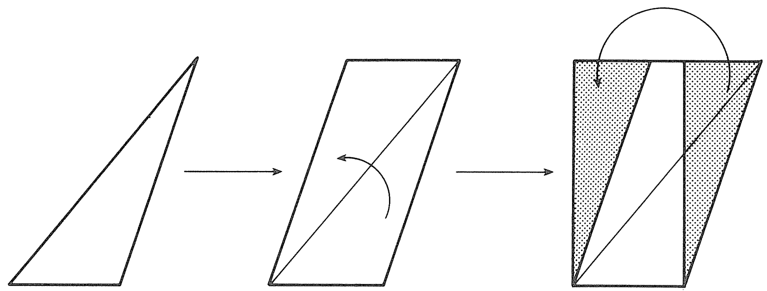
(2) Hilbert and a Chinese view

FUNG-KIT SIU AND MAN-KEUNG SIU

1. Introduction

"Every schoolboy knows" that the area of a triangle with altitude h and base b is equal to $\frac{1}{2}hb$. It is not difficult to explain how the $\frac{1}{2}$ comes in. One way to do this is by means of Fig. 1. Indeed, the ancients might have



already used such dissection methods to find area almost 4000 years ago! But when we come to volume of solids, things are not as easy. The volume of a tetrahedron with altitude h and base of area A is equal to $\frac{1}{3}hA$. Some curious soul may question, sooner or later, where the $\frac{1}{3}$ comes in. An answer such as "It comes out of computation by using calculus" is hardly convincing to someone who does not even know what calculus is. Besides, a heuristic argument which affords more insight is pedagogically much more desirable. In section 2 we shall sample some of these heuristic arguments, borrowing from the wise ancients. However, there seems to be one flaw, and that is the intervention of some kind of infinitesimal argument in certain cases. Can this be avoided? We shall discuss the question in section 3 and find (happily) that our ignorance is not to be blamed!

2. Some heuristic arguments by ancient mathematicians

Figure 1 is a typical illustration of a method frequently employed by the ancients. In this method, known as "*dissection*", a figure is cut up into pieces which are then reassembled to form another figure of known area. Sometimes it is easier to apply a variation known as "*complementation*", in which congruent pieces are added to two figures, one of whose area is known, to obtain congruent figures. Both methods were extended by the ancients to the three-dimensional case with some success. We shall look at a few examples, mostly from Chinese mathematics, as these may be less well known to western readers.

Math. Gazette, 65(1981), 265-291.

"Jiu Zhang Suan Shu" (Nine Chapters on the Mathematical Art), whose date of compilation is usually placed in the latter part of the first century but whose content is believed to be known well before that, is one of the earliest mathematical classics in China [1–3]. Many formulae for the area and volume of various geometric figures appear in prose form in its second and fifth chapters, but neither proofs nor explanations are given. In a commentary written by Liu Hui (third century) an interesting method of dissection is introduced to verify and explain these formulae. The basic idea is to cut up the figure into pieces, each being of standard type known as a "qi" (meaning chessman). Some examples of "qi" frequently used are:

(i) "Li Fang" (= cube or sometimes including rectangular parallelepiped),
(ii) "Jian Du" (= right prism with triangular base),
(iii) "Yang Ma" (= pyramid with rectangular base),
(iv) "Bie Nao" (= tetrahedron with right triangle as base and altitude perpendicular to the base).

