MATHEMATICS MAGAZINE

CONTENTS

VOLUME 44 · MAY 1971 · NUMBER 3

MATHEMATICS MAGAZINE

GERHARD N. WOLLAN, *EDITOR*

EDITORIAL CORRESPONDENCE should be sent to the EDITOR, G. N. WOLLAN, Department of Mathematics, Purdue University, Lafayette, Indiana 47907. Articles should be typewritten and triple-spaced on 8½ by 11 paper. The greatest possible care should be taken in preparing the manuscript, and authors should keep a complete copy. Figures should be drawn on separate sheets in India ink and of a suitable size for photographing.

NOTICE OF CHANGE OF ADDRESS and other subscription correspondence should be sent to the Executive Director, A. B. WILLCOX, Mathematical Association of America, Suite 310, 1225 Connecticut Avenue, N. W., Washington, D. C. 20036.

ADVERTISING CORRESPONDENCE should be addressed to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo (University of Buffalo), Buffalo, New York 14214.

The MATHEMATICS MAGAZINE is published by the Mathematical Association of America at Washington, D. C., bi-monthly except July–August. Ordinary subscriptions are: 1 year $4.00. Members of the Mathematical Association of America and of Mu Alpha Theta may subscribe at the special rate of 2 years for $6.00. Single copies are 80¢.

Second class postage paid at Washington, D. C. and additional mailing offices.

POINT SET GEOMETRY

R. B. KILLGROVE, California State College at Los Angeles

Dedicated to Professor Emeritus Paul H. Daus, Secretary to the MAA Southern California Section 1924 to 1957.

1. Introduction. We wish to organize well-known ideas about geometry and topology from a different point of view. We begin by first reviewing the fruits of the Erlangen program and then produce a simple scheme to unify the descriptions of geometry and topology. We also indicate other known approaches to geometry which cannot fit into our scheme.

2. The extension of the Erlangen program. Consider the permutation group π of E_2, i.e., all the 1–1 transformations from E_2 onto E_2. The continuous members of π are homeomorphisms and thus form the largest subgroup T of π which preserve open sets. (One proof of this uses the Jordan curve theorem.) The subgroup generated by nonsingular matrices and translations, called the collineations, is the largest subgroup K of π which preserves lines. Contrary to Klein's opinion [9], K is also a subgroup of T (Coxeter [4], Artin [1]). The subgroup H generated by expansions, contractions, rotations, and translations is the largest subgroup which preserves circles, or equivalently, perpendicularity. Then H is a subgroup of K. Finally, the subgroup M of π, called the rigid motions, is the largest subgroup which preserves distance. Of course, M is a subgroup of H. For more symmetry, one could extend the plane to the real projective plane, and likewise extend K to the projective group. Likewise, one could extend the plane to the Riemann sphere, and likewise extend H to the fractional linear transformations and conjugate transformations.

Now one starts with "open sets," that which was preserved by T, as the basic concept and one takes certain properties of open sets as axioms for an abstract topology. Then by adding further conditions, also true in E_2, one obtains a metrization, and finally (see Blumenthal [3]), by adding further conditions, one finally returns to E_2.

Similarly, one starts with "line," that which was preserved by K, as the basic concept and takes certain properties of lines, as axioms for an abstract affine plane. Then by adding further conditions, also true in E_2, one obtains a coordinatization (see Artin [1]), and finally by requiring the coordinatizing algebraic system to be the reals, one finally returns to E_2. A similar study for projective planes can be made.

Although such a complete program has not yet been done, for "circle" or "perpendicular," that which was preserved by H, at least there are axioms for an abstract inversive plane (see Dembowski [5]), and there are already studies of configuration theorems for such planes.

Finally, for the subgroup M, one merely points out that the simplest generalization of this is the abstract metric space from which E_2 can be found again by adding the correct conditions.

3. The problem of a unification. In the above review of standard results one notes that the theory of an abstract affine or projective plane starts with a set of points and a class of distinguished subsets called lines. Similarly, the theory of an abstract topology starts with a set of points and a class of distinguished subsets called open sets. The first theory continues by introducing configuration theorems with the goal of attaining a nice coordinatization. The second theory continues by introducing separation axioms with the goal of attaining a metrization. Yet further inspection of the axioms of these two systems as usually given seems to indicate no common ground for the two theories. However, to substantiate such ideas formally one needs some sort of common language. It is this discovery of a common language, no matter how trivial to the reader, which was accomplished only after some seven years. We now present the polished form.

4. Abstract point set geometries. An abstract *point set geometry* is a system (Σ, β, A) where Σ is a nonempty set of points, β is a nonempty class of nonempty sets of Σ called *blocks*, and A is a list of axioms describing the *meeting* and *covering* done by blocks of β.

DEFINITION. *Two sets A and B are said to meet iff $A\cap B$ is nonempty; the set $A\cap B$, if not empty, is called the meet of the sets A and B.*

DEFINITION. *A set A covers a point p iff $p\in A$. A set A covers a set B iff $B\subset A$. A set A covers a family $\mathfrak{F}$ of sets iff for all B, $B\in\mathfrak{F}\Rightarrow B\subset A$.*

DEFINITION. *A set A exactly covers an object (point, set, family) iff there is no $x\in A$ such that $A-\{x\}$ covers the object.*

We now list some axioms describing meeting, M1–M6, and some axioms describing covering, C1–C8.

M1. *No two distinct blocks meet.*
M2. *If two distinct blocks meet, their meet is a point.*
M3. *Every pair of distinct blocks meet, and their meet is a point.*
M4. *If two distinct blocks meet, their meet is at most two points.*
M5. *If two distinct blocks meet, their meet is a block.*
M6. *Every pair of distinct blocks meet, and the meet is a block.*

C1. *For each point p there is a block A which covers p.*
C2. *For each pair of points p, q, there is at least one block A which covers p and covers q.*
C3. *For each pair of points p, q, there is at most one block A which covers p and covers q.*
C4. *For each triple p, q, r of distinct points, there is exactly one block which covers all three.*
C5. *There are four points, no three of which are covered by the same block.*
C6. *There are four points, not covered by the same block.*
C7. *For each nonempty family of blocks, every set covering that family is a block.*
C8. *For each nonempty family of blocks, the set exactly covering that family is a block.*

5. Known point set geometries. Some of the axioms above came from the systems we wished to discuss—namely topology and projective, affine, and inversive planes. Others were axioms easily constructed which turned out to produce well-known examples. For example, the system with $A=(\mathrm{M1}, \mathrm{C1})$, turned out to give a partition of Σ, thus in turn, each such system defines an equivalence relation.

Each system with $A=(\mathrm{M2}, \mathrm{C1})$ is a partial plane (Hall [6]) but not conversely; for example let p be a point, and L, M, N be lines, p incident to L and M. Many block designs (Ryser [11]) satisfy $A=(\mathrm{M4}, \mathrm{C2})$ where points are varieties, and blocks are blocks. "Block" is used in point set geometries, however, in the sense of building blocks made of points.

In any classical geometry, M2, C2, C3, C5 hold. The projective plane can be defined by M3, C2, C5, and the inversive plane can be defined by M4, C4, C6. An abstract topology is given by M5, C1, C8 while a filter (Kelly [8]) is given by (M6, C7).

6. Comparison of geometry and topology. It should not be too surprising, once we have a common language, that the parallel postulates HP and EP below are very much like the separation postulates T0, T1, and T2 below. Unfortunately neither set of postulates improves on the other system; in fact T1 and HP or EP in a topological space forces the topology to be the discrete one.

HP: For each p and each block B not covering p, there is a block C covering p and not meeting B.

HP: For each point p and each block B not covering p, there is exactly one block C covering p and not meeting B.

T0: For each pair of distinct points p, q, there is a block which covers one but not the other.

T1: For each pair of points p, q there are blocks P, Q such that P covers p, Q covers q, P doesn't cover q and Q doesn't cover p.

T2: For each pair of points p, q, there are blocks P, Q, such that P covers p, Q covers q, and P and Q do not meet.

Another set of similarities was known before the development of this language, namely, Sylvester's conjecture proved by Gallai—if n given points are not all on one line, there exists a line containing exactly two of them (in E_2)—which can be thought of as a configuration theorem or as a Kelly-type theorem, the latter being related to the definition of compact sets.

The notion of limit which is so useful in topology, and its applications to analysis, can be interpreted in ordered affine planes by using interior points, since every line through an interior point p meets the set in points on either side of p. But for finite topologies and for finite planes, limit points and order respectively are of little value. This is unfortunate. Geometry, unlike topology, produces finite systems having applications to combinatorial analysis.

7. Alternate approaches in topology and geometry. The abstract metric space which arose from rigid motions is not a point set geometry. Some attempts to obtain a substitute point set geometry for the metric space will be discussed in the next section. An extension to topology by using functions, as one does in defining metric spaces, has been developed by Hammer [7]; this generalized Kuratowski closure approach to topology allows one to develop models from algebra and logic as well as to show applications to numerical analysis. Overlapping with applications of Hammer's extended topology to convexity is Prenowitz' [10] algebraic approach to develop ordered spaces via convexity.

The logical approach to geometry in contrast to the set theory approach is obtained by use of predicates (i.e., relations) in the axiomatic structure. The partial plane mentioned in 5 above is such a structure, and, as was already pointed out, the set theory-point set geometry approach does not always lead to a description of such a system. Bernays [2] did develop a geometry using the perpendicular relation. To the author's knowledge this has not appeared as a point set geometry. Tarski [12] pointed out, in the symposium where Bernays gave his perpendicular relation, that the relation or logical approach and the set theory (—point set geometry) approach yield somewhat different logical characteristics.

8. Abstract unit interval spaces. Among the many invariants which occur from the rigid motions of the plane and which are not invariants for any larger permutation subgroup of the plane is the class of open unit intervals. Clearly open unit intervals are preserved by rigid motions. Also, by overlapping the open unit intervals, we note that lines are preserved whenever open unit intervals are preserved. Thus any permutation subgroup preserving open unit intervals is a subgroup of the collineation group K.

Now let us show this subgroup is M, the subgroup of rigid motions. First, since this subgroup is a subgroup of T, the homeomorphisms, then any transformation of the subgroup fixing two points a unit apart also fixes pointwise the open unit interval having these points as end points. Now let S be a member of the subgroup which sends the origin to (a, b). Then let S_1 be the composition of S followed by the translation sending (a, b) to the origin. Now S_1 is also a member of the subgroup and fixes the origin and sends $(1, 0)$ to $(\cos\theta, \sin\theta)$. Let S_2 be a composition of S_1 followed by a rotation sending $(\cos\theta, \sin\theta)$ to $(1, 0)$. Thus S_2 fixes $(0, 0)$ and $(1, 0)$, and hence fixes circles with these centers with radius 1. Hence their intersections points $(1/2, \sqrt{3}/2)$ and $(1/2, -\sqrt{3}/2)$ are either fixed or interchanged by S_2. If the latter case, let S' be S_2 composed with the reflection about the x-axis, and otherwise let $S' = S_2$. Thus S' fixes points $(0, 0)$, $(1, 0)$, $(1/2, \sqrt{3}/2)$. Suppose S' is the identity. Then by composing S' with appropriate rigid motions we have S is a rigid motion. Thus our strategy is to show S' is the identity.

Clearly the x-axis is fixed, and the line $x = 1/2$ is fixed; since the y-axis is parallel to $x = 1/2$, and goes through the origin, the y-axis is also fixed. Therefore if the points on the two axes are fixed, then S' is the identity. Clearly the integers are fixed on both axes. Thus the lattice points are fixed. For rational p/q, $q > 1$, on the x-axis, note the fixed x-axis intersects the fixed line joining $(0, -1)$ and $(p, q-1)$ in point $(p/q, 0)$. Similarly the rationals are fixed on the y-axis. By continuity, all the points are fixed. Note that we used two dimensions for this proof.

Several axiom schemes have been investigated for this kind of geometry using open unit intervals as the undefined objects. None yet has been discovered which leads to a nice theory. However, since these objects are one-dimensional and proper subsets in one dimension, one might investigate the situation in this case. Surprisingly, we obtain more than just the rigid motions. For example, f defined on the reals as $f(x)=(x-[x])^2+[x]$ is a nonrigid motion which preserves open unit intervals. In fact, it can be shown that any nonrigid motion is $g(x-[x])+[x]$ where g is 1-1 and onto $[0, 1)$. In order to find a suitable invariant for the line for the rigid motions one takes a pair of unit intervals spaced $\sqrt{2}$ distance apart. Since one can approach any real modulo one arbitrarily closely by $m+n\sqrt{2}$ reduced modulo one, for suitable choices of integers m, n, it is easy to see that fixing the origin and one is enough to have the identity in this case.

References

1. E. Artin, Geometric Algebra, Interscience, New York, 1957.

2. Paul Bernays, Die Mannigfaltigkeit der Directiven fur die Gestaltung geometrischer Axiomsysteme, The Axiomatic Method (Editors: Henkin, Suppes, Tarski), North-Holland, Amsterdam, 1959.

3. L. M. Blumenthal, A Modern View of Geometry, Freeman, San Francisco, 1961.

4. H. S. M. Coxeter, The Real Projective Plane, McGraw-Hill, New York, 1949.

5. P. Dembowski, Inversive planes of even order, Bull. Amer. Math. Soc., 69 (1963) 850–854.

6. M. Hall, Jr., Projective Planes and Related Topics, C. I. T., Pasadena, 1954.

7. P. C. Hammer, Extended topology: The continuity concept, this MAGAZINE, 36 (1963) 101–105.

8. J. L. Kelly, General Topology, Van Nostrand, Princeton, 1955.

9. Felix Klein, Elementary Mathematics from an Advanced Standpoint, Geometry, Dover, New York, 1939.

10. Walter Prenowitz, A Contemporary Approach to Classical Geometry, Number 7 of the Herbert Ellsworth Slaught Memorial Papers, Amer. Math. Monthly.

11. H. J. Ryser, Combinatorial Mathematics, The Carus Mathematical Monographs, Number 14, MAA, 1963.

12. Alfred Tarski, What is Elementary Geometry?, The Axiomatic Method, North-Holland, Amsterdam, 1959.

A DIAGRAMMATIC SOLUTION TO "INSTANT INSANITY" PROBLEM

A. P. GRECOS, Université Libre de Bruxelles, Belgium and R. W. GIBBERD, University of Texas at Austin

I. Introduction. We describe a method for obtaining the solutions to the puzzle "Instant Insanity." This puzzle consists of four unit cubes, with their faces colored red, blue, white or green, as shown in Figure 1. The solution involves assembling these cubes into a $1\times1\times4$ prism such that all four colors appear on each of the four lateral faces of the prism.

Recently Brown [1] and Schwartz [2] have discussed solutions to this problem. Their methods reduce the possible number of configurations from 82,944 to 81, but no simple systematic procedure is provided in order to obtain the two

Several axiom schemes have been investigated for this kind of geometry using open unit intervals as the undefined objects. None yet has been discovered which leads to a nice theory. However, since these objects are one-dimensional and proper subsets in one dimension, one might investigate the situation in this case. Surprisingly, we obtain more than just the rigid motions. For example, f defined on the reals as $f(x)=(x-[x])^2+[x]$ is a nonrigid motion which preserves open unit intervals. In fact, it can be shown that any nonrigid motion is $g(x-[x])+[x]$ where g is 1-1 and onto $[0, 1)$. In order to find a suitable invariant for the line for the rigid motions one takes a pair of unit intervals spaced $\sqrt{2}$ distance apart. Since one can approach any real modulo one arbitrarily closely by $m+n\sqrt{2}$ reduced modulo one, for suitable choices of integers m, n, it is easy to see that fixing the origin and one is enough to have the identity in this case.

References

1. E. Artin, Geometric Algebra, Interscience, New York, 1957.

2. Paul Bernays, Die Mannigfaltigkeit der Directiven fur die Gestaltung geometrischer Axiomsysteme, The Axiomatic Method (Editors: Henkin, Suppes, Tarski), North-Holland, Amsterdam, 1959.

3. L. M. Blumenthal, A Modern View of Geometry, Freeman, San Francisco, 1961.

4. H. S. M. Coxeter, The Real Projective Plane, McGraw-Hill, New York, 1949.

5. P. Dembowski, Inversive planes of even order, Bull. Amer. Math. Soc., 69 (1963) 850–854.

6. M. Hall, Jr., Projective Planes and Related Topics, C. I. T., Pasadena, 1954.

7. P. C. Hammer, Extended topology: The continuity concept, this MAGAZINE, 36 (1963) 101–105.

8. J. L. Kelly, General Topology, Van Nostrand, Princeton, 1955.

9. Felix Klein, Elementary Mathematics from an Advanced Standpoint, Geometry, Dover, New York, 1939.

10. Walter Prenowitz, A Contemporary Approach to Classical Geometry, Number 7 of the Herbert Ellsworth Slaught Memorial Papers, Amer. Math. Monthly.

11. H. J. Ryser, Combinatorial Mathematics, The Carus Mathematical Monographs, Number 14, MAA, 1963.

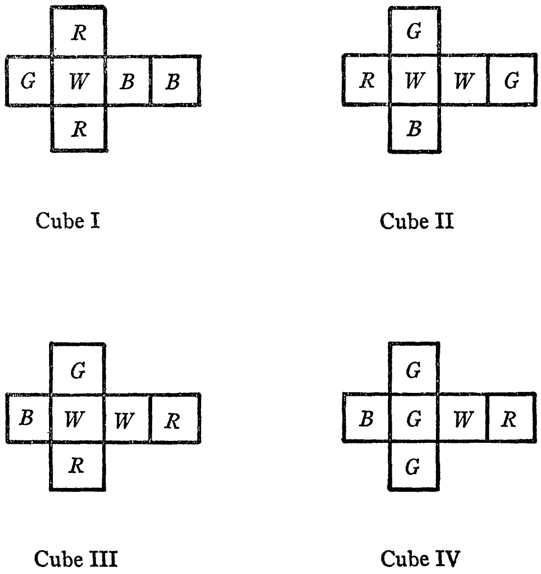
12. Alfred Tarski, What is Elementary Geometry?, The Axiomatic Method, North-Holland, Amsterdam, 1959.

A DIAGRAMMATIC SOLUTION TO "INSTANT INSANITY" PROBLEM

A. P. GRECOS, Université Libre de Bruxelles, Belgium and R. W. GIBBERD, University of Texas at Austin

I. Introduction. We describe a method for obtaining the solutions to the puzzle "Instant Insanity." This puzzle consists of four unit cubes, with their faces colored red, blue, white or green, as shown in Figure 1. The solution involves assembling these cubes into a $1\times1\times4$ prism such that all four colors appear on each of the four lateral faces of the prism.

Recently Brown [1] and Schwartz [2] have discussed solutions to this problem. Their methods reduce the possible number of configurations from 82,944 to 81, but no simple systematic procedure is provided in order to obtain the two



Cube I

Cube II

Cube III

Cube IV

FIG. 1.

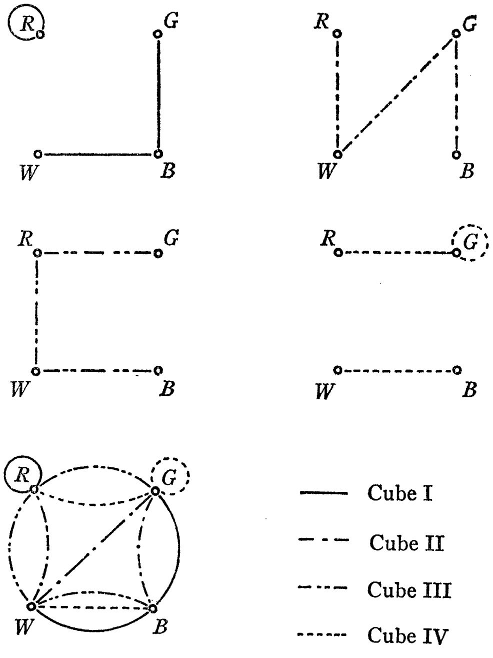


FIG. 2.

possible solutions. The method that we describe here is an improvement on the previous solutions as one can obtain by inspection the two unique solutions, and also the arithmetic manipulations required are replaced by a simple graphical problem. We also use this method to solve the three problems suggested by Brown.

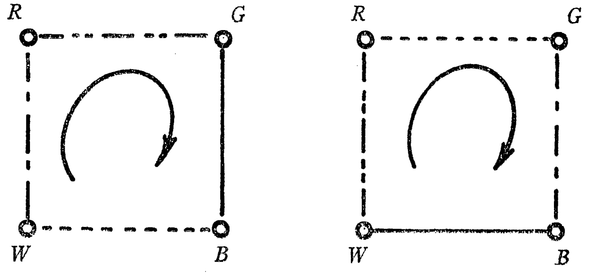
II. Graphical solution. For a discussion of the puzzle, we refer the reader to the original articles by Brown and Schwartz, who show that in order to obtain a solution it is sufficient to characterize the cubes in terms of opposite pairs of faces. In order to obtain a graphical representation of this problem, let us label the vertices of a square by the colors blue, green, red and white. Then joining by a line those vertices of the square which correspond to the colors of the opposite faces of a cube, we obtain a diagrammatic representation of the cubes as shown in Figure 2.

Each cube is specified by three lines (bonds) on the graph. The diagrams representing the four cubes are superimposed on the single graph so as to help represent the solution, as is shown in the next paragraph.

The blocks are to be placed on top of one another to form a $1 \times 1 \times 4$ prism. The exact position of a cube in this prism does not matter as long as we do not change its orientation. A graphical representation of a pair of parallel lateral sides (north-south, say) of the prism can be made by drawing the four bonds representing the four faces, where an arrow is placed on the bond to indicate the orientation of the block. Thus, an arrow on a bond points to the color which is on the south face, say, or the west face when discussing the east-west pair of lateral sides. When representing a side in which four different colors exist, the bonds on the graph will form a closed path (possibly disconnected), such that each colored vertex is passed through only once and each bond belongs to a different cube. We shall call such a path an admissible path. The solution of the "Instant Insanity" problem requires all lateral sides of the prism to have different colors. Therefore, a necessary condition that the problem have a solution is that at least two admissible paths (corresponding to the e-w and n-s sides) can be drawn on the graph. To obtain a solution we must also insure that the two admissible paths do not require two different orientations of the same cube. This condition is expressed by requiring that the paths be distinct or compatible, i.e., that they do not have common bonds.

From the graph of the problem given in Figure 2, we obtain by inspection the two required paths shown in Figure 3 (a, b). A third path which can be obtained, Figure 3′, is neither compatible with (a) because of the line *RW*, nor with (b) because of the line *RG*.

To construct the two solutions we assign a clockwise direction to both paths in Figure 3 (a, b) and we assume that an arrow on a bond indicates a direction "north-south" for path (a) and "east-west" for path (b). The correct solution is now determined by picking up each cube, finding the four exposed faces from Figure 3 and placing it according to the directions determined by the arrows. For instance, the face-pairs of cube 1 which will appear on the lateral sides of the prism are green-blue with green facing "north" and blue "south," and blue-



(a) (b)

FIG. 3.

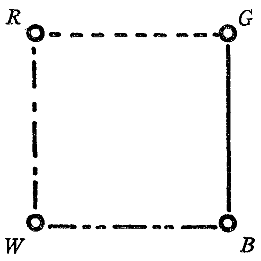


FIG. 3.

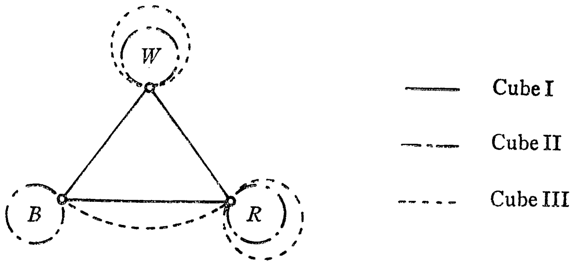


FIG. 4.

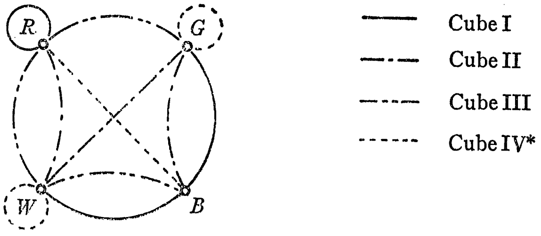


FIG. 5.

white with blue facing "east" and white "west." Similarly, we choose and place the face-pairs of the other cubes. The second solution is obtained by directing one of the paths anticlockwise. Thus we have a systematic method for obtaining the two solutions of the "Instant Insanity" problem.

Using the same method we can easily study the three exercises suggested by Brown [1] (for the coloring of the cubes we refer to the original article).

For the "three cube" problem (Ex. 1 of Reference 1) the corresponding graph is shown in Figure 4. For the "modified four cube" problem (Ex. 2 of Reference 1) the graph is that of Figure 5. The reader may verify that it is impossible to construct two distinct paths in either case, hence solution is impossible for both problems.

III. The six cube problem. The same procedure can be used for a discussion of the "six cube" problem (Ex. 3 of Reference 1). We construct the graph of this problem, shown in Figure 6, using the same rules as for the "four cube" problem. From the structure of the graph we may conclude immediately that the hexagon of Figure 7(a) is an admissible path. Because of the large number of vertices and bonds it is difficult to obtain by inspection any other paths.

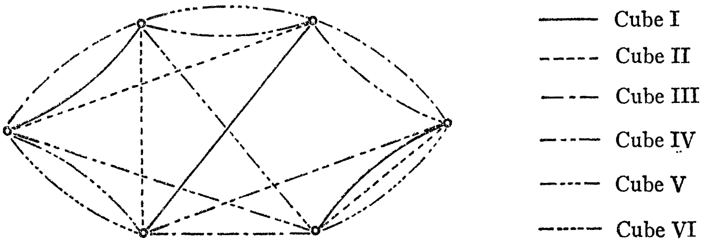


FIG. 6.

A systematic procedure to obtain all admissible paths is the following. First we draw all possible closed paths with each vertex passed through only once. In drawing these paths it is helpful to use tables of graph diagrams [3]. The next step then is to eliminate those paths containing two or more bonds belonging to the same cube. This can be easily done because a certain number of pairs of vertices are connected by a single bond. We finally obtain all the admissible paths which are shown in Figure 7. It is easy to see that only the first three paths are pairwise compatible. Therefore the six cube problem has six and only six solutions because for any pair of compatible hexagons there are two possible relative orientations.

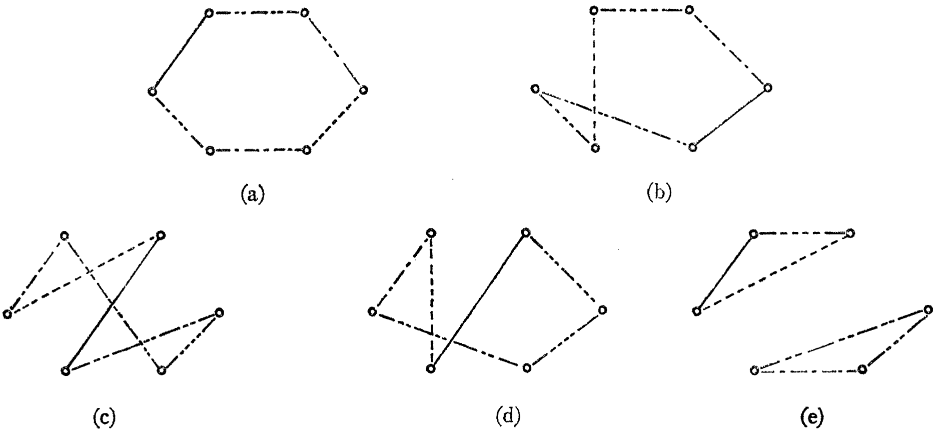


FIG. 7.

