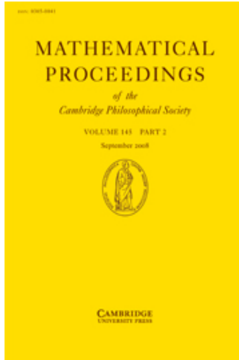


On embedding categories in groupoids

P. T. JOHNSTONE

Mathematical Proceedings of the Cambridge Philosophical Society / Volume 145 / Issue 02 / September 2008, pp 273 - 294
DOI: 10.1017/S0305004108001345, Published online: 19 May 2008

Math. Proc. Camb. Phil. Soc. (2008), **145**, 273 © 2008 *Cambridge Philosophical Society*

doi:10.1017/S0305004108001345 *Printed in the United Kingdom*
First published online 19 May 2008

On embedding categories in groupoids

BY P. T. JOHNSTONE

Department of Pure Mathematics, University of Cambridge, England.

(*Received* 13 *February* 2008)

Abstract

category = monoid with a partial op.

We provide a new, unified approach to the necessary and sufficient conditions found by Mal'cev (1939) and by Lambek (1951) for embeddability of a semigroup in a group, and also show that each provides a necessary and sufficient set of conditions for the embeddability of a category in a groupoid. We show that all such conditions, and more besides, may be derived in a uniform way from a particular class of directed graphs which we call quadrangle clubs, and we prove a number of results (extending those of Mal'cev, Lambek, Bush and Krstić) on which families of quadrangle clubs provide sufficient conditions for embeddability.

Introduction

When can a monoid (that is, a semigroup with 1) be embedded as a submonoid of a group? Equivalently, given that quasivarieties of universal algebras can be characterized by universal Horn axioms (cf. [**4**], section VI.4), we can ask for a set of universal Horn axioms in the language of monoids which are satisfied by all groups, and which collectively imply all such axioms. We recall that a universal Horn axiom is a sequent of the form

$$(\phi_1, \phi_2, \ldots, \phi_n \vdash \psi)$$

where the ϕ_i and ψ are atomic formulae (that is, equations between products of variables); the interpretation of the axiom is that, given any assignment of values to the variables involved which makes all of $\phi_1, \ldots, \phi_n$ true, ψ is necessarily true as well.

For commutative monoids, the answer is well known (and easy to prove): if M is a commutative monoid satisfying the *cancellation law*

$$(ab = ac \vdash b = c)$$

then we may construct its 'group of fractions' in much the same way that one constructs the field of fractions of an integral domain, and prove that it is a group in which M can be embedded. But the cancellation law is also necessary, since it holds in all groups.

For non-commutative monoids, we clearly need both the left cancellation law, as above, and the right cancellation law ($ac = bc \vdash a = b$). But these are not sufficient, as was first shown by A.I. Mal'cev [**8**]. In a subsequent paper [**9**], Mal'cev described an algorithm for generating an infinite sequence of universal Horn axioms, all of which must be satisfied in any monoid embeddable in a group, and which are collectively sufficient for this. In [**10**], he also showed that no finite set of universal Horn axioms could be necessary and sufficient for embeddability. Some time later, J. Lambek [**7**] produced a different, though formally similar, list of Horn axioms.

A question obviously related to that with which we began is 'When can a category be embedded as a subcategory of a groupoid?' Equivalently, we can ask when a category $\mathcal{A}$ admits a faithful functor to a groupoid; for if $F: \mathcal{A} \to \mathcal{G}$ is such a functor, we may replace $\mathcal{G}$ by the full image of F (that is, the category whose objects are those of $\mathcal{A}$, but whose morphisms $A \to B$ are defined to be morphisms $FA \to FB$ in $\mathcal{G}$), and $\mathcal{A}$ is isomorphic to a subcategory of this groupoid. Clearly, once again, the left and right cancellation laws are necessary (in categorical language, every morphism of $\mathcal{A}$ must be both monic and epic), but the case of non-commutative monoids shows that these conditions are not sufficient. So far as the present writer is aware, this question has not been explicitly posed before; however, it comes as no great surprise to find that the lists of Horn axioms produced by Mal'cev and by Lambek, with the variables in them interpreted as ranging over the morphisms of $\mathcal{A}$, are both necessary and sufficient.

Our aim in this paper is to prove this, and also to provide a more explicitly categorical approach to the original results of Mal'cev and Lambek, via a class of directed graphs which we call *quadrangle clubs*. We show that all the Horn axioms generated by Mal'cev's and Lambek's algorithms may be derived in a uniform way from quadrangle clubs, and that all Horn axioms derived from quadrangle clubs must hold in any category embeddable in a groupoid. We also show that not all of the axioms generated by Mal'cev, and not all those generated by Lambek, are necessary; it suffices to restrict to those for which the corresponding quadrangle clubs satisfy a condition which we call *propriety*. However, we shall show that it is not sufficient to restrict to those Horn axioms which are generated both by Mal'cev's algorithm and by Lambek's. As a digression, we shall also present a topos-theoretic characterization of the presheaf toposes corresponding to small categories embeddable in groupoids.

It should be mentioned that comparisons between the axiom-systems of Mal'cev and of Lambek have previously been given, in the semigroup-theoretic case, by G.C. Bush [**1**, **2**] and by A.H. Clifford and G.B. Preston ([**3**], chapter 12). However, we believe that the more geometric viewpoint engendered by the category-theoretic perspective enables us to give more perspicuous proofs of their results. A similar geometric viewpoint was taken by S. Krstić [**6**], again in the semigroup-theoretic context.

I am indebted to Jim Lambek for his perceptive comments on an earlier version of this paper, and in particular for directing my attention to the works of George Bush and Sava Krstić.

1. *Quadrangle Clubs*

Definition 1·1. We define a *quadrangle club* to be a finite directed graph $\mathcal{C}$ with the following structure:

(1) The vertices of $\mathcal{C}$ are partitioned into three subsets $\mathcal{B}, \mathcal{R}, \mathcal{G}$ (called blue, red and green vertices respectively).

(2) There is at most one edge between any pair of vertices of $\mathcal{C}$, and each edge runs either from a blue to a red vertex, or from a red to a green vertex. We refer to these two types of edges as purple and yellow edges, respectively.

(3) We are given a set $\mathcal{Q}$ of subgraphs of $\mathcal{C}$ (called *quadrangles*) of the form

$$\begin{tikzcd} b \arrow[r, "p_1"] \arrow[d, "p_2"'] & r_1 \arrow[d, "y_1"] \\ r_2 \arrow[r, "y_2"'] & g \end{tikzcd}$$

containing one blue, one green and two red vertices, and two purple and two yellow edges, such that every edge of $\mathcal{C}$ (of either colour) belongs to exactly two members of $\mathcal{Q}$. Note that we do not (at this stage) require $\mathcal{Q}$ to contain *all* subgraphs of $\mathcal{C}$ of this form; if it does, we say $\mathcal{C}$ is *proper*.

(4) $\mathcal{C}$ is connected; and the edges at each vertex are connected by quadrangles — that is, for any two edges having a given vertex in common, it is possible to link them by a chain of edges beginning or ending at that vertex, such that any two consecutive edges in the chain belong to a common quadrangle.

(5) (Euler condition) Writing B, R, G and Q respectively for the numbers of blue, red and green vertices and of quadrangles, we have $B + R + G = Q + 2$. (Note that it follows from (3) that the number of purple edges, and the number of yellow edges, are both equal to Q; so this is equivalent to Euler's formula $V + F = E + 2$ for the polyhedron formed by the quadrangles.)

Two quadrangles of $\mathcal{C}$ may have more than a single vertex or a single edge in common; for example, in the (proper) quadrangle club $\mathcal{E}_4$ represented by the diagram

$$\begin{tikzcd} b_1 \arrow[r] \arrow[dr] & r_1 \arrow[r] \arrow[dr] & g_1 \\ b_2 \arrow[r] \arrow[ur] & r_2 \arrow[r] \arrow[ur] & g_2 \end{tikzcd}$$

any two quadrangles have either a pair of purple edges or a pair of yellow edges in common, and in the proper club $\mathcal{D}_3$ represented by

$$\begin{tikzcd} & r_1 \arrow[dr] & \\ b \arrow[ur] \arrow[r] \arrow[dr] & r_2 \arrow[r] & g \\ & r_3 \arrow[ur] & \end{tikzcd}$$

any two quadrangles have one purple and one yellow edge in common. We shall eventually legislate the latter possibility out of existence; in fact, we shall require that no two quadrangles have both a blue and a green vertex in common, a condition which we shall refer to as the *blue-green condition*. Note also that the valency of any red vertex (i.e., the number of edges coming into or out of it) is necessarily even, since condition (3) implies that there are as many purple edges coming in as there are yellow edges going out; thus, once we have imposed the restriction mentioned above, we shall be able to conclude that the valency of any red vertex is at least 4.

Definition 1·2. Let X denote one of the three colours {blue, red, green}. We shall say that a quadrangle club is *X-minimal* if every X vertex has valency $2n_X$, where $n_X = 2$ if X is 'red' and $n_X = 1$ otherwise. We shall also use the term 'Mal'cev quadrangle club' for a red-minimal quadrangle club, and 'Lambek quadrangle club' for a green-minimal one: the reasons for these terms will become apparent later.

In passing, we note that the Lambek condition implies the blue-green condition mentioned earlier: for each green vertex of a Lambek quadrangle club belongs to just two quadrangles, which have two yellow edges in common and hence must have distinct blue vertices.

