Rightarrow \sigma_0$. The quids of this form will be called *stable*.

In the next section we associate a "diagram" with every stable quid and then in Section 3 show that the classes of Mal'cev and Lambek quids are distinguished

AMS Subject Classification (1980): Primary 20M10, 20F32.

in the class of all stable quids by imposing two analogous restrictions on the corresponding diagrams.

Starting from the simple fact that the class of embeddable semigroups is axiomatizable by quids (Section 4) we prove in Section 6 that each one of certain six geometrically characterized sets of stable quids axiomatizes the embeddable semigroups. Two of these sets are considerably smaller than the set of Mal'cev quids, another two are smaller than the set of Lambek quids, so the embedding theorems of Mal'cev and Lambek are obtained as corollaries. The technique we use is the representation by diagrams of the relation "u belongs to the normal closure of $\{u_1, \dots, u_n\}$ in a free group". It is described in Section 5.

A bibliographical note. Chapter 12 of Clifford and Preston's book [**3**] is devoted to the embedding problem of semigroups in groups. For a complete account, history, and references we refer the reader to this book. The simplest exposition of Mal'cev's proof, as revised by Cohn is tu be found in Section VII.3 of [**4**]. The efforts to visualize Mal'cev quids, originally defined in a rather complicated way, started with Tamari [**15**]. Radó [**13**] obtained a geometrical characterization similar to the one given by our Theorem 1. The main reference for geometrical methods in group theory is Lyndon and Schupp [**8**, Chapters III and V]. A recent application of diagrams in semigroup theory is given in Remmers [**14**].

2. Stable quids. In this section we describe a way to visualize stable quids. With every identity $\sigma = (x'y' = x''y'')$ we associate a closed (topological) disc $D(\sigma)$ the boundary of which is subdivided into four *edges* oriented and labelled as depicted on Figure 1. The four *vertices* (i.e. endpoints of edges) of $D(\sigma)$ will be called the *source, sink* and *switches*, according in an obvious way to the orientation of edges incident with them.

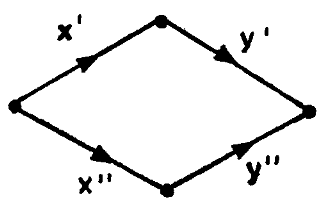

Fig. 1

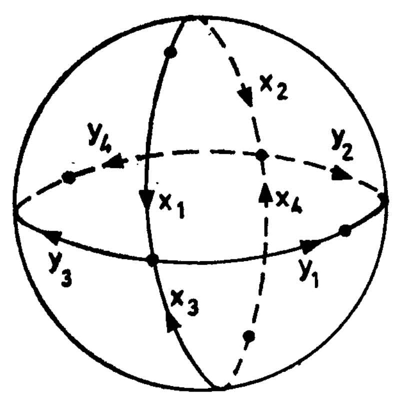

Fig. 2

Suppose now $\Sigma = (\sigma_1 \wedge \dots \sigma_n \Rightarrow \sigma_0)$ is a stable quid. If X is the disjoint union of $D(\sigma_0), D(\sigma_1), \dots, D(\sigma_n)$ then it is easy to see that the quotient space

$\widetilde{X}$ obtained from X by identifying all pairs of equally labelled oriented edges is a closed (not necessarily connected) surface. We will say that Σ is *spherical* whenever $\widetilde{X}$ is a sphere.

The images of edges of discs $D(\sigma_i)$ form an oriented graph in $\widetilde{X}$. The surface $\widetilde{X}$ together with this graph will be called the *diagram associated with* Σ, and denoted by $\Gamma(\Sigma)$. (See Section 5 for a general definition of diagrams).

Example. Figure 2 presents the diagram associated with the famous condition Z (or "quotient condition") of Mal'cev [**9**]

$$x_1y_1 = x_2y_2 \wedge x_3y_1 = x_4y_2 \wedge x_3y_3 = x_4y_4 \Rightarrow x_1y_3 = x_2y_4.$$

From the definition of stable quids it follows that if a vertex of some $D(\sigma_i)$ is identified in $\Gamma(\Sigma)$ with a vertex of $D(\sigma_j)$ then the two vertices are of the same type–sources, sinks, or switches. Therefore, in $\Gamma(\Sigma)$, edges incident with a vertex v either all emanate from v or all terminate at v or alternatively emanate and terminate. We will speak accordingly of sources, sinks, and *switches* of $\Gamma(\Sigma)$ and use also an alternate notation: O-*vertices*, I-*vertices*, and W-*vertices* respectively.

The degree of any switch of $\Gamma(\Sigma)$ is an even number. We will consider only stable quids for which the degree of any switch is ≥ 4. This is not a loss of generality because a switch of degree 2 corresponds to a pair σ_i, σ_j of the form $\sigma_i = (xy = x'y')$, $\sigma_j = (xy = x''y'')$ and by deletmg one of σ_i, σ_j and replacing the other by $x'y' = x''y''$ we obtain a quid which is trivially equivalent to the orginal one. Repeating the process we eventually get a quid with no switches of degree 2.

A stable quid Σ will be called W-*minimal* if all switches of $\Gamma(\Sigma)$ have degree 4. Similarly, it is O-*minimal* (I-*minimal*) whenever all sources (sinks) have degree 2.

Proposition 1. *Σ is a Lambek quid $\leftrightarrow$ Σ is spherical and* O-*minimal.*

This is merely an observation. Those who are familiar with the original definition of any Lambek quids ("polyhedral conditions" of [**7**]) will easily see the equivalence. Those who are not can take Proposition 1 as definition.

Remark. The dual of a stable quid $\Sigma = (\sigma_1 \wedge \ldots \sigma_n \Rightarrow \sigma_0)$ is the quid obtained from Σ by replacing each $\sigma_i = (x_py_q = x_ry_s)$ by $y_qx_p = y_sx_r$. Clearly the dual of any O-minimal quid is an I-minimal quid an conversely. Also, the dual of a W-minimal quid is W-minimal.