As Brown pointed out, because every color appears on every cube, the existence of one solution for this problem implies the existence of at least six distinct ones. A simple proof of this assertion can be given by considering not only the colors of each cube appearing on the lateral sides of the $1 \times 1 \times 6$ prism but also those on the sides facing "up" and "down." The same conclusion can be drawn also from the structure of the graph. The fact that each cube contains all 6 different colors implies that each vertex contains six bonds belonging to the six cubes. If one solution exists, we have shown that two distinct admissible paths must be drawn on the graph. Four of the six bonds meeting at each vertex belong to these paths. Therefore, we are left with two bonds at each vertex. These bonds must also form a closed path because otherwise we should have a vertex with an odd number of bonds meeting there. The path is admissible because each bond belongs to a different cube and it is compatible with either of the previous ones because by construction they have no bonds in common. Hence, there must be three admissible paths and thus 6 solutions.

This example again illustrates the use of graphical methods in handling combinatorial type problems. It may be possible to use this graphical technique to study further generalizations such as, "For how many sets of four colored cubes are there solutions?" The use of graph theory in problems of theoretical physics and applied mathematics is being widely developed and the reader should consult Harary [3, 4] for a general review of the field.

Note added in Proof: A similar discussion of this problem has recently been given by J. V. Deventer in *The Many Facets of Graph Theory*, p. 283, Lecture Notes in Mathematics, Vol. 110 (Springer-Verlag).

The authors would like to thank the referee for helpful comments on the presentation of this work.

References

1. T. A. Brown, A note on "Instant insanity," this MAGAZINE, 41 (1968) 167–169.
2. B. L. Schwartz, An improved solution to "Instant insanity," this MAGAZINE, 43 (1970) 20–23.
3. F. Harary, Graph Theory, Addison-Wesley, Reading, Mass., 1969.
4. ——— (ed.), Graph Theory and Theoretical Physics, Academic Press, New York, 1967.

IDEALS IN COMMUTATIVE DOMAINS

I. SINHA, Michigan State University, and J. B. SRIVASTAVA, IIT, Delhi, India

1. Introduction. The purpose of this note is to develop an equivalence relation on commutative rings such that the equivalence classes give us information about the ideals, notably the prime ideals, of the rings. Specifically, for principal ideal domains, we determine the number of prime-power factors and their indices, in the factorization of an ideal in terms of these equivalence classes.

We do not need any references in this self-contained paper, if the reader recalls rudimentary definitions and properties of commutative rings, ideals, prime-ideals, principal ideal domains, and the unique Prime Factorization Theorem for ideals in such domains. All these are included in most textbooks used for the standard undergraduate introductory course in abstract algebra.

tinct ones. A simple proof of this assertion can be given by considering not only the colors of each cube appearing on the lateral sides of the $1\times 1\times 6$ prism but also those on the sides facing "up" and "down." The same conclusion can be drawn also from the structure of the graph. The fact that each cube contains all 6 different colors implies that each vertex contains six bonds belonging to the six cubes. If one solution exists, we have shown that two distinct admissible paths must be drawn on the graph. Four of the six bonds meeting at each vertex belong to these paths. Therefore, we are left with two bonds at each vertex. These bonds must also form a closed path because otherwise we should have a vertex with an odd number of bonds meeting there. The path is admissible because each bond belongs to a different cube and it is compatible with either of the previous ones because by construction they have no bonds in common. Hence, there must be three admissible paths and thus 6 solutions.

This example again illustrates the use of graphical methods in handling combinatorial type problems. It may be possible to use this graphical technique to study further generalizations such as, "For how many sets of four colored cubes are there solutions?" The use of graph theory in problems of theoretical physics and applied mathematics is being widely developed and the reader should consult Harary [3, 4] for a general review of the field.

Note added in Proof: A similar discussion of this problem has recently been given by J. V. Deventer in *The Many Facets of Graph Theory*, p. 283, Lecture Notes in Mathematics, Vol. 110 (Springer-Verlag).

The authors would like to thank the referee for helpful comments on the presentation of this work.

References

1. T. A. Brown, A note on "Instant insanity," this MAGAZINE, 41 (1968) 167–169.
2. B. L. Schwartz, An improved solution to "Instant insanity," this MAGAZINE, 43 (1970) 20–23.
3. F. Harary, Graph Theory, Addison-Wesley, Reading, Mass., 1969.
4. ——— (ed.), Graph Theory and Theoretical Physics, Academic Press, New York, 1967.

IDEALS IN COMMUTATIVE DOMAINS

I. SINHA, Michigan State University, and J. B. SRIVASTAVA, IIT, Delhi, India

1. Introduction. The purpose of this note is to develop an equivalence relation on commutative rings such that the equivalence classes give us information about the ideals, notably the prime ideals, of the rings. Specifically, for principal ideal domains, we determine the number of prime-power factors and their indices, in the factorization of an ideal in terms of these equivalence classes.

We do not need any references in this self-contained paper, if the reader recalls rudimentary definitions and properties of commutative rings, ideals, prime-ideals, principal ideal domains, and the unique Prime Factorization Theorem for ideals in such domains. All these are included in most textbooks used for the standard undergraduate introductory course in abstract algebra.

2. Preliminaries. Let R be a commutative unitary ring and S be an ideal in R.

DEFINITION 1. *For any $r \in R$, the idealizer of r into S is the set $I_S(r) = \{x \in R \mid xr \in S\}$.*

For example, if $r \in S$, then clearly $I_S(r) = R$ itself since S is an ideal in R. Also $I_S(1) = S$, obviously, while for any $r \in R$, $I_S(r)$ contains the two sided annihilators of r. Also it is clear that $I_S(r) \supseteq S$ for any $r \in R$.

DEFINITION 2. *Given r_1, $r_2 \in R$, we shall say that r_1 is S-equivalent to r_2 and we write $r_1 \underset{S}{\sim} r_2$, if $I_S(r_1) = I_S(r_2)$.*

This is trivially an equivalence-relation on R and hence R is partitioned into S-equivalence-classes.

DEFINITION 3. *The number of distinct S-equivalence-classes of R will be called the class-number of the ideal S of R.*

Observe that the class-number of R itself is 1. It will follow from our analysis that at least for principal ideal domains, this is almost the only case of class-number 1.

The following two theorems show the invariance of class-number under multiplication by units and under certain R-endomorphisms.

THEOREM 1. *If u is a unit in R, then r and ru belong to the same class.*

Proof. $x \in I_S(r)$ implies that $xr \in S$, whence $uxr \in uS \subseteq S$. Thus $x \in I_S(ur)$, or $I_S(r) \subseteq I_S(ur)$. Conversely, let $y \in I_S(ur)$ so that $yur \in S$. Hence $yr \in u^{-1}S \subseteq S$, and we have $y \in I_S(r)$. Thus $I_S(ur) \subseteq I_S(r)$ and equality follows.

Note that in the first part of the proof above, we at once have:

COROLLARY. *For each $x \in R$, $I_S(r) \subseteq I_S(xr)$.*

THEOREM 2. *Let f be an R-endomorphism of R such that $f|_S$, the restriction of f to S is an automorphism. Then $r \underset{S}{\sim} f(r)$.*

Proof. $x \in I_S(r)$ implies that $f(xr) = x \cdot f(r) \in S$. Hence $I_S(r) \subseteq I_S(f(r))$. Conversely, let $y \in I_S(f(r))$. Hence $yf(r) = f(yr) \in S$ so that $yr \in S$ as $f|_S$ is an automorphism of S. Thus $y \in I_S(r)$, so that $I_S(f(r)) \subseteq I_S(r)$ and equality follows.

3. A characterization of prime-ideals. We recall that an ideal S in R is called prime if $xy \in S$ implies $x \in S$ or $y \in S$. The ring R itself satisfies this definition. However R is a prime ideal of R only if $\{0\}$ is the only other prime ideal in R.

We characterize the prime-ideals by means of their class-numbers in the following two theorems:

THEOREM 3. *Class-number of an ideal S is 1 if and only if $S = R$ or $S = \{0\}$.*

Proof. If $S = R$, then $I_S(r) = R$ for any $r \in R$. On the other hand, if the class-number of S is 1, then for any $r \in R$, $I_S(r)$ is the same. But then, for $r \in S$, $I_S(r) = R$, and also $I_S(1) = S$. So if $S \neq \{0\}$, then $S = R$.

Thus apart from these two extreme cases, we have:

THEOREM 4. *An ideal S such that $\{0\}\neq S\neq R$ is a prime if and only if the class-number of S is 2.*

Proof. Let S be prime. We know that for $r\in S$, $I_S(r)=R$. Hence all elements of S belong to one equivalence-class. On the other hand, if $r\notin S$, then $r\cdot z\in S$ implies $z\in S$ as S is prime. Conversely, since $I_S(r)\supseteq S$ already, $I_S(r)=S$ for all $r\notin S$. Thus all elements of R, not in S, belong to one equivalence-class. Since $S\neq R$, these two are distinct and exclusive possibilities. Hence the class-number of S is 2.

Conversely, let the class-number of S be 2. If $r\in S$ then $I_S(r)=R$, while if $r\notin S$, then $I_S(r)=S^*\neq R$ since $1\in R$ and $1\in S^*$ would imply $1\cdot r=r\in S$, contrary to the assumption that $r\notin S$.

Since $I_S(1)=S\neq R$, it follows that $S^*=S$, as, by hypothesis, there are only two distinct equivalence-classes.

This implies that if $r\notin S$, then $zr\in S$ only if $z\in S$. Hence S is prime.

COROLLARY. *For all ideals S such that $\{0\}\neq S\neq R$, the class-number of S is greater than or equal to 3 if S is not prime.*

4. The prime-factorization. Hence onwards let R be a principal-ideal-domain. Then the following three properties for proper prime ideals P, P_1, P_2 will be used below:

I. $z\cdot r\in P^n$, $r\notin P\Rightarrow z\in P^n$.

II. $z\cdot r\in P^n$, $r\in P^i$ but $r\notin P^{i+1}\Rightarrow z\in P^{n-i}$, where $i=1, 2, \cdots, n-1$.

III. $P_1\cdot P_2=P_1\cap P_2$.

The proofs of these follow from the unique prime-factorization theorem (analogous to the properties of integers). We then have:

THEOREM 5. *If P is a proper prime ideal in R and $S=P^n$, then the class-number of S is $n+1$.*

Proof. Suppose $r\notin P$. Then $z\cdot r\in S=P^n\Rightarrow z\in P^n$ by I above. Since $P^n=S\subseteq I_S(r)$ always, so $I_S(r)=P^n$. Now assume that $r\in P^i$ but not in P^{i+1}. Then again by II, $z\cdot r\in S=P^n\Rightarrow z\in P^{n-i}$. Clearly $P^{n-i}\subseteq I_S(r)$ here, so that $I_S(r)=P^{n-i}$. Letting $i=0, 1, 2, \cdots, n$, we have $n+1$ distinct equivalence-classes in R.

THEOREM 6. *If $S=P_1P_2\cdots P_t$ where P_i are distinct proper prime ideals, then the class-number of S is 2^t.*

Proof. For $t=1$, this follows from Theorem 5 above. So let $S_1=P_1P_2\cdots P_{t-1}$ and suppose the class-number of S_1 is $l=2^{t-1}$. Let $r_1, r_2, \cdots, r_l$ be the distinct class-representatives for S_1 in R. We consider two cases:

Case 1. Let $r\notin P_t$ and $r\,\tilde{}_{S_1}\,r_i$ for some unique i. Then $I_{S_1}(r)=I_{S_1}(r_i)$.

Now let $z\in I_S(r)$. So $z\cdot r\in S=S_1\cdot P_t=S_1\cap P_t$. Hence $z\cdot r\in S_1$ and $z\cdot r\in P_t$, so that $z\in I_{S_1}(r)=I_{S_1}(r_i)$ and $z\in P_t$. Thus $I_S(r)\subseteq I_{S_1}(r_i)\cap P_t$. Conversely, $z\in I_{S_1}(r_i)\cap P_t$ clearly implies that $z\cdot r\in S_1\cap P_t=S$ since $I_{S_1}(r_i)=I_S(r)$. Thus for all $r\notin P_t$, we get l idealizers $I_{S_1}(r_i)\cap P_t$, $i=1, 2, \cdots, l$.

Case 2. Let $r \in P_t$ and again $r \tilde{s}_1 r_i$.

This time $z \in I_S(r) \Rightarrow z \cdot r \in S_1 \cdot P_t = S_1 \cap P_t \Rightarrow z \cdot r \in S_1 \Rightarrow z \in I_{S_1}(r) = I_{S_1}(r_i)$. Conversely $z \in I_{S_1}(r_i) \Rightarrow z \cdot r \in S_1$ and $r \in P_t \Rightarrow z \cdot r \in P_t$ so that $zr \in S_1 \cap P_t = S$. Then for each $r \in P_t$ we get the idealizer $I_S(r_i)$. Hence for S, we get $2 \cdot l$ distinct idealizers. Now by induction-hypothesis $l = 2^{t-1}$. Hence the class-number of S is 2^t.

As an illustration one may compute that the 8 classes for $S = P_1P_2P_2$ correspond to the idealizers as follows:

(i) $\{r \mid I_S(r) = P_1\}$.
(ii) $\{r \mid I_S(r) = P_2\}$.
(iii) $\{r \mid I_S(r) = P_3\}$.
(iv) $\{r \mid I_S(r) = P_1 \cap P_2\}$.
(v) $\{r \mid I_S(r) = P_1 \cap P_3\}$.
(vi) $\{r \mid I_S(r) = P_2 \cap P_3\}$.
(vii) $\{r \mid I_S(r) = P_1 \cap P_2 \cap P_3\}$.
(viii) $\{r \mid I_S(r) = R\}$.

THEOREM 7. *If* $S = P_1^{m_1} \cdot P_2^{m_2} \cdots P_t^{m_t}$ *is a proper prime factorization of a proper ideal* S, *then the class-number of* S *is* $(m_1+1)(m_2+1) \cdots (m_t+1)$.

Proof. For $t=1$, the result follows from Theorem 5. Assume that it is true for $S_1 = P_1^{m_1}P_2^{m_2} \cdots P_{t-1}^{m_{t-1}}$. Then by an argument similar to those above, we can show that each $r \in P_t^{m_t}$, we have as many distinct equivalence classes as the class-number of S_1, in each case. Thus the total number of distinct equivalence-classes with respect to S is $(m_t+1) \times$ (the class-number of S_1). This completes the proof by virtue of the induction hypothesis.

COROLLARY. *Two proper ideals in R have the same class-number if and only if they have the same number t of proper prime power factors with the same set* $(m_1, m_2, \cdots, m_t)$ *of indices, occurring in some order.*

SYMMETRIC GROUPOIDS AND RINGS

R. L. DUNCAN and HILDA F. DUNCAN, Pennsylvania State University

Our purpose is to consider the relationship between some of the familiar laws used in defining various algebraic systems and the less familiar laws:

(1) $a(bc) = (ac)b$ (right symmetric law),
(2) $a(bc) = (ba)c$ (left symmetric law).

A groupoid is a set G in which every ordered pair of elements has a uniquely determined product in G. A groupoid is called right (left) symmetric if all its elements satisfy the right (left) symmetric law and is called symmetric if it is both right and left symmetric. Also, a semigroup is an associative groupoid and a group is a semigroup in which the equations $ax = b$ and $xa = b$ are solvable.

The proofs of the parenthetical parts of the following theorems are left to the reader.

Case 2. Let $r\in P_t$ and again $r\tilde{s}_1 r_i$.

This time $z\in I_S(r)\Rightarrow z\cdot r\in S_1\cdot P_t=S_1\cap P_t\Rightarrow z\cdot r\in S_1\Rightarrow z\in I_{S_1}(r)=I_{S_1}(r_i)$. Conversely $z\in I_{S_1}(r_i)\Rightarrow z\cdot r\in S_1$ and $r\in P_t\Rightarrow z\cdot r\in P_t$ so that $zr\in S_1\cap P_t=S$. Then for each $r\in P_t$ we get the idealizer $I_S(r_i)$. Hence for S, we get $2\cdot l$ distinct idealizers. Now by induction-hypothesis $l=2^{t-1}$. Hence the class-number of S is 2^t.

As an illustration one may compute that the 8 classes for $S=P_1P_2P_2$ correspond to the idealizers as follows:

(i) $\{r \mid I_S(r)=P_1\}$.
(ii) $\{r \mid I_S(r)=P_2\}$.
(iii) $\{r \mid I_S(r)=P_3\}$.
(iv) $\{r \mid I_S(r)=P_1\cap P_2\}$.
(v) $\{r \mid I_S(r)=P_1\cap P_3\}$.
(vi) $\{r \mid I_S(r)=P_2\cap P_3\}$.
(vii) $\{r \mid I_S(r)=P_1\cap P_2\cap P_3\}$.
(viii) $\{r \mid I_S(r)=R\}$.

THEOREM 7. *If $S=P_1^{m_1}\cdot P_2^{m_2}\cdots P_t^{m_t}$ is a proper prime factorization of a proper ideal S, then the class-number of S is $(m_1+1)(m_2+1)\cdots(m_t+1)$.*

Proof. For $t=1$, the result follows from Theorem 5. Assume that it is true for $S_1=P_1^{m_1}P_2^{m_2}\cdots P_{t-1}^{m_{t-1}}$. Then by an argument similar to those above, we can show that each $r\in P_t^{m_t}$, we have as many distinct equivalence classes as the class-number of S_1, in each case. Thus the total number of distinct equivalence-classes with respect to S is $(m_t+1)\times$(the class-number of S_1). This completes the proof by virtue of the induction hypothesis.

COROLLARY. *Two proper ideals in R have the same class-number if and only if they have the same number t of proper prime power factors with the same set $(m_1, m_2, \cdots, m_t)$ of indices, occurring in some order.*

SYMMETRIC GROUPOIDS AND RINGS

R. L. DUNCAN and HILDA F. DUNCAN, Pennsylvania State University

Our purpose is to consider the relationship between some of the familiar laws used in defining various algebraic systems and the less familiar laws:

(1) $a(bc)=(ac)b$ (right symmetric law),
(2) $a(bc)=(ba)c$ (left symmetric law).

A groupoid is a set G in which every ordered pair of elements has a uniquely determined product in G. A groupoid is called right (left) symmetric if all its elements satisfy the right (left) symmetric law and is called symmetric if it is both right and left symmetric. Also, a semigroup is an associative groupoid and a group is a semigroup in which the equations $ax=b$ and $xa=b$ are solvable.

The proofs of the parenthetical parts of the following theorems are left to the reader.

THEOREM 1. *Every commutative semigroup is a symmetric groupoid.*

Proof. The commutative and associative laws imply

$$a(bc) = a(cb) = (ac)b \quad \text{and} \quad a(bc) = (ab)c = (ba)c.$$

THEOREM 2. *Every right (left) symmetric commutative groupoid is a commutative semigroup.*

Proof. The commutative law and (1) imply $a(bc) = a(cb) = (ab)c$.

THEOREM 3. *Every right (left) symmetric groupoid with a left (right) identity is a commutative semigroup.*

Proof. If e is a left identity, then (1) implies $bc = e(bc) = (ec)b = cb$ and the desired result follows from Theorem 2.

THEOREM 4. *Every symmetric groupoid with right (left) cancellation is a commutative semigroup.*

Proof. (1) and (2) imply $(ba)b = a(bb) = (ab)b$. Hence right cancellation implies the commutative law and the desired result follows from Theorem 2.

THEOREM 5. *Every right (left) symmetric semigroup with left (right) cancellation is a commutative semigroup.*

Proof. (1) and the associative law imply $a(bc) = (ac)b = a(cb)$.

THEOREM 6. *Every right (left) symmetric groupoid G, in which the equation $ax = b$ ($xa = b$) is solvable in G for all a and b in G, is a commutative group.*

Proof. (1) and the solvability of $ax = b$ imply $ab = a(ax) = (ax)a = ba$ and G is a commutative semigroup by Theorem 2. Also, commutativity implies the solvability of $xa = b$ and the desired result follows.

THEOREM 7. *The center of a right (left) symmetric groupoid G is an ideal in G.*

Proof. Let C be the center of G, i.e., the set of all elements of G which commute with every element of G. If a and b are in G and c is in C, then (1) implies $(cb)a = c(ab) = (ab)c = a(cb)$. Thus cb is in C and $bc = cb$ is in C. Hence C is an ideal. The following corollary now follows from Theorem 2:

COROLLARY. *When regarded as a subgroupoid of G, C is a commutative semigroup.*

THEOREM 8. *The set of all idempotent elements of a symmetric groupoid G is a subgroupoid of G and is contained in the center of G.*

Proof. If a and b are idempotent, then (1) and (2) imply $(ab)(ab) = a[(ab)b] = a[a(bb)] = a(ab) = (aa)b = ab$. Also, $ax = (aa)x = a(ax) = (ax)a = x(aa) = xa$ for all x in G. The following result is a consequence of the corollary to Theorem 7:

COROLLARY. *When regarded as a subgroupoid of G, the set of all idempotent elements of G is a semilattice.*

THEOREM 9. *In a symmetric groupoid a triple product can have at most two values and in a symmetric semigroup every triple product is uniquely determined.*

Proof. It follows easily from (1) and (2) that

$$a(bc) = b(ca) = c(ab) = (cb)a = (ba)c = (ac)b.$$

Also, the application of (1) or (2) to any member of the above set of products yields a product in this set. The same is true of the set of products obtained by shifting the parentheses in each of the above products. The desired results follow since the union of these two sets constitutes the set of all possible triple products formed from the elements a, b and c.

THEOREM 10. *If G is a symmetric groupoid with center C, then G is associative if and only if the product of any two elements of G is in C.*

Proof. Let a and b be arbitrary elements of G. Then ab is in C if and only if $(ab)x=x(ab)$ for all x in G. Also, (1) and (2) imply $(ab)x=a(xb)=(xa)b$. Hence $(ab)x=x(ab)$ if and only if G is associative.

THEOREM 11. *If G is a symmetric groupoid with center C, then C contains at least one element. Also, if C contains only one element, then that element is an idempotent.*

Proof. Let x be an arbitrary element of G. Then (1) and (2) imply $x^2y=(xx)y=x(yx)=(yx)x=y(xx)=yx^2$ for all y in G. Hence x^2 is in C and the first part of the theorem follows. If $x^2=c$ for all x in C, then the second part of the theorem follows on taking $x=c$.

COROLLARY. *Every symmetric groupoid with two elements is a commutative semigroup.*

THEOREM 12. *Every symmetric groupoid with three elements is a commutative semigroup.*

Proof. Let $G=\{a, b, c\}$ be a symmetric groupoid. Then the center C of G contains at least one element by Theorem 11. Suppose $C=\{a\}$. Then $(bc)b=b(bc)$ and $(bc)c=c(bc)$. Hence bc is in C and $bc=a$. Similarly, $(cb)b=b(cb)$ and $(cb)c=c(cb)$. Hence cb is in C and $cb=a$. It follows that $bc=cb$ and G is commutative. But this contradicts the assumption that C contains only one element. Thus C must contain at least two elements, which implies that G is commutative. The desired result now follows from Theorem 2.

THEOREM 13. *Every symmetric groupoid with four elements is a symmetric semigroup.*

Proof. Let $G=\{a, b, c, d\}$ be a symmetric groupoid. Then the center C of G contains at least one element by Theorem 11. Suppose that $C=\{a\}$. If $bc=a$, then Theorem 7 implies $(bc)d=ad=a=da=d(bc)=(bd)c=b(cd)$. If $bc=b$, then $bc=(bc)c=b(cc)=ba=a$. Hence $bc\neq b$. Similarly $bc\neq c$. If $bc=d$, then $(bc)d=a=d(bc)=(bd)c=b(cd)$. Thus $(bc)d=b(cd)$. By Theorems 7 and 9 the associative law holds for the product of any three distinct elements and this result is easily extended to the case of repeated factors.

Now suppose that $C=\{a, b\}$. In this case cd and dc commute with c and d. It follows from Theorem 7 that the product of any two elements of G is in C and G is associative by Theorem 10.

Finally, if C contains more than two elements, then G is commutative and the desired result follows from Theorem 2.

THEOREM 14. *If G is a symmetric semigroup with at least two elements and center C, then C contains at least two elements. Also, if C contains only two elements, then at least one of these elements is an idempotent.*

Proof. It follows from Theorem 10 that C contains at least one element and that G is commutative and $G=C$ if C contains only one element. Hence C must contain at least two elements if G does. Now suppose that C contains only the elements a and b. If a and b are not idempotent, it follows from Theorem 7 that $a^2=b$ and $b^2=a$. Also, this and Theorem 7 imply that ab is idempotent and in C. This contradiction establishes the second part of the theorem.

We now give an example of a symmetric semigroup with four elements which is not commutative. Let $G=\{a, b, c, d\}$ and define $dc=b$ and $xy=a$ in all other cases. Then G is a symmetric semigroup since every triple product is equal to a. Also, $C=\{a, b\}$ and a is the only idempotent element of G. Thus the conclusions of Theorem 14 are the best possible.

Let G be a right (left) symmetric groupoid and consider the cyclic subgroupoid generated by a. Also, let a^n be the product of n a's with an arbitrary distribution of parentheses. Then $a^1=a$ and $a^2=aa$ are uniquely determined. Suppose that a^i is uniquely determined for all $i\leqq n$. Then $a^{n+1}=a^ia^j$ with $i+j=n+1$. If $i=1$, we have $a^{n+1}=aa^n$. If $i>1$, then (1) implies $a^{n+1}=(aa^{i-1})a^j=a(a^ja^{i-1})=aa^{i+j-1}=aa^n$. Hence a^{n+1} is uniquely determined and it follows by induction that a^n is uniquely determined for every positive integer n, i.e., a^n depends only on a and n and not on the distribution of the parentheses. Hence the familiar laws of exponents $a^ma^n=a^{m+n}$ and $(a^m)^n=a^{mn}$ hold. If G is symmetric, it can also be shown that $(ab)^n=a^nb^n$ and $a^mb^n=b^na^m$ for all a and b in G if $m\geqq 2$ and $n\geqq 1$.

We are now in a position to generalize Theorem 8 as follows:

THEOREM 15. *Let G be a symmetric groupoid and let P denote the set of all x in G such that $x^{n(x)}=x$ for some positive integer $n(x)\geqq 2$. Then P is contained in the center of G and, if $n(x)=n$ for all x in P, P is a subgroupoid of G.*

Proof. If $a^n=a$ and $b^n=b$ $(n\geqq 2)$, it follows from the above results that $(ab)^n=a^nb^n=ab$ and $ax=a^nx=xa^n=xa$ for all x in G. The following results are a consequence of the corollary to Theorem 7:

COROLLARY. *When regarded as a subgroupoid of G, the set of all elements of G which satisfy the equation $x^n=x$ is a commutative semigroup.*

COROLLARY. *If $G=P$, then G is a commutative semigroup.*

The nucleus of a groupoid G is the set of all a in G such that $a(xy)=(ax)y$,

$x(ay)=(xa)y$ and $x(ya)=(xy)a$ for all x and y in G. Only those properties of the nucleus needed to establish still another sufficient condition for a symmetric groupoid to be a commutative semigroup will be considered.