We note that if a quadrangle club is X-minimal then the number of X vertices is $Q/2$, so that if it is both X-minimal and Y-minimal (where X and Y are different colours) then the Euler condition implies that it has only two Z vertices, where Z is the third colour. In practice, minimality in two different colours is too strong a restriction, but we shall consider a weaker notion which we call (X, Y)-*sesquiminimality*: we say a club is (X, Y)-sesquiminimal if its X vertices all have valency $2n_X$, and its Y vertices have valency at most $3n_Y$.

Remark 1·3. If we 'geometrically realize' a quadrangle club $\mathcal{C}$ by taking Q copies of the unit square, and identifying their boundaries according to the common edges of the quadrangles in $\mathcal{C}$, the resulting topological space will be a connected 2-manifold, by conditions (3) and (4). Condition (5) then ensures that this manifold is a sphere. Thus, in place of the combinatorial definition above, we could alternatively give a more geometrical one by saying that a quadrangle club is any polyhedral decomposition of a sphere whose faces are all quadrangles, and whose vertices are coloured blue, red and green in such a way that each face has one blue, one green and two red vertices, the red vertices being opposite each other.

For example, the combinatorial quadrangle club $\mathcal{E}_4$ above corresponds to the polyhedral decomposition of the sphere (an octahedron with four missing edges) obtained by placing the two red vertices at the poles, and joining them to two blue and two green vertices placed alternately around the equator; similarly, $\mathcal{D}_3$ corresponds to the decomposition with a blue vertex at the north pole, a green vertex at the south pole and three red vertices equally spaced around the equator. In passing, we note that we may enlarge these examples to improper quadrangle clubs $\mathcal{E}_{2n}$ ($n \geqslant 3$) and $\mathcal{D}_n$ ($n \geqslant 4$) respectively: $\mathcal{E}_{2n}$ has n blue and n green vertices alternately around the equator, and $\mathcal{D}_n$ has n red vertices around the equator. (In each case, all the equatorial vertices are joined to both poles, and the distinguished quadrangles are those which appear in the surface of the sphere.) Note also that the $\mathcal{E}_{2n}$ are exactly the quadrangle clubs which are both blue-minimal and green-minimal.

For future reference, we also mention a family of (proper) quadrangle clubs $\mathcal{W}_{2n}$, $n \geqslant 2$, satisfying the Lambek and Mal'cev conditions: $\mathcal{W}_{2n}$ has two blue vertices (which we can consider to be at the poles), n red and n green vertices placed alternately around the equator, $2n$ purple edges joining each blue vertex to each red vertex, and $2n$ yellow edges joining adjacent vertices around the equator. Again, we note that these are exactly the quadrangle clubs which are both Lambek and Mal'cev (and that $\mathcal{W}_4 \cong \mathcal{E}_4$ is the unique quadrangle club which is minimal in all three colours).

Definition 1·4. Let $\mathcal{C}$ be a quadrangle club, and let $q \in \mathcal{Q}$ be a particular quadrangle of $\mathcal{C}$. By the $(\mathcal{C}, q)$-*quadrangularity axiom* we mean the assertion that, whenever we are given a diagram of shape $\mathcal{C}$ in (that is, a directed-graph homomorphism from $\mathcal{C}$ to) some category $\mathcal{A}$, in which all the quadrangles in $\mathcal{Q}$ except for q are mapped to commutative squares in $\mathcal{A}$, then q is also mapped to a commutative square. It is easy to see that any such assertion may be expressed as a universal Horn axiom in the language of categories. We note that a particular quadrangle club $\mathcal{C}$ may have symmetries which mean that these axioms, for different $q \in \mathcal{Q}$, are not independent; for example, the symmetry groups of the particular examples $\mathcal{E}_{2n}$, $\mathcal{D}_n$ and $\mathcal{W}_{2n}$ described above are all transitive on quadrangles, so that all axioms obtained from one of them are equivalent. Thus all the axioms derived from $\mathcal{E}_4$ are of the form

$$(y_1p_1 = y_2p_2, y_1p_3 = y_2p_4, y_3p_1 = y_4p_2 \vdash y_3p_3 = y_4p_4) .$$

(Strictly speaking, the hypotheses of the Horn axiom should also include equations between the domains and codomains of the y_i and p_j to ensure that the composites which appear, and the equations between them, make sense; but we shall suppress these — they can always be inferred from the equations between composites.)

We shall say that a category $\mathcal{A}$ is $(\mathcal{C}, q)$-*quadrangular* if it satisfies the $(\mathcal{C}, q)$-quadrangularity axiom, and simply $\mathcal{C}$-quadrangular if it is $(\mathcal{C}, q)$-quadrangular for every distinguished quadrangle q of $\mathcal{C}$. More generally, if $\mathcal{S}$ is a set of quadrangle clubs, we shall say $\mathcal{A}$ is $\mathcal{S}$-quadrangular if it is $\mathcal{C}$-quadrangular for every $\mathcal{C} \in \mathcal{S}$; for example if $\mathcal{S}$ is the set of proper quadrangle clubs, we say $\mathcal{A}$ is *properly quadrangular*. If $\mathcal{S}$ is the set of all quadrangle clubs, we say $\mathcal{A}$ is *totally quadrangular*.

In much of this paper, we shall be concerned with finding sets of quadrangle clubs such that the corresponding quadrangularity axioms are necessary and sufficient for embeddability in a groupoid; thus we shall need to compare the strength of different quadrangularity axioms. We shall introduce the notation $((\mathcal{C}, q) \vdash (\mathcal{C}', q'))$ to mean that every $(\mathcal{C}, q)$-quadrangular category is also $(\mathcal{C}', q')$-quadrangular; and $(\mathcal{C} \vdash \mathcal{C}')$ to mean that, for every distinguished quadrangle q' of $\mathcal{C}'$, there exists a quadrangle q of $\mathcal{C}$ such that $((\mathcal{C}, q) \vdash (\mathcal{C}', q'))$.

For future reference, we note:

LEMMA 1·5. *If $\mathcal{A}$ is $\mathcal{E}_4$-quadrangular, then every morphism of $\mathcal{A}$ is both monic and epic.*

Proof. In the sequent displayed in Definition 1·4, if we set $y_1 = y_2$ and take p_1, p_2, y_3 and y_4 to be the appropriate identity morphism, then we obtain

$$(y_1p_3 = y_1p_4 \vdash p_3 = p_4)$$

which is the assertion that every morphism is monic. The proof that every morphism is epic is similar. □

The converse of Lemma 1·5 is false. It is easy to see that if we take $\mathcal{A}$ to be the category freely generated by the directed graph $\mathcal{E}_4$, subject to the relations which say that three of the four quadrangles commute, then every morphism of $\mathcal{A}$ is monic and epic, but the fourth quadrangle does not commute.

In the opposite direction, we have:

LEMMA 1·6. *Any category admitting a faithful functor to a groupoid is totally quadrangular.*

Proof. If $\mathcal{A}$ admits a faithful functor to a groupoid, then, as we observed in the Introduction, we may regard it up to isomorphism as a subcategory of a groupoid. Since the validity of universal Horn axioms is inherited by substructures, it suffices to prove that any groupoid satisfies all such axioms. But if we are given a quadrangle club $\mathcal{C}$ and a diagram of shape $\mathcal{C}$ in a groupoid (or, more generally, a diagram of shape $\mathcal{C}$ in a category $\mathcal{A}$ all of whose edges are isomorphisms of $\mathcal{A}$), and we know that all but one of the quadrangles of $\mathcal{C}$ commute (that is, are mapped to commuting squares in $\mathcal{A}$), we may deduce the commutativity of the remaining quadrangle by 'chasing around the back of the sphere'. For example, for the club $\mathcal{E}_4$, if we are given morphisms p_1, p_2, p_3, p_4 and y_1, y_2, y_3, y_4 (with appropriate domains and codomains) satisfying the conditions $y_1 p_1 = y_2 p_2$, $y_1 p_3 = y_2 p_4$ and $y_3 p_1 = y_4 p_2$, and we know that y_1 and p_2 (at least) are invertible, then we may deduce

$$y_3 p_3 = y_3 y_1^{-1} y_2 p_4 = y_3 p_1 p_2^{-1} p_4 = y_4 p_4 ,$$

as required. □

Remark 1·7. We digress to note that the Euler condition (5) is essential to the proof of Lemma 1·6. It is of course possible to 'quadrangulate' other 2-manifolds according to the rules in 1·1(1)–(4): for example, here is a quadrangulation of the torus, pictured as a square with opposite edges identified.

$$\begin{tikzcd}
b_1 \arrow[r] \arrow[d] & r_1 \arrow[d] & b_2 \arrow[l] \arrow[r] \arrow[d] & r_2 \arrow[d] & b_1 \arrow[l] \arrow[d] \\
r_3 \arrow[r] & g_1 & r_4 \arrow[l] \arrow[r] & g_2 & r_3 \arrow[l] \\
b_3 \arrow[r] \arrow[u] \arrow[d] & r_5 \arrow[u] \arrow[d] & b_4 \arrow[l] \arrow[r] \arrow[u] \arrow[d] & r_6 \arrow[u] \arrow[d] & b_3 \arrow[l] \arrow[u] \arrow[d] \\
r_7 \arrow[r] & g_3 & r_8 \arrow[l] \arrow[r] & g_4 & r_7 \arrow[l] \\
b_1 \arrow[r] \arrow[u] & r_1 \arrow[u] & b_2 \arrow[l] \arrow[r] \arrow[u] & r_2 \arrow[u] & b_1 \arrow[l] \arrow[u]
\end{tikzcd}$$

However, if we are given a diagram of this shape in a groupoid $\mathcal{A}$, in which all but one of the cells commute, the assertion that the final cell commutes is equivalent to saying that the two automorphisms of b_1 obtained by composing along the horizontal and vertical edges of the square commute with each other. And since neither of these automorphisms is required to be the identity, there is no reason in general why they should commute.