3. Mal'cev quids. A considerable amount of notation is necessary to define what Mal'cev quids are. We fix a set of variables $X = \{a_i, b_i, c_i, d_i, A_i, B_i, C_i, D_i | \, i \in \mathbf{N}\}$ in which quids are to be written. We need also another alphabet $Y = \{L_i, L_i^*, R_i, R_i^* | \, i \in \mathbf{N}\}$. A Mal'cev sequence is a word $M = X_1X_2 \ldots X_{2(p+q)}$, $p, q \geq 1$, in the alphabet Y such that

(m_1) The set of letters occurring in M is $\{L_1, L_1^*, \ldots, L_p, L_p^*, R_1, R_1^*, \ldots, R_q, R_q^*\}$ and the occurrence of $L_i(R_j)$ in M precedes the occurrence of $L_i^*(R_j^*)$ for every $i \in \{1, \ldots, p\}$ $(l \in \{1, \ldots, q\})$; .

(m_2) If $L_k(R_k)$ occurs between L_i and $L_i^*(R_j$ and $R_j^*)$ then so does $L_k^*(R_k^*)$.

The *Mal'cev quid* $qi\,(M)$ arising from the Mal'cev sequence $M = X_1X_2\ldots$ $\ldots X_{2(p+q)}$ is defined by

$$qi\,(M) = \left(\bigwedge_{i=1}^{2(p+q)-1} \lambda(X_i) = \rho(X_{i+1}) \Rightarrow \lambda(X_{2(p+q)} = \rho(X_1) \right),$$

where $\lambda(X_i)$ and $\rho(X_i)$ are read from the following table.

X_i	L_i	L_I^*	R_j	R_j^*
$\lambda(X_i)$	d_ia_i	c_ib_i	A_jD_j	B_jC_j
$\rho(X_i)$	c_ia_i	d_ib_i	A_jC_j	B_jC_j

Obviously every Mal'cev quid is stable.

Example. We reproduce from [**3**, p. 311] an example of a Mal'cev quid. The word $M = L_1L_2R_1L_2^*R_2L_3R_2^*L_3^*L_1R_1^*$ is a Mal'cev sequence; from the table we get $qi\,(M)$:

$$\begin{aligned} d_1a_1 = c_2a_2 \wedge d_2a_2 = A_1C_1 \wedge A_1D_1 = d_2b_2 \wedge c_2b_2 = A_2C_2 \wedge A_2D_2 = c_3a_3 \wedge d_3a_3 \\ = B_2D_2 \wedge B_2C_2 = d_3b_3 \wedge c_3b_3 = d_1b_1 \wedge c_1b_1 = B_1D_1 \Longrightarrow B_1C_1 = c_1a_1, \end{aligned}$$

Figure 3 presents $\Gamma(qi\,(M))$ and shows $qi\,(M)$ is W-minimal.

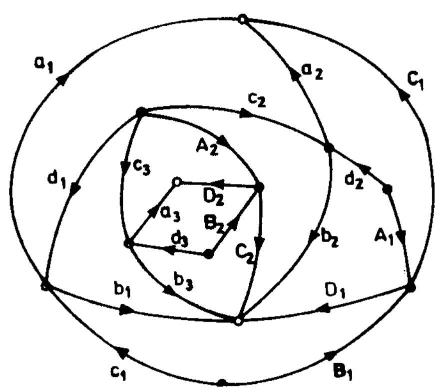

Fig. 3

THEOREM 1. *Σ is a Mal'cev quid $\Leftrightarrow$ Σ is spherical and* W*-minimal.*

Proof. Part 1 (*Rightarrow*). Let $\Sigma = qi\,(M)$ be a Mal'cev quid as in the definition and $\Gamma = \Gamma(\Sigma)$ the diagram associated with it. Switches of Γ are the terminal vertices of edges (labelled by) a_i, b_i, C_j, D_j, i.e. the initial vertices of edges

c_i, d_i, A_j, B_j. It follows immediately from the table defining $qi\,(M)$ that every switch is of degree 4. More precisely, there are $p+q$ switches and edges incident with a switch are either a_i, b_i, c_i, d_i or A_j, B_j, C_j, D_j for some i or j.

It remains only to prove that Σ is spherical; we do it by computing the Euler characteristic. So far we know that Γ has $4(p+q)$ edges, $2(p+q)$ regions and $p+q$ switches. To prove

$$|\{\text{vertices}\}| - |\{\text{edges}\}| + |\{\text{regious}\}| = 2$$

we need to show that the total number of sources and sinks is $p+q+2$.

It suffices to prove $|\{\text{sources}\}| = p+1$; $|\{\text{sinks}\}| = q+1$ will follow by symmetry.

The sources of Γ are initial vertices of edges (labeled by) c_i, d_i, A_j, B_j. Let $Z = \{c_1, d_1, \ldots, c_p, d_p, A_1, B_1, \ldots, A_q, B_q\}$ and let $\sim$ be the equivalence relation on Z such that two elements of Z are equivalent iff the corresponding two edges have the same initial vertex. Obviously, $|\{\text{sources}\}| = |Z/\sim|$.

The relation $\sim$ is generated by "the first symbol of $\lambda(X_i)$ '$\sim$' the first symbol of $\rho(X_{i+1})$", where $i = 1, \ldots, 2(p+q)$ and $i+1$ taken modulo $2(p+q)$.

Let $M_L = Y_1 \ldots Y_{2p}$ be the word obtained from M by deleting all symbols R_j, R_j^*. Let $Z_L = \{c_1, d_1, \ldots, c_p, d_p\}$ and let $\sim$ be the equivalence relation on Z_L generated by "the first symbol of $\lambda(Y_i)$ " $\sim$ "the first symbol of $\rho(Y_{i+1})$", where $i = 1, \ldots, 2p$ and $i+1$ taken modulo $2p$.