THEOREM 16. *The nucleus of a symmetric groupoid G is just the set of all a in G such that $a(xy)=(ax)y$ for all x and y in G.*

Proof. If $a(xy)=(ax)y$, then (1) and (2) imply

$$x(ay) = (ax)y = a(xy) = (xa)y \quad \text{and} \quad x(ya) = (xa)y = x(ay) = (xy)a.$$

THEOREM 17. *If G is a symmetric groupoid with nucleus N, then $G^2 \subset N$.*

Proof. (1) and (2) imply

$$(ab)(xy) = a[(xy)b] = a[x(by)] = a[(bx)y] = [(bx)a]y = [b(ax)]y = [(ab)x]y$$

and the desired result follows from Theorem 16.

THEOREM 18. *Every symmetric groupoid G, for which $G^2=G$, is a commutative semigroup.*

Proof. If G is a symmetric groupoid and $G^2=G$, then Theorem 17 implies that $N=G$ and G is associative. It follows from Theorem 10 that $G=G^2$ is contained in the center of G and G is commutative.

A subgroupoid H of a groupoid G is called normal if $aH=Ha$ for every a in G.

THEOREM 19. *A symmetric groupoid G is a commutative semigroup if and only if every cyclic subgroupoid of G is normal.*

Proof. It is obvious that every subgroupoid of a commutative groupoid is normal. Suppose that the cyclic subgroupoid $H=\{x, x^2, \cdots\}$ of G is normal. Then $aH=\{ax, ax^2, \cdots\}$; $Ha=\{xa, x^2a, \cdots\}$ and $aH=Ha$ for all a in G. Let m be the smallest positive integer such that $ax=x^m a$ and let n be the smallest positive integer such that $xa=ax^n$. If $m>n$, then $m\geqq 2$ and (1) and (2) imply

$$\begin{aligned} ax &= x^m a = ax^m = a(x^{m-n}x^n) = (ax^n)x^{m-n} \\ &= (xa)x^{m-n} = a(xx^{m-n}) = ax^{m-n+1} = x^{m-n+1}a. \end{aligned}$$

Hence $m-n+1\geqq m$ or $n=1$ and $xa=ax$ for every a in G. Similar arguments show that $xa=ax$ if $m\leqq n$ and the desired result follows from Theorem 2 since x is arbitrary.

We conclude with some remarks concerning the role of the symmetric laws in the theory of rings. Adjectives, such as symmetric, apply to a ring R if they apply to the multiplicative groupoid of R. Some of the above results yield sufficient conditions for a symmetric ring to be associative-commutative. Another such result, involving the additive structure of the ring, is the following:

THEOREM 20. *Every symmetric ring R, in which $x^2=0$ implies $x=0$, is an associative-commutative ring.*

Proof. (1) and (2) imply

$$(ab-ba)^2 = (ab)(ab) - (ba)(ab) - (ab)(ba) + (ba)(ba)$$

$$= (ab)(ab) - a[b(ab)] - b[a(ba)] + (ba)(ba)$$
$$= (ab)(ab) - a[(ab)b] - b[(ba)a] + (ba)(ba)$$
$$= (ab)(ab) - (ab)(ab) - (ba)(ba) + (ba)(ba) = 0$$

for all a and b in R. Hence R is commutative if $x^2=0$ implies $x=0$, and the desired result follows from Theorem 2.

The center C of an arbitrary ring R is the set of all c in R for which $cb=bc$; $c(ab)=(ca)b$ and $(ab)c=a(bc)$ for all a and b in R and is an associative-commutative subring of R.

THEOREM 21. *The center C of a right (left) symmetric ring R is just the center of the multiplicative groupoid of R and is an ideal in R.*

Proof. If $cx=xc$ for all x in R, then (1) implies $(ab)c=a(cb)=a(bc)$ and $c(ab)=(cb)a=(bc)a=b(ac)=b(ca)=(ba)c=c(ba)=(ca)b$ for all a and b in R. The desired result now follows from Theorem 7 and the fact that C is a subring of R. The following corollary is a consequence of Theorem 14:

COROLLARY. *If R is a symmetric associative ring with at least two elements, then C contains at least two elements.*

We now give an example of a symmetric associative ring which is not commutative. All that is required for the construction of this example is the usual sum and Cayley (row by column) product of two 3×3 matrices. Let S denote the set of all 3×3 upper matrices over an arbitrary ring T with these binary operations. Now S is closed under matrix multiplication and is noncommutative since

$$\begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & a' & b' \\ 0 & 0 & c' \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & ac' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & a' & b' \\ 0 & 0 & c' \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & a'c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Also, every triple product vanishes and hence (1), (2) and the associative law hold, since

$$\begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & d \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & d \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}.$$

The desired result follows since S is a subring of the ring of all 3×3 matrices over T. If we take T to be the ring of integers modulo two, then the center of S contains only two elements. Thus the conclusion of the above corollary is the best possible.

Any ring whose multiplicative groupoid satisfies the hypotheses of Theorems 4 or 5 (with cancellation of zero excluded) is an integral domain. Also, it follows from Artin's theorem that every symmetric ring is alternative and it follows from previous remarks that every right (left) symmetric ring is power associative. It is also clear that every antiassociative anticommutative ring is symmetric. Examples of such rings may be constructed as follows. Let R be an arbitrary symmetric ring. If we preserve the additive group of R and replace the operation of multiplication ab by the operation of commutation $a \circ b = ab - ba$, it is easily shown that we obtain a ring which is antiassociative and anticommutative.

Since symmetric rings are a generalization of associative-commutative rings, it is natural to ask which theorems concerning associative-commutative rings are true for symmetric rings. A notable example is the binomial expansion for $(a+b)^n$ for $n>2$. It is also easily shown that the radical (set of all nilpotent elements) of a symmetric ring is an ideal.

Reference

1. A. G. Kurosh, Lectures on General Algebra, Chelsea, New York, 1963.

ON A THEOREM OF G. P. BARKER ON TRIANGULAR MATRICES

MIRKO STOJAKOVIC, Gustavus Adolphus College

In a recent article [1] G. P. Barker gave a proof of the following

THEOREM. *If $S_1, S_2, \cdots, S_n$ are (upper) triangular matrices (over a ring R) such that the (i, i) element of S_i is zero, then*

$$S_1S_2 \cdots S_n = 0.$$

We give here a simpler proof of that theorem which may be of some interest for its own sake.

Proof (by induction on n). For $n=1$, there is nothing to be proved since $S_1=0$. For $n=2$, we have

$$S_1 = \begin{bmatrix} 0 & x \\ 0 & y \end{bmatrix}, \qquad S_2 = \begin{bmatrix} u & v \\ 0 & 0 \end{bmatrix},$$

where x, y, u and v are arbitrary elements of the ring R. It is clear that $S_1S_2=0$.

Assume now that the theorem is true for some integer m. Then

$$S_1S_2 \cdots S_mS_{m+1} = TS_{m+1} \quad \text{where } S_1 = \begin{bmatrix} T_1 & a_1 \\ \bar{0} & x_1 \end{bmatrix}, \cdots, S_m = \begin{bmatrix} T_m & a_m \\ \bar{0} & x_m \end{bmatrix}$$

and $T_1, \cdots, T_m$ are $m\times m$ upper triangular matrices such that the (i, i) element of T_i is 0.

Any ring whose multiplicative groupoid satisfies the hypotheses of Theorems 4 or 5 (with cancellation of zero excluded) is an integral domain. Also, it follows from Artin's theorem that every symmetric ring is alternative and it follows from previous remarks that every right (left) symmetric ring is power associative. It is also clear that every antiassociative anticommutative ring is symmetric. Examples of such rings may be constructed as follows. Let R be an arbitrary symmetric ring. If we preserve the additive group of R and replace the operation of multiplication ab by the operation of commutation $a \circ b = ab - ba$, it is easily shown that we obtain a ring which is antiassociative and anticommutative.

Since symmetric rings are a generalization of associative-commutative rings, it is natural to ask which theorems concerning associative-commutative rings are true for symmetric rings. A notable example is the binomial expansion for $(a+b)^n$ for $n>2$. It is also easily shown that the radical (set of all nilpotent elements) of a symmetric ring is an ideal.

Reference

1. A. G. Kurosh, Lectures on General Algebra, Chelsea, New York, 1963.

ON A THEOREM OF G. P. BARKER ON TRIANGULAR MATRICES

MIRKO STOJAKOVIC, Gustavus Adolphus College

In a recent article [1] G. P. Barker gave a proof of the following

THEOREM. *If $S_1, S_2, \cdots, S_n$ are (upper) triangular matrices (over a ring R) such that the (i, i) element of S_i is zero, then*

$$S_1S_2 \cdots S_n = 0.$$

We give here a simpler proof of that theorem which may be of some interest for its own sake.

Proof (by induction on n). For $n=1$, there is nothing to be proved since $S_1=0$. For $n=2$, we have

$$S_1 = \begin{bmatrix} 0 & x \\ 0 & y \end{bmatrix}, \qquad S_2 = \begin{bmatrix} u & v \\ 0 & 0 \end{bmatrix},$$

where x, y, u and v are arbitrary elements of the ring R. It is clear that $S_1S_2=0$.

Assume now that the theorem is true for some integer m. Then

$$S_1S_2 \cdots S_mS_{m+1} = TS_{m+1} \quad \text{where } S_1 = \begin{bmatrix} T_1 & a_1 \\ \bar{0} & x_1 \end{bmatrix}, \cdots, S_m = \begin{bmatrix} T_m & a_m \\ \bar{0} & x_m \end{bmatrix}$$

and $T_1, \cdots, T_m$ are $m \times m$ upper triangular matrices such that the (i, i) element of T_i is 0.

Then

$$T = S_1S_2 \cdots S_m = \begin{bmatrix} 0_m & u_m \\ \bar{0} & x \end{bmatrix} \quad \text{and} \quad S_{m+1} = \begin{bmatrix} A_m & t_m \\ \bar{0} & 0 \end{bmatrix}$$

where 0_m is the zero matrix of order $m\times m$, A_m is an upper triangular matrix of order $m\times m$, u_m and t_m are vector columns of the order $m\times 1$, $\bar{0}$ is the zero row of order $1\times m$ and x is an element of the ring R. The proof is complete.

The same proof is also valid in the partitioned form of the theorem.

Reference

1. G. P. Barker, Triangular matrices and the Cayley-Hamilton theorem, this MAGAZINE, 44 (1971) 34-36.

PACKING OF 14, 16, 17 AND 20 CIRCLES IN A CIRCLE

MICHAEL GOLDBERG, Washington, D. C.

1. **Introduction.** Several contributions have been made recently to the problem of the determination of the minimum diameter D of a circle which can contain n equal nonoverlapping circles of diameter d [1, 2, 3, 4, 5]. The optimum packing has been derived and proved by Pirl [5] for $n\leqq 10$. For $n>10$, conjectured optima have been obtained by Pirl [5] and Kravitz [3, 4]. This note submits improved packings for 14, 16, 17 and 20 circles.

2. **Packing of 14 circles.** Kravitz [3] published a conjectured packing of 14 circles in which $D/d=4.3738$. In a private communication, he submitted an improved packing shown in Figure 1, for which $D/d=4.3328$. However, this arrangement is unstable. By closing the gap at the top of the arrangement shown in Figure 1, one obtains the arrangement of Figure 2 for which $D/d=4.3284$. The angular displacement between the centers of two contacting circles on the perimeter is approximately 34.969°.

3. **Packing of 16 circles.** Kravitz [3] published a conjectured arrangement of 16 circles in which $D/d=4.7013$. He was aware of the fact that this arrangement could be improved since many of the circles were not completely constrained. The arrangement shown in Figure 3 is such an improvement. Each circle is constrained by at least three contacts. The angular separation between the centers of two contacting circles on the perimeter is approximately 32.112°. In this case, $D/d=4.6154$.

4. **Packing of 17 circles.** The only arrangement of 17 circles which has been proposed previously is obtained by the omission of two of the circles from the very efficient arrangement of 19 circles [3, 5]. The arrangement in Figure 4 is new and, furthermore, it is more efficient. The angular separation between the centers of two contacting circles on the perimeter is approximately 30.445°. In this case, $D/d=4.8085$.

Then

$$T = S_1S_2 \cdots S_m = \begin{bmatrix} 0_m & u_m \\ \bar{0} & x \end{bmatrix} \quad \text{and} \quad S_{m+1} = \begin{bmatrix} A_m & t_m \\ \bar{0} & 0 \end{bmatrix}$$

where 0_m is the zero matrix of order $m \times m$, A_m is an upper triangular matrix of order $m \times m$, u_m and t_m are vector columns of the order $m \times 1$, $\bar{0}$ is the zero row of order $1 \times m$ and x is an element of the ring R. The proof is complete.

The same proof is also valid in the partitioned form of the theorem.

Reference

1. G. P. Barker, Triangular matrices and the Cayley-Hamilton theorem, this MAGAZINE, 44 (1971) 34–36.

PACKING OF 14, 16, 17 AND 20 CIRCLES IN A CIRCLE

MICHAEL GOLDBERG, Washington, D. C.

1. Introduction. Several contributions have been made recently to the problem of the determination of the minimum diameter D of a circle which can contain n equal nonoverlapping circles of diameter d [1, 2, 3, 4, 5]. The optimum packing has been derived and proved by Pirl [5] for $n \leqq 10$. For $n > 10$, conjectured optima have been obtained by Pirl [5] and Kravitz [3, 4]. This note submits improved packings for 14, 16, 17 and 20 circles.

2. Packing of 14 circles. Kravitz [3] published a conjectured packing of 14 circles in which $D/d = 4.3738$. In a private communication, he submitted an improved packing shown in Figure 1, for which $D/d = 4.3328$. However, this arrangement is unstable. By closing the gap at the top of the arrangement shown in Figure 1, one obtains the arrangement of Figure 2 for which $D/d = 4.3284$. The angular displacement between the centers of two contacting circles on the perimeter is approximately 34.969°.

3. Packing of 16 circles. Kravitz [3] published a conjectured arrangement of 16 circles in which $D/d = 4.7013$. He was aware of the fact that this arrangement could be improved since many of the circles were not completely constrained. The arrangement shown in Figure 3 is such an improvement. Each circle is constrained by at least three contacts. The angular separation between the centers of two contacting circles on the perimeter is approximately 32.112°. In this case, $D/d = 4.6154$.

4. Packing of 17 circles. The only arrangement of 17 circles which has been proposed previously is obtained by the omission of two of the circles from the very efficient arrangement of 19 circles [3, 5]. The arrangement in Figure 4 is new and, furthermore, it is more efficient. The angular separation between the centers of two contacting circles on the perimeter is approximately 30.445°. In this case, $D/d = 4.8085$.

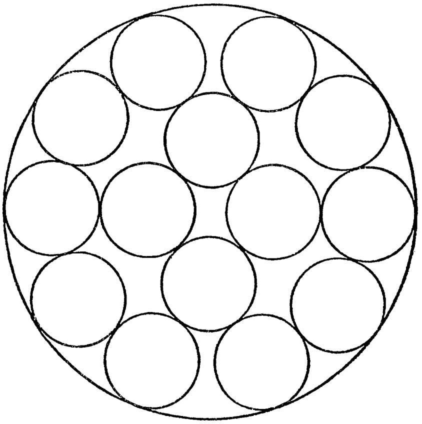


FIG. 1. 14 circles (unstable), $D/d = 4.3328$, Eff. $= 0.8223$

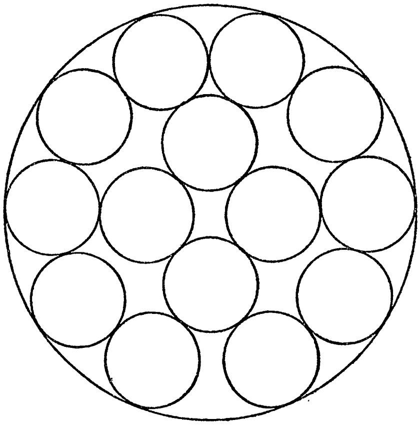


FIG. 2. 14 circles, $D/d = 4.3284$, Eff. $= 0.8240$

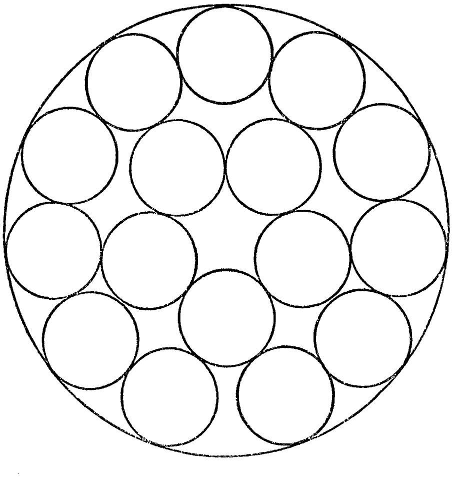


FIG. 3. 16 circles, $D/d = 4.6154$, Eff. $= 0.8282$

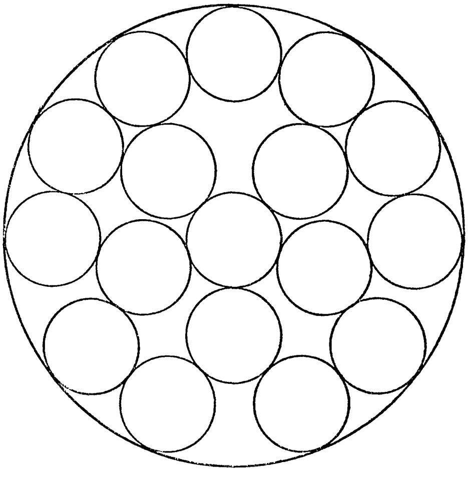


FIG. 4. 17 circles, $D/d = 4.8085$, Eff. $= 0.8107$

5. Packing of 20 circles. Kravitz [3] published a conjectured arrangement of 20 circles for which $D/d = 5.1786$. This consists of a ring of 13 circles within which a loose cluster of 7 circles can be placed. An improved arrangement, shown in Figure 5, has only one loose circle. The angular separation between the centers of contacting circles on the perimeter is approximately 28.078°. In this case, $D/d = 5.1223$.

6. Summary. The efficiency of coverage, designated by ϕ, is the ratio of the total area of the small circles to the area of the large enclosing circle. It is given by the equation

$$\phi = 2\sqrt{3}\, n/\pi(D/d)^2 = 1.10266n/(D/d)^2.$$

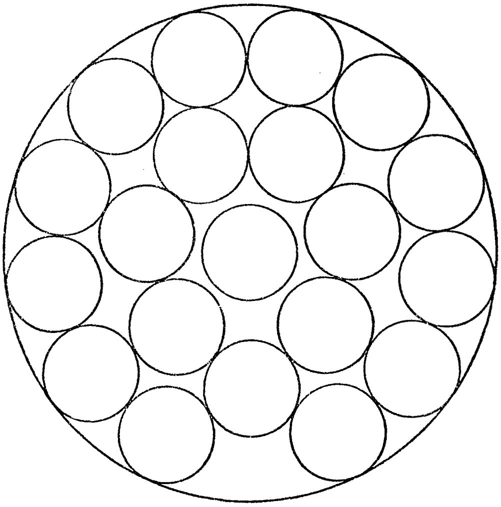


FIG. 5. 20 circles, $D/d = 5.1223$, Eff. $= 0.8405$

The data of the foregoing paragraphs are tabulated in Table 1. The rectangular and polar coordinates of the centers of the circles are given in Tables 2, 3, 4 and 5.

7. General considerations. The principle which led to the foregoing packings is the same as used by the author in the packing of circles on the surface of a sphere [6, 7]. This principle is the use of the least amount of symmetry in making the arrangements. Note that each of the arrangements in Figures 2, 3, 4 and 5 has only one axis of symmetry. The other arrangements of Table 1 have two or more axes of symmetry. When multiple symmetry is imposed, usually for

TABLE 1

Arrangement	D/d	Efficiency	Fig. No.
14 {2,2,2,2,2,2,2}	4.3738	0.8069	[3], Fig. 14
14 {2,2,1,4,1,2,2}	4.3328	0.8223	Fig. 1
14 {2,2,1,2,2,1,2,2}	4.3284	0.8240	Fig. 2
16 {5-ring, 11-ring)	4.7013	0.7982	[3], Fig. 15
16 {1,2,2,2,2,2,1,2,2}	4.6154	0.8282	Fig. 3
17 {modified 19}	4.8637	0.7924	[3], Fig. 17
17 {1,2,2,2,1,2,2,1,2,2}	4.8085	0.8107	Fig. 4
20 {7-cluster, 13-ring}	5.1786	0.8223	[3], Fig. 16
20(2,2,2,2,2,1,2,2,2,1,2}	5.1223	0.8405	Fig. 5

ease of computation, these additional limitations sometimes exclude more efficient packings from consideration.

8. Acknowledgements. The author would like to express his appreciation to Mr. Sidney Kravitz of Dover, New Jersey for the verification of the numerical data in the tables.

TABLE 2

14 Circles, Coordinates of Centers

P	*X*	*Y*	Rho	Theta
1	0.6407	—1.5359	1.6642	292.643°
2	1.4053	−0.8915	1.6642	327.611°
3	1.6625	0.0749	1.6642	2.579°
4	1.3195	1.0142	1.6642	37.548°
5	0.5000	1.5873	1.6642	72.516°
6	−0.5000	1.5873	1.6642	107.484°
7	−1.1319	1.0142	1.6642	142.452°
8	−1.6625	0.0749	1.6642	177.421°
9	−1.4053	−0.8915	1.6642	212.389°
10	−0.6407	−1.5359	1.6642	247.357°
11	0.0000	0.7213	0.7213	90.000°
12	0.0000	−0.7681	0.7681	270.000°
13	0.6674	−0.0234	0.6678	357.990°
14	−0.6674	−0.0234	0.6678	182.010°

TABLE 3

16 Circles, Coordinates of Centers

P	*X*	*Y*	Rho	Theta
1	0.6014	−1.7048	1.8077	289.431°
2	1.4156	−1.1242	1.8077	321.544°
3	1.7967	−0.1997	1.8077	353.658°
4	1.6279	0.7860	1.8077	25.772°
5	0.9610	1.5311	1.8077	57.886°
6	0.0000	1.8077	1.8077	90.000°
7	−0.9610	1.5311	1.8077	122.114°
8	−1.6279	0.7860	1.8077	154.228°
9	−1.7967	−0.1997	1.8077	186.342°
10	−1.4156	−1.1242	1.8077	218.456°
11	−0.6014	−1.7048	1.8077	250.569°
12	0.0000	−0.9058	0.9058	270.000°
13	0.5000	0.6437	0.8151	52.162°
14	0.8027	−0.3094	0.8602	338.922°
15	−0.5000	0.6437	0.8151	127.838°
16	−0.8027	−0.3094	0.8602	211.078°

TABLE 4

17 Circles, Coordinates of Centers

P	X	Y	Rho	Theta
1	1.9037	−0.0444	1.9043	358.664°
2	1.6637	0.9264	1.9043	29.110°
3	0.9649	1.6417	1.9043	59.555°
4	0.0000	1.9043	1.9043	90.000°
5	−0.9649	1.6417	1.9043	120.445°
6	−1.6637	0.9264	1.9043	150.890°
7	−1.9037	−0.0444	1.9043	181.336°
8	0.6944	0.6788	0.9714	44.332°
9	−0.6944	0.6788	0.9714	135.668°
10	0.0000	−0.0403	0.0403	270.000°
11	0.9512	−0.3488	1.0132	339.860°
12	−0.9512	−0.3488	1.0132	200.140°
13	0.0000	−1.0403	1.0403	270.000°
14	1.4811	−1.6469	1.9043	321.056°
15	−1.4811	−1.6469	1.9043	218.944°
16	0.6703	−1.7824	1.9043	290.611°
17	−0.6703	−1.7824	1.9043	249.389°

TABLE 5

20 Circles, Coordinates of Centers

P	X	Y	Rho	Theta
1	0	0	Loose	
2	0.8897	−1.8593	2.0612	295.571°
3	1.6601	−1.2217	2.0612	323.649°
4	2.0397	−0.2966	2.0612	351.727°
5	1.9392	0.6984	2.0612	19.805°
6	1.3823	1.5289	2.0612	47.883°
7	0.5000	1.9996	2.0612	75.961°
8	−0.5000	1.9996	2.0612	104.039°
9	−1.3823	1.5289	2.0612	132.117°
10	−1.9392	0.6984	2.0612	160.195°
11	−2.0397	−0.2966	2.0612	188.273°
12	−1.6601	−1.2217	2.0612	216.351°
13	−0.8897	−1.8593	2.0612	244.429°
14	0.5000	0.9996	1.1177	63.426°
15	0.0000	−1.4027	1.4027	270.000°
16	1.0808	0.1805	1.0966	9.740°
17	0.7704	−0.7651	1.0858	315.198°
18	−0.5000	0.9996	1.1177	116.574°
19	−1.0808	0.1805	1.0966	170.260°
20	−0.7704	−0.7651	1.0858	224.802°

References

1. L. Fejes Tóth, Lagerungen in der Ebene, auf der Kugel und im Raum, Springer, Berlin, 1953.

2. C. A. Rogers, Packing and Covering, Cambridge University Press, New York, 1964.

3. Sidney Kravitz, Packing cylinders into cylindrical containers, this MAGAZINE, 40 (1967) 65–71.

4. ———, Engineering Materials and Design (London), June 1969, 875–876.

5. Udo Pirl, Der Mindesabstand von n in der Einheitskreisscheibe gelegenen Punkten, Math. Nachr., 40 (1969) 111–124.

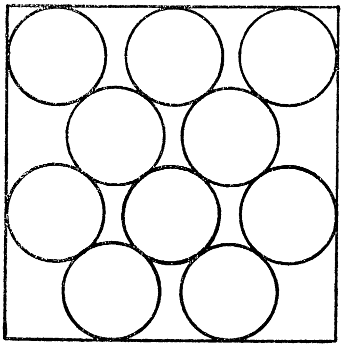
6. Michael Goldberg, Axially symmetric packing of equal circles on a sphere, Ann. Univ. Sci. Budapest. Eötvös. Sect. Math., 10 (1967) 37–48.

7. ——— Part II, Ann. Univ. Sci. Budapest. Eötvös. Sect. Math., 12 (1969) 137–142.

ON THE PACKING OF TEN EQUAL CIRCLES IN A SQUARE

J. SCHAER, University of Calgary

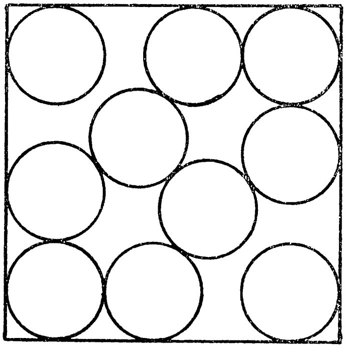
In a recent paper [1] M. Goldberg investigates the packing of n equal circles in a square. He finds good (i.e., dense) packings for $n \leqq 27$. In contrast to the cases $n \leqq 9$ [2], [3] his results are not proven to be best, i.e., densest possible. In fact, already for $n = 10$ the arrangement given by Goldberg (see Figure 1) can easily be improved (see Figure 2). However, Goldberg's arrangement is a structure, whereas in the packing of Figure 2 there are two loose circles; for packing bottles in square boxes, e.g., one would prefer to waste some space in packing in order to save bottles (and their content).



$m = 5/12 \approx .4167$

$\frac{4d}{\pi} \approx .8650$

FIG. 1.



$m = (4+\sqrt{2})(\sqrt{\sqrt{8}+1} - \sqrt{2})/7 \approx .4195$

$\frac{4d}{\pi} \approx .8735$

FIG. 2.

Unfortunately, I have no proof that the packing of Figure 2 is densest possible. So there is, for $n \geqq 10$, still much room for future investigation.