For instance Liu Hui mentioned that three identical "Yang Ma"'s made up one "Li Fang" so that a "Yang Ma" with square base of area a^2 and one side of length a perpendicular to the base has a volume $\frac{1}{3}a^3$. In another instance he mentioned, accompanied by some sort of infinitesimal argument, that a "Li Fang" was composed of two equal "Jian Du"'s, and that a "Jian Du" could be dissected into one "Yang Ma" and one "Bie Nao" with volumes in the ratio 2:1. Again this yields the same formula for the volume of a "Yang Ma". Liu Hui was using the following intuitively clear principle: if in two solids of equal height, the section made by the planes parallel to and at the same distance from their respective bases always bear a constant ratio, then the volume of the two solids bear the same ratio. In another instance he pointed out that π:4 was the ratio of the volume of a sphere to that of another solid in which he was thinking of a pile of circles making up the sphere and a pile of corresponding circumscribing squares making up the "Mou He Fang Gai" (meaning 'closely fitted square lids'). About 250 years later another great Chinese mathematician Zu Chong-Zhi (fifth century), together with his son Zu Xuan, applied Liu's principle in a most ingenious way to compute the volume of the "Mou He Fang Gai" [1, 3, 4]. They announced explicitly the principle (actually a particular case of it) as: if the sections of two solids at equal heights have equal area, then the two solids cannot differ in volume [1, 3]. (This same principle has come to be known as Cavalieri's Principle in the west because it was discovered independently by Cavalieri in 1635.)

3. *Hilbert's third problem*

Readers may have noticed that some kind of infinitesimal consideration is involved in the heuristic arguments on volume presented in section 2,

while nothing of that sort is required in the case of finding area of polygons (which boils down to finding area of triangles). Is that a failing on our part? Or is that the nature of things? Mathematicians had already posed this question in the last century, and it came from no lesser beings than Gauss and Hilbert. We shall take a glimpse of this story in this section, but for technical reason we shall omit all proofs.

During the 1900 International Congress of Mathematicians held at Paris, Hilbert delivered his famous address in which he presented twenty-three problems, which had since influenced much of the present century's mathematical research. Let us quote his third problem in full [5]:

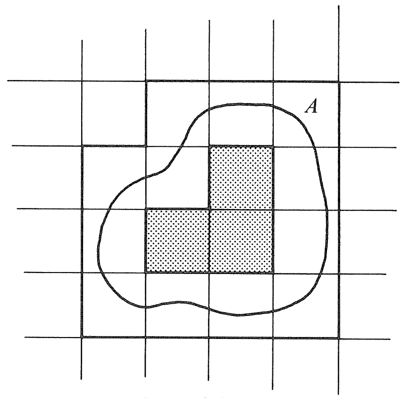
> "In two letters to Gerling, Gauss expresses his regret that certain theorems of solid geometry depend upon the method of exhaustion, i.e., in modern phraseology, upon the axiom of continuity (or upon the axiom of Archimedes). Gauss mentions in particular the theorem of Euclid, that triangular pyramids of equal altitudes are to each other as their bases. Now the analogous problem in the plane has been solved. Gerling also succeeded in proving the equality of volume of symmetrical polyhedra by dividing them into congruent parts. Nevertheless, it seems to me probable that a general proof of this kind for the theorem of Euclid just mentioned is impossible, and it should be our task to give a rigorous proof of its impossibility. This would be obtained, as soon as we succeeded in specifying two tetrahedra of equal bases and equal altitudes which can in no way be split up into congruent tetrahedra, and which cannot be combined with congruent tetrahedra to form two polyhedra which themselves could be split up into congruent tetrahedra."

The last statement has come to be known as *Hilbert's Third Problem*, and is the first amongst the twenty-three problems to be solved, indeed in that very same year by Dehn.

Before we go any further, let us see what happens in the plane. The methods of dissection and complementation mentioned in section 2 are based upon the following seemingly-obvious fact: if a polygon is cut up into finitely many pieces which are then reassembled to form another polygon, then the two polygons have the same area. Let us ask if the converse of this fact is true, i.e. can we always cut up a given polygon into finitely many pieces and reassemble the pieces to form another given polygon of equal area? In about 1832 Bolyai (father of the younger Bolyai who discovered hyperbolic geometry) and Gerwien independently found an affirmative answer to this question. Interested readers can find an account of the Bolyai–Gerwien Theorem in many books, e.g. in [6, 7].