Since $\lambda(R_j)$ and $\rho(R_j)$ ($\lambda(R_j^*)$ and $\rho(R_j^*)$) have the same first symbol it follows that $|Z/\sim| = |Z_L/\sim|$. Now for some i the word $L_i L_i^*$ occurs as a subword in M_L. Therefore d_i the first symbol of both $\lambda(L_i)$ and $\rho(L_i^*)$, constitutes an equivalence class in $Z_L/\sim$. Furthermore, if M_L' denotes the word obtained from M_L by deleting L_i and L_i^* then $|Z_L/\sim| = 1 + |Z_L'/\sim|$, where $Z_L' = Z_L - \{c_i, d_i\}$.

If M' is the sequence obtained from M by deleting L_i, L_i^* then M' is a Mal'cev sequence and $(M')L = M_L'$. Since in case $M_L = L_1 L_1^*$ we have $|Z_L/\sim| = 2$ the desired equality follows by induction.

Part 2 ($\Leftarrow$). Assuming Σ is a spherical W-minimal quid we show that there is a Mal'cev sequence M such that $qi\,(M)$ coincides with Σ up to renaming variables. Let $\Gamma = \Gamma(\Sigma)$, $p = |\{\text{sources of } \Gamma\}| - 1$, $q = |\{\text{sinks of } \Gamma| - 1$. By a simple computation using Euler formula it follows that there are $p+q$ switches, $4(p+q)$ edges, and $2(p+q)$ regions in Γ. An edge of Γ will be called an OW-edge or WI-edge according to the types of vertices incident with it. Two OW-edges (WI-edges) will be called *related* if they are incident with the same switch.

Let Γ_O be the graph consisting of all OW-edges and all vertices incident with them. Then

$$|\{\text{edges of } \Gamma_O\}| - |\{\text{vertices of } \Gamma_O\}| = 2(p+q) - ((p+q) + (p+1)) = q - 1.$$

Since the degree of any switch of Γ_O is 2 it follows that by removing some q pairs of related edges of Γ_O we obtain a tree Γ_O. Vertices of T_0 are all sources of Γ

and some p switches which we denote by $l_1, \dots, l_p$. The remaining q switches we denote by $r_1, \dots, r_q$. Let T_I be the graph consisting of the $2q$ WI-edges incident with $r_1, \dots, r_q$ together with the vertices incident with them. Then

$$|\{\text{edges of } T_I\}| - |\{\text{vertices of } T_I\}| = 2q - (q + |\{\text{sinks of } T_I\}|) \geq -1.$$

If there is a circuit γ in T_I then the two sources incident with a switch of γ belong to different connected components of the complement of γ. Then, since T_O is connected and contains all sources we have $T_O \cap T_I \neq \emptyset$, a contradiction. Therefore, there are no circuits in T_I and so T_I is a union of $k \geq 1$ trees, where

$$-k = |\{\text{edges of } T_I\}| - |\{\text{vertices of } T_I\}|.$$

Comparing with the inequality above we get $k = 1$, so T_1 is a tree containing all sinks of Γ.

Let S be the set of all edges not included in $T_o \cup T_I$. S contains exactly one pair of related edges incident with any switch. The complement of $T_O \cup T_I$ is homeomorphic to an open annulus and every edge of S cuts the annulus without disconnecting it. It follows that there exists a simple closed curve ω contains no vertices of Γ and intersects every region in an interval. We subdivide ω into $2(p+q)$ edges by $2(p+q)$ points, one from each region. Thus every of ω meets exactly one edge of S. (See Figure 4.)

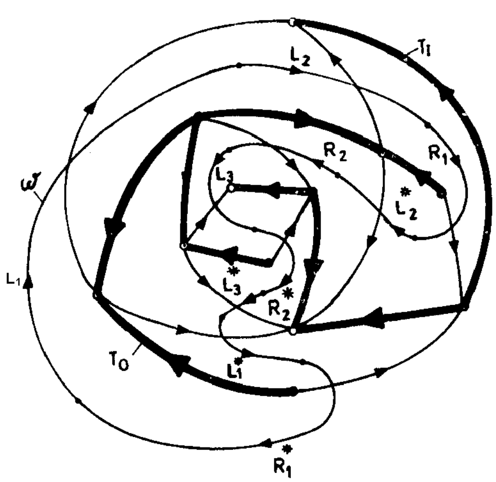

Fig. 4

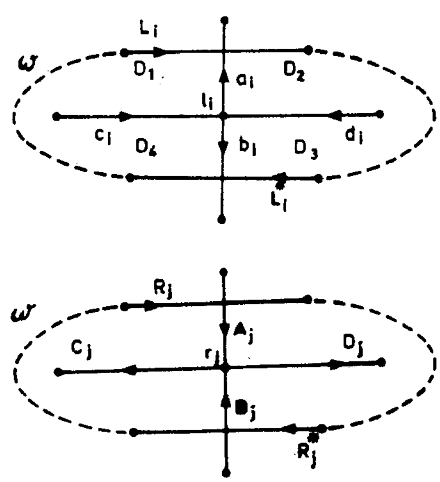

Fig.5

The edges of Γ are labelled by variables involved in Σ. Now we relabel them to see Σ is a Mal'cev quid.

Let ν be the subdivision point on ω which belongs to the region $D(\sigma_0)$, where σ_0 is the consequent identity of the quid Σ. We traverse ω once in one chosen direction starting from ν. For any switch l_1 there are two WI-edges incident with l_1 which belong to S. The edge first met by ω (with respect to the traversing chosen) we label by a_i, the other by b_i. The two corresponding edges of ω we denote respectively by L_i and L_i^*; see Figure 5. Let D_1, D_2, D_3, D_4 be the regions incident with l_i written in such cyclic order that any two adjacent share an edge and that traversing L_i we pass from D_1 to D_2.