References

1. L. Fejes Tóth, Lagerungen in der Ebene, auf der Kugel und im Raum, Springer, Berlin, 1953.

2. C. A. Rogers, Packing and Covering, Cambridge University Press, New York, 1964.

3. Sidney Kravitz, Packing cylinders into cylindrical containers, this MAGAZINE, 40 (1967) 65–71.

4. ———, Engineering Materials and Design (London), June 1969, 875–876.

5. Udo Pirl, Der Mindesabstand von n in der Einheitskreisscheibe gelegenen Punkten, Math. Nachr., 40 (1969) 111–124.

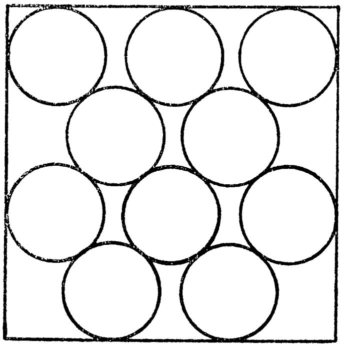
6. Michael Goldberg, Axially symmetric packing of equal circles on a sphere, Ann. Univ. Sci. Budapest. Eötvös. Sect. Math., 10 (1967) 37–48.

7. —— Part II, Ann. Univ. Sci. Budapest. Eötvös. Sect. Math., 12 (1969) 137–142.

ON THE PACKING OF TEN EQUAL CIRCLES IN A SQUARE

J. SCHAER, University of Calgary

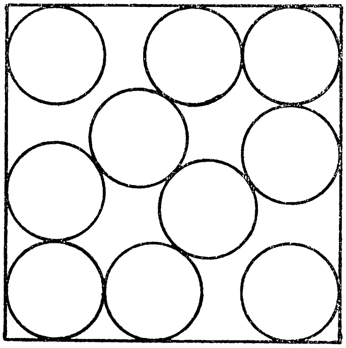
In a recent paper [1] M. Goldberg investigates the packing of n equal circles in a square. He finds good (i.e., dense) packings for $n \leqq 27$. In contrast to the cases $n \leqq 9$ [2], [3] his results are not proven to be best, i.e., densest possible. In fact, already for $n=10$ the arrangement given by Goldberg (see Figure 1) can easily be improved (see Figure 2). However, Goldberg's arrangement is a structure, whereas in the packing of Figure 2 there are two loose circles; for packing bottles in square boxes, e.g., one would prefer to waste some space in packing in order to save bottles (and their content).



$m=5/12\approx.4167$

$\frac{4d}{\pi}\approx.8650$

FIG. 1.



$m=(4+\sqrt{2})(\sqrt{\sqrt{8}+1}-\sqrt{2})/7\approx.4195$

$\frac{4d}{\pi}\approx.8735$

FIG. 2.

Unfortunately, I have no proof that the packing of Figure 2 is densest possible. So there is, for $n \geqq 10$, still much room for future investigation.

References

1. M. Goldberg, The packing of equal circles in a square, this MAGAZINE, 43 (1970) 24–30.
2. J. Schaer and A. Meir, On a geometric extremum problem, Canad. Math. Bull, 8 (1965) 21–27.
3. J. Schaer, The densest packing of nine circles in a square, Canad. Math. Bull, 8 (1965) 273–277.

ANSWERS

A518. The number 121 is a perfect square for every base $n>2$, viz.,

$$(121)_n = n^2 + 2n + 1 = (n+1)^2.$$

A519. The given condition is equivalent to

$$\begin{aligned}\tan C &= -(\tan A + \tan B)/(1 - \tan A \tan B)\\ &= -\tan(A+B).\end{aligned}$$

Hence

$$C = -(A+B) + n\pi.$$

Thus

$$A + B + C = n\pi$$

$n=0, \pm 1, \pm 2 \cdots$.

A520. Let $d(h) = h/\sqrt{1+h} - \log(1+h)$. Observe that $d(0)=0$ and $d(h)>0$ for h sufficiently large, since $h/\sqrt{1+h}$ is asymptotic to $\sqrt{h}$. So if $d(h) \leqq 0$ for some $h>0$, it can only mean that $d'(\alpha)=0$ for some $\alpha>0$. But it is easily checked that $d'=0$ only at 0 and so $d(h)>0$ for all $h>0$.

A521. If x_i is an odd prime, p, then it is easily computed that for each positive integer, r,

$$\begin{aligned}x_{i+r} &= 2^r x_i + (2^r - 1)\\ &= 2^r p + (2^r - 1).\end{aligned}$$

When $r=p-1$, $x_{i+(p-1)} = 2^{p+1}p + (2^{p+1} - 1)$. Since Fermat's theorem guarantees that $p \mid (2^{p+1} - 1)$, it follows that $x_{i+(p-1)}$ is composite. The sequence (x_i) cannot, therefore, consist wholly of primes.

A522. First note that XYZ must be acute. Now let $2X = \pi - R$, $2Y = \pi - S$, and $2Z = \pi - T$. Thus RST is a triangle and

$$\frac{\sin R}{\sin A} = \frac{\sin S}{\sin B} = \frac{\sin T}{\sin C}.$$

Whence $RST \sim ABC$ and

$$2X = \pi - A, \quad 2Y = \pi - B, \quad 2Z = \pi - C.$$

(Quickies on page 178.)

CORN ROWS AND CONVEX CURVES

FOSTER BROOKS, Kent State University, Kent, Ohio

Suppose one has a field of arbitrary shape that is to be planted in some crop such as corn, in uniformly spaced parallel straight rows, how does the average length of the rows vary as a function of the angle of orientation? Clearly such a function is substantially restricted by the conditions of the problem, but in fact it turns out to be less so than might be expected naïvely. The result is that the average row length, when plotted radially in polar coordinates as a function of the orientation angle, gives a convex curve of period π, and conversely every such curve is attainable in this way by proper choice of the shape and size of the field.

In mathematical terms, a "field" may be represented as a bounded connected open planar region R, and a "row" as the intersection of R with a straight line in the plane, passing through R. If R is convex, every row consists of a single interval on the line; otherwise a row may comprise two or more intervals, in which case its length may be taken as the sum (possibly denumerable) of the lengths of the separate pieces. Now let coordinate systems be established in the plane, an xy-cartesian system with some point O of R as origin and a related $r\theta$-polar system with $x=r\cos\theta$ and $y=r\sin\theta$ as usual. Then for a fixed direction θ we can represent the set of all straight lines parallel to each other and normal to the direction θ, by $x\cos\theta+y\sin\theta-p=0$, each choice of the parameter p giving one line of the set. Since O is interior to R, lines with p sufficiently near zero must intersect R, but for p sufficiently large, either positive or negative, they will not, since R is bounded. Thus for each direction θ there exist a least upper bound $p_+(\theta)$ and a greatest lower bound $p_-(\theta)$ of the set of those values of p which give lines that do intersect R. These will be called respectively the *positive extent* and the *negative extent* of R for the direction θ with respect to the origin O. The lines with $p=p_+(\theta)$ and $p=p_-(\theta)$ are the *supporting lines* for R normal to the direction θ. Letting θ vary, the function $r=p_+(\theta)$ will be called the *extent function* of R with respect to the origin O, and the function $d(\theta)=p_+(\theta)-p_-(\theta)=p_+(\theta)+p_+(\theta+\pi)$ will be called the *diameter function* for the region R. The diameter function clearly is independent of the position of the origin, and it is periodic of period π.

Now for each value of θ, and for each p between $p_-(\theta)$ and $p_+(\theta)$, a unique row is determined whose length we denote by $l(\theta, p)$. To compute the average row length for a given θ the row lengths for those values of p giving the uniform row spacing should be added, and the total divided by the number of rows. Idealizing this to its limit as the row spacing approaches zero, we define the *average row length*, denoted by $l(\theta)$, as $l(\theta)=(\int_{p_-(\theta)}^{p_+(\theta)} l(\theta, p)dp)/(p_+(\theta)-p_-(\theta))$. Since the integral appearing here gives the area of R, we have $l(\theta)=(\text{Area of } R)/d(\theta)$. Thus the function $l(\theta)$ is a constant multiple of the reciprocal of the diameter function $d(\theta)$, so that its nature is determined by that of $d(\theta)$. This in turn depends upon the nature of $p_+(\theta)$, which may be described by the following

THEOREM. *If $p_+(\theta)$ is an extent function for a region R as described above, and* *if P is a point on the curve C whose equation in polar coordinates is $r=p_+(\theta)$, then there exists a circle K passing through P and through the origin O such that no point of C lies inside K.*

Proof. Construct the region R, and plot on the same figure the corresponding curve C (Figure 1). At any point P on C let S be the straight line through P perpendicular to OP, i.e., normal to the direction θ. Clearly S is a supporting line for R. As such it has no points in common with the open region R, but it must contain at least one point Q of the boundary of R. Let K be the circle with diameter OQ. This circle passes through P, but no point P' of the curve C can lie inside it, since the angle $OP'Q$ would then exceed $\pi/2$, and this would require the point Q of the boundary of R to lie outside (opposite side from O) of the line through P' perpendicular to OP', a supporting line for R, which would give a contradiction.

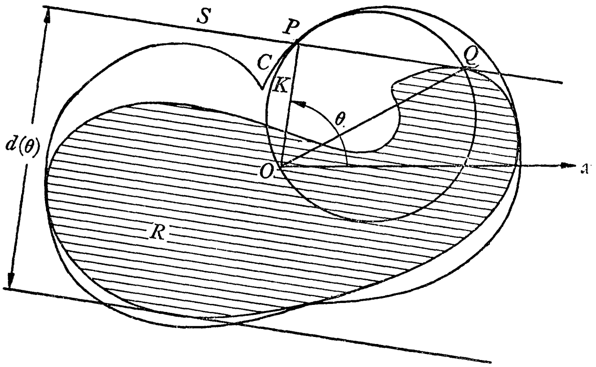


FIG. 1.

A polar function $r=f(\theta)$, $f(\theta)>0$, having the property proved for $p_+(\theta)$ in the above theorem will be called a *reciprocally convex* function. This terminology is used because of the following

THEOREM. *If $r=f(\theta)$ is a reciprocally convex function, then the function $r=1/f(\theta)$ has a polar graph which is a convex curve, and conversely.*

Proof. Under the reciprocal transformation every circle passing through the origin transforms into a straight line. Hence if for every point P on the graph of $r=f(\theta)$ there exists a circle passing through both P and the origin with no point of the graph inside it, then for every corresponding point P^{-1} on the graph of $r=1/f(\theta)$ there exists a straight line through P^{-1} with no points of this graph outside it (opposite side from the origin). Hence the graph of $r=1/f(\theta)$ is a closed curve through each point of which there exists a supporting line, and thus it is a convex curve. The converse follows since all steps above are reversible.

Since the diameter function $d(\theta)$ is the sum of two extent functions $p_+(\theta)$ and $p_+(\theta+\pi)$, both reciprocally convex, we now need the following

THEOREM. *The sum of two reciprocally convex functions is reciprocally convex.*

Proof. Let $f_1(\theta)$ and $f_2(\theta)$ be two reciprocally convex functions, and for θ_0 some fixed value of θ, let P_1 and P_2 be the points $(\theta_0, f_1(\theta_0))$ and $(\theta_0, f_2(\theta_0))$ respectively. Then there exist two circles, K_1 through P_1 and K_2 through P_2, both through the origin, such that no points of the respective graphs lie inside them. Polar equations for these circles are of the form $r=A_1\cos(\theta-\delta_1)$ and $r=A_2\cos(\theta-\delta_2)$, which give by adding their right members an equation of the form $r=A_3\cos(\theta-\delta_3)$, again a circle K_3 through the origin, and one which passes through the point $(\theta_0, f_1(\theta_0)+f_2(\theta_0))$. Since no point of the graph of $r=f_1(\theta)$ lies inside K_1 we have $f_1(\theta)\geqq\max(A_1\cos(\theta-\delta_1), 0)$, with a similar result for $f_2(\theta)$. Thus

$$\begin{aligned} f_1(\theta)+f_2(\theta) &\geqq \max(A_1\cos(\theta-\delta_1), 0)+\max(A_2\cos(\theta-\delta_2), 0) \\ &\geqq A_1\cos(\theta-\delta_1)+A_2\cos(\theta-\delta_2), \end{aligned}$$

and since $f_1(\theta)+f_2(\theta)$ cannot be negative, $f_1(\theta)+f_2(\theta)\geqq\max(A_3\cos(\theta-\delta_3), 0)$. Hence no point on the graph of $r=f_1(\theta)+f_2(\theta)$ can lie inside K_3.

From this theorem it follows directly that $d(\theta)$ is reciprocally convex and hence that $l(\theta)$ is a function whose polar graph is a convex curve. Examples of regions of various shapes, with graphs of their corresponding $d(\theta)$ and $l(\theta)$ functions, are shown in Figure 2.

The theorem just above has another immediate consequence of some interest, namely the following

COROLLARY. *If $g_1(\theta)$ and $g_2(\theta)$ are two positive functions whose polar graphs are convex curves, then their harmonic mean is a function having this same property.*

The converse of the main result above holds also, as stated in the following

THEOREM. *If $r=f(\theta)$ is defined for all θ, positive, bounded, periodic of period π, and its polar graph is a convex curve, then there exists a bounded connected open region R whose average row length function $l(\theta)=f(\theta)$.*

Proof. Let $f(\theta)$ be a function satisfying the hypothesis of the theorem. Let $d(\theta)=1/f(\theta)$, and $p(\theta)=p(\theta+\pi)=\frac{1}{2}d(\theta)$, all of these being reciprocally convex functions. Now for each direction θ we construct two parallel straight lines normal to the direction θ with the origin O midway between them and at distance $p(\theta)$ from each, and we denote by $G(\theta)$ the open strip in the plane lying between these lines. Let R be the intersection of such strips for all directions θ. Clearly R contains the origin O as interior point, and it is a bounded connected set. We now show that for every θ each of the two parallel lines bounding $G(\theta)$ contains at least one point of the closure of R, thus assuring that all such lines are supporting lines of R, and hence that $p(\theta)$ actually is the extent function for R with respect to O. Let θ be fixed, and let P be the point $(\theta, p(\theta))$. Since $p(\theta)$ is a reciprocally convex function, there exists a circle K through P and through O which all points of the graph of $r=p(\theta)$ lie on or outside. Let Q be the point on K diametrically opposite O. Then for no point P' on the graph of $r=p(\theta)$ can the line through P' perpendicular to OP' pass between Q and O. Thus the open segment OQ is contained in every $G(\theta)$, and hence in R, so that Q is a boundary

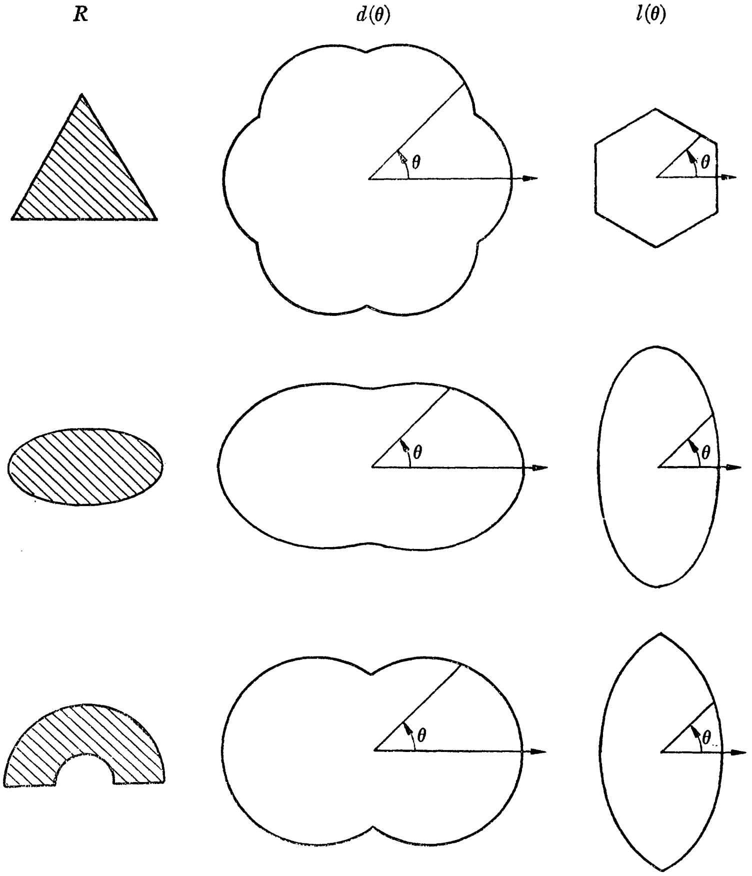


Fig. 2.

point of R. For the region R thus obtained the average row length function $l(\theta)$ is given by $l(\theta) = (\text{Area of } R)/d(\theta) = (\text{Area of } R)\cdot f(\theta)$. If now the entire figure is magnified by a factor k, the area of R appearing in the numerator just above will be multiplied by k^2 whereas $d(\theta)$ in the denominator will be multiplied by k. Hence a value of k exists for which the magnified region has its average row length function equal to $f(\theta)$.

Finally, it may be noted that any two regions having the same convex hull will have identical diameter functions, so that their average row length functions will be proportional. Hence the inclusion or exclusion of holes or dents of any shape that do not extend beyond the convex hull of a region will affect the average row lengths of the region in all directions equally, in the same ratio as the area.

SEQUENCES VS. NEIGHBORHOODS

JOHN H. STAIB, Drexel University

In the teaching of advanced calculus I believe that it is important for the limit of a function to be considered both from a neighborhood point of view and from a sequence point of view. In my own classes I not only demonstrate the equivalence of the two definitions but also strive throughout the course to select proofs that show off first the one approach and then the other. On occasion it seems instructive to give both neighborhood and sequence proofs for the same theorem, especially where the distinctive flavor of each approach is apparent.

This dual treatment of the limit concept is, of course, time consuming. However, I would defend my practice on two grounds: first, both the fundamental nature of the limit concept and its subtlety warrant the emphasis implied by the dual treatment; second, the whole business of "enjoying" a proof seems to be stimulated by giving the class a choice of proofs and then encouraging each student to take a position as to which proof is more elegant, more natural, more clever, and so on.

A familiar classroom example will serve, hopefully, to make my point. Let f be defined on $[0, 1]$ by the formula

$$f(x) = \begin{cases} 0, & x \text{ is irrational} \\ \dfrac{1}{q}, & x = \dfrac{p}{q} \text{ (reduced).} \end{cases}$$

After the immediate observation that f is discontinuous at each rational point, I turn to the problem of proving that f is continuous at each irrational point α. (The following two proofs are essentially in the form that I use at the blackboard.)

A Neighborhood Proof. Given $\epsilon > 0$ we define a *bad rational* to be one whose denominator does not exceed $1/\epsilon$, and then let S be the set of all bad rationals in the interval $(\alpha-1, \alpha+1)$. Thus, S consists of those fractions in $(\alpha-1, \alpha+1)$ having denominator 1, those having denominator 2, and so on through those having denominator $[1/\epsilon]$. Evidently S is a finite set and, consequently, members of S cannot come arbitrarily close to α. It follows that we may choose a δ-neighborhood of α which excludes the members of S.

Now, take x as any number in this neighborhood. There are two cases:

(1) If x is irrational, then

$$|f(x) - f(\alpha)| = |0 - 0| < \epsilon.$$

(2) If x is rational, then x—not being in S—is a *good rational.* That is, $x = p/q$ where q does exceed $1/\epsilon$. Then

$$|f(x) - f(\alpha)| = \left|\frac{1}{q} - 0\right| = \frac{1}{q} < \epsilon.$$

Thus, $|f(x)-f(\alpha)| < \epsilon$ whenever $x \in (\alpha-\delta, \alpha+\delta)$.

A Sequence Proof. Let $\{x_n\}$ be any α-approaching sequence; we wish to prove that the corresponding sequence $\{f(x_n)\}$ has the limit $0=f(\alpha)$. Since $f(x)=0$ for irrationals, we may assume—without loss of generality—that each x_n is rational, say p_n/q_n.

We proceed by contradiction. Suppose that $\lim \{f(p_n/q_n)\} \neq 0$. This means that there exists an $\epsilon>0$ such that $f(p_n/q_n)=1/q_n \geqq \epsilon$ for infinitely many n. Let us call such a p_n/q_n a *bad term* of the sequence $\{p_n/q_n\}$. Now, from $1/q_n \geqq \epsilon$ it follows that $q_n \leqq 1/\epsilon$. Thus, the denominators of our bad terms range over a finite set. Consequently, there exists at least one value, say q^*, that occurs infinitely often as the denominator of a bad term. Therefore, we may take $\{p_{k_n}/q_{k_n}\}$ to be a subsequence of $\{p_n/q_n\}$ such that $q_{k_n}=q^*$ for all n. Then we may write

$$q^*\alpha = \lim\left\{q^* \cdot \frac{p_{k_n}}{q_{k_n}}\right\} = \lim\left\{q_{k_n} \cdot \frac{p_{k_n}}{q_{k_n}}\right\} = \lim\{p_{k_n}\}.$$

But $q^*\alpha$ is irrational and thus cannot be the limit of a sequence of integers.

A PROBABILITY OF MORE HEADS

MURRAY S. KLAMKIN, Scientific Research Staff, Ford Motor Company

In this note we present some nonroutine material which could be used in elementary probability classes. Although the first problem treated is not new (see [1]), it does not appear to be well known.

The first problem we shall consider is a simple coin tossing one whose probability can easily be obtained as a finite double sum by a direct enumeration of all the possible outcomes. We then show how to evaluate this double sum simply by calculating the probability in another way by exploiting the symmetry of the situation. Subsequently, we shall generalize the problem.

PROBLEM. *If A and B toss $n+1$ and n fair coins, respectively, what is the probability P_n that A gets more heads than B?*

First we consider some simple special cases.

1. $n=0$. Obviously $P_0=\frac{1}{2}$.
2. $n=1$. Here the various tosses can be

A	B
HH	H
HT	T
TH	
TT	

Thus, $|f(x)-f(\alpha)| < \epsilon$ whenever $x \in (\alpha-\delta, \alpha+\delta)$.

A Sequence Proof. Let $\{x_n\}$ be any α-approaching sequence; we wish to prove that the corresponding sequence $\{f(x_n)\}$ has the limit $0=f(\alpha)$. Since $f(x)=0$ for irrationals, we may assume—without loss of generality—that each x_n is rational, say p_n/q_n.

We proceed by contradiction. Suppose that $\lim \{f(p_n/q_n)\} \neq 0$. This means that there exists an $\epsilon > 0$ such that $f(p_n/q_n) = 1/q_n \geqq \epsilon$ for infinitely many n. Let us call such a p_n/q_n a *bad term* of the sequence $\{p_n/q_n\}$. Now, from $1/q_n \geqq \epsilon$ it follows that $q_n \leqq 1/\epsilon$. Thus, the denominators of our bad terms range over a finite set. Consequently, there exists at least one value, say q^*, that occurs infinitely often as the denominator of a bad term. Therefore, we may take $\{p_{k_n}/q_{k_n}\}$ to be a subsequence of $\{p_n/q_n\}$ such that $q_{k_n} = q^*$ for all n. Then we may write

$$q^*\alpha = \lim \left\{ q^* \cdot \frac{p_{k_n}}{q_{k_n}} \right\} = \lim \left\{ q_{k_n} \cdot \frac{p_{k_n}}{q_{k_n}} \right\} = \lim \{p_{k_n}\}.$$

But $q^*\alpha$ is irrational and thus cannot be the limit of a sequence of integers.

A PROBABILITY OF MORE HEADS

MURRAY S. KLAMKIN, Scientific Research Staff, Ford Motor Company

In this note we present some nonroutine material which could be used in elementary probability classes. Although the first problem treated is not new (see [1]), it does not appear to be well known.

The first problem we shall consider is a simple coin tossing one whose probability can easily be obtained as a finite double sum by a direct enumeration of all the possible outcomes. We then show how to evaluate this double sum simply by calculating the probability in another way by exploiting the symmetry of the situation. Subsequently, we shall generalize the problem.

PROBLEM. *If A and B toss $n+1$ and n fair coins, respectively, what is the probability P_n that A gets more heads than B?*

First we consider some simple special cases.

1. $n=0$. Obviously $P_0 = \frac{1}{2}$.
2. $n=1$. Here the various tosses can be

A	B
HH	H
HT	T
TH	
TT	

Since there are 4×2 outcomes of which 4 cases are favorable, $P_1 = \frac{1}{2}$.

3. $n=2$. The various tosses can be

A	B
HHH	HH
HHT	HT
HTH	TH
THH	TT
TTH	
THT	
HTT	
TTT	

Since there are 8×4 outcomes of which 16 cases are favorable, $P_2 = \frac{1}{2}$.

At this stage, if one were daring, one might conjecture that $P_n = \frac{1}{2}$ for all n. We now show that this conjecture is indeed valid.

In the general case, the number of different tosses that A and B can get are 2^{n+1} and 2^n, respectively, since each coin can come up in two different ways (heads or tails). If B gets r heads, then A will have to have any number of heads greater than r; i.e., $r+1$, $r+2$, $r+3, \cdots, n+1$. The number of ways that B can get r heads tossing n coins is $\binom{n}{r}$. The number of ways that A can get more than r heads by tossing $n+1$ coins is

$$\binom{n+1}{r+1} + \binom{n+1}{r+2} + \cdots + \binom{n+1}{n+1}.$$

The number of ways that A can have more heads than B if the latter gets r heads is then

$$\binom{n}{r}\left\{\binom{n+1}{r+1} + \binom{n+1}{r+2} + \cdots + \binom{n+1}{n+1}\right\} = \binom{n}{r}\sum_{s=1}^{n+1-r}\binom{n+1}{r+s}.$$

Since r can be $1, 2, \cdots, n$, the total number of ways that A can have more heads than B is obtained by summing the latter expression over r, i.e.,

$$\sum_{r=1}^{n}\binom{n}{r}\sum_{s=1}^{n+1-r}\binom{n+1}{r+s}.$$

Since the total number of different outcomes is $2^{n+1}\cdot 2^n$, the desired probability is given by

$$P_n = \frac{1}{2^{2n+1}}\sum_{r=1}^{n}\sum_{s=1}^{n+1-r}\binom{n}{r}\binom{n+1}{r+s}.$$

Although the sum can be evaluated by the use of two combinatorial identities, we will now calculate P_n in a simpler fashion.

First consider n coins of A versus n coins of B. We then have the following table giving the number of ways A has more, equal or less heads than B, respectively:

Event	*Number of ways*
$A > B$	r
$A = B$	s
$A < B$	r

By symmetry, the number of ways for $A > B$ is the same as for $A < B$ (since the coins are fair). Since $2n$ coins have been tossed, $2r+s=2^{2n}$. Now consider the $(n+1)$th coin of A. If it comes up heads that gives $r+s$ ways for $A>B$; if it comes up tails, that gives r ways more. Then,

$$P_n = \frac{2r+s}{2^{2n+1}} = \frac{1}{2} \cdot$$

For an alternative simple proof, we give a somewhat expanded version of the one of Uspensky [1].