Remark 1·8. On the other hand, we note that the proof of Lemma 1·6 works equally well for any diagram whose shape is a quadrangulation of the sphere, whether or not its vertices can be coloured in such a way as to make it a quadrangle club, provided that its edges can be oriented in such a way that it makes sense to ask whether each face commutes. More generally, if we are given any polyhedral decompositon of the sphere, and any suitable choice of orientations for all its edges, then each of the axioms which says that the commutativity of all but one of the faces forces the commutativity of the last face must hold in any category embeddable in a groupoid. We shall return to this point in the next section, where we consider triangulations.

We next dispose of the improper quadrangle clubs.

LEMMA 1·9. *Given any quadrangle club $\mathcal{C}$, there exists a finite set $\mathcal{S}$ of proper quadrangle clubs such that $\mathcal{S}$-quadrangularity implies $\mathcal{C}$-quadrangularity. In particular, a category is totally quadrangular iff it is properly quadrangular.*

Proof. Suppose $\mathcal{C}$ is improper, and let

$$\begin{array}{ccc} b & \xrightarrow{p_1} & r_1 \\ {\scriptstyle p_2}\downarrow & & \downarrow{\scriptstyle y_1} \\ r_2 & \xrightarrow{y_2} & g \end{array}$$

be a 'phantom quadrangle' of $\mathcal{C}$ (that is, one not occurring in the distinguished set $\mathcal{Q}$). If we regard $\mathcal{C}$ as a quadrangulation of a sphere, then the boundary of this phantom quadrangle will divide the surface of the sphere into two discs D_1, D_2. We may form two new quadrangle clubs $\mathcal{C}_1$, $\mathcal{C}_2$ by saying that $\mathcal{C}_i$ contains (the vertices and edges of) all those distinguished quadrangles of $\mathcal{C}$ which lie in the disc D_i, together with the given phantom quadrangle.

Now a typical Horn axiom derived from $\mathcal{C}$ will have the form

$$(\phi_1, \phi_2, \ldots, \phi_n, \psi_1, \ldots, \psi_{m-1} \vdash \psi_m)$$

where the ϕ_i and ψ_j respectively denote the assertions that the quadrangles lying in D_1 and in D_2 commute. But from $\mathcal{C}_1$ we may obtain the Horn axiom

$$(\phi_1, \phi_2, \ldots, \phi_n \vdash y_1 p_1 = y_2 p_2),$$

and from $\mathcal{C}_2$ we obtain

$$(y_1 p_1 = y_2 p_2, \psi_1, \ldots, \psi_{m-1} \vdash \psi_m) .$$

Clearly, the validity of the first displayed sequent above follows from that of the other two.

Of course, the new quadrangle clubs $\mathcal{C}_1$ and $\mathcal{C}_2$ need not be proper. But they certainly contain fewer phantom quadrangles than $\mathcal{C}$; so by iterating this process we may arrive at a finite set $\mathcal{S}$ of proper quadrangle clubs with the property in the statement of the Lemma. □

Remark 1·10. By a slight extension of the above argument, we may fulfil a promise made earlier, and eliminate from our consideration all quadrangle clubs where two distinct quadrangles have both a blue and a green vertex in common. For if q_1 and q_2 are two such quadrangles, then they can have at most one purple and one yellow edge in common (otherwise they would be identical), so by taking the 'other' pair of edges from each we obtain a quadrangle which must be phantom unless there are only three quadrangles in the club altogether, i.e. unless $\mathcal{C}$ is the club $\mathcal{D}_3$ described earlier. But the axioms which we obtain from $\mathcal{D}_3$ are of the form

$$(ab = cd, cd = ef \vdash ab = ef) ,$$

and they are vacuously true in any category. Thus we conclude that *a category is totally quadrangular iff it is properly blue-green quadrangular*, i.e. satisfies the Horn axioms arising from proper quadrangle clubs satisfying the blue-green condition.

2. *Quadrangle Clubs and Triangle Clubs*

To many people, the notion of a triangulation of a 2-manifold is more familiar than that of a quadrangulation which we introduced in the last section: it is therefore appropriate to ask whether one could work with triangulations instead of quadrangulations. The answer is easily seen to be yes:

Definition 2·1. Given a quadrangle club $\mathcal{C}$ satisfying the blue-green condition, we define its *associated triangle club* $\widehat{\mathcal{C}}$ to be the directed graph obtained from $\mathcal{C}$ by adding a directed edge (which we think of as being coloured turquoise) from the blue vertex to the green vertex of each quadrangle in the distinguished set $\mathcal{Q}$.

We shall not give a formal combinatorial definition of a triangle club, along the lines of Definition 1·1: the reader may easily formulate one for himself if he wishes. Geometrically, it may be described as a triangulation of the 2-sphere (in the usual sense of that term, cf. [**11**], 2.3.12) whose vertices are coloured blue, red or green in such a way that each 2-simplex has one vertex of each colour. (The colouring of vertices of course determines the orientations, as well as the colouring, of the 1-simplices.) It is not hard to see that a triangulation of the 2-sphere admits such a colouring iff the valency of each vertex is even. (We note, incidentally, that in the passage from $\mathcal{C}$ to $\widehat{\mathcal{C}}$ the valency of every blue or green vertex is doubled, whereas that of each red vertex is unchanged.)

If $\mathcal{C}$ is a blue-green quadrangle club, it is clear that a diagram of shape $\mathcal{C}$ in a category, in which all but at most one of the quadrangles in $\mathcal{Q}$ commute, may be extended (in either one or two ways) to a diagram of shape $\widehat{\mathcal{C}}$ in which at most one of the oriented 2-simplices in $\widehat{\mathcal{C}}$ is not mapped to a commutative triangle. Conversely, any diagram of shape $\widehat{\mathcal{C}}$ may be restricted to one of shape $\mathcal{C}$, again with the same commutativity conditions. Thus, if we define 'triangularity axioms' analogously with the quadrangularity axioms of the last section, and extend the notation $(\mathcal{C} \vdash \mathcal{C}')$ to cover triangle clubs as well as quadrangle clubs, we have both $(\widehat{\mathcal{C}} \vdash \mathcal{C})$ and $(\mathcal{C} \vdash \widehat{\mathcal{C}})$. Since we already saw in Remark 1·10 that it is sufficient to restrict our attention to blue-green quadrangle clubs, it follows that we could equally well work with triangle clubs:

PROPOSITION 2·2. *For any set $\mathcal{S}$ of quadrangle clubs, there is a set $\widehat{\mathcal{S}}$ of triangle clubs such that a category is $\mathcal{S}$-quadrangular if and only if it is '$\widehat{\mathcal{S}}$-triangular', i.e. satisfies all Horn axioms derived from triangle clubs in $\widehat{\mathcal{S}}$.*

Our reasons for preferring quadrangle clubs are two-fold. First, the axioms are simpler to state: in passing from $\mathcal{C}$ to $\widehat{\mathcal{C}}$, we increase the number of variables in the corresponding Horn axioms by 50%, and more than double the number of hypotheses in each axiom. But, more importantly, the notion of propriety which we defined for quadrangle clubs does not seem to translate very naturally into the language of triangle clubs.

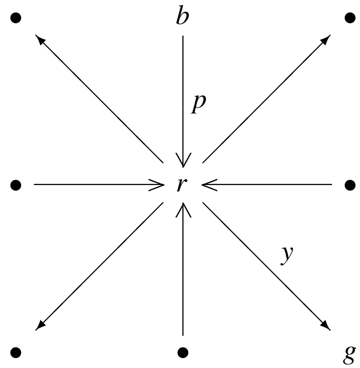
To illustrate this, let us remark that, if a triangulation of the sphere admits a vertex colouring making it into a triangle club, it actually admits six such colourings, corresponding to the six possible permutations of the three colours; thus we may obtain several bijective operations on the set of all blue-green quadrangle clubs, by passing from a quadrangle club to its associated triangle club, permuting the colours of its vertices, and then deleting the turquoise edges (that is, the edges which have become turquoise in the new colouring). One of these operations, corresponding to the transposition which interchanges blue and green, is simply dualization; but the others are less obvious, and they do not respect propriety. For example, if we take the proper quadrangle club $\mathcal{W}_{2n}$ defined in the previous section, and apply the transposition which interchanges red and green, we obtain the club $\mathcal{E}_{2n}$, which is improper for $n \geqslant 3$.

3. *Minimization*

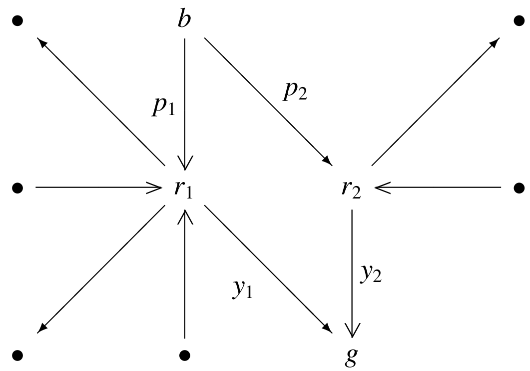
In this section, we describe operations on quadrangle clubs which enable us to restrict our attention to ones which are minimal in the sense of Definition 1·2 (and even sesquiminimal). We begin with red-minimality (that is, the Mal'cev condition).