It easily follows that traversing L_i^* we pass from D_3 to D_4. We label the common edge of D_1 and $D_4 D$ by c_1 and the common edges of D_2 and D_3 by d_1. Similarly we label edges incident with vertices r_j by A_j, B_j, C_j, D_j; see Figure 5.

Writing the $2(p+q)$ edges of ω in the order we traverse them traversing ω from ν we obtain a word M which clearly satisfies (m_1). Since ω wraps once around T_O it follows that component of $\omega - (L_i \cup L_i^*)$ contains the edge L_j^* whenever it contains L_j. Therefore M is a Mal'cev sequence. $\Sigma = qi\,(M)$ follows immediately we from the way how we relabeled edges of Γ.

As an application of Theorem 1 we can easily describe the quids which are both Mal'cev and Lambek.

COROLLRY. (Clifford and Preston [**3**, Theorem 12.21]). *Let M be a Mal'cev sequence. Then $qi\,(M)$ is a Lambek quid if and only if M is of the form $L_1 \dots L_m R L_m^* \dots L_1^* L_{m+1} \dots L_n R_1^* L_n^* \dots L_{m+1}^*$ with $n \geq 1$, $0 \leq m \leq n$.*

Proof. Let $\Gamma = \Gamma(\Sigma)$ where Σ is both Mal'cev and Lambek quid. Then Γ is both O- and W-minimal. This immediately forces Γ to have only two sinks and the same number $n \leq 1$ of sources and switches; see Figure 6. In order to see which Mal'cev sequences correspond to diagrams of this form notice that the choice of trees T_O and T_I is unique up to a cyclic symmetry of the diagram. Depending on the choice of the region at which one starts traversing the separating circle ω, the method described in the proof of Theorem 1 gives sequences $M_{m,n} = L_1 \dots L_m\, R_1\, L_m^* \dots L_1^*\, L_{m+1} \dots L_n\, R_1^*\, L_n^* \dots L_{m+1}^*$. Moreover, by symmetry of the diagram, the quids $qi\,(M_{m,n})$ and $qi\,(M_{m',n})$ are obtainable from each other by renaming variables.

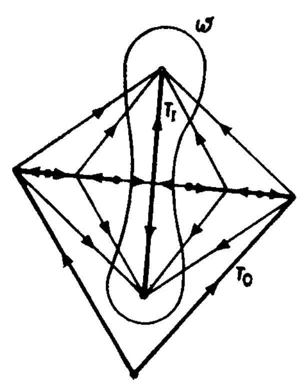

Fig. 6

4. Quids and the embedding problem. The question of finding necessary and sufficient conditions for embeddability of semigroups in groups is the question of axiomatizing the class of embeddable semigroups. Not being closed under taking quotients this class fails to be a variety. But it is a quasi-variety, i.e. axiomatizable by quids. This follows from a general theorem of Mal'cev, see [**12**, p. 216]. For reader's convenience we supply below a simple proof of the special case we are interested in.

We recall that a *semigroup identity* (or *s-identity*) is a formula of the form $x_1 \dots x_m = y_1 \dots y_n$, where x_i and y_j are variables. A *group identity* (*g-identity*) is a formula of the form $x_1^{\varepsilon_1} \dots x_m^{\varepsilon_m} = y_1^{\eta_1} \dots y_n^{\eta_n}$ where ε_i, η_j are integers. An *s-quid* (*g-quid*) is a formula $\sigma_1 \wedge \dots \wedge \sigma_n \Rightarrow \sigma_0$, where $\sigma_0, \sigma_1, \dots, \sigma_n$ are s-identities (g-identities).

Suppose now a semigroup S is given by presentation, i.e. the set $\{x_i \mid i \in I\}$ of generators subject to defining relations $\{u_j = v_j \mid j \in J\}$. We define $G(S)$ to be the group with the same presentation. The map $x_i \mapsto x_i$ induces a semigroup homomorphism $\alpha : S \to G(S)$. It is readily seen that every semigroup homomorphism from S into a group factors through α. Consequently, S is embeddable in a group iff α is an embedding.

Proposition 2. *A semigroup is embeddable in a group if and only if it satisfies all s-quids which are satisfied by any group.*

Proof. Necessity is obvious. To prove sufficiency assume S is not embeddable in any group. Then the homomorphism a above is not an embedding and so there exist two words u and v in letters $\{x_i \mid i \in I\}$ such that $u \neq v$ in S but $u = v$ in $G(S)$. If F denotes the free group freely generated by $x_i \mid i \in I\}$ then uv^{-1} belongs to the normal subgroup of F generated by elements $u_j v_j^{-1}$, $j \in J$. Hence uv^{-1} is a product of conjugates in F of $U_j v_j^{-1}$, $j \in J_0$ with $J_0 \subseteq J$ finite. Considering x_i as variables it follows that the g-quid

$$\bigwedge_{j \in J_0} u_j v_j^{-1} = 1 \Rightarrow uv^{-1} = 1$$

is true on every group. Thus the s-quid

$$\bigwedge_{j \in J_0} u_j = v_j \Rightarrow u = v$$

is true on every group but not on S,

5. Diagrams. We recall that a g-quid $\Sigma = (u_1 = v_1 = v_1 \wedge \dots \wedge u_n = v_n \Rightarrow u_0 = v_0)$ is true on every group iff $u_0 v_1^{-1} \in \{u_1 v_1^{-1}, \dots, u_n v_n^{-1}\}^F$ where F is the free group freely generated by the set of variables involved in Σ and A^F denotes the normal subgroup of F generated by the set $A \subseteq F$. Since the relation $u_0 \in \{u_1, \dots, u_n\}^F$ can be visualized by means of diagrams ("cancellation diagrams", "van Kampen diagrams") we devote this section to describing basic facts about diagrams and how they are connected with quids, especially s-quids. For more details the reader should consult [**8**, Section V. 1] or [**14**].