Let A_1 and B_1 denote the number of heads tossed by A and B, respectively, and let A_2 and B_2 denote the respective number of tails. Then

$$A_1 + A_2 = n + 1,$$
$$B_1 + B_2 = n.$$

The probability of $A_1 > B_1$ is the same as the probability of $n+1-A_2>n-B_2$ or $A_2 \leqq B_2$. Since the coins are all fair, there will be no change if we interchange heads with tails, i.e.,

$$\text{Prob.}\{A_1 > B_1\} = \text{Prob.}\{A_2 \leqq B_2\} = \text{Prob.}\{A_1 \leqq B_1\}.$$

Since the two inequalities $A_1>B_1$ and $A_1 \leqq B_1$ exhaust all the possible outcomes,

$$\text{Prob.}\{A_1 > B_1\} + \text{Prob.}\{A_1 \leqq B_1\} = 1.$$

Whence,

$$\text{Prob.}\{A_1 > B_1\} = \tfrac{1}{2}.$$

One generalization of the previous problem is to have A and B toss $n+m$ and n coins, respectively. If $P(m, n)$ denotes the probability that A has more heads than B, then proceeding in the same way as before, we find that

$$P(m, n) = \frac{1}{2^{2n+m}} \sum_{r=0}^{n} \sum_{s=1}^{n+m} \binom{n}{r}\binom{n+m}{r+s}.$$

It is to be noted that $\binom{n+m}{i}=0$ if $i>n+m$.

To evaluate the summation over r, we first determine the coefficient of t^s on both sides of the identity

$$\frac{(1+t)^{m+2n}}{t^n} = (1+t)^{m+n}(1+1/t)^n.$$

This is given by

$$\binom{2n+m}{s+n} = \sum_{r=0}^{n} \binom{n}{r}\binom{n+m}{s+r}.$$

Substituting back in P, gives

$$P(m, n) = \frac{1}{2^{2n+m}} \sum_{s=1}^{m+n} \binom{2n+m}{s+n} = \frac{1}{2^{2n+m}} \sum_{t=0}^{m+n-1} \binom{2n+m}{t}.$$

To reduce the number of terms in the latter summation, we use the identity

$$\sum_{t=0}^{2n+m} \binom{2n+m}{t} = 2^{2n+m}.$$

This gives for special cases that

$$P(0, n) = \frac{1}{2} - \frac{1}{2^{2n+1}}\binom{2n}{n},$$

$$P(1, n) = \frac{1}{2},$$

$$P(2, n) = \frac{1}{2} + \frac{1}{2^{2n+3}}\binom{2n+2}{n+1}.$$

For general m, we have

1. $m = 2a > 0$,

$$P(m, n) = \frac{1}{2} - \frac{1}{2^{2n+2a+1}} \left\{ \binom{2n+2a}{n+a} - 2 \sum_{t=n+a}^{n+2a-1} \binom{2n+2a}{t} \right\},$$

2. $m = 2a+1 > 1$,

$$P(m, n) = \frac{1}{2} + \frac{1}{2^{2n+2a+1}} \sum_{t=n+a+1}^{n+2a} \binom{2n+2a+1}{t}.$$

If the probability of getting a head and tail for each coin is p and q, respectively, then

$$P(m, n) = \frac{1}{2^{2n+m}} \sum_{r=0}^{n} \sum_{s=1}^{m+n} \binom{n}{r}\binom{m+n}{r+s} p^{2r+s} q^{m+2n-r-2s}.$$

It is doubtful that this double sum can be reduced to a "simple" single one.

Reference

1. J. V. Uspensky, Introduction to Mathematical Probability, McGraw-Hill, New York, 1937, pp. 38, 59.

THE LAW OF SINES AND LAW OF COSINES FOR POLYGONS

R. B. KERSHNER, The Johns Hopkins University

In connection with my investigations of polygons [1] I had need of a convenient analytic formulation of the restrictions imposed on a set of numbers by the fact that they were the angles and sides of a convex polygon. I found appropriate expressions in the form of straightforward generalizations of the law of sines and law of cosines for triangles. These laws are so elegant, useful, and elementary that they can hardly be new but a modest search has failed to discover them in the literature. Accordingly I am here making them available. I hope that others may find them as useful as I have.

The laws will be stated and proved here for pentagons and hexagons, thus illustrating the slight difference between the even and odd cases; the extension to any number of sides is completely obvious. Let the angles of a pentagon be A, B, C, D, E successively and the sides a, b, c, d, e in such a way that a and b are the sides of angle A. Place this pentagon on the Cartesian plane as in Figure 1, so that side a occupies the interval from $(0, 0)$ to $(a, 0)$. Now the inclination of side b is clearly the exterior angle $\pi-A$. The direction of side c is obtained from that of side b by a further rotation through $\pi-B$, hence side c has inclination $2\pi-(A+B)$; similarly the inclination of side d is $3\pi-(A+B+C)$, the inclination of side e is $4\pi-(A+B+C+D)$, and the inclination of side a is $5\pi-(A+B+C+D+E)=2\pi$. The last equality expresses the very well known fact that the sum of the interior angles of a pentagon is 3π, i.e.,

$$A + B + C + D + E = 3\pi. \tag{1}$$

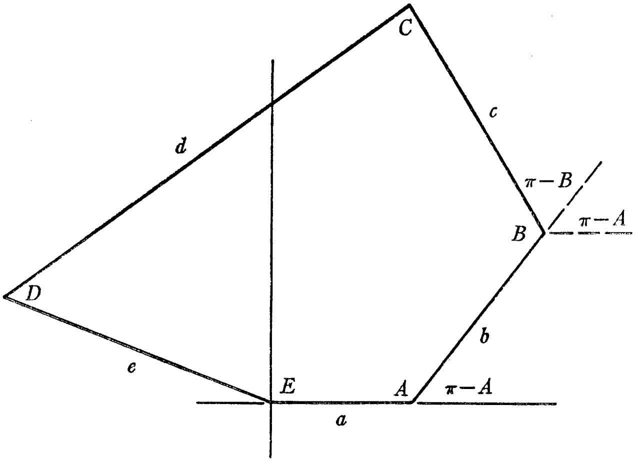


FIG. 1.

The fact that the pentagon (Figure 1) is a closed figure implies that the sum of the horizontal projections of the five sides is zero, i.e.,

$$\begin{aligned} &a + b\cos(\pi - A) + c\cos(2\pi - (A + B)) \\ &\qquad + d\cos(3\pi - (A + B + C)) + e\cos(4\pi - (A + B + C + D)) = 0. \end{aligned} \tag{2}$$

Using (1) and elementary trigonometry, (2) yields

$$a - b \cos A + c \cos(A + B) = e \cos E - d \cos(D + E). \tag{3}$$

In the same way it is seen that the sum of the vertical projections of the five sides in Figure 1 is also zero, i.e.,

$$\begin{aligned} &b \sin(\pi - A) + c \sin(2\pi - (A + B)) + d \sin(3\pi - (A + B + C)) \\ &\qquad + e \sin(4\pi - (A + B + C + D)) = 0, \end{aligned} \tag{4}$$

or, from (1),

$$b \sin A - c \sin(A + B) = e \sin E - d \sin(D + E). \tag{5}$$

Squaring each of the equations (3) and (5) and adding yields, immediately,

$$\begin{aligned} &a^2 + b^2 + c^2 - 2ab \cos A - 2bc \cos B + 2ac \cos(A + B) \\ &\qquad = d^2 + e^2 - 2de \cos D. \end{aligned} \tag{6}$$

The equation (5) is easily seen to be a direct generalization of the ordinary law of sines for triangles (allowing for the difference between the nomenclature scheme of Figure 1 and the usual nomenclature for a triangle). It clearly can be written in five different ways by relabelling the sides, namely,

Law of sines for pentagons.

(a) $b \sin A - c \sin(A + B) = e \sin E - d \sin(D + E)$

(b) $c \sin B - d \sin(B + C) = a \sin A - e \sin(E + A)$

(c) $d \sin C - e \sin(C + D) = b \sin B - a \sin(A + B)$

(d) $e \sin D - a \sin(D + E) = c \sin C - b \sin(B + C)$

(e) $a \sin E - b \sin(E + A) = d \sin D - c \sin(C + D)$

Like the law of sines for triangles, each of these equations involves all but one of the sides and all but one of the angles of the polygon.

In the same way equation (6) is a direct generalization of the law of cosines for triangles. It also can be written in five different ways, namely,

Law of cosines for pentagons.

(a) $a^2 + b^2 + c^2 - 2ab \cos A - 2bc \cos B + 2ac \cos(A + B)$
$= d^2 + e^2 - 2de \cos D$

(b) $b^2 + c^2 + d^2 - 2bc \cos B - 2cd \cos C + 2bd \cos(B + C)$
$= e^2 + a^2 - 2ea \cos E$

(c) $c^2 + d^2 + e^2 - 2cd \cos C - 2de \cos D + 2ce \cos(C + D)$
$= a^2 + b^2 - 2ab \cos A$

(d) $d^2 + e^2 + a^2 - 2de \cos D - 2ea \cos E + 2da \cos(D + E)$
$= b^2 + c^2 - 2bc \cos B$

(e) $$e^2 + a^2 + b^2 - 2ea \cos E - 2ab \cos A + 2eb \cos(E + A) = c^2 + d^2 - 2cd \cos C$$

Like the law of cosines for triangles each of these equations involves all of the sides and all but two of the angles of the polygon.

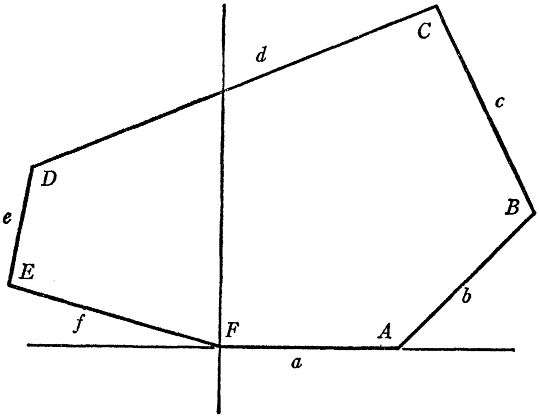


FIG. 2.

The derivation of the equivalent laws for a hexagon is now quite obvious. Let the hexagon be represented by Figure 2. Then the fact that the horizontal and vertical projections add to zero can be written as

$$a - b \cos A + c \cos(A + B) = f \cos F - e \cos(E + F) + d \cos(D + E + F) \tag{7}$$

and

$$b \sin A - c \sin(A + B) = f \sin F - e \sin(E + F) + d \sin(D + E + F). \tag{8}$$

Squaring (7) and (8) and adding gives

$$a^2 + b^2 + c^2 - 2ab \cos A - 2bc \cos B + 2ac \cos(A + B) = d^2 + e^2 + f^2 - 2de \cos D - 2ef \cos E + 2df \cos(D + E). \tag{9}$$

Equation (8) will be called the law of sines for hexagons and may be written in six different ways, namely,

Law of sines for hexagons.

(a) $b \sin A - c \sin(A + B) = f \sin F - e \sin(E + F) + d \sin(D + E + F)$

(b) $c \sin B - d \sin(B + C) = a \sin A - f \sin(F + A) + e \sin(E + F + A)$

(c) $d \sin C - e \sin(C + D) = b \sin B - a \sin(A + B) + f \sin(F + A + B)$

(d) $e \sin D - f \sin(D + E) = c \sin C - b \sin(B + C) + a \sin(A + B + C)$

(e) $f \sin E - a \sin(E + F) = d \sin D - c \sin(C + D) + b \sin(B + C + D)$

(f) $a \sin F - b \sin(F + A) = e \sin E - d \sin(D + E) + c \sin(C + D + E)$

Again, equation (9) will be called the law of cosines for hexagons. In this case, because of symmetry the six different formulations consist of three identical pairs so that there are only three forms of the law of cosines for hexagons, namely,

Law of cosines for hexagons.

$$\begin{aligned}\text{(a)}\quad & a^2+b^2+c^2-2ab\cos A-2bc\cos B+2ac\cos(A+B)\\ &\qquad = d^2+e^2+f^2-2de\cos D-2ef\cos E+2df\cos(D+E)\\ \text{(b)}\quad & b^2+c^2+d^2-2bc\cos B-2cd\cos C+2bd\cos(B+C)\\ &\qquad = e^2+f^2+a^2-2ef\cos E-2fa\cos F+2ea\cos(E+F)\\ \text{(c)}\quad & c^2+d^2+e^2-2cd\cos C-2de\cos D+2ce\cos(C+D)\\ &\qquad = f^2+a^2+b^2-2fa\cos F-2ab\cos A+2fb\cos(F+A)\end{aligned}$$

Again, the equalities in the law of sines involve all but one side and all but one angle, those in the law of cosines involve all sides and all but two angles.

As stated earlier, the generalization of these laws to those for polygons of any number of sides is straightforward.

Reference

1. R. B. Kershner, On paving the plane, Amer. Math. Monthly, 75 (1968) 839–844.

QUADRATIC RESIDUES IN GF (p^2)

REINALDO E. GIUDICI, Universidad Santa Maria, Valparaiso, Chile

1. Introduction. In [1] N. R. Hardman and J. H. Jordan have studied the distribution of quadratic residues in fields of order p^2, where p is a prime and $p\equiv 3 \pmod 4$.

In this paper we will eliminate the above restriction and study quadratic residues in $GF(p^2)$ [3, p. 446] for any odd prime p. For this purpose, $GF(p^2)$ is considered as a normal extension of the finite field $GF(p)$ and, as is well known, it can be conveniently written in the form:

$$GF(p^2) = \{a+b\theta \mid \theta^2 = g;\ a, b \in GF(p)\},$$

where g is a primitive root (mod p) [3, p. 447].

As is expected, some of the theorems stated in [1] turn out to hold also in the case $p\equiv 1$ (mod 4); but others change when passing from $p=4k+3$ to $p=4k+1$.

2. Generators; the symbol $\chi(\alpha)$. Let λ be a fixed generator of the cyclic multiplicative group $GF^*(p^2) = GF(p^2) - \{0\}$. Then, $GF^*(p^2) = [\lambda] = \{1, \lambda, \lambda^2, \cdots, \lambda^{p^2-2}\}$ and λ^{p+1} is a primitive root of p.

DEFINITION 1. *Let $\alpha \in GF^*(p^2)$. If the congruence $\eta^2 \equiv \alpha \pmod p$ has a solution*

Again, equation (9) will be called the law of cosines for hexagons. In this case, because of symmetry the six different formulations consist of three identical pairs so that there are only three forms of the law of cosines for hexagons, namely,

Law of cosines for hexagons.

$$\text{(a)}\quad a^2+b^2+c^2-2ab\cos A-2bc\cos B+2ac\cos(A+B)$$
$$=d^2+e^2+f^2-2de\cos D-2ef\cos E+2df\cos(D+E)$$
$$\text{(b)}\quad b^2+c^2+d^2-2bc\cos B-2cd\cos C+2bd\cos(B+C)$$
$$=e^2+f^2+a^2-2ef\cos E-2fa\cos F+2ea\cos(E+F)$$
$$\text{(c)}\quad c^2+d^2+e^2-2cd\cos C-2de\cos D+2ce\cos(C+D)$$
$$=f^2+a^2+b^2-2fa\cos F-2ab\cos A+2fb\cos(F+A)$$

Again, the equalities in the law of sines involve all but one side and all but one angle, those in the law of cosines involve all sides and all but two angles.

As stated earlier, the generalization of these laws to those for polygons of any number of sides is straightforward.

Reference

1. R. B. Kershner, On paving the plane, Amer. Math. Monthly, 75 (1968) 839–844.

QUADRATIC RESIDUES IN $GF(p^2)$

REINALDO E. GIUDICI, Universidad Santa Maria, Valparaiso, Chile

1. Introduction. In [1] N. R. Hardman and J. H. Jordan have studied the distribution of quadratic residues in fields of order p^2, where p is a prime and $p\equiv 3 \pmod 4$.

In this paper we will eliminate the above restriction and study quadratic residues in $GF(p^2)$ [3, p. 446] for any odd prime p. For this purpose, $GF(p^2)$ is considered as a normal extension of the finite field $GF(p)$ and, as is well known, it can be conveniently written in the form:

$$GF(p^2)=\{a+b\theta \mid \theta^2=g;\ a,b\in GF(p)\},$$

where g is a primitive root (mod p) [3, p. 447].

As is expected, some of the theorems stated in [1] turn out to hold also in the case $p\equiv 1$ (mod 4); but others change when passing from $p=4k+3$ to $p=4k+1$.

2. Generators; the symbol $\chi(\alpha)$. Let λ be a fixed generator of the cyclic multiplicative group $GF^*(p^2)=GF(p^2)-\{0\}$. Then, $GF^*(p^2)=[\lambda]=\{1,\lambda,\lambda^2,\cdots,\lambda^{p^2-2}\}$ and λ^{p+1} is a primitive root of p.

DEFINITION 1. *Let $\alpha\in GF^*(p^2)$. If the congruence $\eta^2\equiv\alpha(\text{mod }p)$ has a solution*

$\eta \in GF^*(p^2)$ then α is said to be a quadratic residue (Q.R.) *in the $GF^*(p^2)$. If the above congruence has no solution, then it is said to be a quadratic nonresidue* (Q.N.R.) *in the $GF^*(p^2)$.*

DEFINITION 2. *Let $\chi(\alpha)$ denote the symbol defined by*

$$\chi(\alpha) = \begin{cases} +1 & \textit{if } \alpha \textit{ is a } \text{Q.R. } \textit{in the } GF^*(p^2), \\ -1 & \textit{if } \alpha \textit{ is a } \text{Q.N.R. } \textit{in the } GF^*(p^2), \\ 0 & \textit{if } \alpha \notin GF^*(p^2). \end{cases}$$

Now we can have a generalization of Euler's criterion which is similar to Theorem 1 of [1].

THEOREM 1. *If $p \nmid \alpha$ then $\chi(\alpha) \equiv \alpha^{(p^2-1)/2} (\mathrm{mod}\ p)$.*

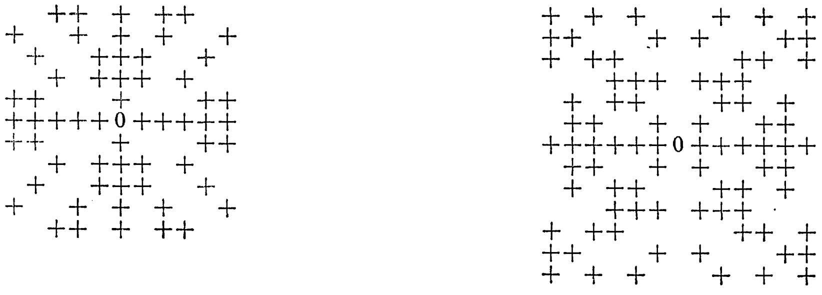
Proof. If $p \nmid \alpha$, then $\alpha \in [\lambda]$. That is, $\alpha = \lambda^t$ with $0 \leqq t \leqq p^2 - 2$. Now α is a Q.R. if t is even and Q.N.R. if t is odd. Let us form

$$\alpha^{(p^2-1)/2} = \lambda^{t(p^2-1)/2} = (\lambda^{p+1})^{t(p-1)/2} = (g^s)^{t(p-1)/2}$$

where $(s, p-1) = 1$. Since $g^{(p-1)/2} \equiv -1 \pmod{p}$ and s is odd we have $\alpha^{(p^2-1)/2} \equiv (-1)^t (\mathrm{mod}\ p)$. That is,

$$\chi(\alpha) \equiv \alpha^{(p^2-1)/2} \pmod{p}.$$

COROLLARY. $\chi(\alpha\beta) = \chi(\alpha)\chi(\beta)$.



$p = 11,\ g = 2$ $p = 13,\ g = 2$

FIG. 1

3. Symmetry. In order to study the distribution of the quadratic residues in $GF(p^2)$ one can plot the quadratic residues in a coordinate system taking as the X axis the p values of a and as the Y axis the p values of $b\theta$ as is indicated in Figure 1. Of course, when we speak of symmetry we have in mind the special representation $a + b\theta$ with $|a| < p/2$ and $|b| < p/2$. The next lemma tells us that the elements of the X axis are always quadratic residues. But the elements of the Y axis are Q.R. if $p \equiv 3 \pmod{4}$ and Q.N.R. if $p \equiv 1 \pmod{4}$.

LEMMA 1. *If* $a \in GF^*(p) = GF(p) - \{0\}$, *then*

$$\text{(a) } \chi(a) = 1.$$

$$\text{(b) } \chi(a\theta) = \begin{cases} 1, & \text{if } p \equiv 3 \pmod 4. \\ -1, & \text{if } p \equiv 1 \pmod 4. \end{cases}$$

Proof. (a) By using Theorem 1 and Fermat's little theorem, we have

$$\chi(a) = a^{(p^2-1)/2} \equiv (a^{p-1})^{(p+1)/2} \equiv 1 \pmod p.$$

Hence: $\chi(a) = 1$.

Proof. (b) By using Corollary 1, Theorem 1, Lemma 1, (a) and the fact that $\theta^2 = g$, we have, $\chi(a\theta) = \chi(a)\chi(\theta) = \chi(\theta)$. Now,

$$\chi(\theta) = \theta^{(p^2-1)/2} = \left[(\theta^2)^{(p-1)/2}\right]^{(p+1)/2} = \left(g^{(p-1)/2}\right)^{(p+1)/2}$$
$$\equiv (-1)^{(p+1)/2} \qquad \text{(Euler's criterion).}$$

The next theorem tells us that the quadratic residues are always symmetric with respect to the X axis, the Y axis and the origin.

THEOREM 2. $\chi(a+b\theta) = \chi(-a+b\theta) = \chi(a-b\theta) = \chi(-a-b\theta)$.

Proof. The case $a = b = 0$ is trivially true. In any other case, $a+b\theta = -(-a-b\theta)$ and $-a+b\theta = -(a-b\theta)$; Lemma 1 tells us that

$$\chi(a + b\theta) = \chi(-a - b\theta) \quad \text{and} \quad \chi(-a + b\theta) = \chi(a - b\theta).$$

On the other hand, $\chi(a + b\theta)\chi(a - b\theta) = \chi(a^2 - b^2g)$. Therefore, $\chi(a+b\theta)\chi(a-b\theta) = 1$ by Lemma 1.

Hence: $\chi(a+b\theta) = \chi(a-b\theta)$.

It is interesting to observe that there is no symmetry with respect to the diagonals as it happens in Theorem 2 of [1]. Also it is interesting to see that the symmetries change when another primitive root is used. Indeed one can observe permutations of rows and columns. The following theorem gives us a criterion about the behavior of the symmetry when the primitive root g is replaced by g^{-1}.

THEOREM 3.

$$\chi(a + b\theta) = \begin{cases} \chi(-b + a\bar{\theta}), & \textit{if } p \equiv 3 \pmod 4, \\ -\chi(-b + a\bar{\theta}), & \textit{if } p \equiv 1 \pmod 4, \end{cases}$$

where $\theta^2 = g$ *and* $\bar{\theta}^2 = g^{-1}$.

Proof. Since $(\theta\bar{\theta}^2) = 1$, then $\bar{\theta} = \pm\,\theta^{-1}$. Hence, $-b + \theta\bar{\theta} = -b \pm a\theta^{-1} = (-b\theta \pm a)\theta^{-1}$. Therefore, $\chi(-b+a\bar{\theta}) = \chi(-b\theta \pm a)\chi(\theta)^{-1} = \chi(a+b\theta)\chi(\theta)^{-1}$ (Theorem 2).

By using Lemma 1 we have the proof.

The above theorem says that replacing g by g^{-1}, one observes a rotation of

$\pi/2$ if $p\equiv 3 \pmod 4$, and a rotation of $\pi/2$ followed by a substitution of Q.R.'s by Q.N.R.'s and vice-versa, if $p\equiv 1 \pmod 4$.

4. The number of residues per line. We have two cases. If $p\equiv 3 \pmod 4$ in each line other than the axes there is exactly one more nonresidue than there are residues. If $p\equiv 1 \pmod 4$ in each line other than the axes there is exactly the same number of residues and nonresidues. This result can be stated in the following way:

THEOREM 4. *If $b\neq 0$ and $c\neq 0$, then*

$$\sum_{a=1}^{(p-1)/2} \chi(a+b\theta) = \sum_{d=1}^{(p-1)/2} \chi(c+d\theta) = \begin{cases} -1, & \text{if } p\equiv 3 \pmod 4, \\ 0, & \text{if } p\equiv 1 \pmod 4. \end{cases}$$

The proof follows the same lines as that given in [1] for the case $p\equiv 3 \pmod 4$.

Finally, we can have a connection between the symbol χ and the familiar Legendre symbol.

THEOREM 5. *$\chi(a+b\theta)=((a^2-b^2g)/p)$, where $(\;/\,p)$ denotes the Legendre symbol.*

Proof. The case $a=b=0$ is trivial. Let $a+b\theta\in GF^*(p^2)$. Then, $a+b\theta=\lambda^t$. Since $GF(p^2)$ is a normal extension of $GF(p)$ it is well known that $a-b\theta=\lambda^{pt}$. Therefore, $a^2-b^2g=(a+b\theta)(a-b\theta)=(\lambda^{p+1})^t$. That is, $a^2-b^2g=g^{st}$, where $(s,\,p-1)\equiv 1$. Hence,

$$((a^2-b^2g)/p) \equiv (a^2-b^2g)^{(p-1)/2} \equiv (g^{st})^{(p-1)/2} \equiv (g^{(p-1)/2})^{st}.$$

Since, $g^{(p-1)/2}\equiv -1 \pmod p$ and s is odd, we have $((a^2-b^2g)/p)=(-1)^t$.

On the other hand, $a+b\theta=\lambda^t$. Then, $\chi(a+b\theta)=\chi(\lambda^t)=(\chi(\lambda))^t$. Since λ is a generator of $GF^*(p^2)$, λ is a Q.N.R. Therefore $\chi(\lambda)=-1$. That is, $\chi(a+b\theta)=(-1)^t$.

A consequence of Theorems 4 and 5 is the evaluation of the following character sum:

COROLLARY 2. *If $b\neq 0$, then*

$$\sum_{a=1}^{(p-1)/2} ((a^2-b^2g)/p) = \begin{cases} -1, & \text{if } p\equiv 3 \pmod 4, \\ 0, & \text{if } p\equiv 1 \pmod 4. \end{cases}$$

The above corollary can also be obtained from the Jacobsthal sum

$$\sum_{x=0}^{p-1} ((x^2+c)/p) = -1, \qquad \text{if } c\not\equiv 0 \pmod p \quad [2].$$

The author is grateful to Dr. Roberto Frucht for his valuable suggestions and to the student Luis Salinas for his work at the Computer Center of the Santa María University.

Research for this paper was supported by Chilean Law 11575 under contract G 719G.D.45 a.

References

1. N. R. Hardman and J. H. Jordan, The distribution of quadratic residues in fields of order p^2, this MAGAZINE, 42 (1969) 12–17.

2. E. Jacobsthal, Uber die darstellung der primazahalen der form $4n+1$ als summe zweier quadrate, J. Reine Angew. Math., 132 (1907)238–245.

3. Garrett Birkhoff and Saunders MacLane, A Survey of Modern Algebra, Macmillan, New York 1953.

A NOTE ON THE GEOMETRY OF ZEROS OF POLYNOMIALS

C. A. LONG, Bowling Green State University

It is known that the zeros of a polynomial, $P(z)$, or any other complex function of a complex variable may be thought of as minimum points on its modulus surface [1, p. 292]. The modulus surface of P is defined as follows: With each point z in the complex plane, associate the value of the modulus, $|P(z)|$, of the polynomial at that point (the modulus of the complex number $w=x+iy$ is the real number $|w|=\sqrt{x^2+y^2}$). We may now consider this real number $|P(z)|$ to be associated with the real and imaginary parts of the complex number z, and form the graph of this function of two real variables. The graph of this function is the modulus surface. Figure 1 shows the modulus surface of z^2+z+1 with minimum points at the zeros, $-\frac{1}{2}\pm\sqrt{3}/2i$. When the zeros are interpreted in this way, it is interesting to conjecture as to how these minimum points will move as certain changes are made in the original polynomial.