LEMMA 3·1. *For any quadrangle club $\mathcal{C}$ whose red vertices all have valency at least* 4*, there exists a Mal'cev quadrangle club $\widetilde{\mathcal{C}}$ such that $(\widetilde{\mathcal{C}} \vdash \mathcal{C})$. In particular, a category is totally quadrangular iff it is Mal'cev quadrangular.*

Proof. Suppose $\mathcal{C}$ is a quadrangle club having a red vertex r of valency $2n$ for some $n > 2$ (and none of valency 2). Then by 'doubling up' the vertex r and a suitable pair of edges (p, y) we may create a new club $\mathcal{C}_1$ in which r has been replaced by two vertices of valencies 4 and $2n - 2$. The pictures below illustrate the case $n = 4$:



may be replaced by



Moreover, any diagram of shape $\mathcal{C}$ in a category $\mathcal{A}$ may be regarded as a diagram of shape $\mathcal{C}_1$ in which the pairs (p_1, p_2) and (y_1, y_2) have each been mapped to the same morphism of $\mathcal{A}$ (so that the additional quadrangle of $\mathcal{C}_1$ automatically commutes). Thus the validity of any Horn axiom derived from $\mathcal{C}$ follows from that of one derived from $\mathcal{C}_1$. By repeating this process enough times, we will eventually arrive at a Mal'cev quadrangle club $\widetilde{\mathcal{C}}$ with the property in the first statement of the Lemma. The second statement follows from this and the observation in Remark 1·10 that we may eliminate quadrangle clubs having red vertices of valency 2. $\square$

The 'blowing-up' process in the foregoing proof may destroy the blue-green property, since the vertices b and g which we picked out may already have some other quadrangle of $\mathcal{C}$ in common. However, we note:

LEMMA 3·2. *A Mal'cev quadrangle club is proper iff it satisfies the blue-green condition.*

Proof. One direction follows from Remark 1·10 (note that the club $\mathcal{D}_3$ does not satisfy the Mal'cev condition). Conversely, suppose $\mathcal{C}$ is a Mal'cev quadrangle club containing a 'phantom quadrangle'

$$\begin{tikzcd} b \arrow[r, "p_1"] \arrow[d, "p_2"] & r_1 \arrow[d, "y_1"] \\ r_2 \arrow[r, "y_2"] & g \end{tikzcd}$$

Since the vertex r_1 has valency 4, the edges p_1 and y_1 must also belong to a 'real quadrangle' lying in the surface of the sphere; and similarly for p_2 and y_2. Thus we deduce that there are two 'real quadrangles' of $\mathcal{C}$ sharing a common blue vertex b and a common green vertex g. $\square$

Moreover, if $\mathcal{C}$ is a Mal'cev quadrangle club containing a 'phantom quadrangle' as above, and we use the technique of Lemma 1·9 to 'split' it into two clubs $\mathcal{C}_1$ and $\mathcal{C}_2$, each of the latter will have a single red vertex of valency 2 (and all other red vertices of valency 4), so that if we further 'split off' a copy of $\mathcal{D}_3$ from each we shall arrive at a pair of Mal'cev quadrangle clubs $\mathcal{C}_1'$, $\mathcal{C}_2'$. Thus we may conclude:

PROPOSITION 3·3. *A category is totally quadrangular iff it is properly Mal'cev quadrangular.*

Next, we consider green-minimality, i.e. the Lambek condition. Here we need a 'subdivision' rather than a 'blowing-up' process:

LEMMA 3·4. *For any quadrangle club $\mathcal{C}$ satisfying the blue-green condition, there exists a Lambek quadrangle club $\overline{\mathcal{C}}$ such that ($\overline{\mathcal{C}} \vdash \mathcal{C}$). In particular, total quadrangularity is equivalent to Lambek quadrangularity.*

Proof. We obtain $\overline{\mathcal{C}}$ from $\mathcal{C}$ on 'drawing in the diagonals' of all its quadrangles, replacing each green vertex g by a red vertex $[g]$ at the same point, and adding a new green vertex at the mid-point of each yellow edge of $\mathcal{C}$. The following picture illustrates the case of a green vertex of valency 4:

$$\begin{tikzcd} b_1 \arrow[r] \arrow[d] & r_1 \arrow[d] & b_2 \arrow[l] \arrow[d] \\ r_2 \arrow[r] & g & r_3 \arrow[l] \\ b_3 \arrow[r] \arrow[u] & r_4 \arrow[u] & b_4 \arrow[l] \arrow[u] \end{tikzcd}$$

is replaced by

$$
\begin{tikzcd}
b_1 \arrow[rr] \arrow[dd] \arrow[ddrr] & & r_1 \arrow[d] & & b_2 \arrow[ll] \arrow[dd] \arrow[ddll] \\
& & g_1 & & \\
r_2 \arrow[r] & g_2 & {[g]} \arrow[l] \arrow[r] \arrow[u] \arrow[d] & g_3 & r_3 \arrow[l] \\
& & g_4 & & \\
b_3 \arrow[rr] \arrow[uu] \arrow[uurr] & & r_4 \arrow[u] & & b_4 \arrow[ll] \arrow[uu] \arrow[uull]
\end{tikzcd}
$$

A diagram of shape $\mathcal{C}$ in a category, in which at most one of the quadrangles fails to commute, may be extended to a diagram of shape $\overline{\mathcal{C}}$ with at most one non-commuting quadrangle, by mapping each edge with domain $[g]$ to an identity morphism, and each edge with codomain $[g]$ to the diagonal of the corresponding square in the diagram of shape $\mathcal{C}$ (or, if that square is the non-commuting one, to one of the two possible composites). Hence every quadrangularity axiom derived from $\mathcal{C}$ follows from one derived from $\overline{\mathcal{C}}$; but $\overline{\mathcal{C}}$ clearly has the Lambek property. The second assertion of the Lemma follows, once again, from the fact that we may disregard quadrangle clubs which fail to satisfy the blue-green condition (Remark 1·10). $\square$

It is not hard to see that the subdivision process of Lemma 3·4, unlike the construction of 3·1, cannot destroy propriety or the blue-green property — recall that we have already observed that every Lambek quadrangle club satisfies the blue-green condition. But in any case, the splitting process of Lemma 1·9 cannot destroy the Lambek property — if the 'split' involves a green vertex of valency n, the latter is replaced by green vertices of valencies $p+1$ and $q+1$ where $p+q=n$, so that if $n=2$ we must have $p=q=1$. Thus we may deduce:

PROPOSITION 3·5. *A category is totally quadrangular iff it is properly Lambek quadrangular.*

We should mention that Bush [**1**] showed that each Horn axiom derived using Mal'cev's algorithm is a consequence of a single axiom derived using Lambek's algorithm, and vice versa: this may be viewed as a version of our Lemmas 3·1 and 3·4. However, his argument was entirely different, and depended on the proofs by Mal'cev and Lambek that their axiom-systems suffice for embeddability in a group.

Finally in this section, we consider sesquiminimality. To achieve this, we need to use a more drastic subdivision process, which was developed by Krstić [**6**]. Here we begin by replacing a (blue-green) quadrangle club $\mathcal{C}$ by its associated triangle club $\widehat{\mathcal{C}}$, as defined in the previous section, and we then form the barycentric subdivision (in the usual sense of topologists, [**11**], 2.5.7) of its underlying simplicial complex. The latter can then be made into a triangle club in six different ways: if the three colours are X, Y, Z in some order, we apply colour Z to each vertex of the original simplicial complex, colour X to the mid-point of each edge and colour Y to the centre of each 2-simplex. It is clear that if we do this, each X vertex has valency 4 and each Y vertex has valency 6; so that if we then convert it back into a quadrangle club, by deleting the turquoise edges, the result will be (X, Y)-sesquiminimal. We shall denote this quadrangle club by $\widehat{\mathcal{C}}_{X,Y}$.

In the four cases when Y is 'blue' or 'green', it is then straightforward to verify that, if we have a diagram of shape $\widehat{\mathcal{C}}$ in a category $\mathcal{A}$ with at most one non-commuting triangle, we may extend it to a diagram of shape $\widehat{\mathcal{C}}_{X,Y}$ with at most one non-commuting quadrangle: each edge of $\widehat{\mathcal{C}}_{X,Y}$ is mapped either to the image of one of the edges of the triangle in which it lies, or to an identity morphism, in such a way that two of the three quadrangles into which a given triangle is subdivided trivially commute, and the third commutes if and only if the triangle did so.

Unfortunately, this argument cannot be made to work in the cases when Y is 'red'. But we may obtain the same result in these cases by combining the construction above with that of Lemma 3·4. For example, to achieve (green,red)-sesquiminimality, we first replace our given club $\mathcal{C}$ by $\widehat{\mathcal{C}}_{\mathrm{red,green}}$, and then apply the construction of Lemma 3·4 to the latter. In the resulting club, the red vertices which derive from red vertices of $\widehat{\mathcal{C}}_{\mathrm{red,green}}$ still have valency 4, as they did in $\widehat{\mathcal{C}}_{\mathrm{red,green}}$; and those which appear as replacements for green vertices of $\widehat{\mathcal{C}}_{\mathrm{red,green}}$ have valency 6. So we may conclude:

LEMMA 3·6. *Let X and Y be any two distinct members of the set* {*blue, red, green*}. *For any blue-green quadrangle club $\mathcal{C}$, there exists an (X, Y)-sesquiminimal quadrangle club $\widetilde{\mathcal{C}}_{X,Y}$ such that $(\widetilde{\mathcal{C}}_{X,Y} \vdash \mathcal{C})$. In particular, total quadrangularity is equivalent to (X, Y)-sesquiminimal quadrangularity.*

As we have previously observed, the splitting process by which we reduce an arbitrary quadrangle club to a set of proper clubs cannot increase the valency of vertices of any given colour. Thus we may once again strengthen Lemma 3·6 by adding the word 'proper' to its conclusion.