A *diagram* Γ is a collection of *vertices, edges*, and *regions*, where vertices and edges form a connected finite oriented graph $\Gamma^{(1)}$ in the 2-sphere and regions are connected components of the complement of $\Gamma^{(1)}$. Thus, edges and regions are (homeomorphic to) open intervals and open discs respectively. We use the notation $\iota(e)$, $\tau(e)$, e^{-1} for the initial vertex, the terminal vertex and the inverse edge of the edge e. A *path* is word $e_1^{\varepsilon_1}\dots e_n^{\varepsilon_n}$, $n \leq 1$ where $\iota(e)$ are edges, $\varepsilon_i \in \{\pm 1\}$ and $\tau(e_i^{\varepsilon_i}) = \iota(e_{i+1}^{\varepsilon_{i+1}})$, assuming $\iota(e^{-1} = \tau(e)$ and $\tau(e^{-1} = \iota(e)$. The path above is *positive* if $\varepsilon_1 = \dots = \varepsilon_n = 1$ it is a *cycle* if $\iota(e_1^{\varepsilon_1}) = \tau(e_n^{\varepsilon_n})$.

The boundary ∂D of any region D of Γ consists of vertices and edges incident with D and is a connected subgraph of $\Gamma_{(1)}$. Moreover there is a cycle δD which involves all edges of ∂D and is such that traversin δD D stays all the time on the same (left or right) side. Such δD called a *boundary cycle* of D; it is unique up to cyclic permutations and taking inverses.

A *labeling* $\varphi(e)$ of Γ amounts to assigning to every edge e of Γ a variable $\varphi(e) \in X$ is a set of variables. Labelling extends multiplicatively to all paths in Γ, so that the label of the path $e_1^{\varepsilon_1}\dots e_n^{\varepsilon_n}$ is $\varphi(e_1)^{\varepsilon_1}\dots\varphi(e_n)^{\varepsilon_n}$ – a group word over X.

If F is the free group with X a set of free generators then we, may think of labels of paths in Γ as being elements of F. The following Propositions show that labelled diagrams provide a geometrical interpretation of the relation $u \in \{\nu_1, \dots, \nu_n\}^F$.

PROPOSITION 3. *Let φ be a labeling of a diagram Γ and suppose a boundary cycle δD is chosen for every region D of Γ. Then*

$$\varphi(\delta D) \in \{\varphi(\delta D) |\, E \text{ is a region of } \Gamma,\ E \neq D\}^F$$

PROPOSITION 4. *Let $u, v_1, \dots, v_n \in F$ and $u \in \{v_1, \dots, v_n\}^F$. Then there exists a diagram Γ, a region D and a labelling φ of Γ such that for suitable choices of boundary cycles*

(i) *$u = \varphi(\delta D)$ and*

(ii) *for erery region $E \neq D$ of Γ $\varphi(\delta E) = v_i$ for some i.*

The Propositions above are due to van Kampen [**6**] and are fundamentals of a powerful method in combinatorial group theory as developed from 1966 by Lyndon, Schupp, and others. Essentially they are Lemma V.1.2 and Theorem V.1.1 of [**8**]. The main notational difference is that diagrams are planar in [**8**] while we need them to be spherical in this work. The two concepts obviously amount to the same thing by a stereographic projection.

Now, by the remark made in the first sentence of this section, Propositions 3 and 4 establish a connection between labelled diagrams and g-quids which are true on all groups. To interpret s-quids we are led to the following definition of s-diagrams. Namely, an *s-diagram* is a diagram in which every region has a boundary cycle of the form $e_1 \dots e_m f_n^{-1} \dots f_1^{-1}$, $m, n \geq 1$. (We should note that these diagrams need not be s-diagrams in the sense of [**14**]; the corresponding notion in [**14**] is "two-sided map with all regions two-sided".)

Suppose Γ is an s-diagram and φ a universal labeling, that is a labeling which assigns different variables to different edges of Γ. With every region D of Γ with $\delta D = e_1 \dots, e_m f_n^{-1} \dots f_1^{-1}$ we associate the identity $i(D) = (\varphi(e_1) \dots \varphi(e_m) = \varphi(f_1) \dots \varphi(f_n))$, so that for every region D we have an associated s-quid

$$qi\,(\Gamma, D) = (\bigwedge_{E \neq D} i(E) \Rightarrow i(D)),$$

where the conjunction is taken over all regions of Γ different from D. By Proposition 3, $qi\,(\Gamma, D)$ is true on every group.

To express the dependence of quids we introduce the following notation. If Σ and Θ are s-quids we write $\Sigma \vdash \Theta$ ($\Sigma \vdash_c \Theta$) whenever Θ is true on every semigroup (cancellative semigroup) on which Σ is true.

Proposition 5. *If Σ is an s-guid which is true on every group then there exists an s-diagram Γ and a region D of it such that $qi\,(\Gamma, D) \vdash \Sigma$.*

Proof. Let $\Sigma = (u_1 = \nu_1 \wedge \dots \wedge u_n = nu_n \Rightarrow u_0 = \nu_0)$. Then by Proposition 4 there is a labeled diagram Γ and a region D of Γ such that $u_0 \nu_0^{-1}$ is the label of δD and for every region $E \neq D$ of Γ δE is labeled by some $u_i \nu_i^{-1}$, $i \neq 0$. Clearly Γ is an s-diagram and Σ is obtained from $qi\,(\Gamma, D)$ by renaming (and possibly equating) some variables, so Σ is an obvious consequence of $qi\,(\Gamma, D)$.