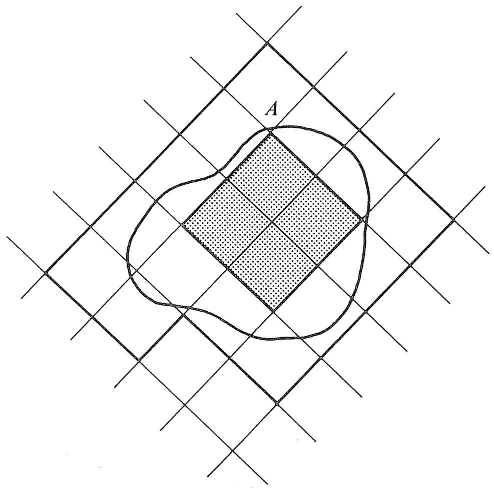
Hilbert's Third Problem queries the validity of the three-dimensional analogue of the Bolyai–Gerwien Theorem. What Dehn proved in 1900 is the surprising fact (perhaps not so surprising to Hilbert who suspected a negative answer all along) that one can never dissect a regular tetrahedron into finitely many pieces and reassemble them to form a cube of equal volume. What relevance does this result have concerning a theory of area and volume? Even more fundamentally, what is area? and what is volume? In elementary and high schools we usually appeal to students' geometric intuition for an explanation and mean by the area (volume) of a figure the number of unit squares (unit cubes) making up the figure. Although this is admittedly an unsatisfactory explanation (e.g. it is not clear how squares can make up a circle), it is wise to let things stand at this somewhat vague but adequate state, for we shall soon see that it is not at all easy to define area and volume in a way comprehensible to an average school student but at the same time mathematically rigorous. In fact, it is already not easy to do so for the concept of length, but we shall assume in the sequel that we can take that for granted.

Let us adopt a down-to-earth approach to begin with. Given a figure A, we superimpose upon it a grid consisting of unit squares. Suppose 3 squares lie within A and 15 squares enclose A, then the area $m(A)$ of A



$3 \leqslant m(A) \leqslant 15$

(a)



$4 \leqslant m(A) \leqslant 18$

(b)

will satisfy $3 \leqslant m(A) \leqslant 15$ (see Fig. 2(a)). We now refine the original grid to a new one in which each square is further divided into 100 smaller squares. Suppose 824 such smaller squares lie within A and 917 such smaller squares enclose A, then the area $m(A)$ of A will satisfy $8{\cdot}24 \leqslant m(A) \leqslant 9{\cdot}17$. By repeating this refinement procedure we can squeeze $m(A)$ in between tighter and tighter bounds. Now, let us forget about the intuitive meaning of $m(A)$. Suppose at the nth stage ($n = 0, 1, 2, \ldots$, where the 0th stage refers to the use of the original grid of unit squares) there are $a(n)$ squares lying within A and $b(n)$ squares enclosing A, then a little bit of analysis will tell us that

(i) the sequence $\{a(n)/10^{2n}\}$ converges to a limit $\underline{m}(A)$,
(ii) the sequence $\{b(n)/10^{2n}\}$ converges to a limit $\overline{m}(A)$,
(iii) $\underline{m}(A) \leqslant \overline{m}(A)$.

If A is a figure which is reasonably nice so that it deserves to have an "area", then in fact $\underline{m}(A) = \overline{m}(A)$, in which case we shall honour A with the decoration that "A is measurable" and denote this common value by $m(A)$ called the "area of A". Thus we can define the notion of area by the above procedure, which however leaves a few things to be desired. It is not clear a priori what may happen if one chooses another grid of unit squares to begin with (see Fig. 2(b)). Perhaps one may not have $\underline{m}(A) = \overline{m}(A)$ as before, or even if one does, the common value may not be the same as the $m(A)$ obtained previously! In any case, some proof will be needed to wash away such ambiguities. The usual way to resolve this difficulty is to adopt an axiomatic approach which we now set out to describe. We first extract the following four basic properties from the "area function" $m(A)$.