As the coefficients of the real polynomial, $a_nz^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0$, are varied over the reals, the zeros take on corresponding complex values. Since the zeros of a polynomial are continuous functions of the coefficients (indeed, regular functions [2, p. 121]), as one of the coefficients a_i is varied about a fixed base value b_i, we obtain a continuous curve C_i in the complex plane as the trace or path of the zeros. The curves so obtained will intersect at the zeros of the base polynomial, $b_nz^n+b_{n-1}z^{n-1}+\cdots+b_1z+b_0$. We prove the following surprising result about these curves:

THEOREM. *Let the real base polynomial $b_nz^n+\cdots+b_0$ have n distinct zeros and let $z_0=re^{i\theta}$ be one of these zeros. Let C_j be the curve through z_0 formed by allowing the coefficient of z^j to vary and by tracing the path of the zeros of the resulting polynomials. The angle between consecutive curves, C_j, C_{j+1} is constant and equal to θ at their intersection point z_0.*

Proof. Consider the equation $P(a, z) = a_nz^n+\cdots+a_0 = 0$ with $a=(a_0, a_1, \cdots, a_n)$ as defining z implicitly as a function of $a_0, a_1, \cdots, a_n$, in a neighborhood of the point $b=(b_0, b_1, \cdots, b_n)$ at which it has value z_0. For the curve C_j, all except the coefficient of z_j have fixed values b_i. Thus on C_j, z is a function of the coefficient a_j only, and the tangent angle to C_j at z_0 is the argument of the complex number $\partial z/\partial a_j$ when evaluated at $a_j=b_j$. (Note that a_j is the curve parameter and we wish to evaluate the slope of the curve at the point where this parameter has value b_j, i.e., at the point z_0.) From the chain rule we have for C_j:

2. E. Jacobsthal, Uber die darstellung der primazahalen der form $4n+1$ als summe zweier quadrate, J. Reine Angew. Math., 132 (1907)238–245.

3. Garrett Birkhoff and Saunders MacLane, A Survey of Modern Algebra, Macmillan, New York 1953.

A NOTE ON THE GEOMETRY OF ZEROS OF POLYNOMIALS

C. A. LONG, Bowling Green State University

It is known that the zeros of a polynomial, $P(z)$, or any other complex function of a complex variable may be thought of as minimum points on its modulus surface [1, p. 292]. The modulus surface of P is defined as follows: With each point z in the complex plane, associate the value of the modulus, $|P(z)|$, of the polynomial at that point (the modulus of the complex number $w=x+iy$ is the real number $|w|=\sqrt{x^2+y^2}$). We may now consider this real number $|P(z)|$ to be associated with the real and imaginary parts of the complex number z, and form the graph of this function of two real variables. The graph of this function is the modulus surface. Figure 1 shows the modulus surface of z^2+z+1 with minimum points at the zeros, $-\frac{1}{2}\pm\sqrt{3}/2i$. When the zeros are interpreted in this way, it is interesting to conjecture as to how these minimum points will move as certain changes are made in the original polynomial.

As the coefficients of the real polynomial, $a_nz^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0$, are varied over the reals, the zeros take on corresponding complex values. Since the zeros of a polynomial are continuous functions of the coefficients (indeed, regular functions [2, p. 121]), as one of the coefficients a_i is varied about a fixed base value b_i, we obtain a continuous curve C_i in the complex plane as the trace or path of the zeros. The curves so obtained will intersect at the zeros of the base polynomial, $b_nz^n+b_{n-1}z^{n-1}+\cdots+b_1z+b_0$. We prove the following surprising result about these curves:

THEOREM. *Let the real base polynomial $b_nz^n+\cdots+b_0$ have n distinct zeros and let $z_0=re^{i\theta}$ be one of these zeros. Let C_j be the curve through z_0 formed by allowing the coefficient of z^j to vary and by tracing the path of the zeros of the resulting polynomials. The angle between consecutive curves, C_j, C_{j+1} is constant and equal to θ at their intersection point z_0.*

Proof. Consider the equation $P(a,z)=a_nz^n+\cdots+a_0=0$ with $a=(a_0,a_1,\cdots,a_n)$ as defining z implicitly as a function of $a_0,a_1,\cdots,a_n$, in a neighborhood of the point $b=(b_0,b_1,\cdots,b_n)$ at which it has value z_0. For the curve C_j, all except the coefficient of z_j have fixed values b_i. Thus on C_j, z is a function of the coefficient a_j only, and the tangent angle to C_j at z_0 is the argument of the complex number $\partial z/\partial a_j$ when evaluated at $a_j=b_j$. (Note that a_j is the curve parameter and we wish to evaluate the slope of the curve at the point where this parameter has value b_j, i.e., at the point z_0.) From the chain rule we have for C_j:

$$\frac{\partial P}{\partial b_0}\,\frac{\partial b_0}{\partial a_j}+\cdots+\frac{\partial P}{\partial b_{j-1}}\,\frac{\partial b_{j-1}}{\partial a_j}+\frac{\partial P}{\partial a_j}\,\frac{\partial a_j}{\partial a_j}+\frac{\partial P}{\partial b_{j+1}}\,\frac{\partial b_{j+1}}{\partial a_j}+\cdots$$

$$+\frac{\partial P}{\partial b_n}\,\frac{\partial b_n}{\partial a_j}+\frac{\partial P}{\partial z}\,\frac{\partial z}{\partial a_j}=0,$$

i.e.,

$$\frac{\partial P}{\partial a_j}\cdot 1+\frac{\partial P}{\partial z}\,\frac{\partial z}{\partial a_j}=0 \qquad \text{(since the } b_i\text{'s are fixed)}$$

and

$$\frac{\partial z}{\partial a_j}=\frac{-\partial P/\partial a_j}{\partial P/\partial z}$$

$$=\frac{-z^j}{nb_nz^{n-1}+\cdots+(j+1)b_{j+1}z^j+ja_jz^{j-1}+(j-1)b_{j-1}z^{j-2}+\cdots+b_1}.$$

In a neighborhood of z_0, $(\partial P/\partial z)\neq 0$, and when we evaluate this form at $a_j=b_j$ with $z_0=re^{i\theta}$, we have (in polar form):

$$\frac{\partial z}{\partial a_j}(b,\,z_0)=\frac{-z_0{}^j}{b_nnz_0^{n-1}+\cdots+b_1}=\frac{-(re^{i\theta})^j}{se^{i\alpha}}=te^{i(j\theta-\alpha)}$$

where θ and α are fixed by each base zero z_0. Thus the tangent angle for C_j is $j\theta-\alpha$ and the angle of intersection of C_j and C_{j+1} is simply the difference of the arguments computed for C_{j+1} and C_j, i.e., θ.

The example z^2+z+1 of Figure 1 is particularly interesting in that the base

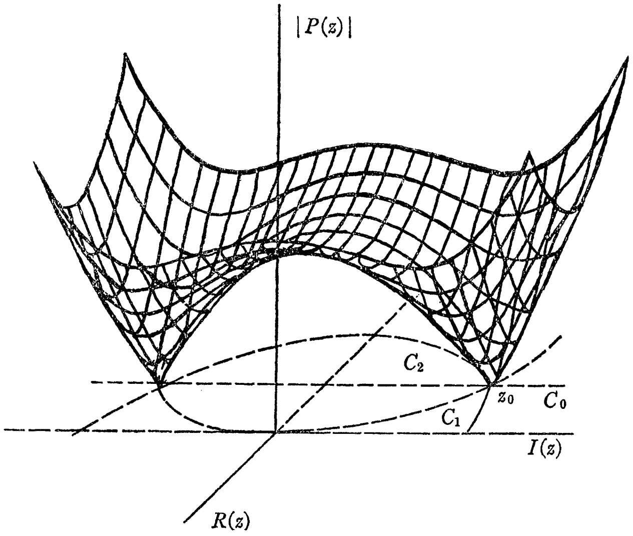


FIG. 1. Modulus surface of $P(z)=z^2+z+1$.

curves turn out to be lines and circles as indicated in Figure 2. In this case with $z_0 = e^{2\pi i/3}$, we see that the angle between consecutive curves is $2\pi/3$, with the curves being $C_0: x = -\frac{1}{2}$, $C_1: x^2+y^2=1$ and $C_3: (x+1)^2+y^2=1$.

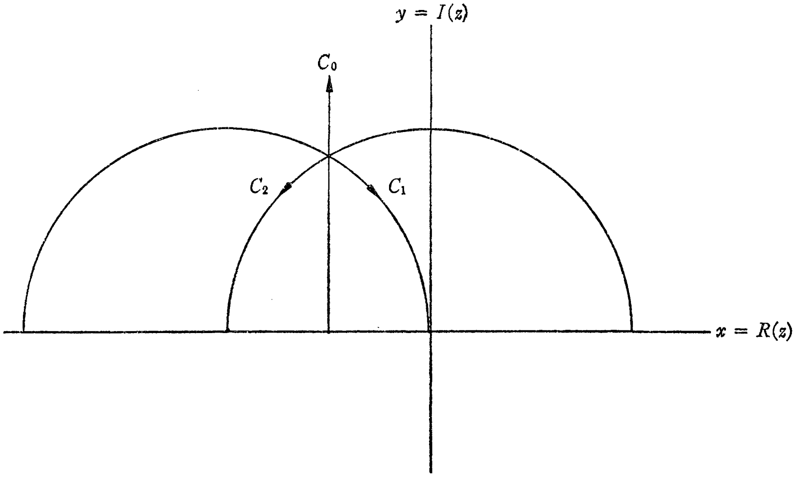


FIG. 2.

References

1. M. Marden, Geometry of Polynomials, Math. Surveys No. 3, Amer. Math. Soc., Providence, R. I., 1966.
2. K. Knopp, Theory of Functions, vol. 2, Dover, New York, 1952.

A CAUCHY CONDITION FOR FUNCTIONS

D. P. SCHAWE, SUNY at Plattsburgh

The purpose of this paper is to define a condition for functions from subsets of $\boldsymbol{R}$ into $\boldsymbol{R}$ which specializes to the Cauchy condition in the case of real sequences. We shall also prove a theorem to demonstrate the expected connection between the condition for a function to be Cauchy at a point and its convergence to a finite limit there. Hafstrom has proved a theorem fully equivalent to the one to follow in [1, p. 245]. The proof, however, follows a separate treatment of real sequences and uses the convergence of Cauchy sequences in the argument. The development here is meant to furnish a general theory of limits of real functions which includes sequences and which provides access to the Cauchy condition without a separate treatment of sequences.

We find a setting for such a development in the *extended real number system* $\boldsymbol{R}^*$ which is based on the following definitions.

DEFINITIONS.

(i) $\boldsymbol{R}^* = \boldsymbol{R} \cup \{-\infty, \infty\}$.

curves turn out to be lines and circles as indicated in Figure 2. In this case with $z_0=e^{2\pi i/3}$, we see that the angle between consecutive curves is $2\pi/3$, with the curves being $C_0: x=-\frac{1}{2}$, $C_1: x^2+y^2=1$ and $C_3: (x+1)^2+y^2=1$.

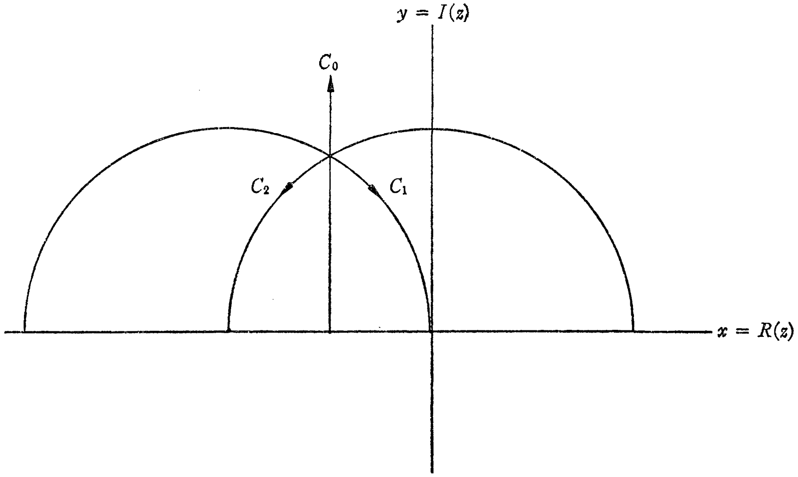


FIG. 2.

References

1. M. Marden, Geometry of Polynomials, Math. Surveys No. 3, Amer. Math. Soc., Providence, R. I., 1966.
2. K. Knopp, Theory of Functions, vol. 2, Dover, New York, 1952.

A CAUCHY CONDITION FOR FUNCTIONS

D. P. SCHAWE, SUNY at Plattsburgh

The purpose of this paper is to define a condition for functions from subsets of R into R which specializes to the Cauchy condition in the case of real sequences. We shall also prove a theorem to demonstrate the expected connection between the condition for a function to be Cauchy at a point and its convergence to a finite limit there. Hafstrom has proved a theorem fully equivalent to the one to follow in [1, p. 245]. The proof, however, follows a separate treatment of real sequences and uses the convergence of Cauchy sequences in the argument. The development here is meant to furnish a general theory of limits of real functions which includes sequences and which provides access to the Cauchy condition without a separate treatment of sequences.

We find a setting for such a development in the *extended real number system* R^* which is based on the following definitions.

DEFINITIONS.
(i) $R^*=R\cup\{-\infty, \infty\}$.

(ii) *Let* $\epsilon > 0$, *then*

$$\begin{aligned} N[a, \epsilon] &= \{x \in \boldsymbol{R} \mid |a - x| < \epsilon\}, \qquad \textit{for } a \in \boldsymbol{R}; \\ N^*[a, \epsilon] &= N[a, \epsilon] - \{a\}, \qquad \textit{for } a \in \boldsymbol{R}; \\ N^*[\infty, \epsilon] &= \{x \in \boldsymbol{R} \mid \epsilon < x\} \quad \textit{and} \\ N^*[-\infty, \epsilon] &= \{x \in \boldsymbol{R} \mid x < -\epsilon\}. \end{aligned}$$

(iii) *Let* $M \subset \boldsymbol{R}^*$ *and* $p \in \boldsymbol{R}^*$. *Then* p *is a cluster point of* M *if for each* $\epsilon > 0$ *there exists* $x \in M$ *such that* $x \in N^*[p, \epsilon]$.

(iv) *Let* $M \subset \boldsymbol{R}$, $f: M \to \boldsymbol{R}$ *and* p *be a cluster point of* M. *Then* $\lim_{x \to a} f(x) = F \in \boldsymbol{R}$ *if for each* $\epsilon > 0$, *there exists* $\delta(\epsilon) > 0$ *such that* $f(x) \in N[F, \epsilon]$ *whenever* $x \in N^*[p, \delta(\epsilon)] \cap M$.

For the sake of brevity, we include here only the definitions necessary for a theory of finite limits. For definitions sufficient to include infinite limits see [1, p. 227]. The inclusion of neighborhoods of ∞ in (ii) makes ∞ a cluster point of the set $\boldsymbol{N}$ of natural numbers. The use of neighborhoods in the definition (iv) of limit allows for limits at ∞ which includes limits of sequences. We now state the main definition.

DEFINITION. *Let* $M \subset \boldsymbol{R}$, $f: M \to \boldsymbol{R}$ *and* $p \in \boldsymbol{R}^*$ *be a cluster point of* M. *The function* f *is said to be Cauchy at* p *if for each* $\epsilon > 0$ *there exists* $\delta(\epsilon) > 0$ *and* $x(\epsilon) \in \boldsymbol{R}$ *such that* $f(x) \in N[x(\epsilon), \epsilon]$ *whenever* $x \in N^*[p, \delta(\epsilon)] \cap M$.

Only brief consideration of the definitions of the neighborhoods involved should be needed to convince the reader that in the case of sequences, Cauchy at ∞ specializes to the usual Cauchy condition. We now state the main theorem.

THEOREM. *Let* $M \subset \boldsymbol{R}$, $f: M \to \boldsymbol{R}$ *and* $p \in \boldsymbol{R}^*$ *be a cluster point of* M. *The following two statements are equivalent*:

(1) *The function* f *is Cauchy at* p.

(2) *The function* f *has a finite limit at* p.

Proof. To show (2)→(1), let $F = \lim_{x \to p} f(x)$. Then, for each $\epsilon > 0$, let $x(\epsilon) = F$ and the definition of convergence guarantees that f is Cauchy at p.

To show (1)→(2), let $0 < \epsilon < 1$. Since f is Cauchy at p, there exists $\delta(\epsilon) > 0$ and $x(\epsilon) \in \boldsymbol{R}$ such that $f(x) \in N[x(\epsilon), \epsilon]$ whenever $x \in N^*[p, \delta(\epsilon)] \cap M$. Consider the set $K = \{N[x(\epsilon), \epsilon] \mid 0 < \epsilon < 1\}$. Since p is a cluster point of M, there exists

$$x \in N^*[p, \delta(\epsilon')] \cap N^*[p, \delta(\epsilon'')] \cap M$$

for each ϵ' and ϵ'' in (0, 1). Then

$$f(x) \in N[x(\epsilon'), \epsilon'] \cap N[x(\epsilon''), \epsilon''].$$

Therefore K is a collection of pairwise overlapping open intervals with a bounded union.

For each ϵ in (0, 1), let $y(\epsilon)$ be the left end point of $N[x(\epsilon), \epsilon]$. Then $L = \{y(\epsilon) \mid 0 < \epsilon < 1\}$ is a bounded subset of $\boldsymbol{R}$ with least upper bound $F \in \boldsymbol{R}$. The real number F is the candidate for the $\lim_{x \to p} f(x)$.

To show $\lim_{x\to p} f(x)=F$, let $\epsilon>0$. Since $F=\text{lub } L$, there exists ϵ' in (0, 1) such that $y(\epsilon')\leqq F$ and $F-y(\epsilon')<\epsilon/3$. Therefore

$$\text{(a)} \qquad |y(\epsilon')-F| < \epsilon/3.$$

There exists ϵ'' in (0, 1) such that $\epsilon''<\epsilon/6$. Since $N[x(\epsilon'), \epsilon']$ and $N[x(\epsilon''), \epsilon'']$ overlap, $y(\epsilon')<y(\epsilon'')+\epsilon/3$ or $y(\epsilon')-y(\epsilon'')<\epsilon/3$. Since $F=\text{lub } L$, $y(\epsilon'')\leqq F<y(\epsilon')+\epsilon/3$ or $y(\epsilon'')-y(\epsilon')<\epsilon/3$. Therefore

$$\text{(b)} \qquad |y(\epsilon'')-y(\epsilon')| < \epsilon/3.$$

Let $x\in N^*[p, \delta(\epsilon'')]\cap M$. Then $f(x)\in N[x(\epsilon''), \epsilon'']$ and

$$\text{(c)} \qquad |f(x)-y(\epsilon'')| < \epsilon/3.$$

Adding (a), (b) and (c) obtains $|f(x)-F|<\epsilon$. Therefore $f(x)\in N[F, \epsilon]$ and $\lim_{x\to a} f(x)=F$, completing the proof.

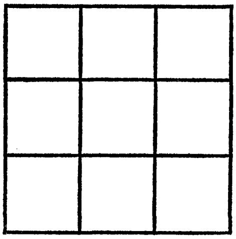
Reference

1. J. E. Hafstrom, Introduction to Analysis and Algebra, Saunders, Philadelphia, 1967.

A COMBINATORIAL PROOF THAT $\sum k^3=(\sum k)^2$

ROBERT G. STEIN, California State College, San Bernardino

How many rectangles can you find in this picture?



A little counting will show that there are just thirty-six rectangles here. In general, for any positive integer n, a picture of an $n\times n$ square ruled into unit squares will have $(1+2+3+\cdots+n)^2=[n(n+1)/2]^2$ rectangles in it. Giving two separate counting arguments for this yields a combinatorial proof of the identity

$$\text{(1)} \qquad 1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2,$$

which relates "squares," "triangle numbers," and "cubes" in a most interesting way. (The usual proof of this, by mathematical induction, is quite unenlightening. Toeplitz, in his wonderful *Calculus, A Genetic Approach*, gives an interesting old Arabic proof, so the proof given here is the third known to this writer.)

To show $\lim_{x\to p} f(x) = F$, let $\epsilon > 0$. Since $F = \text{lub } L$, there exists ϵ' in $(0, 1)$ such that $y(\epsilon') \leqq F$ and $F - y(\epsilon') < \epsilon/3$. Therefore

$$\text{(a)} \qquad | y(\epsilon') - F | < \epsilon/3.$$

There exists ϵ'' in $(0, 1)$ such that $\epsilon'' < \epsilon/6$. Since $N[x(\epsilon'), \epsilon']$ and $N[x(\epsilon''), \epsilon'']$ overlap, $y(\epsilon') < y(\epsilon'') + \epsilon/3$ or $y(\epsilon') - y(\epsilon'') < \epsilon/3$. Since $F = \text{lub } L$, $y(\epsilon'') \leqq F < y(\epsilon') + \epsilon/3$ or $y(\epsilon'') - y(\epsilon') < \epsilon/3$. Therefore

$$\text{(b)} \qquad | y(\epsilon'') - y(\epsilon') | < \epsilon/3.$$

Let $x \in N^*[p, \delta(\epsilon'')] \cap M$. Then $f(x) \in N[x(\epsilon''), \epsilon'']$ and

$$\text{(c)} \qquad | f(x) - y(\epsilon'') | < \epsilon/3.$$

Adding (a), (b) and (c) obtains $|f(x) - F| < \epsilon$. Therefore $f(x) \in N[F, \epsilon]$ and $\lim_{x\to a} f(x) = F$, completing the proof.

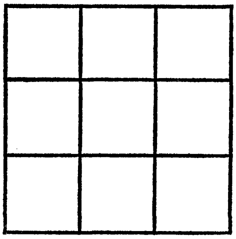
Reference

1. J. E. Hafstrom, Introduction to Analysis and Algebra, Saunders, Philadelphia, 1967.

A COMBINATORIAL PROOF THAT $\sum k^3 = (\sum k)^2$

ROBERT G. STEIN, California State College, San Bernardino

How many rectangles can you find in this picture?



A little counting will show that there are just thirty-six rectangles here. In general, for any positive integer n, a picture of an $n \times n$ square ruled into unit squares will have $(1+2+3+ \cdots +n)^2 = [n(n+1)/2]^2$ rectangles in it. Giving two separate counting arguments for this yields a combinatorial proof of the identity

$$\text{(1)} \qquad 1^3 + 2^3 + 3^3 + \cdots + n^3 = (1 + 2 + 3 + \cdots + n)^2,$$

which relates "squares," "triangle numbers," and "cubes" in a most interesting way. (The usual proof of this, by mathematical induction, is quite unenlightening. Toeplitz, in his wonderful *Calculus, A Genetic Approach*, gives an interesting old Arabic proof, so the proof given here is the third known to this writer.)

For the first counting method, notice that the rectangles in question have length $n-a$ and width $n-b$, where a and b are integers between 0 and n. How many such $n-a$ by $n-b$ rectangles are there in our $n \times n$ square? Each such rectangle may be translated $a+1$ units in one direction and $b+1$ in the perpendicular direction, to $(a+1)(b+1)$ positions in all. If $a \neq b$, a 90° rotation of one of our $n-a$ by $n-b$ rectangles produces another such rectangle which is not a translate of the original, and this in turn has $(a+1)(b+1)$ translates. Thus if $a \neq b$, there are just $2(a+1)(b+1)$ rectangles $n-a$ by $n-b$ in the $n \times n$ square. If $a=b$, there are $(a+1)^2$ such rectangles, since rotation by 90°, in the case of a square, does not yield a new rectangle. Then the number of rectangles whose shortest side is $n-a$ is

$$\begin{aligned}(a+1)^2 &+ 2(a+1)a + 2(a+1)(a-1) + \cdots + 2(a+1)(1)\\ &= (a+1)^2 + 2(a+1)[a + (a-1) + (a-2) + \cdots + 1]\\ &= (a+1)^2 + 2(a+1)\left[\frac{a(a+1)}{2}\right] = (a+1)^3.\end{aligned}$$

Since each rectangle has as its shortest side just one of the numbers 1, 2, 3, $\cdots$, n, we may count the rectangles in the $n \times n$ square as

$$\sum_{a=0}^{n-1} (a+1)^3.$$

We now count the rectangles a second way. There are just n^2 points which could serve as the lower left vertex of a rectangle, namely all the points in the picture of the ruled $n \times n$ square where lines meet, except those on the upper or righthand boundaries. If we consider a point c units to the left of the upper right corner of the big square and d units below it, we see this is the lower left corner of just cd of the rectangles we are trying to count. Thus in all there are $\sum_{c=1}^{n} \sum_{d=1}^{n} cd$ rectangles, or $(1+2+3 \cdots +n)^2$, which proves (1).

The second counting method generalizes easily to count the number of rectangles in a ruled $n \times m$ rectangle as

$$(1+2+3+\cdots+n)(1+2+3\cdots+m) = \frac{n(n+1)}{2}\,\frac{m(m+1)}{2}$$

but because the symmetry of n and m is lost, this is no longer an elegant sum of cubes.

A REMARK ON NONASSOCIATIVE BINARY OPERATIONS

PHILIP H. ANDERSON, Montclair State College

Associativity and commutativity of binary operations, in general, are independent properties. This remark is about a condition sufficient for a nonassociative binary operation to be noncommutative.

For the first counting method, notice that the rectangles in question have length $n-a$ and width $n-b$, where a and b are integers between 0 and n. How many such $n-a$ by $n-b$ rectangles are there in our $n\times n$ square? Each such rectangle may be translated $a+1$ units in one direction and $b+1$ in the perpendicular direction, to $(a+1)(b+1)$ positions in all. If $a\neq b$, a 90° rotation of one of our $n-a$ by $n-b$ rectangles produces another such rectangle which is not a translate of the original, and this in turn has $(a+1)(b+1)$ translates. Thus if $a\neq b$, there are just $2(a+1)(b+1)$ rectangles $n-a$ by $n-b$ in the $n\times n$ square. If $a=b$, there are $(a+1)^2$ such rectangles, since rotation by 90°, in the case of a square, does not yield a new rectangle. Then the number of rectangles whose shortest side is $n-a$ is

$$\begin{aligned}(a+1)^2&+2(a+1)a+2(a+1)(a-1)+\cdots+2(a+1)(1)\\ &=(a+1)^2+2(a+1)[a+(a-1)+(a-2)+\cdots+1]\\ &=(a+1)^2+2(a+1)\left[\frac{a(a+1)}{2}\right]=(a+1)^3.\end{aligned}$$

Since each rectangle has as its shortest side just one of the numbers 1, 2, 3, $\cdots$, n, we may count the rectangles in the $n\times n$ square as

$$\sum_{a=0}^{n-1}(a+1)^3.$$

We now count the rectangles a second way. There are just n^2 points which could serve as the lower left vertex of a rectangle, namely all the points in the picture of the ruled $n\times n$ square where lines meet, except those on the upper or righthand boundaries. If we consider a point c units to the left of the upper right corner of the big square and d units below it, we see this is the lower left corner of just cd of the rectangles we are trying to count. Thus in all there are $\sum_{c=1}^{n}\sum_{d=1}^{n}cd$ rectangles, or $(1+2+3\cdots+n)^2$, which proves (1).