Remark. The diagrams $\Gamma(\Sigma)$ associated in Section 2 with stable quids meet all requirements to be s-diagrams but one-they need not be spherical. A natural question of whether a stable quid is true on every group is partially answered by Proposition 3: it is true if it is spherical. With a bit of additional considerations one can prove the converse also holds, so a *stable quid is true on every group iff it is spherical.*

6. Axiomatizations of the class of embeddable semigroups. If Γ_1 and Γ_2 are s-diagrams we write $\Gamma_1 \vdash \Gamma_2$ whenever for every region D of Γ_2 there exists a region D_1 of Γ_1 such that $qi\,(\Gamma_1, D_1) \vdash qi\,(\Gamma_2, D_2)$. Also, $\Gamma_1 \vdash\dashv \Gamma_2$ whenever both $\Gamma_1 \vdash \Gamma_2$ and $\Gamma_2 \vdash \Gamma_1$ hold. $\Gamma_1 \vdash_c \Gamma_2$ and $\Gamma_1 \vdash\dashv_c \Gamma_2$ are defined analogously.

Suppose D is region of an s-diagram Γ and α a positive path on δD. Let Γ' be the diagram obtained by subdividing the region D by an edge e connecting the initial and the terminal vertex of α. Γ' is clearly an s-diagram (see Figure 7) and we have

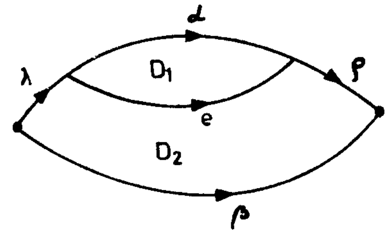

Fig. 7

Lemma 1. $\Gamma' \vdash\dashv_c \Gamma$.

Proof. Let $\delta D = \lambda\alpha\rho\beta^{-1}$ where $\lambda, \alpha, \rho, \beta$ are positive paths and λ, ρ possibly empty. Then in Γ' the region D is replaced by two regions D_1 and D_2 with $\delta D_1 = \alpha e^{-1}$, $\delta D_2 = \lambda e\rho\beta^{-1}$.

Let E be a region of Γ', $E \neq D_1, D_2$. Then E is a region of Γ as well and we have $qi\,(\Gamma, E) = (\psi \wedge lur\,\nu \Rightarrow \Theta)$ and $qi\,(\Gamma', E) = (\psi \wedge lxr\ = \nu \wedge x = u \Rightarrow \Theta)$, where ψ is the conjunction of the identities $i(D')$, $D' \neq D_1, D_2, E$, and $\Theta = i(E)$ and x, u, ν, l, r are labels of $e, \alpha, \beta, \lambda, \rho$ respectively. The specialization of $qi\,(\Gamma', E)$ obtained by replacing every occurrence of the variable x by u is tautologically equivalent with $qi\,(\Gamma, E)$, so $qi\,(\Gamma', E) \vdash qi\,(\Gamma, E)$. The converse $qi\,(\Gamma', E) \vdash qi\,(\Gamma, E)$ is also true because $lxr\ = \nu \wedge x = u \Rightarrow lur\ = \nu$.

For the remaining cases we write $qi\,(\Gamma', D_1) = (\psi \wedge lxr\ = \nu \Rightarrow x = u)$, $qi\,(\Gamma', D_2) = (\psi \wedge x = u \Rightarrow lxr\ = \nu)$ and $qi\,(\Gamma, D) = (\psi \Rightarrow lur\ = \nu)$. By the same kind of argument we have $qi\,(\Gamma, D) \vdash qi\,(\Gamma', D_2)$ and $qi\,(\Gamma, D) \vdash_c qi\,(\Gamma', D_1)$. Notice that we do not need cancellation if both λ and ρ are empty, so, if e connects the source vertex of D with the sink vertex of D we have $\Gamma' \vdash\dashv \Gamma$.

Lemma 2. *Let ν be a source or a sink in an s-diagram Γ and $e_1, \ldots, e_n$ $(n \geq 1)$ all edges incident with ν. Let Γ' be the diagram obtained by collapsing some of these edges to the point ν. If Γ' is an s-diagram then $\Gamma' \vdash_c \Gamma$.*

Proof. We prove the Lemma assuming ν is a sink; the other case follows by symmetry.

Let x_i be the label of e_i, $_1, \ldots, D_n$ regions of Γ incident with ν and $i(D_i) = (u_i x_i = \nu_i x_{i+1})$, $i = 1, \ldots, n$, where $i+1$ is taken modulo n and some words u_i, ν_j are possibly empty.

There is a 1—1 correspondence $D \leftrightarrow D'$ between regions of Γ and Γ'. Define $\varepsilon_i = 0$ if e_i is contracted, otherwise $\varepsilon_i = 1$. Then $i(D') = i(D)$ for $D \neq D_1, \ldots, D_n$ and i $(D_i') = (u_i x_i^{\varepsilon_i} = \nu_i x_{i+1}^{\varepsilon_{i+1}})$, $i = 1, \ldots, n$ where $x_i^1 = x_i$ and $x_i^0 =$ "emty word". (Notice that the assumption that Γ' is an s-diagram implies $u_i \nu_{i-1} \neq \nu_i x_{i+1}^{\varepsilon_{i+1}}$) whenever $\varepsilon = 0$.)

Replacing every occurrence of x_i in $qi\,(\Gamma, D)$ by $x_i^{\varepsilon_i} x$, where x is a new variable, we obtain a quid Θ such that $qi\,(\Gamma, D) \vdash \Theta$. The occurrence of $i(D_i)$

in $qi\,(\Gamma, D)$ corresponds to the occurence of $i'(D_i) = (u_i x_i^{\varepsilon_i} x = \nu_i x_{i+1}^{\varepsilon_{i+1}} x)$ in Θ. Now $i'(D_i)$ and $i(D_i')$ are equivalent on every cancellative semigroup whence $\Theta \vdash\dashv_c qi\,(\Gamma', D')$ and the Lemma follows.

