PATRICK SUPPES

FINITISM IN GEOMETRY*

Developments in the foundations of mathematics in the twentieth century have been dominated by a concern to provide an adequate conceptual basis for classical analysis, as well as classical number theory. The attitude of Frege, Hilbert, Russell and Brouwer, along with others who have followed them, have had a continual focus on providing a proper conceptual foundation for classical parts of pure mathematics. What is remarkable about this history is how little attention has been paid to finitism in the application of mathematics. Attention to various applied traditions would have supported a much more radical finitism than was possible, as long as the viewpoint was that of providing an adequate basis for classical elementary number theory. Even the weakest systems of recursive arithmetic have closure conditions that lead to constructions far more complex than are needed in any standard applications. The passions for closure and for completeness are in themselves wonderful intellectual passions, but they run counter to the drive for solutions of applied problems by limited means, ones that are computationally feasible. This line of thought easily leads to finitism in the applications of mathematics to the real world.

Those whose focus is pure mathematics might immediately respond to my emphasis on finitism by saying that there is already a considerable emphasis on finitism in pure geometry, and by then citing the extensive and deep research on finite geometries in the last half century. It is also natural to think of this work as an extension in many ways of Felix Klein's Erlangen program, that is, the approach to geometry that emphasizes, above all, the structure of the group of automorphisms of the geometry. The study of automorphisms is also an important way of studying finite geometries, for given a finite geometry, the group of automorphisms is a finite group. Problems abound. Given, for example, a finite group, a natural problem is to determine its finite geometry. A magisterial survey up until the mid-sixties of the twentieth century is to be found in Dembowski (1968). In the last thirty years there has also been a substantial body of work, and a number of open problems of a purely mathematical sort have been solved, especially in the relation of finite geometries to finite simple groups.

Erkenntnis **54:** 133–144, 2001.

However, this vast body of research is not the topic of this paper. I have in mind more elementary mathematical problems, ones that are focused on applications. What are the applications I have in mind? I would begin with the mathematical methods used by architects since ancient times. As early as the sixth century B.C., architects were applying geometry to detailed questions of construction. It was not that they were simply mechanically using the elementary geometry familiar from the philosophical and mathematical tradition of that time, but they were using detailed geometrical ideas to provide visual illusions. These methods of providing for visual illusions led to what are called in the architectural literature, then and now, "refinements". A principal example is *entasis,* which is a geometric construction that makes the vertical outline of marble columns convex rather than linear. Here is what Vitruvius (1960) has to say about this construction.

> These proportionate enlargements are made in the thickness of columns on account of the different heights to which the eye has to climb. For the eye is always in search of beauty, and if we do not gratify its desire for pleasure by a proportionate enlargement in these measures, and thus make compensation for ocular deception, a clumsy and awkward appearance will be presented to the beholder. With regard to the enlargement made at the middle of columns, which among the Greeks is called εντασις, at the end of the book a figure and calculation will be subjoined, showing how an agreeable and appropriate effect may be produced by it. (p. 86)

Entasis is, of course, a famous example, but there are many more; the most extensive extant written record is Vitruvius' Roman treatise, written in the time of Augustus. The kind of detailed applications to be found in Vitruvius are not exceptional but standard. It is a justified lament that the book on architecture written by Ictinus, architect of the Parthenon, is now lost. The important point in the present context is that, long before Ictinus and long after Vitruvius, architects were trained in geometry and how to apply it in a detailed way to the construction of buildings. As I like to put it, these architects were not proving theorems, but doing constructions (*problems* in the language of Euclid). The constructions were not all just simple and obvious, in fact, the lost drawing of Vitruvius' diagram of the construction of entasis was more complicated in its mathematical content than what can be found in most well-known architectural treatises today, but certainly not more so than the mathematics back of modern computer programs for architectural design. More than a thousand years after Vitruvius, in Palladio's famous four books of architecture, the Greek tradition of proportion, rather than direct numerical computation, continued to dominate the detailed thinking of architects as may be seen in many places in Palladio's exposition. In many ways, the computer resources now available are encouraging once again the emphasis on proportion.

Beyond entasis and other visual illusions, the most important mathematical or geometrical discovery already used by architects of the fifth century was the discovery of perspective, traditionally attributed to the painter Agatharchus, who designed the background for a tragedy by Aeschylus and wrote a book about scene painting, as a consequence of this experience. Mathematical investigation of the theory of perspective began as early as Democritus and Anaxagoras. This is what Vitruvius had to say about this early work on perspective, writing half a millenium later in Rome.

Agatharcus, in Athens, when Aeschylus was bringing out a tragedy, painted a scene, and left a commentary about it. This led Democritus and Anaxagoras to write on the same subject, showing how, given a centre in a definite place, the lines should naturally correspond with due regard to the point of sight and the divergence of the visual rays, so that by this deception a faithful representation of the appearance of buildings might be given in painted scenery, and so that, though all is drawn on a vertical flat facade, some parts may seem to be withdrawing into the background, and others to be standing out in front. (p. 198)

What is important about this variety of applications of geometry, including, of course, the important case of perspective, is that the methods and results are highly finitistic in character, just as are, in fact, most, if not all, the results of Greek geometry.

So, if we look at the applications of geometry in this long and important architectural tradition, which includes the thorough development of the mathematical theory of perspective and also of projective geometry, one is naturally struck by the tight interaction between the finitistic character of the mathematics and the familiar applications.

The applications of numerical methods in ancient times have the same finitistic character, but I do not try to consider the details here.

At a certain point in the nineteenth century, nonfinitistic methods entered classical analysis and were extended, in some sense, to geometry, due to several different forces at work. I would stress especially the demand for representation theorems in complete form, because the representations of geometrical spaces came to be formulated so as to be isomorphic to a given Cartesian product of the real number system, the product itself depending upon the dimension of the space. Such a search for representation theorems is not at all a part of the Greek tradition in much earlier times. One way of putting the matter, in terms of many different discussions of the foundations of mathematics, is that such representation theorems demand a representation of the actual infinite. It is this representation of the actual infinite, or, to put it in more pedantic terms, sets of infinite cardinality, which are mapped from the geometrical space onto a Cartesian product of the real numbers isomorphically preserving the structure of the geometrical space. The stronger sense of nonfinitistic methods involving the axiom of choice, or its equivalent, is a still further remove from the finitism of applications I am focusing on here.

It is a separate question to what extent the parts of classical analysis actually used in applications of mathematics in physics and elsewhere can avoid commitments not only to any results using the axiom of choice, but, also, any sets of an actually infinite character. My own conviction is that one can go the entire distance, or certainly almost the entire distance, in a purely finitistic way, but this is not what I focus on here.

1. QUANTIFIER-FREE AXIOMS AND CONSTRUCTIONS

One requirement of very strict constructive methods is that the axioms of the theory in question should be quantifier-free and thus avoid all purely existential assumptions. Such existential demands are replaced by specific constructions, which are either primitive or definable, in a quantifier-free way, in terms of the given primitive constructions. Often, in traditional formulations of geometry, existential axioms are innocuous, in the sense that they can easily be replaced by essentially equivalent quantifier-free axioms which use a finite number of given constants.

2. TEMPORAL SEQUENCE

An aspect of applications ignored in standard axiomatic formulations of constructions is the temporal sequence in which any actual constructions are performed. The usual mathematical notation for operations or relations does not provide a place for indicating temporal order. In probability theory, however, such temporal matters are of great importance in developing the theory or applications of stochastic processes. Whether time is treated as continuous or discrete, the time of possible occurrence of events or random variables is indicated by subscripts, t, t', t_1, etc., for continuous time and $n, n', n+1$, etc., for discrete time. When the time of possible occurrence is not stipulated as part of the theoretical framework, then, in principle, a new random variable is required to describe this uncertainty.

Such probabilistic considerations lie outside the framework of geometry being considered here, although they could naturally be included if the actual fact of errors in real constructions were incorporated. Such matters are of undoubted importance in applications, but, only for the sake of simplicity, will be ignored here. This is because of the much greater complexity of the axioms required to include from the beginning the conceptual apparatus for dealing with the inevitable errors of approximation in real constructions. For examples of the complexity of such axioms, see Suppes et al. (1989, Ch. 16).

What we can easily use in the notation for constructions are subscripts to show temporal sequence. This provides an easy way of showing whether or not two constructions require the same number of steps. Here is a simple example from affine geometry using only the two constructions of finding the midpoint, i.e., bisecting a line segment, and doubling, in one of two directions, a line segment. In customary fashion, line segments are referred to in the intuitive discussion, but only the two points that are endpoints of a segment are required formally.

In this example I formulate the axioms of construction informally.

PROBLEM. Given three noncollinear points α_0, β_0 and γ_0, construct a parallelogram with adjacent sides $\alpha_0\beta_0$ and $\alpha_0\gamma_0$.

Construction:

1. Construct the midpoint a_1 of $\beta_0\gamma_0$.
2. Construct the point a_2 that doubles the segment $\alpha_0 a_1$ in the direction of a_1. The figure $\alpha_0\gamma_0 a_2\beta_0$ is the desired parallelogram (see Figure 1).

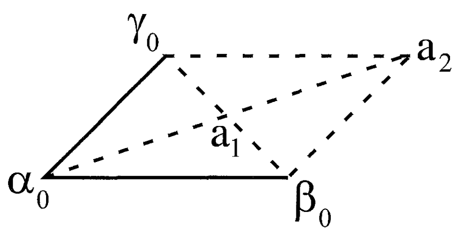

Figure 1. Constructed parallelogram.

3. AFFINE AXIOMS

I now turn to axioms for the affine plane constructions that can be made by bisecting and doubling the distances between pairs of points. The axioms are quantifier-free, as would be expected from my earlier remarks. I have, in stating the axioms, dropped the notation for temporal order of construction, but the appropriate subscripts can be added, if desired, when presenting constructions, in an obvious manner. The subscripts are less useful in stating the axioms.

The axioms have trivial models, for the spirit is that the necessary distinct points are given as data at the beginning of each construction. This point is clarified later in the definition of a finite configuration. A few elementary definitions and theorems follow the axioms, but no proofs are given here, except the representation theorem for introducing coordinates. (More theorems and proofs are to be found in Suppes (2000), although in a different notation.) To avoid closure conditions on operations, the bisection construction is formally expressed as a relation $B(ab, c)$, with the intuitive meaning that c is the constructed point bisecting segment ab. Correspondingly, $D(ab, c)$ is the doubling construction in the direction b. Intuitively, the position of c is shown in Figure 2. Also needed as a primitive ternary relation is that of linearity. Intuitively $L(abc)$ if a, b and c all lie on a line.

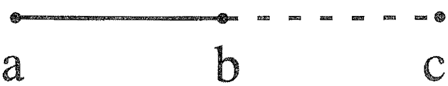

Figure 2. The doubling construction.

First, the axioms involving just linearity.

(L1) If $a = b$, $a = c$ or $b = c$, then $L(abc)$,

(L2) If $a \neq b$, $L(abp)$, $L(abq)$ and $L(abr)$, then $L(pqr)$.

Next, the axioms just involving bisection. Note that the first axiom asserts that when points a and b, not necessarily distinct, are given, then the constructed point is unique.

(B1) If $B(ab, c)$ and $B(ab, c)'$ then $c = c'$. (Uniqueness)

(B2) $B(aa, a)$. (Idempotency)

(B3) If $B(ab, c)$ then $B(ba, c)$. (Commutativity)

(B4) If $B(ab, c)$, $B(de, f)$, $B(ad, g)$, $B(be, h)$, $B(cf, i)$ and $B(gh, j)$, then $i = j$. (Bicommutativity)

(B5) If $B(ab, c)$ and $B(ab', c)$ then $b = b'$. (Cancellation)

Axiom B4 illustrates well the simplification of understanding introduced by using algebraic operation notation. This axiom then has the much simpler form:

$$(a \oplus b) \oplus (d \oplus e) = (a \oplus d) \oplus (b \oplus e).$$

As should be clear from what I said already, I use the awkward relation-style notation to make the restriction to finite configurations transparent later.

The four axioms for the doubling construction are the following.

(D1) If $D(ab, c)$ and $D(ab, c')$, then $c = c'$. (Uniqueness)

(D2) If $D(ab, c)$ and $D(ba, c)$, then $a = b$. (Antisymmetry)

(D3) If $D(ab, c)$ and $D(ab', c)$, then $b = b'$. (Left cancellation)

(D4) If $D(ab, c)$ and $D(a'b, c)$ then $a = a\prime$. (Right cancellation)

Finally, there are three axioms using more than one relation.

(BD) If $D(ab, c)$ then $B(ac, b)$. (Reduction)

(LB) If $B(ab, c)$ then $L(abc)$. (Linearity of Bisection)

(LBL) If $B(ab, d)$, $B(bc, e)$, $B(ac, f)$ and $L(def)$ then $L(abc)$. (Linearity of Midpoints)

I introduce two conditional definitions to reintroduce the simplifying operational notation.

DEFINITION 1. If $B(ab, c)$ then $a \oplus b = c$.

DEFINITION 2. If $D(ab, c)$ then $a * b = c$.

In other words, $\oplus$ is the conditional operation of bisecting and $*$ of doubling. In the following theorems and definitions, which use the operational notation, the conditions required by Definitions 1 and 2 to employ the operations $\oplus$ and $*$ are assumed satisfied in all cases, but without explicit statement.

4. THEOREMS

First, I summarize in one theorem the elementary properties of collinearity.

THEOREM 1. (Collinearity)

(i) If $L(abc)$ then L holds for any permutation of abc.
(ii) $L(aba)$.
(iii) If $a \neq b$, $L(abc)$ and $L(abd)$ then $L(acd)$.
(iv) If $p \neq q$, $L(abp)$, $L(abq)$ and $L(pqr)$ then $L(abr)$.

Szmielew (1983) points out that (i), (ii) and (iii) of Theorem 1 are equivalent to Axioms L2 and L3 – actually a weaker form of (i), namely, if $L(abc)$ then $L(bac)$.

THEOREM 2. $(a \oplus b) \oplus c = (a \oplus c) \oplus (b \oplus c)$ (Self-distributivity).

THEOREM 3. If $a \oplus b = a$ then $a = b$.

The next theorem shows that reduction and linearity hold for doubling.

THEOREM 4. $a * (a \oplus b) = b$ and $L(ab(a * b))$.

Any three noncollinear points "form" a triangle, so we may define the ternary relation T of triangularity as the negation of L.

DEFINITION 3. $T(abc)$ iff it is not the case $L(abc)$.

In the next definition the quaternary relation P defined has the intuitive meaning that four points standing in this relation form a parallelogram (thus "P" for "parallelogram").

DEFINITION 4. $P(abcd)$ *iff* $T(abc)$ & $a \oplus c = b \oplus d$.

This definition characterizes parallelograms, which do the work "locally" of parallel lines in standard nonconstructive affine geometry. The triangularity or nonlinearity condition eliminates degeneracy. The important condition is that a convex quadrilateral *abcd* (see Figure 3) is a parallelogram if and only if the midpoints of the diagonals *ac* and *bd* coincide, which is a familiar property of parallelograms, but also sufficient for the definition. Concave quadrilaterals violate the midpoint condition.

I now define informally what a construction is in terms of the concepts already introduced. Given three noncollinear points α_0, β_0 and γ_0, a *construction* is a finite sequence of terms $(C_1, \ldots, C_n)$, with each term C_k, $k = 1, \ldots, n$ being an ordered triple:

(i) whose first member is a point a_k provided a_k is neither α_0, β_0, γ_0 nor a previously constructed point;

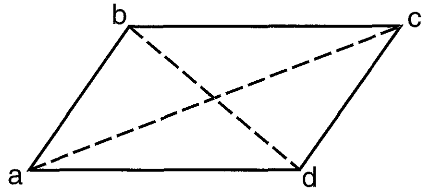

Figure 3.

(ii) whose second member is a pair of points (a_i, a_j) which are either $\alpha_0, \beta_0, \gamma_0$ or points previously constructed in the sequence, i.e., $i, j < k$,;

(iii) and whose third member is either 'B' standing for use of bisection or 'D' for doubling such that the operation shown is applied to the pair of points (a_i, a_j) to construct a_k.

Thus, the earlier construction of a parallelogram with adjacent sides $\alpha_0\beta_0$ and $\alpha_0\gamma_0$ is the sequence:

$$((a_1, \beta_0\gamma_0, B), (a_2, \alpha_0 a_1, D)).$$

Note that a construction requires the construction of a new point at every step. The constructive method of verifying that a point has not been previously introduced is by introducing coordinates inductively at each step to make a numerical test possible. In applications this theoretical coordinate check is ordinarily replaced by a quick visual inspection of the construction at each stage.

5. ANALYTIC REPRESENTATION THEOREM

Given three noncollinear points $\alpha_0, \beta_0, \gamma_0$ and a construction $C = (C_1, \ldots, C_n)$ based on these three points, then the pair (φ_1, φ_2) of rational numerical functions as defined below provide an analytic representation of the points $(a_1, \ldots, a_k)$ constructed by C, where:

(i) $\varphi_1(\alpha_0) = 0$, $\varphi_2(\alpha_0) = 0$, $\varphi_1(\beta_0) = 1$, $\varphi_2(\beta_0) = 0$, $\varphi_1(\gamma_0) = 0$, $\varphi_2(\gamma_0) = 1$,

(ii) $\varphi_i(a \oplus b) = \frac{\varphi_i(a)+\varphi_i(b)}{2}$, $i = 1, 2$, (bisecting),

(iii) $\varphi_i(a * b) = 2\varphi_i(b) - \varphi_i(a)$, $i = 1, 2$, (doubling).

Outline of Proof. By induction: For $n = 1$, if $C_1 = (a_1, \alpha_0\beta_0, B)$, then

$$\varphi_1(a_1) = \frac{0+1}{2} = \frac{1}{2},$$

$$\varphi_2(a_1) = \frac{0+0}{2} = 0.$$

And similarly for the remaining 8 possible cases of C_1.

If the representation holds for $1, \ldots, n-1$, then, although there are many possible C_n's, for the given construction C, C_n is unique, with the coordinates inductively defined up to $n-1$ being used to check that the point a_n constructed is new. So if the general term is of the form $C_n = (a_n, a_i a_j, B), i, j < n$,

$$\varphi_1(a_n) = \frac{\varphi_1(a_i) + \varphi_1(a_j)}{2}$$

$$\varphi_2(a_n) = \frac{\varphi_2(a_i) + \varphi_2(a_j)}{2},$$

and if the general form is $C_n = (a_n, a_i a_j, D), i, j < n$, then

$$\varphi_1(a_n) = 2\varphi_1(a_j) - \varphi_1(a_i),$$

$$\varphi_2(a_n) = 2\varphi_2(a_j) - \varphi_2(a_i).$$

As is evident, the method of inductively assigning numerical coordinates is obvious and so is the proof, which is how it should be for the standard geometric constructions, as opposed to the necessarily lengthy and complicated proofs for the affine, projective or Euclidean representation of the entire real plane or three-dimensional space. A rigorous example of such a proof is developed over many pages in Borsuk and Szmielew (1960).

6. ANALYTIC INVARIANCE THEOREM

There is also a finitistic affine invariance theorem that holds for the introduction of coordinates, which I shall state without proof.

Invariance Theorem. Let C be a construction based on three noncollinear points α_0, β_0 and γ_0, and let (φ_1, φ_2) be coordinate functions as defined above. Let

$$\varphi_1' = a\varphi_1 + b, \;\; a \neq 0$$

$$\varphi_2' = c\varphi_2 + d, \;\; c \neq 0,$$

then (φ_1', φ_2'), is a coordinate representation of C with the origin possibly changed (if $b \neq 0$ or $d \neq 0$) and with affine changes of scale unless $a = 1$

or $c = 1$, where now $\varphi_1'(\alpha_0) = b$, $\varphi_2'(\alpha_0) = d$, $\varphi_1(\beta_0) = a + b$, $\varphi_2(\beta_0) = d$, $\varphi_1(\gamma_0) = b$ and $\varphi_2(\gamma_0) = c + d$.

It should be obvious that for each particular construction something more is needed, namely, the kind of proof, so familiar in Euclid, that the construction actually carries out what was intended as the construction. Constructions are like proofs in that each step can be valid, but what is constructed, like what is proved, is not for what was given as the problem or theorem. So accompanying each construction there should be a matching proof.

Second, we in general want for constructions an "independence-of-path" theorem. If we are given a finite configuration of points we can prove the configuration has certain properties independent of the particular construction used to generate it.

None of these additional proof requirements for particular constructions take us outside the finitistic framework already described. In Suppes (2000) an additional trapezoid construction is introduced to permit the easy, finitistic construction of any rational number.

7. ALGEBRAIC PROBLEMS

The emphasis on introducing coordinates via constructions simplifies drastically the analytic representation theorem. A purely algebraic approach with a general representation for finite models of the axioms requires additional algebraic identities as axioms and a more complicated representation theorem. Examples of the sorts of identities that are needed are the following, which is not complete:

$$\begin{aligned} (a * b) * b &= a, \\ (a * b) \oplus (b * a) &= a \oplus b, \\ (a * b) * (a \oplus b) &= b * a. \end{aligned}$$

It is easy to check that the intended coordinate representation requires that these be satisfied. It is also apparent that these and similar identities represent redundant steps in constructions, which are eliminated by the earlier requirement that each point constructed be a new one. These identities show well enough why the restricted nature of constructions drastically simplifies the theory, which should be the aim of any theory intended to be used in a variety of applications, as has historically been the case for the classical geometric constructions that preceded by many centuries the theory of real numbers.

8. FINAL REMARK

My main philosophical point that I have tried to illustrate in this article is that the foundations of many parts of applied mathematics can be given a simple finitistic formulation. Some efforts in this direction for the more complicated case of applications of analysis are to be found in joint articles with Rolando Chuaqui or Richard Sommer listed in the references.

NOTE

* It is a pleasure to dedicate this article to Willy Essler in celebration of his 60th birthday. I am grateful to Richard Sommer for several useful comments.

REFERENCES

Borsuk, K. and Szmielew, W.: 1960, *Foundations of Geometry*, North Holland, Amsterdam.

Chuaqui, R. and Suppes, P.: 1990, 'An Equational Deductive System for the Differential and Integral Calculus', in P. Martin-Lof and G. Mints (eds.), *Lecture Notes in Computer Science, Proceedings of COLOG-88 International Conference on Computer Logic*, Springer-Verlag, Berlin, pp. 25–49.

Chuaqui, R. and Suppes, P.: 1995, 'Free-variable Axiomatic Foundations of Infinitesimal Analysis: A Fragment with Finitary Consistency Proof', *The Journal of Symbolic Logic* **60**, 122–159.

Dembowski, P.: 1968, *Finite Geometries*, Springer-Verlag, New York.

Sommer, R. and Suppes, P.: 1996, 'Finite Models of Elementary Recursive Nonstandard Analysis', *Notas de la Sociedad Matematica de Chile* **15**, 73–95.

Sommer, R. and Suppes, P.: 1997, 'Dispensing with the Continuum', *Journal of Mathematical Psychology* **41**, 3–10.

Suppes, P.: 2000, 'Quantifier-free Axioms for Constructive Affine Plane Geometry', *Synthese* (to appear).

Suppes, P. and Chuaqui, R.: 1993, 'A Finitarily Consistent Free-variable Positive Fragment of Infinitesimal Analysis', *Notas de Logica Matematica* **38**, 1–59.

Suppes, P., Krantz, D. H., Luce, R. D., and Tversky, A.: 1989, *Foundations of Measurement*, Vol. II, Academic Press, San Diego, CA.

Szmielew, W.: 1983, *From Affine to Euclidean Geometry: An Axiomatic Approach*, PWN-Polish Scientific Publishers, Warszawa, Poland, and D. Reidel, Dordrecht, Holland.

Vitruvius: 1960, *The Ten Books on Architecture*. Translated by Morris Hicky Morgan Dover Publications, Inc., New York (Originally published by Harvard University Press in 1914.)

Center for the study of Language and Information
Stanford University
Stanford, CA 94305-4115
USA
E-mail: Suppes@csli.stanford.edu