One might ask whether it is possible to find a sufficient set of quadrangle clubs in which the valencies of all three colours of vertices are bounded above. In fact it is not hard to answer this, using the results at our disposal. If we take a (green,blue)-sesquiminimal quadrangle club $\mathcal{C}$ and apply the blowing-up process of Lemma 3·1 to reduce the valencies of its red vertices, it is easy to see that, for any given red vertex, each of its neighbour vertices does not need to be involved more than once, so that the valency of a blue or green vertex will at most double when all red vertices are dealt with. Thus we have:

PROPOSITION 3·7. *Let $\mathcal{S}$ be the set of proper quadrangle clubs in which each red vertex has valency* 4, *each green vertex has valency at most* 4 *and each blue vertex has valency at most* 6. *Then $\mathcal{S}$-quadrangularity is equivalent to total quadrangularity.*

It seems unlikely that this result is best possible, but we have not been able to improve on it. (A 'best conceivable' result would have red, blue and green valencies bounded by 2ρ, β and γ respectively where $1/\rho + 1/\beta + 1/\gamma = 1$; for any triple with $1/\rho + 1/\beta + 1/\gamma > 1$ would, in conjunction with the Euler condition, impose an upper bound on the total number of quadrangles — for example, putting $(\rho, \beta, \gamma) = (2, 3, 5)$ would yield $Q + 2 \geqslant Q/2 + Q/3 + Q/5 = 31Q/30$, so that $Q \leqslant 60$. And this contradicts Corollary 7·2 below.)

4. *From Mal'cev Bracketings to Quadrangle Clubs*

Mal'cev's solution to the problem of when a semigroup is embeddable in a group was originally published in [**9**]; an accessible account can be found in Cohn's book [**4**]. We shall give a brief account here, largely following Cohn's notation. Note first that we may assume without loss of generality that our semigroup is a monoid (i.e., has an identity element), since we may always adjoin an identity to a semigroup which does not possess one.

We consider the monoid G freely generated by the elements of our given monoid M, together with new elements p^+ and p^- for each element p of M, subject to relations which say that the elements of M multiply as they do in M (and in particular the identity element of M remains the identity of G), and the equations $p^-p=1$ and $pp^+=1$ hold for every $p \in M$. Clearly, in G we have $p^+=p^-pp^+=p^-$ for any p, so that G is in fact the group freely generated by M (that is, the reflection of M in the subcategory of groups); the reason for introducing separate symbols to denote the left and right inverses of p is that it simplifies the solution of the word problem for this presentation.

Clearly, two elements x, y of M become equal in G iff it is possible to connect them by a sequence $(x=w_0, w_1, w_2, \ldots, w_n=y)$ of words in the symbols p, p^+, p^-, where for each i the step from w_{i-1} to w_i is of one of the forms: replace a product of elements of M by another product having the same value in M, insert or delete a pair p^-p, or insert or delete a pair pp^+. We denote the insertion or deletion of a pair p^-p by an opening or closing round bracket, respectively, and the insertion or deletion of a pair pp^+ by an opening or closing square bracket; rewritings of products in M are not denoted by any symbol. Thus the sequence generates a sequence of brackets such as ([ )], in which each opening round bracket is matched by a later closing round bracket (since any symbol p^- which is introduced must eventually be removed), and similarly for square brackets. We shall call such a sequence a *Mal'cev bracketing*.

We may assume that,between the introduction of a pair p^-p and its later removal, no changes take place in the part of the word which lies to the left of p^-, since any such changes could have been postponed until after p^- had been removed. Similarly, we may assume that all changes take place to the left of any p^+ while it is present. (In this connection, we define the *active region* of a word to be the part of the word between the last occurrence of a p^- (or the beginning of the word if no p^- is present) and the first occurrence of a p^+ (or the end of the word if no p^+ is present); thus the convention just described may be expressed by saying that all changes take place within the active region of the current word.) With this convention, it is possible to reconstruct the entire sequence of words, up to renaming of elements of M, from the bracketing. For example, the bracketing ([ )] yields the sequence

$$(x, ab, ap^-pb, ap^-cd, ap^-cqq^+d, ap^-peq^+d, aeq^+d, fqq^+d, fd, y)$$

where $x=ab$, $pb=cd$, $cq=pe$, $ae=fq$ and $fd=y$ are equations which hold in M. (Note that there is no loss of generality in assuming that, for example, qq^+ is inserted in the middle of a product cd; if we wanted (say) to insert it adjacent to p^-, we could take c to be the identity element of M.) Thus we conclude that if M embeds in G, then the validity of the equations $pb=cd$, $cq=pe$ and $ae=fq$ (for any assignment of values to a, b, c, d, e, f, p and q) must imply $ab=fd$. (Note that this is, apart from changes of notation, exactly the Horn axiom which we derived from the quadrangle club $\mathcal{E}_4$ in Definition 1·4.) In this way, each bracketing yields a universal Horn axiom which must be satisfied if M is to be embeddable in a group. Conversely, if M satisfies all such Horn axioms, then it is embeddable in a group; for in any sequence $(x=w_0, w_1, \ldots, w_n=y)$ as above, x and y will already be equal in M.

The 'active region' convention of the previous paragraph ensures that no two pairs of brackets of the same shape can be 'interleaved'; that is, in a bracketing such as (([ )]), the second (is paired with the first), since the pair q^-q to whose introduction it corresponds must occur after the p^- that was introduced by the first (, and must therefore be eliminated before p^-p. So we do not need to indicate explicitly how the brackets are paired. We note that the bracketing ( ) corresponds to the condition that $pb = pc$ implies $ab = ac$, from which we may derive the left cancellation law ($pb = pc \ \vdash \ b = c$) by setting $a = 1$; and similarly [ ] yields the right cancellation law. However, these cancellation laws may also be derived from the Horn axiom corresponding to ([ )], as we saw in the proof of Lemma 1·5. Thus we may omit the two 2-term bracketings; indeed, we may adopt the further convention of considering only bracketings in which each pair of round brackets is separated by at least one (opening or closing) square bracket, and vice versa. For a bracketing of the form $\cdots ( ) \cdots$ will correspond to a sequence of words of the form

$$(\ldots, vbw, vp^- pbw, vp^- pcw, vcw, \ldots)$$

(where v and w are words possibly involving q^-'s and r^+'s, respectively), and if the cancellation laws hold in our monoid then we will necessarily have $b = c$ — so that the displayed part of the sequence of words might as well have been omitted.

A slightly subtler point is that we can dispense with bracketings in which the first and last bracket are paired with each other. To see this, let us analyse an example: the bracketing ([( ])). This corresponds to the sequence of words

$$(ab, ap^- pb, ap^- cd, ap^- cqq^+ d, ap^- efq^+ d, ap^- er^- rfq^+ d, ap^- er^- gqq^+ d, ap^- er^- gd, ap^- er^- rh, ap^- eh, ap^- pj, aj)$$

and hence to the Horn axiom

$$(pb = cd, cq = ef, rf = gq, gd = rh, eh = pj \vdash ab = aj)\,.$$

But the bracketing obtained from this one by omitting the outermost pair of brackets would yield the sequence of words obtained from the one above by omitting the first two and last two terms, and deleting the initial ap^- from the rest; thus it would yield the Horn axiom

$$(cq = ef, rf = gq, gd = rh \vdash cd = eh)$$

whence the hypotheses of the earlier Horn axiom would yield $pb = cd = eh = pj$ and the left cancellation law would yield $b = j$. So the condition obtained from the bracketing ([( ])) follows from that obtained from [( ]) plus the cancellation laws.

We shall call a bracketing *nontrivial* if it satisfies the conditions that each pair of round brackets is separated by at least one square bracket, each pair of square brackets is separated by at least one round bracket, and the first and last brackets are not paired with each other. We shall also say that a bracketing is *connected* if it cannot be broken into two sub-bracketings; that is, there is no point within the sequence such that there are as many closing as opening brackets, of both shapes, before that point. Clearly, if we have a bracketing which is not connected, the Horn axiom which we get from it will be of the form

$$(\phi_1, \ldots, \phi_n, cd = ef, \psi_1, \ldots, \psi_n \vdash ab = gh)$$

(where the ϕ_i and ψ_j denote equations), and the conditions obtained from the two sub-bracketings are

$$(\phi_1, \ldots, \phi_n \vdash ab = cd) \quad \text{and} \quad (\psi_1, \ldots, \psi_m \vdash ef = gh),$$

so that the axiom derived from the disconnected bracketing follows from those derived from its sub-bracketings. Thus we can strengthen Mal'cev's result, mentioned above:

PROPOSITION 4·1. *A monoid is embeddable in a group if and only if it satisfies all universal Horn axioms derived from connected nontrivial Mal'cev bracketings.*

(We shall shortly improve this result further by introducing a slightly stronger notion of connectedness.)

It is easy to see that Mal'cev's solution of the word problem for the group freely generated by a monoid works equally well for the groupoid freely generated by a category. Thus we may obtain from each Mal'cev bracketing a Horn axiom which must be satisfied by any category embeddable in a groupoid, and such that the satisfaction of all such axioms is sufficient to ensure such embeddability, by interpreting the variables as ranging over morphisms of the category, and assigning domains and codomains to them in such a way that these coincide only when forced to do so by the definability of composites, or by equations between such composites. We next aim to show that, for nontrivial bracketings, these Horn axioms are precisely those derivable from quadrangle clubs satisfying the Mal'cev condition.

To explain this, we shall consider another example, namely the bracketing ([ ) ( ]). This yields the sequence of words

$$(\underline{ab}, ap^-\underline{pb}, ap^-\underline{cd}, ap^-\underline{cq}q^+d, ap^-\underline{pe}q^+d, \underline{ae}q^+d, \underline{fg}q^+d, fr^-\underline{rg}q^+d, fr^-\underline{hq}q^+d,$$
$$fr^-\underline{hd}, fr^-\underline{rj}, \underline{fj})$$

in which we have emphasized the active region of each word by underlining it. From this we obtain the Horn axiom

$$(pb = cd, cq = pe, ae = fg, rg = hq, hd = rj \vdash ab = fj)\,.$$

We note that each variable appears twice, in two different equations. (The nontriviality conditions are exactly what we need to ensure that a variable does not appear on both sides of the same equation. For if p appears together with p^-, then it is present in the equations immediately after the opening round bracket which corresponds to its introduction, and immediately before the closing one which corresponds to its elimination; but these two brackets cannot be adjacent. Similarly if q appears together with q^+. For any other variable, its two appearances in equations are immediately before an opening bracket and immediately after the matched closing bracket; and these could only be the same equation if the first and last brackets were matched.)

Moreover, each variable appears either as the left factor of a product in both cases (in which case we consider the edge it denotes to be coloured yellow), or as the right factor in both cases (in which case the corresponding edge is coloured purple). In particular, if a variable p is introduced as part of a pair $p^- p$, it is necessarily a yellow edge; if as part of a pair pp^+, it is purple. For the other letters, if one is introduced as the left half of the active region, then it is yellow; if as the right half, then it is purple.

We may similarly colour the vertices (objects) which occur as the domains and codomains of the edges, as follows. Immediately before each opening bracket, we introduce a new red vertex, which is the domain of p if we are introducing a pair $p^- p$, or the codomain of q if we are introducing a pair qq^+. Thus, in the particular sequence above, the initial red vertex is the domain of a, the codomain of b and the domain of p; and it must also occur as the codomain of e (the variable introduced immediately before p is eliminated again). Similarly, the red vertex associated with q is the domain of c, the codomain of d, the codomain of q and the domain of h (the variable introduced just before q is eliminated). We have one blue vertex present at the start (the domain of b) and one new one is introduced at each opening square bracket, to be the domain of the morphism which is introduced along with its right inverse. Similarly, there is one green vertex (the codomain of a) present at the start, and one is introduced at each opening round bracket.

Thus, if we start from a bracketing with ρ pairs of round brackets and σ pairs of square ones, we obtain a diagram with $B = \sigma + 1$ blue vertices, $G = \rho + 1$ green vertices, and $R = \rho + \sigma$ red ones. The number of quadrangles (and of purple and yellow arrows) in the diagram is of course equal to the number of equations (including the conclusion) in the Horn axiom; since the number of hypotheses is easily seen to be one less than the total number of brackets, we therefore have $Q = 2(\rho + \sigma)$. It is thus clear that the Euler condition $B + R + G = Q + 2$ of 1.1(5) is satisfied, as is the Mal'cev condition $Q = 2R$.

It remains only to verify the connectivity conditions of 1.1(4). The connectedness of the whole graph is clear, since each variable denoting an edge appears in an equation immediately before or immediately after it has been introduced into the current word, and so by tracing back through these equations we can link it to one of the edges present at the beginning. The connectivity condition on the edges at a red vertex is clear (we have already seen that there are just four edges at each such vertex); for blue vertices it is also clear, since each time a new edge is introduced with a given blue vertex as domain it is linked by an equation to the previous purple edge with the same domain (for example, d and j are the edges having the same domain as b, and they are introduced via the equations $pb = cd$ and $hd = rj$), and similarly for green vertices.

Thus we may conclude:

LEMMA 4·2. *Every universal Horn axiom obtained from a nontrivial Mal'cev bracketing may also be obtained from a quadrangle club satisfying the Mal'cev condition.*

COROLLARY 4·3. *A category admits a faithful functor to a groupoid iff it is Mal'cev quadrangular.*

In the converse direction, we claim that each universal Horn axiom derived from a Mal'cev quadrangle club may also be derived from a Mal'cev bracketing. We shall not give a detailed proof of this, but we sketch how the bracketing may be produced. Suppose that the diagram below represents part of a Mal'cev quadrangle club (in which, for the moment, the labels attached to the edges should be ignored), and that the leftmost cell in the diagram is the one corresponding to the conclusion of the Horn axiom under consideration.

$$\begin{tikzcd}
 & & \bullet & \\
\bullet \arrow[r, "b"] \arrow[d] & \bullet \arrow[d, "a"] \arrow[ur, "g"] & \bullet \arrow[l, "p"'] \arrow[r, "f"] \arrow[d, "d"] & \bullet \arrow[d, "e"] \arrow[ul, "r"'] \\
\bullet \arrow[r] & \bullet & \bullet \arrow[l, "c"] \arrow[r, "q"] & \bullet
\end{tikzcd}$$

We label two edges of this cell a and b, as shown, and proceed to 'drag' the composite ab across the other quadrangles as follows (attaching labels to the edges of the quadrangles as we go, and writing p^+ or p^- as appropriate when we need to traverse an edge p in the wrong direction):

$$(ab, app^+b, cdp^+b, cq^-qdp^+b, cq^-efp^+b, cq^-er^-rfp^+b,$$
$$cq^-er^-gpp^+b, cq^-er^-gb, \dots)$$

At every odd step of the process, we either introduce a pair pp^+ or q^-q at the red vertex corresponding to the middle of the 'active region' of our current word, where p or q is one of the edges at that vertex not yet labelled, or else we eliminate such a pair if there is one; at even steps, we rewrite the active region of the word by going the other way round a quadrangle. Because the directed graph is finite, the process must eventually terminate with all symbols of the form p^+ or q^- eliminated; when this happens, we must have arrived at a two-letter word representing the other two edges of the quadrangle from which we started. The resulting sequence of words is clearly derivable from a Mal'cev bracketing, and the universal Horn axiom derived from it is exactly that derived from the quadrangle club. (It will be seen that there is a good deal of choice about the order in which we 'use' the quadrangles of our club; thus several different bracketings may give rise to the same axiom.)

Since we know that we do not need to consider all Mal'cev quadrangle clubs, but only proper ones, it is natural to ask what property of a Mal'cev bracketing corresponds to propriety of the corresponding quadrangle club. The answer is a slight strengthening of the notion of connectedness which we introduced before Proposition 4·1. We recall from Lemma 3·2 that propriety of Mal'cev quadrangle clubs is equivalent to the blue-green condition; if this condition fails for the club generated by a Mal'cev bracketing, then there must be two points in the corresponding sequence of words (counting the first and last word as the same point) where the beginning and end of the active region are in the same place: that is, either there is a point in the middle of the sequence at which the active region is the whole word (which corresponds to disconnectedness as we defined it earlier) or there are two distinct points in the sequence such that the p^+'s and/or q^-'s present at the first point are exactly those present at the second point. Accordingly, we declare a Mal'cev bracketing to be *strongly connected* if there are not two distinct points in the sequence (other than the beginning and the end) such that every opening bracket between them is paired with a closing bracket between them, and vice versa. Thus we may conclude:

COROLLARY 4·4. *A category admits a faithful functor to a groupoid iff it satisfies all universal Horn axioms derived from strongly connected nontrivial Mal'cev bracketings.*

5. *Topos-theoretic Interlude*

As is well known, many properties of a small category $\mathcal{A}$ are reflected in topos-theoretic properties of the corresponding topos of presheaves $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$. We mention two examples which are relevant to our present considerations: all references in this section are to the book [**5**].

PROPOSITION 5·1. *Let $\mathcal{A}$ be a small category.*

(i) *Every morphism of $\mathcal{A}$ is epic iff* $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$ *admits a localic geometric morphism to a Boolean topos.*
(ii) *Every morphism of $\mathcal{A}$ is monic iff* $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$ *is an étendue (i.e., admits a surjective local homeomorphism from a topos localic over* **Set***).*

Proof. See C5.4.4 and C5.2.4 respectively. □

In this section, our purpose is to add a new result to the list; recall that a geometric morphism $f : \mathcal{F} \to \mathcal{E}$ between toposes is said to be *essential* if the inverse image functor $f^* : \mathcal{E} \to \mathcal{F}$ has a left adjoint $f_!$, as well as its usual right adjoint f_*.

PROPOSITION 5·2. *Let $\mathcal{A}$ be a small category. Then $\mathcal{A}$ is (properly Mal'cev) quadrangular iff* $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$ *admits an essential localic geometric morphism to a Boolean topos.*

Proof. One direction is easy: If $\mathcal{A}$ is quadrangular, then by the results of the last section we have a faithful functor $F : \mathcal{A} \to \mathcal{B}$ where $\mathcal{B}$ is a groupoid. The topos $[\mathcal{B}^{\mathrm{op}}, \mathbf{Set}]$ is Boolean by A1·4·12; and the geometric morphism $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}] \to [\mathcal{B}^{\mathrm{op}}, \mathbf{Set}]$ induced by F is essential by A4·1·4 and localic by A4·6·2(c).

Conversely, suppose we have an essential localic morphism $f : [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}] \to \mathcal{E}$ where $\mathcal{E}$ is Boolean. We observe first that we may assume without loss of generality that f is surjective (i.e., f^* is faithful); for if we form the image factorization

$$[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}] \xrightarrow{\ g\ } \mathcal{E}' \longrightarrow \mathcal{E}$$

of f, then $\mathcal{E}'$ is Boolean (A4·5·22), and g is still (localic and) essential. (Although $f_!$ does not in general factor through the inclusion $\mathcal{E}' \to \mathcal{E}$, we may obtain a left adjoint for g^* by composing $f_!$ with the associated sheaf functor $\mathcal{E} \to \mathcal{E}'$.)

Now, since $f_!$ has a right adjoint which preserves epimorphisms and coproducts, a standard argument shows that it preserves indecomposable projectives; hence, for every object a of $\mathcal{A}$, the representable functor $Y(a) = \mathcal{A}(-, a)$ is mapped by $f_!$ to an indecomposable projective. But a similar argument using the fact that f^* is faithful shows that $f_!$ preserves separating families; so $\mathcal{E}$ has a separating family of indecomposable projectives. Hence by C2·2·20, $\mathcal{E}$ is a presheaf topos $[\mathcal{B}^{\mathrm{op}}, \mathbf{Set}]$; and by A1·4·12, $\mathcal{B}$ is a groupoid. Moreover, since f is essential, it must be induced by a functor $F : \mathcal{A} \to \mathcal{B}$ (A4·1·5), and since f is localic this functor must be faithful (A4·6·9). So $\mathcal{A}$ is quadrangular. □

It is not clear what, if any significance the condition of admitting an essential localic morphism to a Boolean topos has for a general (Grothendieck) topos. It does not imply the property of being an étendue (since not every Boolean topos is an étendue). A topos admits an essential localic morphism to **Set** iff it is the topos of sheaves on a locally connected locale (C1·5·9); but, for morphisms with more general codomains, local connectedness is a stronger property than that of being essential.

6. *From Lambek Polyhedra to Quadrangle Clubs*

In [**7**], Lambek gave an alternative, more geometric, solution to the question 'when is a monoid embeddable in a group?'. Although Lambek's paper does not mention the word 'category' (still less 'groupoid'), it is clear that he thought of the problem in categorical terms, in that the paper contains diagrams of arrows labelled by elements of the monoid, whose commutativity is intended to represent the validity of equations between products in the monoid. It is thus even easier for his method than it was for Mal'cev's to see that precisely the same argument will yield an answer to the question 'when is a category embeddable in a groupoid?'.

Lambek's construction starts from an arbitrary Eulerian polyhedron, that is a decomposition of the 2-sphere into polygonal faces (which may have an arbitrary number $\geqslant 2$ of sides; similarly, vertices may have any valency $\geqslant 2$). Suppose the decomposition has V vertices, E edges and F faces, with $V + F = E + 2$. Although he does not describe it in exactly this way, Lambek constructs from this polyhedron a quadrangle club as follows: place a blue vertex at each vertex of the given polyhedron, a green vertex at the mid-point of each edge, and a red vertex at the centre of each face. Then draw purple edges to the centre of each face from each of its vertices, and yellow edges to the mid-point of each edge from the centre of the face on either side of it. The resulting quadrangle club clearly has V blue vertices, F red vertices and E green ones, and $2E$ quadrangles (since each quadrangle has a half-edge of the original polyhedron as its 'diagonal'). It therefore satisfies the Lambek condition that each green vertex has valency 2.

Conversely, it is easy to see that every quadrangle club satisfying the Lambek condition is obtained in this way from an Eulerian polyhedron. Given such a quadrangle club $\mathcal{C}$ in geometric form, we may reconstruct the polyhedron by taking the blue vertices of $\mathcal{C}$ as vertices of the polyhedron, and, for each pair of quadrangles having a common green vertex, drawing an edge to connect their blue vertices . (For example, if we apply this process to the quadrangle club $\mathcal{E}_4$, we obtain a 'dihedron' having two vertices, two edges and two 2-sided faces. Similarly, for $\mathcal{E}_{2n}$ we obtain a decomposition with two n-sided faces, and for $\mathcal{W}_{2n}$ we obtain one with n two-sided faces.)

By using a different solution of the word problem (for the group freely generated by a monoid) from the one employed by Mal'cev, Lambek was able to show that a monoid is embeddable in a group if and only if it satisfies all the universal Horn axioms derived from these quadrangle clubs. Since we already know that Lambek quadrangularity is equivalent to Mal'cev quadrangularity, we do not need to repeat the proof here; but we remark that it is easy to see that his method extends readily from monoids to categories.

Moreover, in the course of the proof, Lambek uses only polyhedra whose vertices all have valency 3, except for the 'dihedron' giving rise to $\mathcal{E}_4$. Thus his method actually proves that the set of (green, blue)-sesquiminimal quadrangle clubs yields a sufficient set of axioms for embeddability in a groupoid: of course, this is no surprise, given Lemma 3·6.

As we remarked earlier, not every Lambek quadrangle club is proper, but if we apply the decomposition process of Lemma 1·9 to an improper Lambek quadrangle club, the smaller clubs which we construct will all have the Lambek property, since we cannot increase the valency of any green vertex. It is easy to verify that a polyhedral decomposition of the sphere gives rise to a proper Lambek quadrangle club if and only if it has the property that, whenever two faces have a common edge, they have no more than two common vertices. Thus we may conclude that the set of Horn axioms derived from Lambek polyhedra satisfying this condition is sufficient for embeddability.

7. *Insufficient Sets of Conditions*

In [**10**], Mal'cev proved that no finite set of Horn axioms in the language of semigroups could be equivalent to embeddability in a group. The corresponding result for categories and groupoids is much easier to prove, because we do not have to worry about 'unwanted' composites: we give a brief sketch of the proof here.

Given a proper quadrangle club $\mathcal{C}$ (other than $\mathcal{D}_3$), let $\mathcal{A}(\mathcal{C})$ denote the category freely generated by $\mathcal{C}$ subject to the requirement that all the quadrangles in $\mathcal{Q}$ commute. It is easy to see that, in order to enlarge $\mathcal{C}$ to $\mathcal{A}(\mathcal{C})$, all we need to do is to add in identity morphisms at all the vertices of $\mathcal{C}$, plus composites for each pair (y, p) where y is a yellow edge of $\mathcal{C}$ and p is a purple edge whose codomain matches the domain of y. Propriety, and the commutativity conditions we have imposed, ensure that we add at most one morphism from any blue vertex of $\mathcal{C}$ to any green vertex, so that $\mathcal{A}(\mathcal{C})$ is a poset. Moreover, if we regard the objects, non-identity morphisms and commutative triangles of $\mathcal{A}(\mathcal{C})$ as forming a 2-dimensional simplicial complex, the geometric realization of this complex is still homotopy equivalent to a 2-sphere, since it deformation-retracts onto the geometric realization of the triangle club $\widehat{\mathcal{C}}$ of section 2.

Given a particular quadrangle q of $\mathcal{C}$, we write $\mathcal{B}(\mathcal{C}, q)$ for the category obtained by deleting the requirement that q should commute from the presentation of $\mathcal{A}(\mathcal{C})$; once again it is easy to see that the underlying directed graph of $\mathcal{B}(\mathcal{C}, q)$ differs from that of $\mathcal{A}(\mathcal{C})$ only in having two distinct morphisms from the blue to the green vertex of q. (We write (t, u) for this parallel pair.) Thus $\mathcal{B}(\mathcal{C}, q)$ is the universal example of a category for which the $(\mathcal{C}, q)$-quadrangularity axiom fails, and its geometric realization is obtained from that of $\mathcal{A}(\mathcal{C})$ simply by 'puncturing' the latter along the diagonal of the distinguished quadrangle q. (If the symmetry group of $\mathcal{C}$ is transitive on quadrangles, so that all quadrangularity axioms derived from it are equivalent, we shall simply write $\mathcal{B}(\mathcal{C})$ rather than $\mathcal{B}(\mathcal{C}, q)$.)

LEMMA 7·1. *$\mathcal{B}(\mathcal{C}, q)$ is $\mathcal{C}_1$-quadrangular for any quadrangle club $\mathcal{C}_1$ having fewer quadrangles than $\mathcal{C}$.*

Proof. Let $F : \mathcal{C}_1 \to \mathcal{B}(\mathcal{C}, q)$ be a directed-graph morphism mapping all but one of the quadrangles of $\mathcal{C}_1$ to commutative squares. The composite of F with the quotient map $\mathcal{B}(\mathcal{C}, q) \to \mathcal{A}(\mathcal{C})$ maps all the quadrangles of $\mathcal{C}_1$ to commutative squares, and so extends to a functor $\widetilde{F} : \mathcal{A}(\mathcal{C}_1) \to \mathcal{A}(\mathcal{C})$. In particular, since the triangle club $\widehat{\mathcal{C}_1}$ is a sub-directed-graph of $\mathcal{A}(\mathcal{C}_1)$, we obtain a directed-graph morphism $\widehat{F} : \widehat{\mathcal{C}_1} \to \mathcal{A}(\mathcal{C})$.

Since $\mathcal{C}_1$ has fewer quadrangles than $\mathcal{C}$, there is some 2-simplex of (the underlying simplicial complex of) $\widehat{\mathcal{C}}$ which is not in the image of $\widehat{F}$. It follows that $\widehat{F}$ is null-homotopic; in other words, its image, considered as a subspace of the geometric realization of $\mathcal{A}(\mathcal{C})$, is contractible. But this ensures that the image of F cannot contain both members of the pair (t, u): for if we take the geometric realization of $\mathcal{A}(\mathcal{C})$, puncture it along the diagonal of q, and then remove the interior of some 2-simplex of $\widehat{\mathcal{C}}$, the loop formed by going around the two sides of the puncture is not null-homotopic. But if the images under F of the two ways around the 'last' quadrangle of $\mathcal{C}_1$ are not t and u in some order, then they must be equal. □

COROLLARY 7·2. *No finite set of quadrangularity axioms can be equivalent to total quadrangularity.*

Proof. Given a finite set $\mathcal{S}$ of quadrangle clubs, let N be an upper bound for the number of quadrangles in any member of $\mathcal{S}$, and let $\mathcal{C}$ be a proper quadrangle club having more than N quadrangles (for example, we could take $\mathcal{C} = \mathcal{W}_{2N}$). Then, for any quadrangle q of $\mathcal{C}$, $\mathcal{B}(\mathcal{C}, q)$ is $\mathcal{S}$-quadrangular but not $(\mathcal{C}, q)$-quadrangular. □

COROLLARY 7·3. *Let $\mathcal{S}$ be the set of quadrangle clubs which are both blue-minimal and green-minimal. Then $\mathcal{S}$-quadrangularity is not equivalent to total quadrangularity.*

Proof. As we have already observed, if a quadrangle club is both blue- and green-minimal then it can have only two red vertices, and hence it must be $\mathcal{E}_{2n}$ for some n. But we have also observed that $\mathcal{E}_{2n}$ is improper for $n > 2$, and if we split it into proper clubs then we obtain only copies of ($\mathcal{D}_3$ and) $\mathcal{E}_4$. So $\mathcal{S}$-quadrangularity is equivalent to $\mathcal{E}_4$-quadrangularity. □

It is also true that the quadrangle clubs which are simultaneously red- and green-minimal (that is, satisfy both the Mal'cev and Lambek conditions) do not yield a sufficient set of axioms. The proof of this requires a bit more work; for $\mathcal{W}_{2n}$ is proper for all n. It is not hard to see that $(\mathcal{W}_{2n+2} \vdash \mathcal{W}_{2n})$ for all n, by extending a diagram of shape $\mathcal{W}_{2n}$ to one of shape $\mathcal{W}_{2n+2}$ in which one pair of yellow edges with common codomain is mapped to an identity morphism; but Lemma 7·1 shows that the converse does not hold. Nevertheless, we have:

PROPOSITION 7·4. *Let $\mathcal{S}$ be the set of quadrangle clubs which are both Mal'cev and Lambek. Then $\mathcal{S}$-quadrangularity is not equivalent to total quadrangularity.*

Proof. Let $\mathcal{B} = \mathcal{B}(\mathcal{M}_6)$ be the universal example of a category which is not $\mathcal{M}_6$-quadrangular (where $\mathcal{M}_6$, as before, denotes the dual of $\mathcal{W}_6$). We claim that $\mathcal{B}$ is $\mathcal{W}_{2n}$-quadrangular for all n.

To prove this, we shall need to introduce notation for the objects and morphisms of $\mathcal{B}$. We shall denote the vertices of $\mathcal{M}_6$ by $\{b_1, b_2, b_3, r_1, r_2, r_3, g_1, g_2\}$, with a yellow edge y_{ij} from r_j to g_i for all $i \leqslant 2$ and $j \leqslant 3$, and purple edges p_{ij} from b_j to r_i when $i = j$ or $i \equiv j + 1 \pmod 3$. The composite $y_{ij}p_{jk}$ in $\mathcal{B}$ will be denoted t_{ik}, except for the composite $y_{12}p_{21}$ which will be denoted u_{11} (whereas $y_{11}p_{11} = t_{11} \neq u_{11}$). Similarly, we shall denote the vertices of $\mathcal{W}_{2n}$ by $\{b_1, b_2, r_1, \ldots, r_n, g_1, \ldots, g_n\}$, with purple edges p_{ij} for all i and j and yellow edges y_{ij} if $i = j$ or $i \equiv j - 1 \pmod n$. We assume that the quadrangle of $\mathcal{W}_{2n}$ corresponding to the conclusion of the Horn axiom under consideration is

$$\begin{tikzcd} b_1 \arrow[r, "p_{11}"] \arrow[d, "p_{21}"'] & r_1 \arrow[d, "y_{11}"] \\ r_2 \arrow[r, "y_{12}"] & g_1 \end{tikzcd}$$

As before, if $F : \mathcal{W}_{2n} \to \mathcal{B}$ is a directed-graph morphism for which the Horn axiom fails, then the composites $F(y_{11})F(p_{11})$ and $F(y_{12})F(p_{21})$ must be t_{11} and u_{11} in some order, and we assume without loss of generality that it is the order indicated. Unfortunately, we cannot deduce from this that $F(y_{11}) = y_{11}$ and $F(p_{11}) = p_{11}$, since one of them could be t_{11} and the other an identity morphism. However, we can conclude that $F(b_1) = b_1$, from which it follows that $F(r_i)$ is in the set $\{b_1, r_1, r_2, g_1, g_2\}$ for all $i \leqslant n$, since these are the only objects of $\mathcal{B}$ to which b_1 admits morphisms (whereas, in $\mathcal{W}_{2n}$, b_1 maps to every red vertex).

Suppose $F(r_i) \in \{r_2, g_1, g_2\}$ for all i. Then each $F(p_{i1})$ must be one of t_{11}, u_{11}, p_{21} or t_{21}. Moreover, $F(p_{11}) = t_{11}$ and $F(p_{21})$ must be either p_{21} or u_{11}. And since the span $(F(p_{i1}), F(p_{i+1,1}))$ must be completable to a commutative square for $i \geqslant 2$, we deduce inductively that $F(p_{i1}) \in \{p_{21}, u_{11}, t_{21}\}$ for all $i \geqslant 2$. But for $i = n$ this contradicts the requirement that $(F(p_{n1}), F(p_{11}))$ should also be completable to a commutative square. A similar argument applies if $F(r_i)$ never takes the values b_1 or r_2.

Hence we must either have $F(r_i) = b_1$ for some i, or both r_1 and r_2 must appear among the $F(r_i)$. In either case we may deduce that $F(b_2) = b_1$, since $F(b_2)$ must also admit morphisms to all the $F(r_i)$. This in turn forces $F(p_{i1}) = F(p_{i2})$ for all i such that $F(r_i) \neq g_1$. But since the quadrangle with vertices b_2 and g_1 is mapped to a commutative square in $\mathcal{B}$, and that with vertices b_1 and g_1 is not, we may assume without loss of generality that $F(r_1) = g_1$, $F(p_{11}) \neq F(p_{12})$ and $F(p_{21}) = F(p_{22})$. We now prove inductively that $F(p_{i1}) = F(p_{i2})$ for all $i \geqslant 2$: suppose this holds for some particular i, but fails for $i + 1$.

Then necessarily $F(r_{i+1}) = g_1$, so $F(g_i)$ must also be g_1, and $F(y_{i,i+1}) = 1$. But then

$$F(p_{i+1,1}) = F(y_{ii})F(p_{i1}) = F(y_{ii})F(p_{i2}) = F(p_{i+1,2}) .$$

Applying this argument once more with i and $i+1$ replaced by n and 1 yields $F(p_{11}) = F(p_{12})$, a contradiction which completes the proof. □

A proof of Proposition 7·4, in the semigroup-theoretic context, was given by Bush [**2**]. However, Bush's example is more complicated than ours: it has 24 generators, whereas ours (if presented as a semigroup rather than a category) has 12. (Bush's example is the semigroup reflection of $\mathcal{B}(\mathcal{C})$, where $\mathcal{C}$ is the Lambek quadrangle club corresponding to a tetrahedron.)

REFERENCES

[**1**] G. C. BUSH. The embedding theorems of Mal'cev and Lambek, *Canad. J. Math.* **15** (1963), 49–58.

[**2**] G. C. BUSH. The embeddability of a semigroup—conditions common to Mal'cev and Lambek, *Trans. Amer. Math. Soc.* **157** (1971), 437–448.

[**3**] A. H. CLIFFORD and G. B. PRESTON. *The Algebraic Theory of Semigroups*, volume 2, A.M.S. Surveys no. 7 (Amer. Math. Soc., 1967).

[**4**] P. M. COHN. *Universal Algebra*, Harper & Row, 1965 (second edition published by D. Reidel, 1981).

[**5**] P. T. JOHNSTONE. *Sketches of an Elephant: a Topos Theory Compendium*, vols. 1–2 (Oxford Univ. Press, 2002).

[**6**] S. KRSTIĆ. Embedding semigroups in groups: a geometrical approach, *Publ. Inst. Math. (N.S.)* **38** (52) (1985), 69–82.

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

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

[**9**] A.I. MAL'CEV. On the embedding of associative systems in groups, I (Russian), *Mat. Sbornik* **6** (48) (1939), 331–336.

[**10**] A. I. MAL'CEV. On the embedding of associative systems in groups, II (Russian), *Mat. Sbornik* **8** (50) (1940), 251–264.

[**11**] C. R. F. MAUNDER. *Algebraic Topology* (van Nostrand Reinhold, 1970; republished by Cambridge University Press, 1980).