We say that an s-diagram is *triangular* whenever every its region has a boundary cycle of length 3. The diagram is *stable* if it is triangular and every its vertex is either a source a sink, or a switch. An edge of a stable diagram is either an OI-, OW-, or WI-*edge*, according to the types of vertices incident with it. Observe that the three edges incident with any region of a stable diagram are of three different types and every vertex has even degree which is ≥ 4 in cases the vertex is a source or a sink. As explained in Section 2, switches of degree 2 are redundant in a sense and so there is no loss of generality in the assumption that all vertices of a stable diagram have degree ≥ 4. For $X \in \{\mathrm{O}, \mathrm{I}, \mathrm{W}\}$ we define *X-minimal diagrams* to be those in which every X-vertex has degree 4.

If Σ is a spherical stable quid and $\Gamma' = \Gamma(\Sigma)$ the corresponding diagram (Section 2) then the diagram Γ' obtained by subdividing every region D of Γ by an edge connecting the source vertex with the sink vertex of D is stable. Conversely, the diagram Γ obtained from a stable diagram Γ' by removing all its OI-edges is $\Gamma(\Sigma)$ for some stable quid Σ. By Lemma 1 we have $\Gamma' \vdash\dashv \Gamma$ and so the notions "stable spherical quid" and "quid of the form $qi\,(\Gamma, D)$ with Γ stable" coincide. Also, with the notation as above, Σ is an X-minimal quid iff Γ' is an X-minimal diagram.

Remark. Let Γ' be the diagram obtained from a stable diagram Γ by reversing the orientation of every OW-edge. Then $(\Gamma')' = \Gamma$ and $\Gamma \to \Gamma'$ is a bijection between the sets of O-minimal and W-minimal quids. In view of the results of Sections 2 and 3 this induces a bijection between the sets of Lambek and Mal'cev quids.

Suppose now Γ is an s-diagram. For any $X, Y \in \{\mathrm{O}, \mathrm{I}, \mathrm{W}\}$, $X \neq Y$ we construct the s-diagram ΓXY, the XY-subdivision of Γ in the following way.

1) Subdivide every edge e of Γ by a new vertex $\nu_e \in e$ into two (for a moment non-oriented) edges.

2) For every region D of Γ take a new vertex $\nu_D \in D$. If $\delta D = e_1 \ldots e_m e_{m+1}^{-1} \ldots \ \ldots e_{m+n}^{-1}$ then subdivide D into $2(m+n)$ triangular regions by $2(m+n)$ edges connecting ν_D with vertices ν_e, ν, where e, ν are incident with D, so that every new triangular region is incident with one vertex of Γ, one vertex ν_e, and one vertex ν_D .

3) Orient the diagram just obtained so that every vertex ν_e is an X-vertex and every vertex ν_D is a Y-vertex. (It is easily seen that this can always be done in a unique way.)

Lemma 3. $\Gamma_{XY} \vdash_c \Gamma$.

Proof. We prove $\Gamma_{WO} \vdash_c \Gamma$. Mutatis mutandis the same proof applies to all other cases. The idea is to describe intermediate diagrams $\Gamma_1, \Gamma_2, \Gamma_3$ and using Lemmas 1 and 2 to prove $\Gamma_{XO} \vdash_c \Gamma_1 \vdash_c \Gamma_3 \vdash_c \Gamma$. The transformation of a region of Γ in passing from Γ via $\Gamma_3, \Gamma_2, \Gamma_1$ to Γ_{WO} is shown on Figure 8.

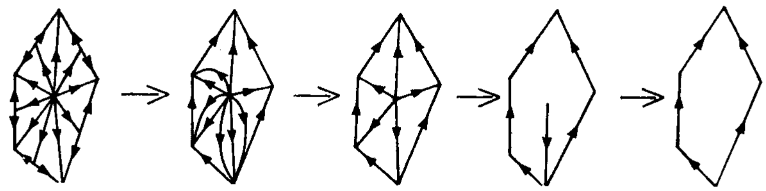

Fig. 8

Let e^+ and e^- denote the two edges of Γ_{WO} obtained by subdivision of the edge e of Γ, where e^+ is the edge oriented "in accordance" with the orientation of e. Γ_1 is obtained by contraction of all edges e^-. Contracting successively at each sink of Γ_{WO} we get $\Gamma_{WO} \vdash_c \Gamma_1$, by Lemma 2.

Now the edges of Γ_{WO} which connect vertices ν_D with vertices ν_e become multiple edges in Γ_1. Using Lemma 1 we can remove them all obtaining that way Γ_2 with $\Gamma_1 \vdash\dashv \Gamma_2$.

The vertex set of Γ_2 is the vertex set of Γ plus vertices ν_D. Removal of all vertices ν_D and all edges incident with them would result in the original diagram Γ. Let ν'_D denote the source vertex of the region D of Γ. Let Γ_3 be the diagram obtained by removing all edges emanating from ν_D but one ending at ν'_D (for every D). The removals can be done in succession so that Lemma 1 applies at each stage and we have $\Gamma_2 \vdash\dashv \Gamma_3$.

Finally, Γ is obtained by collapsing all edges ("spines") connecting ν_D with ν'_D. Obviously (or by Lemma 2) $\Gamma_3 \vdash\dashv \Gamma$, finishing the proof.

If $X, Y \in \{\mathrm{O}, \mathrm{I}, \mathrm{W}\}$, $X \neq Y$ we define XY-minimal diagrams to be those stable diagrams which are X-minimal and have the degree of every Y-vertex≤ 6. A quid Σ will be called XY-minimal if $\Sigma = qi\,(\Gamma, D)$ for some XY-minimal Γ.

Theorem 2. *$X, Y \in \{O, I, W\}$, $X \neq Y$. If Σ is an s-quid which is true on every group then there exists an XY-minimal quid Σ_{XY} such that $\Sigma_{XY} \vdash_c \Sigma$. Consequently, the set of all XY-minimal quids together with the cancellation laws axiomatizes the class of semigroups embeddable in a group.*

Proof. Everything has been done for this proof. By Proposition 5 $qi\,(\Gamma, D) \vdash \Sigma$ for some s-diagram Γ' and a region D of it. Let Γ be the diagram obtained by triangulating every region of Γ' by a number of edges emanating from the source vertex of that region (see Figure 9). Then $\Gamma \vdash \Gamma'$ by Lemma 1. By Lemma 3 $\Gamma_{XY} \vdash_c \Gamma$ and so $qi\,(\Gamma_{XY}, E) \vdash_c \Sigma$ for some region E of Γ_{XY}. Since Γ is triangular it follows that $qi\,(\Gamma_{XY}, E)$ is XY-minimal, completing the proof.

Corollary. (a) (*Mal'cev* [**10**], *Bush* [**2**]). *For every s-quid Σ which is true on all groups there exists a Mal'cev quid Σ_M and a Lambek quid Σ_L such that $\Sigma_M \vdash_c \Sigma$ and $\Sigma_L \vdash_c \Sigma$.*

(b) (*Mal'cev* [**10**], *Lambek* [**7**]). *A semigroup is embeddable in a group if and only if it is cancellative and satisfies either all Mal'cev quids or all Lambek quids.*

Remarks. 1. This is to justify the "stable" notation. Suppose Γ is a triangular s-diagram. Figure 10 presents what changes one is allowed to perform on two adjacent regions of Γ to obtain a new diagram Γ' such that $\Gamma' \vdash_c \Gamma$. The stable diagrams are precisely those to which the changes above cannot be applied. Moreover, it can be proved that every triangular diagram in which there are no positive cycles can be transformed by a finite number or such to a stable diagram.

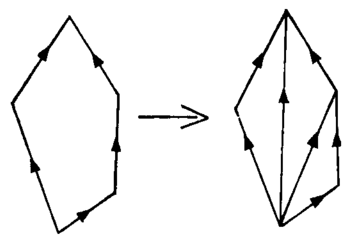

Fig. 9

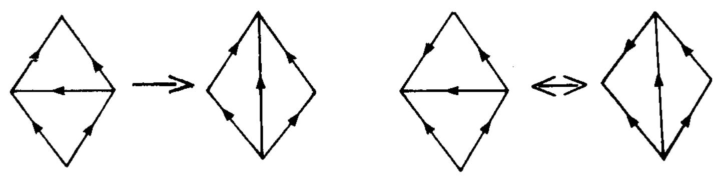

Fig. 10

2. Every semigroup S can be canonically embedded in a monoid (semigroup with identity) S^1 so that S is embeddable in a group iff S^1 is. Thus the embedding problem for semigroups is very close to that for monoids. It can be proved that a set of stable quids axiomatizes the class of embeddable monoids iff the same set with the cancellation laws added axiomatizes the class of embeddable semigroups.

3. As we have already mentioned, the class of emheddable semigroups is not finitely axiomatizable. This brings importance to the task of finding simple criteria which guarantee embeddability. An example is the embedding theorem of Adjan [**1**] for semigroups given by certain presentations; a transparent proof is given hy Remmers [**14**].

Another well-known example is the theorem of Doss [**5**] which asserts the embeddability of semigroups satisfying a certain first-order property (left quasi-regular semigroups). Doss's result generalizes some previously known criteria and is proved by checking that all Mal'cev quids are true on the semigroups in question. We note here that it is possible to give a geometrical variant of the proof in [**5**] which is much less computational.

Akowledgements. I would like to thank Prof. S. Prešić for turning my to the problem of embedding semigroups in groups and his generousity supplying me with an account of relevant facts of universal algebra. I would also like to thank Prof. J. McCool for reading parts of the manuscript and valuable criticism.

REFERENCES

[1] S.I. Adjan, *Defining relations and algorithms problems for groups and semigroups*, Proc. Steklov Inst. Math. **85** (1966).

[2] G.C. Bush, *The embedding theorems of Malcev and Lambek*, Canad J. Math. **15**(1963), 49–58.

[3] A.H. Clifford and G.B. Preston, *The Algebraic Theory of Semigroups, vol. 2*, Math. Surveys 7, Amer. Math. Soc., Providence, Rhode Island, 1977.

[4] P.M. Cohn, *Universal Algebra*, Harper and Row, New York-London, 1965.

[5] R. Doss, *Sur l'immersion d'un semi-groupe dans un groupe*, Bull. Sci. Math. **(2)72**(1948), 139–150.

[6] E.R. van Kampen, *On some lemmas in the theory of groups*, Amer. J. Math. **55**(1933), 268–273.

[7] J. Lambek, *The immersibility of a semigroup into a group*, Canad. J. Math. **3**(1951), 34–43.

[8] R.C. Lyndon and P.E. Schup, *Combinatorial Group Theory*, Springer-Verlag, Berlin-Heidelberg-New York, 1977.

[9] A.I. Ma1'cev, *On the immersion of an algebraic ring into a field*, Math. Ann. **113**(1937), 686–691.

[10] A.I. Mal'cev, *On the immersion of associative systems in groups*, Mat. Sbornik (N. S.) **6**(1939), 331–336.

[11] A.I. Mal'cev, **On the immersion of associative systems in groups II**, ibid. **8**(1940), 241–264.

[12] A.I. Mal'cev, *Algerbaic Systems*, Die Grundlehren der Math. Wissenschaften, Bd. 192, Sprin-ger-Verlag, Berlin-Heidelberg-New York, 1973.

[13] F. Radó, *Einbettung eines Halbgewebes in ein reguläres Gewebe und eines Halbgruppoids in eine Gruppe*, Math. Z. **89**(1965), 395–410.

[14] J.H. Remmers, *On the geometry of semigroup presentations*, Adv. in Math. **36**(1980), 283–296.

[15] D. Tamari, *Monoïdes préordonnés et chains de Malcev*, Bull. Soc. Math. France **82**(1954), 53–96.

Matematicki Institut
Knez Mihailova 35
Beograd, Jugoslavija

(Received 22 01 1985)