(A1) $m(A) \geqq 0$ for any measurable figure A.

(A2) $m(A)$ is additive, i.e. if A_1, A_2 are measurable figures with no common interior point, then $A_1 \cup A_2$ is measurable and $m(A_1 \cup A_2) = m(A_1) + m(A_2)$.

(A3) $m(A)$ is invariant under translation, i.e. if A' is obtained from A by translation, then $m(A) = m(A')$.

(A4) $m(A) = 1$ if A is a (fixed) unit square.

The beauty of this extraction lies in the fact (which requires a technical proof) that there exists uniquely a real-valued function m defined on the set of all measurable figures such that m satisfies (A1)–(A4). Hence, there is a unique grand scheme of assigning a quantitative measure to reflect our intuitive notion of area for every reasonably nice figure. The procedure outlined at the beginning of this paragraph is one constructive realisation of such a scheme. In other words, we can define the notion of area by (A1)–(A4), which are regarded as axioms.

In school we usually take for granted the formula for the area of a rectangle, viz. $A = ab$ where a, b are the sides. In the axiomatic approach, this fact becomes a theorem which requires a proof. Although we shall not give the proof, we would like to point out that (A1) comes into play in a crucial way. One reason why (A1) is so crucial in the proof, and indeed in the whole theory, is due to its role in the proof of a general principle which is the basis of the so-called (modified) *method of exhaustion*: suppose A is a measurable figure and $A_1, A_2, \ldots; B_1, B_2, \ldots$ are measurable figures with $A_n \subset A \subset B_n$ $(n = 1, 2, \ldots)$. If the sequence $\{m(B_n \backslash A_n)\}$ converges to zero, then the sequence $\{m(A_n)\}$ converges to $m(A)$. Now that we have explained the significance of (A1) we simply remark that the two axioms (A2), (A3) taken together form the basis of the methods of dissection and complementation.

We can define in a similar way the notion of volume $v(A)$ of a three-dimensional figure A by four similar axioms, prove in a similar fashion that there exists uniquely a function v defined on the set of all measurable figures such that v satisfies the axioms, and that a rectangular parallelepiped with sides a, b, c has volume abc. Suppose we assume the formula for the volume of a rectangular parallelepiped: can we find the volume of any polyhedron by elementary means? If we suspect the answer to the question is "no", we could start looking for two tetrahedra with congruent bases and equal altitudes which cannot be dissected into congruent pieces, and this is precisely what is asked for in Hilbert's Third Problem. Dehn's solution to this problem means that, for some strange reason, the axiom that $v(A) \geqq 0$ is needed not just in finding the volume of a rectangular parallelepiped, but once more in finding the volume of a polyhedron. This is quite unlike the role played by (A1) in a theory of area of polygons, where (A1) is used only once. In conclusion we can say that there is no way to avoid infinitesimal argument of some kind in treating the volume of polyhedra, even if we are willing to accept the formula for the volume of a rectangular parallelepiped on faith.

References

1. Y. Li and S. R. Du, *A concise history of mathematics in ancient China*. Zhonghua (1963). (in Chinese)
2. J. Needham, *Science and civilization in China*, Vol. 3, Chap. 19. Cambridge University Press (1959).
3. B. C. Qian, *A history of mathematics in China*. Science Press (1964). (in Chinese)
4. T. Kiang, An old Chinese way of finding the volume of a sphere, *Mathl Gaz.* **56**, 88–91 (May 1972).
5. D. Hilbert, Mathematical problems, *Bull. Am. math. Soc.* **8**, 437–479 (1902). (English translation by M. W. Newson.)

6. V. Boltianskii, *Hilbert's third problem*. Winston and Sons (1978). (English translation by R. A. Silverman.)
7. H. Eves, *A survey of geometry*. Allyn and Bacon (1972).

FUNG-KIT SIU
and MAN-KEUNG SIU

School of Education and Department of Mathematics, The University of Hong Kong