The second counting method generalizes easily to count the number of rectangles in a ruled $n\times m$ rectangle as

$$(1+2+3+\cdots+n)(1+2+3\cdots+m)=\frac{n(n+1)}{2}\,\frac{m(m+1)}{2}$$

but because the symmetry of n and m is lost, this is no longer an elegant sum of cubes.

A REMARK ON NONASSOCIATIVE BINARY OPERATIONS

PHILIP H. ANDERSON, Montclair State College

Associativity and commutativity of binary operations, in general, are independent properties. This remark is about a condition sufficient for a nonassociative binary operation to be noncommutative.

Let S be a nonempty set and $*$ a nonassociative binary operation on S.

DEFINITION. *$*$ is entirely nonassociative if and only if for all a, b and c in S $(a * b) * c \neq a * (b * c)$.*

Example. Let S be the set of positive integers greater than 2. The binary operation $*$ defined by $a * b = a^b$ for a and b in S is entirely nonassociative.

THEOREM. *If $*$ is an entirely nonassociative binary operation on a nonempty set set S, then*

(1) $*$ is noncommutative;

(2) there are no idempotent elements with respect to $*$ in S;

(3) there is no left (right) identity for $*$ in S;

(4) there are no left (right) zero elements with respect to $*$ in S.

[z in S is a left (right) zero element with respect to $*$ if $z * a = z$ $(a * z = z)$ for all a in S.]

Proof. (1) Let a be an element of S. $(a * a) * a \neq a * (a * a)$ by hypothesis. Thus, $*$ is noncommutative.

(2) Suppose a is an idempotent element in S. $(a * a) * a \neq a * (a * a)$ by hypothesis. But $a * a = a$ since a is an idempotent element. A contradiction results.

(3) The definition of a left (right) identity implies it is idempotent. From (2) it follows there is no left (right) identity with respect to $*$ in S.

(4) The definition of a left (right) zero element implies it is idempotent. From (2) it follows there are no left (right) zero elements with respect to $*$ in S.

BOOK REVIEWS

EDITED BY D. ELIZABETH KENNEDY, University of Victoria

Materials intended for review should be sent to: Professor D. Elizabeth Kennedy, Department of Mathematics, University of Victoria, Victoria, British Columbia, Canada.

Reviews of texts at the freshman-sophomore level based upon classroom experience will be welcomed by the Book Review Editor.

A boldface capital C in the margin indicates a classroom review.

Algebraic Number Theory. By Serge Lang. Addison-Wesley, Reading, Mass., 1970. xi+354 pp. $14.50.

Modern algebra has two principal sources, both initiated during the first half of the 19th Century. One path leads through the Gaussian integers, cyclotomy, and the ideal theories of Kummer (Fermat's Last Theorem), Kronecker (his *Jugendtraum*), Dedekind, Hilbert, and E. Artin, while the other path leads through the so-called double algebra (the complex field), W. R. Hamilton's quaternions and related hypercomplex systems, and nonnumerical systems such as Boolean algebras. The first path emphasizes number theory with a transition from unique factorization to nonunique factorization and, finally, to some modified form of unique factorization. The second path emphasizes

Let S be a nonempty set and $*$ a nonassociative binary operation on S.

DEFINITION. *$*$ is entirely nonassociative if and only if for all a, b and c in S* $(a * b) * c \neq a * (b * c)$.

Example. Let S be the set of positive integers greater than 2. The binary operation $*$ defined by $a * b = a^b$ for a and b in S is entirely nonassociative.

THEOREM. *If $*$ is an entirely nonassociative binary operation on a nonempty set set S, then*

(1) *$*$ is noncommutative;*

(2) *there are no idempotent elements with respect to $*$ in S;*

(3) *there is no left (right) identity for $*$ in S;*

(4) *there are no left (right) zero elements with respect to $*$ in S.*

[*z in S is a left (right) zero element with respect to $*$ if $z * a = z$ $(a * z = z)$ for all a in S.*]

Proof. (1) Let a be an element of S. $(a * a) * a \neq a * (a * a)$ by hypothesis. Thus, $*$ is noncommutative.

(2) Suppose a is an idempotent element in S. $(a * a) * a \neq a * (a * a)$ by hypothesis. But $a * a = a$ since a is an idempotent element. A contradiction results.

(3) The definition of a left (right) identity implies it is idempotent. From (2) it follows there is no left (right) identity with respect to $*$ in S.

(4) The definition of a left (right) zero element implies it is idempotent. From (2) it follows there are no left (right) zero elements with respect to $*$ in S.

BOOK REVIEWS

EDITED BY D. ELIZABETH KENNEDY, University of Victoria

Materials intended for review should be sent to: Professor D. Elizabeth Kennedy, Department of Mathematics, University of Victoria, Victoria, British Columbia, Canada.

Reviews of texts at the freshman-sophomore level based upon classroom experience will be welcomed by the Book Review Editor.

A boldface capital C in the margin indicates a classroom review.

Algebraic Number Theory. By Serge Lang. Addison-Wesley, Reading, Mass., 1970. xi+354 pp. $14.50.

Modern algebra has two principal sources, both initiated during the first half of the 19th Century. One path leads through the Gaussian integers, cyclotomy, and the ideal theories of Kummer (Fermat's Last Theorem), Kronecker (his *Jugendtraum*), Dedekind, Hilbert, and E. Artin, while the other path leads through the so-called double algebra (the complex field), W. R. Hamilton's quaternions and related hypercomplex systems, and nonnumerical systems such as Boolean algebras. The first path emphasizes number theory with a transition from unique factorization to nonunique factorization and, finally, to some modified form of unique factorization. The second path emphasizes algebraic systems which do *not* satisfy various number axioms such as the commutative and associative laws. These two paths are not entirely distinct.

The book under review emphasizes results on the first path including some recent developments. Clearly, this tome is not an undergraduate textbook; indeed, it will provide challenging reading for "most" graduate students who fancy that they understand some Galois theory and some measure theory. Since exercises and problems are virtually nonexistent, the monograph will be best utilized as (1) collateral reading in a graduate course on commutative ring theory and its applications, (2) a subject-matter guide for a graduate level seminar on global class field theory over number fields (rather than, say, over function fields), and (3) supplementary reading on the topics of functional equations and Tauberian theorems. The title of this concise little book is slightly misleading. A more accurate title might be *Topics in algebraic and analytic number theory.*

The book opens with discussions of prime ideal structure to include local rings (but without mention of the result that every regular local ring is a unique factorization domain), integral closure (every unique factorization domain is integrally closed), the Chinese remainder theorem, finite Galois extensions, and Noetherian rings and their relations to Dedekind rings. Completions are studied to include complete Dedekind rings. Attention is given to cyclotomic fields (including Gaussian sums and the law of quadratic reciprocity), since class field theory, which occupies the middle third of the book, has many of its proofs governed by the proof paradigms for cyclotomic fields. The basic propositions for ideles and adeles (multiplicative and additive constructions, respectively) are developed. As a preliminary to attacking class field theory, the author digresses to study elementary properties of Dirichlet series and L-series, going so far as the usual Eulerian product for Dedekind's zeta-function for a number field, but without mention of the analogous result for a function field in one variable. An interesting aspect of the author's development of class field theory is his use of little known results of Herbrand, who may be better known in mathematical logic for his deduction theorem. The last third of monograph is devoted to analytic number theory. Two proofs of the functional equation are given. One follows Hecke (the Poisson summation formula) and the other follows Tate (the adelic form of the Poisson formula). Convexity results are used to obtain a charming little proof of Hadamard's classical three circle theorem. Another novel feature of this *Topica* is its presentation of Tate's thesis, including a concise proof of the graduate student's nemesis, the Riemann-Roch theorem.

Whether the author is discussing the geometry of numbers, using the Cyrillic alphabet, or only punctuating a proof ("oh miracle!"), there is always a refreshing twist involved. To this reviewer, the author's use of analysis throughout the last third of the book seems to be a challenge to eliminate the *need* for analytical tools in otherwise algebraic results. In any case, this fine monograph is not for undergraduates unless they want sleepless nights.

A. A. MULLIN

QED

"Now given this . . .", "If given that . . .",
Night after night the dreamer sat
And showed what he would do if he
Were shown some generosity.

MARLOW SHOLANDER

algebraic systems which do *not* satisfy various number axioms such as the commutative and associative laws. These two paths are not entirely distinct.

The book under review emphasizes results on the first path including some recent developments. Clearly, this tome is not an undergraduate textbook; indeed, it will provide challenging reading for "most" graduate students who fancy that they understand some Galois theory and some measure theory. Since exercises and problems are virtually nonexistent, the monograph will be best utilized as (1) collateral reading in a graduate course on commutative ring theory and its applications, (2) a subject-matter guide for a graduate level seminar on global class field theory over number fields (rather than, say, over function fields), and (3) supplementary reading on the topics of functional equations and Tauberian theorems. The title of this concise little book is slightly misleading. A more accurate title might be *Topics in algebraic and analytic number theory.*

The book opens with discussions of prime ideal structure to include local rings (but without mention of the result that every regular local ring is a unique factorization domain), integral closure (every unique factorization domain is integrally closed), the Chinese remainder theorem, finite Galois extensions, and Noetherian rings and their relations to Dedekind rings. Completions are studied to include complete Dedekind rings. Attention is given to cyclotomic fields (including Gaussian sums and the law of quadratic reciprocity), since class field theory, which occupies the middle third of the book, has many of its proofs governed by the proof paradigms for cyclotomic fields. The basic propositions for ideles and adeles (multiplicative and additive constructions, respectively) are developed. As a preliminary to attacking class field theory, the author digresses to study elementary properties of Dirichlet series and L-series, going so far as the usual Eulerian product for Dedekind's zeta-function for a number field, but without mention of the analogous result for a function field in one variable. An interesting aspect of the author's development of class field theory is his use of little known results of Herbrand, who may be better known in mathematical logic for his deduction theorem. The last third of monograph is devoted to analytic number theory. Two proofs of the functional equation are given. One follows Hecke (the Poisson summation formula) and the other follows Tate (the adelic form of the Poisson formula). Convexity results are used to obtain a charming little proof of Hadamard's classical three circle theorem. Another novel feature of this *Topica* is its presentation of Tate's thesis, including a concise proof of the graduate student's nemesis, the Riemann-Roch theorem.

Whether the author is discussing the geometry of numbers, using the Cyrillic alphabet, or only punctuating a proof ("oh miracle!"), there is always a refreshing twist involved. To this reviewer, the author's use of analysis throughout the last third of the book seems to be a challenge to eliminate the *need* for analytical tools in otherwise algebraic results. In any case, this fine monograph is not for undergraduates unless they want sleepless nights.

A. A. MULLIN

QED

"Now given this . . .", "If given that . . .",
Night after night the dreamer sat
And showed what he would do if he
Were shown some generosity.

MARLOW SHOLANDER

PROBLEMS AND SOLUTIONS

EDITED BY ROBERT E. HORTON, Los Angeles Valley College

ASSOCIATE EDITOR, J. S. FRAME, Michigan State University

Readers of this department are invited to submit for solution problems believed to be new that may arise in study, in research, or in extra-academic situations. Problems may be submitted from any branch of mathematics and ranging in subject content from that accessible to the talented high school student to problems challenging to the professional mathematician. Proposals should be accompanied by solutions, when available, and by any information that will assist the editor. Ordinarily, problems in well-known textbooks should not be submitted.

The asterisk (*) will be placed by the problem number to indicate that the proposer did not supply a solution. Readers' solutions are solicited for all problems proposed. Proposers' solutions may not be "best possible" and solutions by others will be given preference.

Solutions should be submitted on separate, signed sheets. Figures should be drawn in India ink and exactly the size desired for reproduction.

Send all communications for this department to Robert E. Horton, Los Angeles Valley College, 5800 Fulton Avenue, Van Nuys, California 91401.

To be considered for publication, solutions should be mailed before November 15, 1971.

PROBLEMS

796. *Proposed by Charles W. Trigg, San Diego, California.*

M. Adman Amdam hoped to make the Olympic team in one of the distance events, but he needed a lot of preparatory training,

$$SO/HE = .RANRANRAN \cdots .$$

Each letter in the cryptarithm uniquely represents a positive digit in the scale of nine. Find the only solution less than one-half, and hence more likely to represent his chances of making the team.

797. *Proposed by Frank J. Papp, University of Lethbridge, Alberta, Canada.*

Determine the critical points and relative extrema, if any, of the two functions $f(x_1, \cdots, x_n)$ and $g(x_1, \cdots, x_n)$ for $n=1, 2, 3 \cdots$ where

$$f(x_1, \cdots, x_n) = \det(a_{ij}) \; g(x_1, \cdots, x_n) = \det(b_{ij})$$

with

$$a_{ij} = \begin{cases} 1 & \text{if } i \neq j \quad \text{or if } i = j = 1 \\ 1 + x_{i-1} & \text{if } i = j = 2, 3 \cdots n + 1 \end{cases}$$

and

$$b_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 1 + x_i & \text{if } i = j = 1, 2 \cdots n \end{cases}$$

798. *Proposed by Peter A. Lindstrom, Genesee Community College, New York.*

Show that $\lim_{x\to\infty} (\prod_{p \leq x} p)^{1/x} = e$ where $\prod_{p \leq x} p$ is the product of prime integers that are less than or equal to x.

799. *Proposed by N. J. Kuenzi, Wisconsin State University at Oshkosh.*

A population consists of two distinct types of items, n items of type I and m items of type II. Items are selected randomly one by one without replacement until the kth type I item has been selected, $1 \leqq k \leqq n$. If x_k is the trial on which the kth type I item is selected, find $p[x_p = x]$ where $x = k,\ k+1 \cdots k+m$ and find the expected value of x_k.

800. *Proposed by David Singmaster, University of London, England.*

In *A property of third order gnomon-magic squares*, this MAGAZINE, 1970, 70, a 3×3 array is called gnomon-magic if the four 2×2 subarrays obtained by removing a 5-element gnomon all have the same sum. Show that a gnomon-magic 3×3 array has its two diagonal sums equal. Does this extend to higher orders?

801. *Proposed by Simeon Reich, Israel Institute of Technology, Haifa, Israel.*

Let x_i be the distances of an interior point of a triangle $A_1A_2A_3$ from the side opposite A_i, $i = 1, 2, 3$ and let r be the inradius of the triangle. Prove or disprove that

$$\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} \geqq \frac{3}{r}$$

802. *Proposed by Erwin Just, Bronx Community College, New York.*

Let p, q and r be distinct odd primes and k, m and t be positive integers. If $p = 2q^k + 1$, $n = r^t$ and $x = p^m$, prove that neither $x^n + y^n = z^n$ nor $x^n - y^n = z^n$ has solutions in positive integers.

SOLUTIONS

A Swinging Alphametic

768. [September, 1970] *Proposed by J. A. H. Hunter, Toronto, Canada.*

Solve the alphametic

$$\begin{array}{cccc} & & & A\\ & & G & O\\ & & G & O\\ & G & A & L\\ \hline L & O & O & K\end{array}$$

Solution by André Boileau, Université du Québec à Montréal.

First we suppose that

(1) Different letters represent distinct digits (so the base used $\geqq$ the number of distinct letters $= 5$).

(2) $L > 0$ (because if not, L would have been omitted).

By (1) we have $A + O + O + L \leqq (n-2) + 2(n-1) + (n-3) = 4n - 7$ (where n is the base used) and so the carrying over R_1 is less than 4. Also by (1) we have $G + G + A + R_1 \leqq 2(n-1) + (n-2) + 3 = 3n - 1$ and $R_2 \leqq 2$. At last $G + R_2 \leqq (n-1)$

$+2=n+1$, so $L=R_3\leqq 1$. But (2) gives us $L>0$ and we conclude that $L=1$. And because $n\leqq G+R_2\leqq n+1$ we have that $O=G+R_2-n\leqq 1$. But (1) implies that $O\neq L=1$, and therefore $O=0$.

Because $G\leqq n-1$ and $G+R_2\geqq n$ we have $R_2\neq 0$. And because $R_2\leqq 2$, there are two possible cases:

(a) $R_2=1$, so $G+R_2=Ln+O=n$ and therefore $G=n-1$; but then $A+O+O+L=A+1\leqq(n-2)+1=n-1$ and therefore $R_1=0$; also $G+G+A+R_1=2(n-1)+A=R_2n+O=n$ and therefore $A=2-n$ which is impossible because (1) implies $n\geqq 5$.

(b) $R_2=2$; we have $A\neq n-1$ (because if not, $A+O+O+L=R_1n+K=n$ and so $K=0$ which is impossible by (1) for $O=0$); therefore $R_1=0$ (because $A+O+O+L\leqq(n-3)+1=n-2$) and finally $G+G+A+R_1=2(n-2)+A=R_2n+O=2n$ so $A=4$ and $K=A+1=5$. Therefore, for a base $n\geqq 6$, we have:

$$\begin{array}{cccc} & & & A \\ & & G & O \\ & & G & O \\ & G & A & L \\ \hline L & O & O & K \end{array} \qquad \begin{array}{cccc} & & & 4 \\ & & n-2 & 0 \\ & & n-2 & 0 \\ & n-2 & 4 & 1 \\ \hline 1 & 0 & 0 & 5 \end{array}$$

For the base ten this yields

$$\begin{array}{cccc} & & & 4 \\ & & 8 & 0 \\ & & 8 & 0 \\ & 8 & 4 & 1 \\ \hline 1 & 0 & 0 & 5 \end{array}$$

Also solved by A. N. Aheart, West Virginia State College; Merrill Barnebey, Wisconsin State University at LaCrosse; A. J. Berlau, Hartsdale, New York; W. G. Brady, Slippery Rock State College; R. L. Breisch, Pennsylvania State University; G. L. Britton, Wisconsin State University at West Bend; M. J. Brown, Northern Kentucky State College; C. P. Campbell, West Virginia State College; Howard Carter, Delaware State College; the following students from the high school class at Cincinnati Country Day School: Jeff Spain, Lawrence Williams, Michael Pogue, Kim Derrick, Tom Feige, Thomas Lee, Carl Linder, H. Riehle, David Hunter and Lloyd Miller; Barbara A. Connolly, Mt. Saint John Academy; R. J. Cormier, DeKalb, Illinois; Mickey Dargitz, Ferris State College; J. A. Dossey, Normal, Illinois; Ragnar Dybnik, Tingvoll, Norway; George Fabian, Park Forest, Illinois; Rosalie Farrand, Castilleja High School; Gordon A. Findlay, Wellington, New Zealand; Richard Gibbs, Hiram Scott College; M. G. Greening, University of New South Wales, Australia; Morton Goldberg, Broome Technical Community College; Louise S. Grinstein, New York City; Ned Harrell, Menlo-Atherton High School; Mary R. Hesselgrave, Mt. Saint Mary College; J. M. Howell, Littlerock, California; Yul J. Inn, University of California at Riverside; Mary R. Ireson, North Tazewell, Virginia; Paul Johnson, University of California at Los Angeles; Rick Johnson, University of South Carolina; Jerome Kochek, Gary, Indiana; Alfred Kohler, Long Island University; Lew Kowarski, Morgan State College; Jim Lackritz, Bucknell University; J. F. Leetch, Bowling Green State University; H. R. Leifer, Pittsburgh, Pennsylvania; P. A. Lindstrom, Genesee Community College; Beatriz Margolis, Universidad Nacional de la Plata, Argentina; Carl P. McCarty, LaSalle College; E. P. McCravy, Midlands Technical Education Center; J. V. Michalowicz, Catholic University of America;

J. W. Milsom, Butler County Community College; Steve Morphey, University of British Columbia; C. B. Myers, Austin Peay State University; G. A. Novacky, Jr., University of Pittsburgh; William Nuesslein, Madison, Wisconsin; Joseph O'Rourke, St. Joseph's College; A. J. Patsche, Rock Island, Illinois; B. J. Portz, Creighton University; Simeon Reich, Israel Institute of Technology; L. A. Ringenberg, Eastern Illinois University; E. F. Schmeichel, Itasca, Illinois; Abraham Schwartz, Armonk, New York; D. R. Simpson, Fairbanks, Alaska; Romesh Singh, Arthur, Ontario, Canada; Jackie Steele, South Virginia Community College; Paul Sugarman, Massachusetts Institute of Technology; Jim Tattersall, Attleboro, Massachusetts; G. C. Thompson, Binghamton, New York; C. W. Trigg, San Diego, California; Wolf R. Umbach, Braunschweig, West Germany; John R. Ventura, Jr., New Bedford, Massachusetts; R. F. Wardrop, Central Michigan University; Ronald Whiffen, Queens College; K. M. Wilke, Topeka, Kansas; Gene Zirkel, Nassau Community College; and the proposer.

Prime Decomposition

769. [September, 1970] *Proposed by Charles W. Trigg, San Diego, California.*

A triangular number is composed of nine distinct digits in the decimal system. When it is sectioned into three triads, each triad is prime. Find the number and show it to be unique.

Solution by Nigel F. Nettheim, Toronto, Ontario, Canada.

The triangular number $T(n)=\frac{1}{2}n(n+1)$ has nine digits, so that $14142\leqq n\leqq 44720$. Let

$$n \equiv 2000j + 200k + 20l + m,$$

where $0\leqq m\leqq 19$, $0\leqq l\leqq 9$, $0\leqq k\leqq 9$, $7\leqq j\leqq 22$, and write $T(n)\equiv T(j, k, l, m)$. Then we have

$$T(j, k, l+1, m) - T(j, k, l, m) = (21+2n)\cdot 10$$

so that the last digits of the numbers $T(n)$ repeat in cycles of 20. Since the last triad is prime, the last digit is 1, 3, 7 or 9 so that $m\in\{1, 2, 6, 13, 17, 18\}$. Also

$$T(j+1, k, l, m) - T(j, k, l, m) = (2001+2n)\cdot 1000 \qquad (*)$$

so that the last triads repeat in cycles of 2000. Since, moreover, for $1\leqq n\leqq 1998$ we have

$$T(1999-n) - T(n) = (1999-2n)\cdot 1000,$$

the last triads, for given j, occur symmetrically about the central value $n=2000j+999\frac{1}{2}$. Hence to determine all the values of $(\cdot, k, l, m)$ which yield a prime last triad with distinct digits, it is only necessary to examine $5\cdot 10\cdot 6=300$ cases and use the above-noted symmetry. This procedure yields exactly $77\cdot 2=154$ sets, namely $(\cdot, 0, 2, 13), \cdots (\cdot, 9, 7, 6)$ or, setting $j=0$ for convenience, $n=53, \cdots 1946$.

We proceed to examine the middle triad in each case for each possible value of j; from (*) it follows that the middle triad is increased by $(2n+1)$ as j is incremented by 1. Bearing in mind the requirement of distinct digits and that the first triad also be prime, we find the unique solution: $T(36161)=653827041$.

Also solved by John A. Dossey, Normal, Illinois; Alfred Kohler, Long Island University, New York; E. F. Schmeichel, College of Wooster, Ohio; and the proposer.

Equivalent Integral Expressions

770. [September, 1970] *Proposed by John E. Hafstrom, California State College at San Bernardino,*

Given $f(x)$ continuous on $[a, b]$. Prove that numbers c and d exist such that for $a<c<d<b$ we have

$$\frac{\int_a^b f(x)dx}{b-a} = \frac{\int_c^d f(x)dx}{d-c}$$

Solution by R. P. Boas, Jr., Northwestern University.

With $F(x)=\int_a^x f(t)dt$, the problem states that there are points c, d, with $a<c<d<b$, such that

$$\frac{F(b)-F(a)}{b-a} = \frac{F(d)-F(c)}{d-c}, \tag{1}$$

In other words, that F has at least one chord parallel to the chord through $(a, F(a))$ and $(b, F(b))$. This is geometrically intuitive whenever F is continuous (not necessarily an integral); indeed it is clear that there must be many such chords.

For a formal proof (with F merely continuous), we first reduce the problem to the case of a function that vanishes at a and b (as in the conventional proof of the mean-value theorem). Consider

$$g(x) = F(x) - F(a) - (x-a)\frac{F(b)-F(a)}{b-a}\cdot$$

Then $g(b)=g(a)=0$ and (1) follows if we show that there are c, d, with $a<c<d<b$, such that $g(d)=g(c)$, (i.e., that g has a horizontal chord with neither end at a or b). If $g(x)\equiv 0$ there is nothing to prove. Otherwise, g has either a positive maximum or a negative minimum between a and b, say at h. Then g takes every value between 0 and $g(h)$ on (a, h) and also on (h, b), so we can take c and d to be any pair of points, one in each of these intervals, where g takes equal values.

Also solved by Walter Blumberg, New Hyde Park, New York; Derrill J. Bordelon, Naval Underwater Systems Center, Newport, Rhode Island; Frank A. Chimenti, SUNY at Fredonia, New York; Ellis Detwiler, Adams, New York; William F. Fox, Moberly Junior College, Missouri; Michael Goldberg, Washington, D.C.; M. G. Greening, University of New South Wales, Australia; Phillip M. Kannan, Sweet Briar College, Virginia; Lew Kowarski, Morgan State College, Maryland; Norbert J. Kuenzi, Oshkosh, Wisconsin; J. F. Leetch, Bowling Green State University, Ohio; Douglas Lind, Stanford University; David E. Manes, State University College, Oneonta, New York; Beatriz Margolis, Universidad Nacional de la Plata, Argentina; Michael J. Martino, Poughkeepsie, New York; Carl P. McCarty, LaSalle College, Pennsylvania; Steve Morphey, University of British Columbia; P. G. Pantelidakis, Phoenix, Arizona; Simeon Reich, Israel Institute of Technology, Haifa, Israel; Ron Reitz, University of Minnesota; Kenneth A. Ribet, Harvard University; E. F. Schmeichel, College of Wooster, Ohio; William Squire, West Virginia University; and the proposer. One unsigned solution was received.

A Property of Finite Fields

771. [September, 1970] *Proposed by Douglas Lind, Cambridge University, England.*

Show that the sum of the elements of a finite field of more than two elements must be zero.

Solution by Joseph V. Michalowicz, Catholic University of America.

A finite field contains p^n elements, where p is a prime. These elements are precisely the zeros of the polynomial $x^{p^n}-x$. Hence the sum of the elements of a finite field with $p^n>2$ elements is the first elementary symmetric function of this polynomial, which is zero, since the coefficient of the x^{p^n-1} term is zero.

Also solved by Walter Blumberg, New Hyde Park, New York; R. L. Breisch, Pennsylvania State University; R. J. Cormier, DeKalb, Illinois; Gordon A. Findlay, Wellington, New Zealand; William F. Fox, Moberly Junior College, Missouri; Richard A. Gibbs, Hiram Scott College, Nebraska; Michael Goldberg, Washington, D. C.; M. G. Greening, University of New South Wales, Australia; Don Heller, Carnegie-Mellon University, Pennsylvania; Mary Hesselgrave, Mt. Saint Mary College, New York; Billy F. Hobbs, Pasadena College, California; John E. Homer, Jr., Lisle, Illinois; Yul J. Inn, University of California at Riverside; Phillip M. Kannan, Sweet Briar College, Virginia; Alfred Kohler, Long Island University, New York; J. F. Leetch, Bowling Green State University, Ohio; Henry S. Lieberman, Waban, Massachusetts (four solutions); David E. Manes, State University College, Oneonta, New York; Arthur Marshall, Madison, Wisconsin; Carl P. McCarty, LaSalle College, Pennsylvania; Edwin P. McCravy, Midlands Technical Education Center, South Carolina; Edward Moylan, Ford Motor Company, Dearborn, Michigan; William Nuesslein, Madison, Wisconsin; Simeon Reich, Israel Institute of Technology, Haifa, Israel; Kenneth A. Ribet, Harvard University; Henry J. Ricardo, Yeshiva University; E. F. Schmeichel, College of Wooster, Ohio; Paul Sugarman, Massachusetts Institute of Technology; Jim Tattersall, Attleboro, Massachusetts; Al White, St. Bonaventure University, New York; Kenneth M. Wilke, Topeka, Kansas; Qazi Zameeruddin, K. M. College, Delhi, India; and the proposer.

A Cyclotomic Polynomial

772. [September, 1970] *Proposed by Erwin Just, Bronx Community College.*

Let p be a prime and $\{\omega_i\}$, $i=1, 2, 3, \cdots p-1$, be the primitive pth roots of unity. If a set of rational numbers, $\{r_1, r_2, \cdots, r_{p-1}\}$, is chosen so that $\sum_{i=1}^{p-1} r_i\omega_i$ is rational, prove that $r_1=r_2=\cdots=r_{p-1}$.

Solution by M. G. Greening, University of New South Wales, Australia.

If $\sum_{i=1}^{p-1} r_i\omega_i=q\in Q$, then $g(x)=q-\sum_{i=1}^{p-1} r_ix^i$ divides the cyclotomic polynomial $1+\sum_{i=1}^{p-1} x^i$. We cannot have $\deg g(x)<p-1$ as $1+\sum_{i=1}^{p-1} x^i$ is irreducible in $Q[x]$. Consequently $r_1=r_2=\cdots=r_{p-1}=-q$.

Also solved by Walter Blumberg, New Hyde Park, New York; Douglas Lind, Stanford University; Simeon Reich, Israel Institute of Technology, Haifa, Israel; E. F. Schmeichel, College of Wooster, Ohio; and the proposer.

Another Triangle Property

773. [September, 1970] *Proposed by Norman Schaumberger, Bronx Community College.*

Let M be an arbitrary point not necessarily in the plane of triangle $A_1A_2A_3$.

If B_i is the midpoint of the side opposite A_i prove

$$\sum_{i=1}^{3} MA_i^2 - \sum_{i=1}^{3} MB_i^2 = \tfrac{1}{3}\sum_{i=1}^{3} A_iB_i^2.$$

I. *Solution by Romesh Singh, Arthur District High School, Ontario, Canada.*

Choose rectangular axes such that $\Delta A_1A_2A_3$ is in the XY-plane with A_2 at the origin and A_3 on the X-axis. Then the coordinates of A_1, A_2, A_3 may conveniently be represented by : $A_1(a, b, 0)$, $A_2(0, 0, 0)$ and $A_3(c, 0, 0)$. The coordinates of the respective midpoints are: $B_1(c/2, 0, 0)$, $B_2(a+c/2, b/2, 0)$ and $B_3(a/2, b/2, 0)$. Now if the coordinates of M are (u, v, w), then the desired equality follows immediately by the application of the distance formula

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

II. *Solution by Murray S. Klamkin, Ford Scientific Laboratory, Dearborn, Michigan.*

We prove a more general result, i.e., if $A_1, A_2, \cdots A_{n+1}$ denote any $(n+1)$ points in any E^r and if B_j $(j=1, 2, \cdots n+1)$ denotes the centroid of all the A_i's with the exception of A_j, then for any arbitrary point M

$$(1) \qquad \sum_{i=1}^{n+1} MA_i^2 - \sum_{i=1}^{n+1} MB_i^2 = \frac{n-1}{n+1}\sum_{i=1}^{n+1} A_iB_i^2.$$

Let $\mathbf{A}_i$, $\mathbf{B}_i$, $\mathbf{M}$ denote vectors from the centroid of all the A_i's to A_i, B_i, and M, respectively. Then

$$\sum_{i=1}^{n+1} \mathbf{A}_{n+1} = 0 \quad \text{and} \quad -\mathbf{B}_i = \mathbf{A}_i/n.$$

The l.h.s. of (1) is now

$$\sum_{i=1}^{n+1} (\mathbf{M} - \mathbf{A}_i)^2 - \sum_{i=1}^{n+1} (\mathbf{M} + \mathbf{A}_i/n)^2$$

or

$$\frac{n^2-1}{n^2}\sum_{i=1}^{n+1} \mathbf{A}_i^2.$$

Since the r.h.s. of (1) is now

$$\frac{n-1}{n+1}\sum_{i=1}^{n+1} \mathbf{A}_i^2(1 + 1/n)^2,$$

identity (1) follows. The proposed problem corresponds to the special case $n=2$.

Also solved by Walter Blumberg, New Hyde Park, New York; Wray G. Brady, Slippery Rock State College, Pennsylvania; L. Carlitz, Duke University; Howard Carter, Delaware State College; Mannis Charosh, Brooklyn, New York; R. J. Cormier, DeKalb, Illinois; Mickey Dargitz, Ferris

State College, Michigan; Ragnar Dybnik, Tingvoll, Norway; Michael Goldberg, Washington, D. C.; M. G. Greening, University of New South Wales, Australia; Yul J. Inn, University of California at Riverside; Alfred Kohler, Long Island University, New York; H. R. Leifer, Pittsburgh, Pennsylvania; Steven R. Morphey, University of British Columbia; Simeon Reich, Israel Institute of Technology, Haifa, Israel; E. F. Schmeichel, College of Wooster, Ohio; Jim Tattersall, Attleboro, Massachusetts; Theodore Teichmann, KMS Technology Center, San Diego, California; and the proposer. One unsigned solution was received.

Another Triangle Inequality

774. [September, 1970] *Proposed by A. W. Walker, Toronto, Canada.*

If (a, b, c) are the lengths of the sides of any triangle, show that:

$$3\left(\frac{a^2}{b^2}+\frac{b^2}{c^2}+\frac{c^2}{a^2}\right)-(a^2+b^2+c^2)\left(\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}\right)\geqq 0.$$

Solution by L. Carlitz, Duke University.

The stated inequality is equivalent to

$$3(a^4c^2+b^4a^2+c^4b^2)\geqq(a^2+b^2+c^2)(b^2c^2+c^2a^2+a^2b^2)$$

or

$$3(a^4c^2+b^4a^2+c^4b^2)\geqq a^4c^2+b^4a^2+c^4b^2+a^2c^4+b^2a^4+c^2b^4+3a^2b^2c^2.$$

This can be written as

$$3(a^4c^2+b^4a^2+c^4b^2-a^2c^4+b^2a^4+c^2b^4)$$
$$+(a^4c^2+b^4a^2+c^4b^2+a^2c^4+b^2a^4+c^2b^4)\geqq 6a^2b^2c^2.$$

Since (by the theorem on the arithmetic and geometric means)

$$a^4c^2+b^4a^2+c^4b^2+a^2c^4+b^2a^4+c^2b^4\geqq 6a^2b^2c^2,$$

it suffices to prove that

$$(*)\qquad \left(\sum a^4b^2-6a^2b^2c^2\right)^2\geqq 9(a^2-b^2)^2(b^2-c^2)^2(c^2-a^2)^2.$$

Now put

$$p=\sum a^2,\quad q=\sum a^2b^2,\quad r=a^2b^2c^2.$$

Since (by the formula for the discriminant of a cubic)

$$(a^2-b^2)^2(b^2-c^2)^2(c^2-a^2)^2=-27r^2+18pqr-4p^3r-4q^3+p^2q^2,$$

(*) becomes

$$(**)\qquad 81r^2+9q^3+9p^3r\geqq 45pqr+2p^2q^2.$$

In the next place

$$4[81r^2+9p(p^2-5q)r+9q^3-2p^2q^2]$$
$$=[18r+p(p^2-5q)]^2+36q^3-8p^2q^2-p^2(p^2-5q)^2$$
$$=[18r+p(p^2-5q)]^2-(p^2-4q)(p^2-3q)^2.$$

Since

$$4q - p^2 = 4 \sum a^2b^2 - (\sum a^2)^2 = 2 \sum a^2b^2 - \sum a^4 = 16K^2 \geqq 0,$$

where K denotes the area of the given triangle, it follows that

$$81r^2 + 9p(p^2 - 5q)r + 9q^3 - 2p^2q^2 \geqq 0.$$

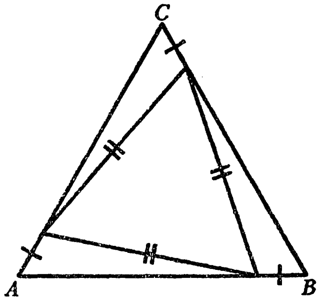
This evidently proves (**).

Also solved by Michael Goldberg, Washington, D. C. (two solutions); M. G. Greening, University of New South Wales, Australia; Steven R. Morphey, University of British Columbia; Simeon Reich, Israel Institute of Technology, Haifa, Israel; E. F. Schmeichel, College of Wooster, Ohio; and the proposer.

Further Comment on Problem 754

754. [March, November, 1970, and January, 1971] *Proposed by NSF Class at University of California at Berkeley.*

Show that the triangle ABC is equilateral.



Further comments by John F. Rigby, University College of South Wales, and Monmouthshire, Cardiff, Wales.

This is a continuation of my previous comment in this MAGAZINE [1], consisting of further counterexamples and theorems that answer various questions posed in [1]. The basic definitions are repeated below; this makes these comments more self-contained, but the numbering of figures, theorems, etc., is continued from [1]. We use the notation "$[LMN]$" to mean "M lies between L and N on the line LN." The basic formulae of hyperbolic trigonometry can be found in any standard textbook of non-Euclidean geometry.

DEFINITIONS. *Let $ABCD \cdots$ be an n-gon in the Euclidean or hyperbolic plane, and let $U, V, W \cdots$ be points on the lines $AB, BC, CD \cdots$ such that $AU = BV = CW = \cdots$. If $[AUB]$, $[BVC]$, $[CWD] \cdots$ then the n-gon $UVW \cdots$ is regularly inscribed in the first way in $ABCD \cdots$* (e.g., see Figure 19). *If $[ABU]$ etc., then $UVW \cdots$ is regularly inscribed in the second way* (e.g., see Figure 11 [1]). *If $[UAB]$ etc., then $UVW \cdots$ is regularly inscribed in the third way* (e.g., see Figure 23).

LEMMA 6. *In Figure 17 in the hyperbolic plane, where the length a is fixed, and x and y are functions of ϕ,*

$$\frac{dx}{d\phi}=\frac{\sinh x}{\tan \psi} \quad \textit{and} \quad \frac{dy}{d\phi}=\frac{\sinh x}{\sin \psi}\cdot$$

Proof. $\tanh a=\tanh x \cos \phi$, so $0=\operatorname{sech}^2 x \cos \phi(dx/d\phi)-\tanh x \sin \phi$. Hence

$$\frac{dx}{d\phi}=\sinh x \cosh x \tan \phi = \sinh x \cosh x \frac{\tanh y}{\sinh a}$$

$$= \sinh x(\cosh y \cosh a)\frac{\tanh y}{\sinh a} = \sinh x \frac{\sinh y}{\tanh a} = \frac{\sinh x}{\tan \psi}\cdot$$

The other result is proved similarly.

LEMMA 7. *Let WJV be a right-angled triangle in either the Euclidean or the hyperbolic plane (Figure 18) with $\angle JVW > \angle JWV=\alpha$, and let $\angle JVR=\alpha$. Then there exists θ $(0<\theta<\alpha)$ such that $WG=WC$.*

Proof. The proof given here is valid only in the hyperbolic plane, but a similar proof can easily be given in the Euclidean plane.

Regard θ as a variable angle, and write $WG=x$, $CJ=y$, $WC=z$, $WR=a$, $VJ=b$, $VR=c$, $VW=d$. Let r be the ray from V parallel to JW. Then

$$\text{(i)} \qquad \frac{dx}{d\theta}=-\frac{dx}{d\phi}=\frac{\sinh a}{\tan \beta} \quad \text{when } \theta=0 \qquad \text{(by Lemma 6).}$$

Also

$$\text{(ii)} \qquad \frac{dz}{d\theta}=\frac{dy}{d\phi}=\frac{dy}{d\omega}=\frac{\sinh c}{\sin \beta} \quad \text{when } \phi=0 \qquad \text{(by Lemma 6).}$$

Now

$$\frac{\sinh a}{\tan \beta}\Big/\frac{\sinh c}{\sin \beta}=\frac{\sin \gamma}{\sin \alpha}\cos \beta$$

(using the sine rule in triangle VWR) $=\sin \gamma \cosh b<\sin(\angle JVR)\cosh b=1$. Hence from (i) and (ii)

$$\text{(iii)} \qquad dx/d\theta < dz/d\theta \quad \text{when } \theta=0.$$

Write $f(\theta)=x-z$. Then $f(0)=a-a=0$, $f'(0)<0$ (from (iii)), and $f(\alpha)=d-WJ>0$. Since f is continuous we deduce that $x-z=f(\theta)=0$ for some θ between 0 and α.

(The result $(dx/d\theta)/(dy/d\theta)=(\sin \gamma/\sin \alpha)\cos \beta$ (when $\theta=0$) is also true in the Euclidean plane.)

COUNTEREXAMPLE 5. *Let $UVW \cdots$ be any regular $2n$-gon in the Euclidean or hyperbolic plane with obtuse angles. Then $UVW \cdots$ can be regularly inscribed in the first way in a nonregular polygon.*

Proof. Figure 19 illustrates the case $n=3$, but the same method holds for any $n\geqq 3$. Let the feet of the perpendiculars from U, V, $W \cdots$ to YZ, ZU, $UV \cdots$ be H, I, $J \cdots$.

(a) Suppose $\angle JVW > \angle JWV = \alpha$. Then HU meets IV at Q, say, between I and V; IV meets JW at R, say, between J and W, etc. Applying Lemma 7 to the triangle WJV, let θ be the angle between 0 and α such that $WG = WC$. By congruent triangles, $BV = GW$, so $AU = BV = CW = DX = \cdots$. Hence $UVW \cdots$ is regularly inscribed in the first way in $ABCD \cdots$ but $ABCD \cdots$ is not regular.

(b) Suppose $\angle JVW \leqq \angle JWV$. We then have Figure 20. Let PUQ, $QVR \cdots$ bisect $\angle IUV$, $\angle JVW \cdots$; then $PQR \cdots$ is a regular $2n$-gon and U, V, $W \cdots$ are the midpoints of PQ, QR, $RS \cdots$. Since $WV > FV$, the method of Counterexample 4[1] can be used to find a nonregular polygon $ABC \cdots$ in which $UVW \cdots$ is regularly inscribed in the first way.

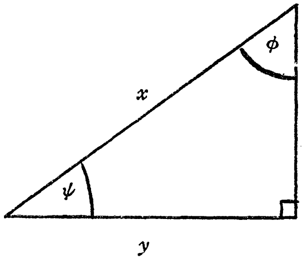


FIG. 17.

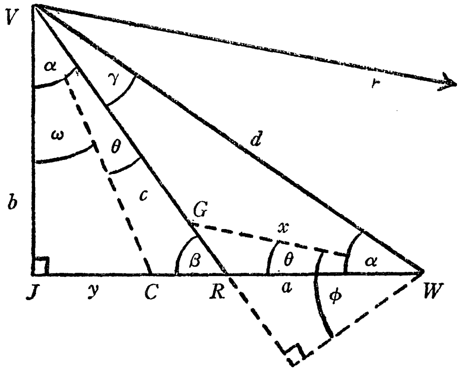


FIG. 18.

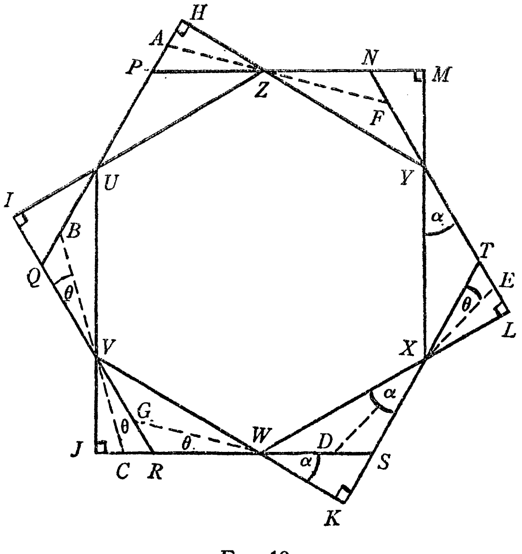


FIG. 19.

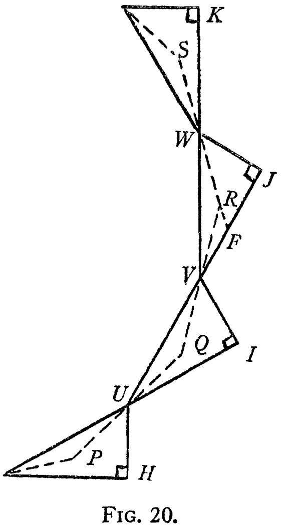


FIG. 20.

This second method is valid as long as $\angle JVW \leqq 2\angle JWV$, so the two methods described in (a) and (b) overlap; they give different counterexamples in general, but when $n=3$ in the Euclidean plane they both give Counterexample 2[1].

COUNTEREXAMPLE 6. *Let $UVW \cdots$ be any regular 2n-gon in the hyperbolic plane with acute angles. Then $UVW \cdots$ can be regularly inscribed in the second way in a nonregular polygon.*

Note. This counterexample supersedes Counterexample 3.

Proof. Figure 21 illustrates the case $n=3$, but the same method holds for any $n \geqq 2$. (The figure is distorted; the true size of the angle α is so small as to make an accurate figure impractical in the Poincaré angle-preserving model.) As in Counterexample 5, UH is the perpendicular from U to YZ, etc. Applying Lemma 7 to the triangle WJV, let θ be the angle between 0 and α such that $WG=WC$. By congruent triangles, $BV=GW$, so $AU=BV=CW=DX=\cdots$. Hence $UVW\cdots$ is regularly inscribed in the second way in $ABCD\cdots$ but $ABCD\cdots$ is not regular.

THEOREM 3. *Suppose that the regular n-gon $UVW\cdots$ ($n \geqq 4$) is regularly inscribed in the third way in $ABCD\cdots$ in the Euclidean or hyperbolic plane. Then $ABCD\cdots$ is regular.*

Proof. A partial proof (briefly repeated here) has already been given in the proof of Theorem 2(b)[1], where the angle β in Figures 23 or 24 (the smallest of the angles $\alpha, \beta, \gamma, \delta \cdots$) *gives rise to* the regular polygon $PQR\cdots$. We shall call the polygon $PQR\cdots$ in Figure 23 a *direct* polygon, to distinguish it from the *reversed* polygon $PQR\cdots$ in Figure 24. (In Figures 23 and 24 we have $[UAB]$, $[VBC]$ etc., and in Figure 23 we have $[UPQ]$, $[VQR]$ etc., while in Figure 24 we have $[UQP]$, $[VRQ]$ etc.) If $PQR\cdots$ has obtuse or right angles (which will always happen in the Euclidean plane), then $BV \leqq QV=RW \leqq CW$; but $BV=CW$ so we must have equality everywhere. Thus $Q=B$, $R=C$, and $\alpha=\beta=\gamma$. We can now prove that $\beta=\gamma=\delta$, etc., so $ABCD\cdots$ is regular. This proof holds in both Figures 23 and 24, and also when $PQR\cdots$ shrinks to a point.

If $PQR\cdots$ has acute angles, we must distinguish between direct and reversed polygons.

(a) Suppose that the smallest angle β gives rise to a direct acute-angled polygon $PQR\cdots$ (Figure 23), then all the other angles $\alpha, \gamma, \delta\cdots$ give rise to larger direct regular polygons, which must therefore have acute angles also. Suppose that $\alpha, \beta, \gamma\cdots$ are not all equal, and that without loss of generality α is the greatest angle and that α is strictly greater than β. (Note that β no longer necessarily denotes the smallest angle; this change of viewpoint avoids any awkward notation.) We then have $\gamma>\beta$, since $CW=BV<QV=RW$ and $\angle QRW$ is obtuse. Then $VB+BQ=VQ=WR<WC+CR=VB+CR$, so $BQ<CR$. Also $FP<BQ$ (this is easily proved as in Figure 22) so $FP<CR$. Hence $\alpha<\gamma$ since $UP=WR$ and $\angle UPF=\angle WRC$, contradicting the assumption that α is the greatest angle. Hence $\alpha, \beta, \gamma \cdots$ are all equal.

(b) If the smallest angle β gives rise to a reversed acute-angled regular polygon $PQR\cdots$ (Figure 24), we require

$$\text{(i)} \qquad WC = VB \leqq VQ = WR.$$

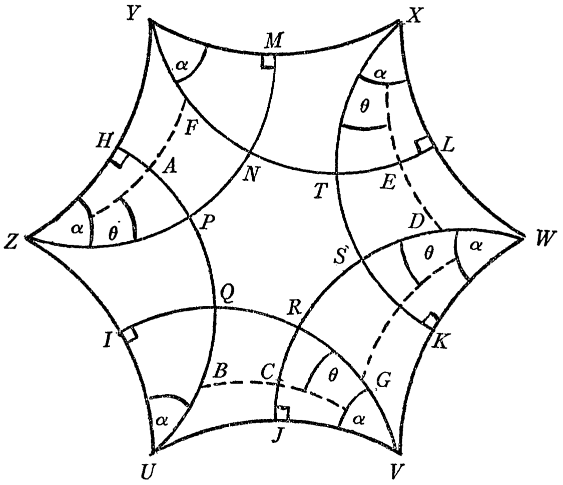


FIG. 21.

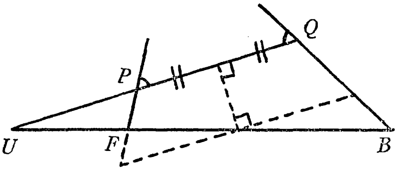


FIG. 22.

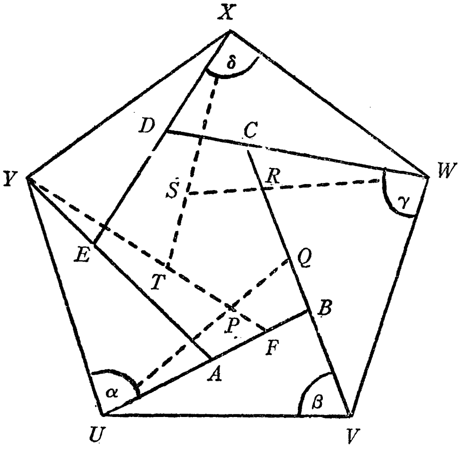


FIG. 23.

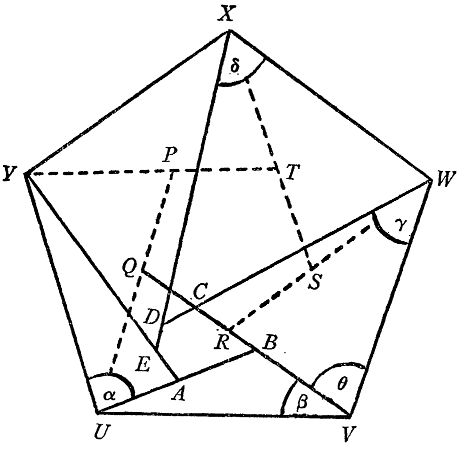


FIG. 24.

It is easy to show that nonadjacent sides of an acute-angled regular polygon do not meet, so $\angle RQW < \angle RQP = \angle QRW$. Hence $\angle RQW$ is acute also, and

$$\text{(ii)} \qquad WQ > WR = VQ.$$

Hence, using (i) also, C lies between Q and R (or else $C=R$), so $\gamma < \angle QWV < \theta$ (from (ii)). Hence $WC > VC = VB + BC = WC + BC$, a contradiction.

These results give the answers to questions (b) and (c) at the end of [**1**] and to the question in the paragraph before Counterexample 4 of [**1**]. I have still found no counterexamples for $(2n+1)$-gons when $n \geqq 2$.

Reference

1. J. F. Rigby, Comment on Problem 754, this MAGAZINE, 44 (1971) 45–53.

QUICKIES

From time to time this department will publish problems which may be solved by laborious methods, but which with the proper insight may be disposed of with dispatch. Readers are urged to submit their favorite problems of this type, together with the elegant solution and the source, if known.

Q518. Does there exist a number in base 10 which is a perfect square for every base $n > 2$?

[*Submitted by Warren Page*]

Q519. Characterize all triples of angles A, B, C for which

$$\tan A + \tan B + \tan C = \tan A \tan B \tan C$$

[*Submitted by Benjamin L. Schwartz*]

Q520. If $h > 0$, then $h/\sqrt{1+h} > \log(1+h)$.

[*Submitted by E. F. Schmeichel*]

Q521. Define a sequence of integers (x_i), $i = 1, 2, 3 \cdots$ such that x_1 is prime and for each i, $x_{i+1} = 2x_i + 1$. Is it possible that each member of the sequence is prime?

[*Submitted by Erwin Just*]

Q522. Determine all triangles XYZ satisfying

$$\frac{\sin 2X}{\sin A} = \frac{\sin 2Y}{\sin B} = \frac{\sin 2Z}{\sin C},$$

where ABC is a given triangle.

[*Submitted by Murray S. Klamkin*]

(Answers on page 140.)

References

1. M. Goldberg, The packing of equal circles in a square, this MAGAZINE, 43 (1970) 24–30.

2. J. Schaer and A. Meir, On a geometric extremum problem, Canad. Math. Bull, 8 (1965) 21–27.

3. J. Schaer, The densest packing of nine circles in a square, Canad. Math. Bull, 8 (1965) 273–277.

ANSWERS

A518. The number 121 is a perfect square for every base $n>2$, viz.,

$$(121)_n = n^2 + 2n + 1 = (n+1)^2.$$

A519. The given condition is equivalent to

$$\begin{aligned}\tan C &= -(\tan A + \tan B)/(1 - \tan A \tan B)\\ &= -\tan(A+B).\end{aligned}$$

Hence

$$C = -(A+B) + n\pi.$$

Thus

$$A + B + C = n\pi$$

$n=0,\ \pm 1,\ \pm 2 \cdots.$

A520. Let $d(h) = h/\sqrt{1+h} - \log(1+h)$. Observe that $d(0)=0$ and $d(h)>0$ for h sufficiently large, since $h/\sqrt{1+h}$ is asymptotic to $\sqrt{h}$. So if $d(h) \leqq 0$ for some $h>0$, it can only mean that $d'(\alpha)=0$ for some $\alpha>0$. But it is easily checked that $d'=0$ only at 0 and so $d(h)>0$ for all $h>0$.

A521. If x_i is an odd prime, p, then it is easily computed that for each positive integer, r,

$$\begin{aligned}x_{i+r} &= 2^r x_i + (2^r - 1)\\ &= 2^r p + (2^r - 1).\end{aligned}$$

When $r=p-1$, $x_{i+(p-1)} = 2^{p+1}p + (2^{p+1}-1)$. Since Fermat's theorem guarantees that $p \mid (2^{p+1}-1)$, it follows that $x_{i+(p-1)}$ is composite. The sequence (x_i) cannot, therefore, consist wholly of primes.

A522. First note that XYZ must be acute. Now let $2X=\pi-R$, $2Y=\pi-S$, and $2Z=\pi-T$. Thus RST is a triangle and

$$\frac{\sin R}{\sin A} = \frac{\sin S}{\sin B} = \frac{\sin T}{\sin C}.$$

Whence $RST \sim ABC$ and

$$2X = \pi - A,\quad 2Y = \pi - B,\quad 2Z = \pi - C.$$

(Quickies on page 178.)

MATHEMATICS MAGAZINE, Vol. 44, No. 3, May-June, 1971

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN