Mathematical Spectrum

A magazine for students and teachers of mathematics
in schools, colleges and universities,
and for everyone interested in mathematics

Volume 41 2008/2009 Number 1

- Sudoku Grids
- Henges, Heel Stones, and Analemmas
- Mathematics and Music

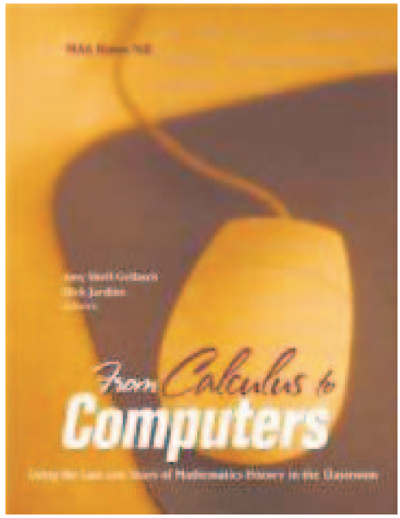
Mathematical Spectrum is a magazine for students and teachers in schools, colleges and universities, as well as the general reader interested in mathematics. It is published by the Applied Probability Trust, a non-profit-making organisation established in 1963 with the support of the London Mathematical Society. The object of the Trust is the encouragement of study and research in the mathematical sciences.

One volume of *Mathematical Spectrum* is published in each British academic year and consists of three issues, which appear in September, January and May.

Articles published in *Mathematical Spectrum* deal with the entire range of mathematical disciplines (pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, biomathematics). Both expository and historical material may be included, as well as elementary research and information on educational opportunities and careers in mathematics. There are also sections devoted to problems, to mathematics in the classroom and to computing.

From the Editor

Thinking Historically



Think how you learned Calculus, say. Perhaps by considering the slope of a chord to a graph and what happens to the extended chord as its two end-points tend to one another. Or by considering speed as the rate of change of distance. So you learned the Differential Calculus. You then considered the area under a curve and were introduced to the Integral Calculus. The two may even have been linked together in The Fundamental Theorem of Calculus. But where did these ideas come from? Did someone wake up one morning and there they were in his, or her, head? Or have they reached this point by a process of evolution? How did they begin, and who started it all? Experience is that such questions are seldom asked in the classroom. Maybe they come later when you have grasped the theory as it is today. After all, there is a process of 'tidying up' and simplifying, a sieving process whereby what is not of the essence gradually falls by the wayside. Try reading translations of Euler's pioneering account of the Bridges of Königsberg problem, for example, a forerunner of modern Topology and Graph Theory. Then look at how it is taught today.

But, if we never look at the history of a subject, do we lose an important sense of perspective? The authors of the volume *From Calculus to Computers: Using the Last 200 Years of Mathematics History in the Classroom* think so, and so, presumably, does the Mathematical Association of America, who published it. The contributors teach Mathematics or Computing in universities and colleges in America, so it is written in an American context. They present various case-studies showing how they have introduced historical work into their teaching. A useful source of ideas for university teachers.

The great Lagrange wrote in 1781 (quoted on page 93):

> It appears to me also that the mine [of mathematics] is already very deep and that unless one discovers new veins it will be necessary sooner or later to abandon it. Physics and chemistry now offer the most brilliant riches and easier exploitation; also our century's taste appears to be entirely in this direction and it is not impossible that the chairs of geometry in the Academy will one day become what the chairs of Arabic presently are in the universities.

It hasn't happened yet, as evidenced by the articles on modern Cryptography in this volume. Like Lagrange, we cannot see into the future, but we can at least be aware of our past.

'Is Mathematics discovered or invented?' Discuss!

Reference

1 Amy Shell-Gellasch and Richard Jardine (eds), *From Calculus to Computers: Using the Last 200 Years of Mathematics History in the Classroom* (Mathematical Association of America, Washington, DC, 2005).

Fundamental Transformations of Sudoku Grids

AVIV ADLER and ILAN ADLER

The Sudoku puzzle can be described as follows. Given a 9×9 grid with nine distinct 3×3 subgrids, we say that an assignment of the numerals $1, 2, \ldots, 9$ to the 81 cells of the grid is called a *valid pattern* if every row, column, and subgrid contains all nine numerals (see figure 1 for an example of a valid pattern).

In reference 1 Felgenhauer and Jarvis enumerated the number of valid patterns. In a follow-up article (reference 2), Russell and Jarvis restricted the enumeration to what they called *essentially different patterns*. For that purpose they introduced a list of several transformations of the cells of the grid such as reflections, rotations, and a few others that transform any valid pattern; they considered two patterns to be essentially the same if one can be obtained from the other by a sequence of these transformations. A natural question that can be asked is whether there exist other transformations besides those introduced in reference 2 that should be considered as leaving two patterns essentially the same. In this article we show that any conceivable transformation that leaves a valid pattern essentially the same can be constructed as a finite sequence of the transformations listed in reference 2, thus validating the completeness of the enumeration performed in reference 2 as the precise number of essentially different valid Sudoku patterns.

We denote a given grid by A, where A_{ij} denotes the cell in row i and column j (see figure 2). A particular valid pattern can then be presented by assigning the appropriate numerals to the cells. So, for example, the valid pattern in figure 1 is $A_{11} = 5$, $A_{12} = 2$, and so on.

Note that permuting the numerals (e.g. exchanging 1 and 2) in a valid pattern leaves it essentially the same. In order to eliminate the dependence on particular numeral choices, we say that a partition $S_1, S_2, \ldots, S_9$ of the 81 cells in a grid is *valid* if, for $i = 1, 2, \ldots, 9$,

1. each S_i contains nine of the cells in the grid,
2. every row, column, and subgrid of A contains one, and only one, element of S_i.

5	2	3	8	1	6	7	4	9
7	8	4	5	9	3	1	2	6
6	9	1	4	7	2	8	3	5
2	3	9	1	4	5	6	8	7
4	5	7	2	6	8	9	1	3
1	6	8	9	3	7	2	5	4
3	4	2	7	8	9	5	6	1
9	1	5	6	2	4	3	7	8
8	7	6	3	5	1	4	9	2

Figure 1

A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}	A_{17}	A_{18}	A_{19}
A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}	A_{27}	A_{28}	A_{29}
A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{36}	A_{37}	A_{38}	A_{39}
A_{41}	A_{42}	A_{43}	A_{44}	A_{45}	A_{46}	A_{47}	A_{48}	A_{49}
A_{51}	A_{52}	A_{53}	A_{54}	A_{55}	A_{56}	A_{57}	A_{58}	A_{59}
A_{61}	A_{62}	A_{63}	A_{64}	A_{65}	A_{66}	A_{67}	A_{68}	A_{69}
A_{71}	A_{72}	A_{73}	A_{74}	A_{75}	A_{76}	A_{77}	A_{78}	A_{79}
A_{81}	A_{82}	A_{83}	A_{84}	A_{85}	A_{86}	A_{87}	A_{88}	A_{89}
A_{91}	A_{92}	A_{93}	A_{94}	A_{95}	A_{96}	A_{97}	A_{98}	A_{99}

A

Figure 2

A_{63}	A_{43}	A_{53}	A_{73}	A_{93}	A_{83}	A_{33}	A_{23}	A_{13}
A_{61}	A_{41}	A_{51}	A_{71}	A_{91}	A_{81}	A_{31}	A_{21}	A_{11}
A_{62}	A_{42}	A_{52}	A_{72}	A_{92}	A_{82}	A_{32}	A_{22}	A_{12}
A_{68}	A_{48}	A_{58}	A_{78}	A_{98}	A_{88}	A_{38}	A_{28}	A_{18}
A_{67}	A_{47}	A_{57}	A_{77}	A_{97}	A_{87}	A_{37}	A_{27}	A_{17}
A_{69}	A_{49}	A_{59}	A_{79}	A_{99}	A_{89}	A_{39}	A_{29}	A_{19}
A_{64}	A_{44}	A_{54}	A_{74}	A_{94}	A_{84}	A_{34}	A_{24}	A_{14}
A_{65}	A_{45}	A_{55}	A_{75}	A_{95}	A_{85}	A_{35}	A_{25}	A_{15}
A_{66}	A_{46}	A_{56}	A_{76}	A_{96}	A_{86}	A_{36}	A_{26}	A_{16}

$\Phi_1(A)$

A_{11}	A_{15}	A_{14}	A_{13}	A_{12}	A_{16}	A_{19}	A_{18}	A_{17}
A_{28}	A_{23}	A_{26}	A_{27}	A_{24}	A_{21}	A_{29}	A_{22}	A_{25}
A_{33}	A_{36}	A_{32}	A_{38}	A_{34}	A_{39}	A_{37}	A_{31}	A_{35}
A_{47}	A_{46}	A_{42}	A_{45}	A_{48}	A_{44}	A_{43}	A_{41}	A_{49}
A_{58}	A_{51}	A_{55}	A_{54}	A_{57}	A_{59}	A_{52}	A_{53}	A_{56}
A_{68}	A_{69}	A_{67}	A_{61}	A_{64}	A_{63}	A_{66}	A_{62}	A_{65}
A_{77}	A_{78}	A_{72}	A_{73}	A_{76}	A_{79}	A_{74}	A_{75}	A_{71}
A_{86}	A_{84}	A_{83}	A_{81}	A_{87}	A_{88}	A_{82}	A_{89}	A_{85}
A_{94}	A_{98}	A_{91}	A_{95}	A_{92}	A_{96}	A_{97}	A_{93}	A_{99}

$\Phi_2(A)$

Figure 3

So, in figures 1 and 2, the partition of the grid corresponding to the given pattern is

$$S_1 = \{A_{11}, A_{24}, A_{39}, A_{46}, A_{52}, A_{68}, A_{77}, A_{83}, A_{95}\} \quad \text{(the component corresponding to 5)},$$
$$S_2 = \{A_{51}, A_{72}, A_{23}, A_{34}, A_{45}, A_{86}, A_{97}, A_{18}, A_{69}\} \quad \text{(the component corresponding to 4)},$$

and so on. (It should be noted that throughout this article the ordering of elements within a set is arbitrary. For example, $\{x, y, z\}$ is the same as $\{y, x, z\}$.)

Given a grid A, let Φ be a permutation of the cells in A. Thus, Φ is a one-to-one transformation of the cells of A to another 9×9 grid, $\Phi(A)$. That is, each of the cells A_{ij} in A is transformed into a cell in $\Phi(A)$ such that no two cells in A are transformed into the same cell in $\Phi(A)$. Figure 3 depicts two examples of such permutations.

Given a subset S of cells in A, we say that $\Phi(S)$ is the *image* of the cells in S under the transformation Φ. We call Φ a *fundamental transformation* if, for every valid partition $S_1, S_2, \ldots, S_9$ of A, $\Phi(S_1), \Phi(S_2), \ldots, \Phi(S_9)$ is a valid partition of $\Phi(A)$. Note that both

(a)

8	9	7	2	6	5	1	4	3
1	2	4	3	8	9	6	7	5
6	3	5	4	7	1	9	8	2
5	8	1	6	9	7	3	2	4
2	6	9	5	4	3	8	1	7
4	7	3	1	2	8	5	6	9
9	1	2	7	3	6	4	5	8
3	4	6	8	5	2	7	9	1
7	5	8	9	1	4	2	3	6

(b)

8	6	2	7	9	5	3	4	1
7	4	9	6	3	1	5	2	8
5	1	3	8	4	2	9	6	7
3	7	8	9	2	6	1	5	4
1	2	4	5	8	7	6	9	3
6	9	5	4	1	3	8	7	2
4	5	1	2	6	8	7	3	9
2	8	6	3	7	9	4	1	5
9	3	7	1	5	4	2	8	6

Figure 4 (a) Φ_1 applied to the Sudoku in figure 1, (b) Φ_2 applied to the Sudoku in figure 1.

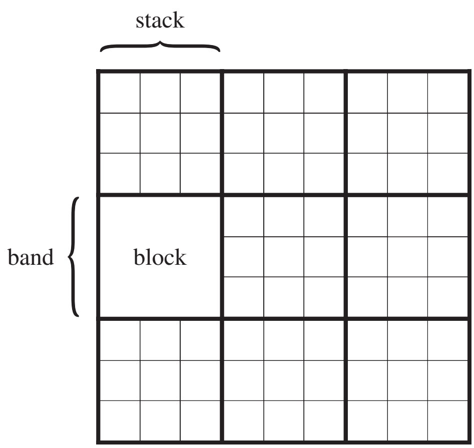


Figure 5

transformations in figure 3 when applied to the particular pattern in figure 1 result in valid partitions (see figure 4). However, as we will later demonstrate, while Φ_1 is a fundamental transformation, Φ_2 is not.

Following the terminology in reference 2, we say that a *block* refers to one of the 3×3 subgrids within the 9×9 grid, a *stack* consists of three blocks in a vertical 9×3 subgrid, and a *band* consists of three blocks in a horizontal 3×9 subgrid (see figure 5).

As discussed in reference 1, there are several well-known simple fundamental transformations that we shall call *elementary* here. Below is the list of the elementary transformations to be considered in this article. (A sixth transformation mentioned in reference 2, rotation, can be obtained by using reflection followed by permutations of stacks and columns.)

1. Permute the three stacks.

2. Permute the three bands.

3. Permute the three columns within a stack.

4. Permute the three rows within a band.

5. Reflection with respect to the main NW-to-SE diagonal (similar to transposing a matrix).

Clearly, any transformation constructed by applying sequentially a series of fundamental transformations is fundamental. We denote the transformation resulting from applying any transformation Φ followed by another transformation ($\bar{\Phi}$, say) as $\bar{\Phi}\Phi$ and the inverse transformation of Φ by Φ^{-1} (that is, $\Phi^{-1}\Phi(A) = A$). Our main result is that any fundamental transformation can be constructed by applying a finite sequence of elementary transformations. Specifically, we prove the following theorem.

Theorem 1 *Let Φ be a fundamental transformation. Then, there exists a sequence of elementary transformations $\Phi_1, \Phi_2, \ldots, \Phi_m$ such that $\Phi = \Phi_m\Phi_{m-1}\cdots\Phi_1$.*

Given a fundamental transformation Φ and a grid A, let $B = \Phi(A)$. We will show that it is always possible to identify a sequence of elementary transformations $\Psi_1, \Psi_2, \ldots, \Psi_m$ such that $\Psi_m\Psi_{m-1}\cdots\Psi_1(B) = A$. Since, as can be easily verified, for any of the five elementary transformations there exists an inverse elementary transformation of the same kind, by applying $\Psi_1^{-1}\Psi_2^{-1}\cdots\Psi_m^{-1}$ on both sides of the preceding equation we get $B = \Psi_1^{-1}\Psi_2^{-1}\cdots\Psi_m^{-1}(A)$, which will prove theorem 1. We shall first prove several preliminary results.

Two cells in a grid A are said to be *dependent within* A if they share a row, a column, or a block within A. Conversely, two cells in a grid A are said to be *independent within* A if they do not share a row, a column, or a block within A. Note that all the cells in any of the components of a valid partition are mutually independent. Thus, given a grid A and a nonfundamental transformation Φ, it means that there exist two independent cells (s and t, say) in A such that $\Phi(s)$ and $\Phi(t)$ are dependent within $\Phi(A)$. Next we prove the converse of the preceding statement.

Lemma 1 *Given a grid A and a transformation Φ, suppose that there exist two independent cells (s and t, say) in A for which $\Phi(s)$ and $\Phi(t)$ are dependent within $\Phi(A)$. Then Φ is nonfundamental.*

Lemma 1 follows directly from the following result.

Lemma 2 *If two cells are independent within A, then there exists a valid partition of A such that these two cells belong to the same component of the partition.*

We can prove lemma 2 by showing that for every valid pattern with two different numerals ('x' and 'y', say) in any two independent cells, we can obtain another valid pattern with numeral 'x' in both cells. Consider the grid in figure 1. Suppose that we want to construct another valid pattern in which the two independent cells A_{11} and A_{54} both are assigned the numeral 5. (In the example, $A_{11} = 5$ and $A_{54} = 2$.) The idea is to exchange the content of cell A_{54} with that of the cell in its block that contains 5 (A_{46} in the grid of figure 1). Now simply switch rows 4 and 5 and then switch columns 4 and 6, placing 5 in A_{54} as desired. The point is that the new pattern is valid since we used only elementary transformations. Thus, we obtain the following theorem.

Theorem 2 *Given a grid A, a transformation Φ is fundamental if and only if, for every pair s, t of independent cells within A, the pair $\Phi(s)$, $\Phi(t)$ is independent within $\Phi(A)$.*

Thus, by theorem 2, Φ_2 (as given in figure 3) is not fundamental since A_{11}, A_{47} is an independent pair of cells within A whereas $\Phi_2(A_{11})$, $\Phi_2(A_{47})$ is a dependent pair of cells within $\Phi_2(A)$ (since both cells appear in the fifth row of $\Phi_2(A)$). Next, we prove a key property of fundamental transformations.

Lemma 3 *Let A be a given grid and Φ a fundamental transformation.*

(a) *Either (i) or (ii) holds.*

(i) *For every row R (that is, the set of all the cells in the row) in A, $\Phi(R)$ is a row in $\Phi(A)$; for every column C in A, $\Phi(C)$ is a column in $\Phi(A)$.*

(ii) *For every row R in A, $\Phi(R)$ is a column in $\Phi(A)$; for every column C in A, $\Phi(C)$ is a row in $\Phi(A)$.*

(b) *If B is a block in A, then $\Phi(B)$ is a block in $\Phi(A)$.*

Proof A key observation in the proof is that a set of nine mutually dependent cells in a grid is necessarily a row, a column, or a block. Now, since the number of pairs of independent cells in a grid A is fixed, we have, by theorem 2, that a transformation Φ is fundamental if and only if every dependent pair of cells is transformed to a dependent pair of cells. Consequently, if S is a set of nine mutually dependent cells in a grid A and Φ is a fundamental transformation, then $\Phi(S)$ is necessarily a row, a column, or a block in $\Phi(A)$. Thus, given a row R and a column C in A and since Φ is a fundamental transformation, we have that each of $\Phi(R)$ and $\Phi(C)$ is a row, column, or block in $\Phi(A)$. However, the union of R and C contains 17 cells so the union of $\Phi(R)$ and $\Phi(C)$ also contains 17 cells. But the union of a block and a row (or a column) contains either 15 or 18 cells while the union of two distinct rows (or two distinct columns, or two distinct blocks) contains 18 cells. This proves part (a) of the lemma. Moreover, there are 27 distinct sets of nine mutually dependent sets in A (as well as in $\Phi(A)$), namely the rows, columns, and blocks, which given (a) proves (b).

An immediate consequence of lemma 3 is the following corollary.

Corollary 1 *Given a stack S and a band N in a grid A and Φ a fundamental transformation, the following holds.*

(i) *If lemma 3(a)(i) holds, then $\Phi(S)$ is a stack in $\Phi(A)$ and $\Phi(N)$ is a band in $\Phi(A)$.*

(ii) *If lemma 3(a)(ii) holds, then $\Phi(S)$ is a band in $\Phi(A)$ and $\Phi(N)$ is a stack in $\Phi(A)$.*

Finally, we are ready to prove theorem 1. As was stated earlier, we shall prove theorem 1 by showing that we can transform $B = \Phi(A)$ to A by a sequence of elementary transformations. We shall proceed by introducing several steps where each step is designed to transform a particular subgrid of B.

Proof of theorem 1 We split the proof into three steps.

1. According to lemma 3, all the rows in B are transformations of either rows or columns of A. If it is the latter, reflect the entries of B along the main NW-to-SE diagonal to obtain B_1.

2. Given (as established in step 1) that every row in A is a row in B_1, every column in A is a column in B_1, and by corollary 1, we get that for any band N in A, $\Phi(N)$ is a band in B_1, and for any stack S in A, $\Phi(S)$ is a stack in B_1. Now, permute the bands and stacks in B_1 to correspond to the bands and stacks in A to give B_2.

3. Finally within each band (or stack) of B_2, permute the rows (or columns) to correspond to those in A to give B_3.

Step 3 completes the proof, as now $B_3 = A$.

We conclude the article with two remarks.

Remark 1 In a recent paper (reference 3), Herzberg and Murty discussed the question regarding the number of valid patterns in any n-*Sudoku*, where the grid has $n^2 \times n^2$ cells with n^2 $(n \times n)$ blocks and the task is to fill up the grid with numerals $1, 2, \dots, n$ such that each numeral appears once and only once in each row, column, and block. In particular, they listed the elementary transformations as those leading to what they call *equivalent* Sudoku patterns. It can be easily confirmed that all the results presented here can be extended directly to any n-Sudoku, thus establishing that the definition of equivalency in reference 3 is justified as it covers all the transformations that preserve valid patterns.

Remark 2 Note that as a corollary to theorem 1, it is possible to generate all the fundamentally valid patterns from an arbitrary valid pattern by apply the six available permutations to each of the eight elementary transformations involving permutations and the two possibilities for the reflection, leading to $6^8 \times 2$ valid patterns ($6^8 \times 2 \times 9! = 1\,218\,998\,108\,160$ if valid patterns with the numerals permuted are considered). However, this number may exceed the actual number of essentially different valid patterns since it is conceivable that two distinct fundamental transformations, $\Phi_1(A)$ and $\Phi_2(A)$, transform a valid partition A to the same partition (that is, $\Phi_1(A) = \Phi_2(A)$). It is interesting to note that the number above is extremely close (within 0.005%) to 1 218 935 174 261, the actual number of valid patterns as computed in reference 2.

Acknowledgment This paper is dedicated to the memory of Professor David Gale, who introduced us to the mathematics of Sudoku and provided us with valuable comments and advice.

References

1 B. Felgenhauer and F. Jarvis, Mathematics of sudoku I, *Math. Spectrum* **39** (2006/2007), pp. 15–22.

2 E. Russell and F. Jarvis, Mathematics of sudoku II, *Math. Spectrum* **39** (2006/2007), pp. 54–58.

3 A. M. Herzberg and M. R. Murty, Sudoku squares and chromatic polynomia, *Notices Amer. Math. Soc.* **54** (2007), pp. 708–717.

Aviv Adler *and* ***Ilan Adler*** *is a son–father team who love puzzles, running, and biking. Ilan is a professor of Industrial Engineering and Operations Research at the Berkeley campus of the University of California. Aviv is a Junior at the College Preparatory High School in Oakland.*

Some Irrationals do not have Special Names

JOZEF DOBOŠ

The proof that $\sqrt{2}$ is irrational appears in textbooks as a standard introduction to proof by contradiction. Numbers like $\sqrt{2}$ fall into a special category of irrational numbers of the form $\sqrt[n]{a}$, where a is not a perfect nth power. These numbers can be seen as the solution to $x^n = a$. Other irrational numbers can be easily obtained as solutions to polynomial equations. For example, $\sqrt{2}+\sqrt{3}$ and $2\cos(\pi/7)$ are solutions to $x^4-10x^2+1=0$ and $x^3+x^2-2x-1=0$ respectively. The irrationality of these numbers follows from the rational root theorem (see reference 1). Yet more irrational numbers arise when you solve nonalgebraic equations. Usually they do not have special names. The aim of this article is to present an elementary approach to irrationality of roots of some charming equations of this type.

Example 1 (See reference 2.) Evidently, $x = 3$ is a solution to

$$3^x = x^3. \tag{1}$$

But there is another real number that satisfies this equation and is irrational.

It existence follows from Bolzano's theorem, which says that a continuous function cannot change its sign without going through the zero. (This result is also known as the mean value theorem.) Put

$$f(x) = 3^x - x^3.$$

Since $f(2) > 0$ while $f(\frac{5}{2}) < 0$, and f is a continuous function, there must be some number x between 2 and $\frac{5}{2}$ with the property $f(x) = 0$.

Since for each real number x we have $3^x > 0$, (1) is equivalent to

$$\frac{\ln x}{x} = \frac{\ln 3}{3}.$$

Since

$$\left(\frac{\ln x}{x}\right)' = \frac{1-\ln x}{x^2},$$

the function $\ln x/x$ is increasing on the interval $(0, \mathrm{e})$ and it is decreasing for $x > \mathrm{e}$.

Therefore, (1) has exactly one root in the interval $(0, \mathrm{e})$, denoted by z, say, and it has exactly one root larger than e, namely $x = 3$, as we know.

Now we show that z is irrational by contradiction. Assume that

$$z = \frac{p}{q},$$

where p and q are positive integers and p/q is in lowest terms. Substituting p/q for x into (1) gives

$$3^{p/q} = \left(\frac{p}{q}\right)^3,$$

$$3^p = \left(\frac{p}{q}\right)^{3q},$$

$$3^p q^{3q} = p^{3q}. \tag{2}$$

Hence, p^{3q} is divisible by 3. This means that p must be as well. Substituting $p = 3r$, where r is a positive integer, into (2) gives

$$3^{3r} q^{3q} = (3r)^{3q}$$

$$q^{3q} = 3^{3(q-r)} r^{3q}. \tag{3}$$

Since $z < 3$, we have $r < q$, which yields that the right-hand side of (3) is divisible by 3. So q^{3q} is divisible by 3. This means that q must be as well. But this also means that p and q are both divisible by 3, so p/q was not in lowest terms, which is a contradiction. The numerical value of z is $2.478\,052\,680\,288\ldots$.

Example 2 (See references 3 and 4.) It is clear that $x = \frac{1}{2}$ and $x = \frac{1}{4}$ are solutions to

$$a^{a^x} = x, \tag{4}$$

where $a = \frac{1}{16}$. But there is another real number that satisfies this equation and is irrational.

Since $a^x = x$ implies $a^{a^x} = x$, we start with the equation

$$a^x = x. \tag{5}$$

Put

$$f(x) = a^x - x.$$

Since $f(\frac{1}{4}) = \frac{1}{4} > 0$ while $f(\frac{1}{2}) = -\frac{1}{4} < 0$, and f is a continuous function, there must be some number x between $\frac{1}{4}$ and $\frac{1}{2}$ with the property $f(x) = 0$. The reason for this number being unique is that the function f is decreasing (as a sum of two decreasing functions). Therefore, (5) has exactly one root, denoted by z, say.

Now we show by contradiction that z is irrational. Assume that

$$z = \frac{p}{q},$$

where p and q are positive integers and p/q is in lowest terms. Substituting p/q for x into (5) gives

$$a^{p/q} = \frac{p}{q},$$

or

$$q^q = 2^{4p} p^q. \tag{6}$$

So q^q is even. This means that q must be even as well. Substituting $q = 2r$, where r is a positive integer, into (6) gives

$$\begin{aligned}(2r)^{2r} &= 2^{4p} p^{2r},\\ r^{2r} &= 2^{4p-2r} p^{2r}.\end{aligned} \tag{7}$$

Since $z > \frac{1}{4}$, we have $4p > 2r$, which yields that the right-hand side of (7) is even. So r^{2r} is even. This means that r must be even as well.

Substituting $r = 2s$, where s is a positive integer, into (7) gives

$$\begin{aligned}(2s)^{4s} &= 2^{4p-4s} p^{4s},\\ 2^{8s-4p} s^{4s} &= p^{4s}.\end{aligned} \tag{8}$$

Since $z < \frac{1}{2}$, we have $4p < 8s$, which yields that the left-hand side of (8) is even. So p^{4s} is even. This means that p must be even as well.

So p and q are both even which is a contradiction to our assumption that they have no common factors. The numerical value of z is $0.364\,249\,889\,784\ldots$.

Finally, we show that (4) has no other roots. Since for each real number x we have $a^x > 0$, (4) is equivalent to

$$f(x) = 0, \tag{9}$$

where

$$f(x) = 16^x \ln x + \ln 16.$$

Then

$$f'(x) = \frac{16^x}{x} g(x),$$

where

$$g(x) = x \ln x \ln 16 + 1.$$

It is not difficult to verify that the function $g(x)$ is decreasing on $(0, 1/\mathrm{e})$ and increasing for $x > 1/\mathrm{e}$. Since $g(\frac{1}{4}) = g(\frac{1}{2}) > 0$ and $g(1/e) < 0$, there are positive real numbers a and b such that $g(x) < 0$ for each $x \in (a, b)$ and $g(x) > 0$ for each $x \in (0, a) \cup (b, \infty)$. Therefore, the function $f(x)$ is decreasing on the interval $[a, b]$ and it is increasing on the intervals $(0, a]$ and $[b, \infty)$. This yields that (9) can have at most three roots.

References

1 http://www.math.ou.edu/~jalbert/putnam/putnam9.pdf
2 http://kvant.mccme.ru/1983/10/p31.htm
3 http://kvant.mccme.ru/1980/01/tri_tochki_tri_tochki_tri_toch.htm
4 http://kvant.mccme.ru/1990/05/ob_odnom_zamechatelnom_uravnen.htm

Jozef Doboš *received his PhD in Mathematics from the Comenius University in Bratislava, Slovakia, in 1986. He is currently a full professor at the P. J. Šafárik University in Košice, Slovakia. Prior to that he taught mathematics at the Technical University in Košice for more than 25 years. His research interests are in real functions theory and teaching mathematics.*

Henges, Heel Stones, and Analemmas

JOHN D. MAHONY

Introduction

There are various henges in existence around the world, some old – for example Stonehenge, UK – and some new – for example Stonehenge Aotearoa, New Zealand. An understanding of heel stones and their positions around a henge may be appreciated from figure 1. In figure 1, six heel stones are shown distributed around a henge centre to mark the apparent positions of sunrise/sunset (as seen from the centre of the henge) at the equinoxes and the solstices. The solstice positions mark the extremes of the Sun's apparent motion during a year and they are usually established by measurement with reference to distant star sets. The heel stone axis thus appears to swing through an angle $2\Psi_{max}$ during the course of one year. It is a simple matter to measure the magnitude of the swing but, since the phenomenon involves primarily only the Sun and the Earth, it should be possible to obtain a simple formula that will determine an approximate value for Ψ at any point of the Earth's orbit, not just at the solstices where it is a maximum.

Results from the formula should be in reasonable agreement with known facts.

1. At Stonehenge (UK, Northerly latitude circa $51.2°$) the apparent magnitude of the swing at the solstices is about $80°$ (see reference 1, pp. 109–110).

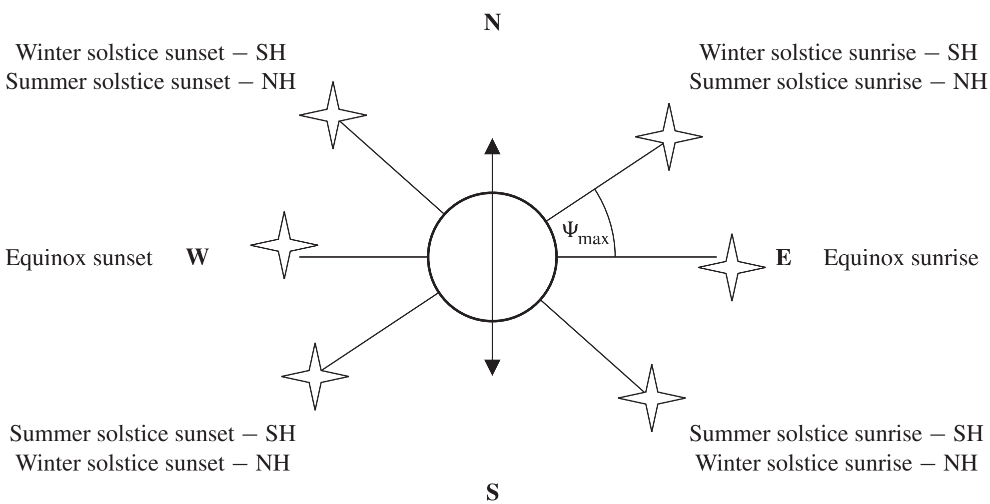


Figure 1 Henge showing six heel stones on Earth that mark the apparent position of sunrise/sunset at the solstices and equinoxes with reference to an east/west axis. The stars denote heel stones; 'NH' and 'SH' denote 'northern hemisphere' and 'southern hemisphere' respectively.

2. At Stonhenge Aotearoa (New Zealand, Southerly latitude circa $41.1°$) the apparent magnitude of the solstice swing appears to be about $63°$ (see reference 2). Photographs illustrating relevant points of interest for this henge are shown in figures 2 and 7.

In parallel with the question of heel stones on a henge, there is the wider issue concerning the *analemma*, which it is also appropriate to address, since the associated sums are but extensions of those proposed for the heel stone problem. Additionally, the interested reader might like to look at reference 3, and, for the practically minded, reference 4 makes for a rewarding read and shows how analemmatic dials and sundials may be constructed. These two henge related issues are discussed separately below.

Heel stones

To keep matters simple, it is assumed that the Earth is a perfect sphere, that its orbit around the Sun is circular, planar, and stable, and that the ratio of the Earth's radius to its distance to the Sun is small enough to be neglected in the formulation. The inclination of the Earth's polar axis to its orbital plane is required and this is about $23.43°$.

The problem may be addressed by considering a ray emanating from the Sun to strike a glancing (tangential) blow at a point on some latitude circle on the Earth. The latitude circle runs from west to east and so the problem of determining the angle of sunrise at a point on the Earth is that of determining the angle between the tangential ray from the Sun to the Earth point and a tangential ray along the latitude circle at that point. Thus, to address the problem further, it is necessary to establish a suitable geometry for the problem. This may be appreciated from figures 3 and 4. Figure 3 shows the Earth–Sun orbital geometry, and figure 4 shows the angles associated with a point at some latitude on the Earth, subject to certain limiting assumptions.

Figure 2 A glimpse of three heel stones can be seen through various windows formed by the uprights and the lintels of Stonehenge Aotearoa, New Zealand.

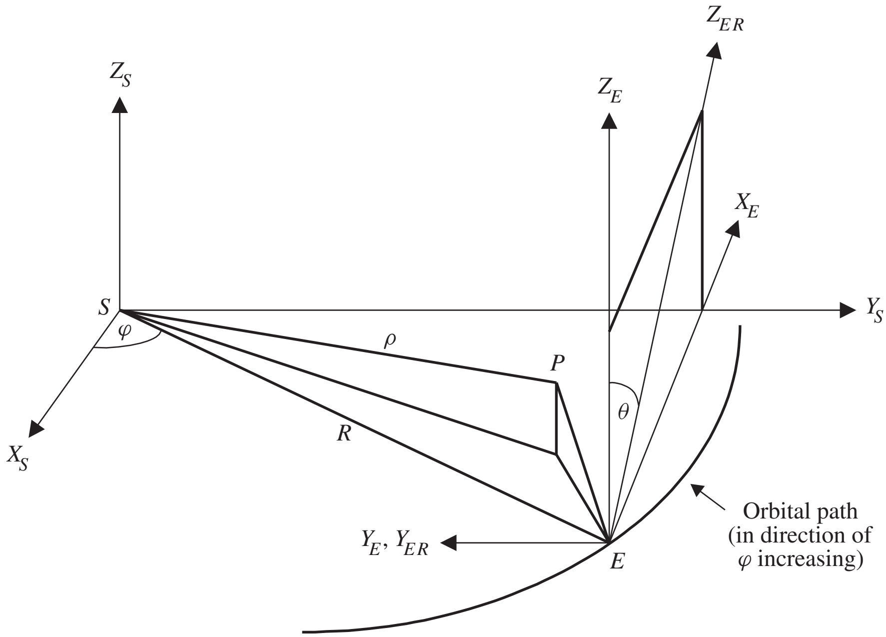


Figure 3 Earth–Sun orbital geometry.

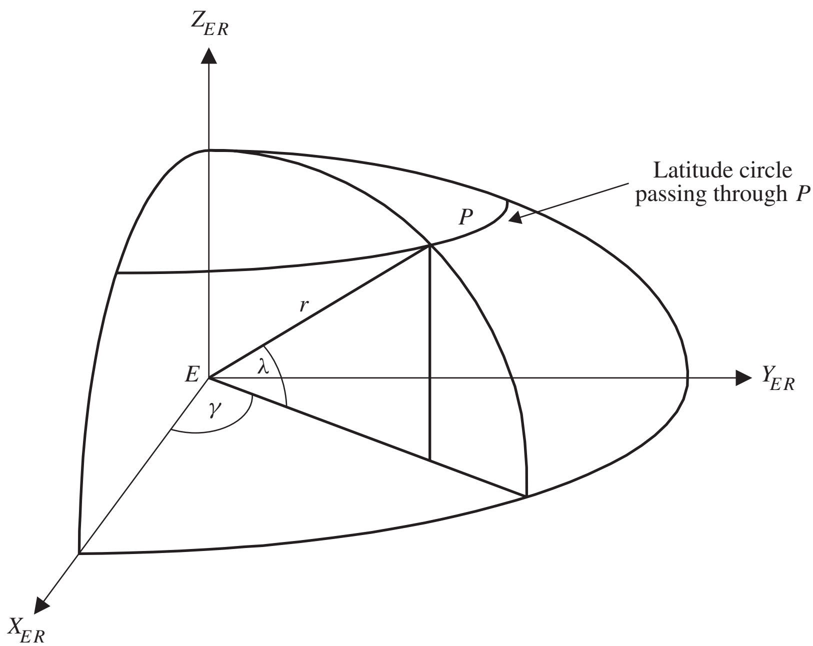


Figure 4 Angles associated with a point at some latitude on the Earth (showing only the first quadrant).

In particular, the following is assumed.

(i) The Earth is assumed to be spherical, of radius $r = 6378.15\,\text{km}$, and its orbit around the Sun is assumed to be planar and circular, of orbital radius $R = 149.6 \times 10^6\,\text{km}$ (note that the ratio $r/R = 0.000\,042\,634\,7$ is very small).

(ii) With reference to figure 3, the centre of the Sun lies at the point S, which is the origin of a fixed, right-handed, rectangular, Cartesian coordinate system (X_S, Y_S, Z_S). The Earth's centre, at the point E and at the distance R from the Sun's centre, subtends an angle φ as shown in this system. The orbital plane is the plane containing the X_S-axis and the Y_S-axis and, for the angle φ in this plane, φ is $0°$ at the first solstice, $90°$ at the first equinox, $180°$ at the second solstice, and $270°$ at the second equinox. The distance from the Sun to the point P is denoted by ρ.

(iii) Again, with reference to figure 3, the Earth's centre also lies at the common origin of two right-handed, rectangular Cartesian coordinate systems that move with the Earth as it orbits the Sun. The first of these is the (X_E, Y_E, Z_E) system, which is always parallel, as shown, to the (X_S, Y_S, Z_S) system. The second of these is the (X_{ER}, Y_{ER}, Z_{ER}) system, which is simply a rotation of the (X_E, Y_E, Z_E) system about the Y_E-axis through an angle θ ($\theta \approx 23.43°$) that represents the tilt of the Earth's axis of rotation with respect to the orbital plane. The $+Z_{ER}$-axis is thus the axis of rotation of the Earth in the sense of rotation of a right-handed screw directed along it.

(iv) The point P on the Earth subtends angles λ (latitude) and γ (azimuth) as shown in the coordinate system of figure 4.

(v) There is a comparatively small change in φ whilst the Earth makes one complete revolution and so it will be assumed that all points on the same latitude see sunrise and sunset at more or less the same value of φ.

The above assumptions are not unreasonable for the proposed calculation.

Having established a suitable geometry, we can in the first instance write down the coordinates of the point P in the (X_{ER}, Y_{ER}, Z_{ER}) system as

$$\begin{aligned} X_{ER}^P &= r\cos\lambda\cos\gamma, \\ Y_{ER}^P &= r\cos\lambda\sin\gamma, \\ Z_{ER}^P &= r\sin\lambda. \end{aligned} \tag{1}$$

Also, the coordinates of the Sun in the (X_E, Y_E, Z_E) system are

$$\begin{aligned} X_E^S &= R\cos\varphi, \\ Y_E^S &= R\sin\varphi, \\ Z_E^S &= 0. \end{aligned} \tag{2}$$

Thus, in view of the axis rotation it is possible to write down the coordinates of the Sun in the (X_{ER}, Y_{ER}, Z_{ER}) system, using the usual axis-rotation formulae, as

$$\begin{aligned} X_{ER}^S &= \cos\theta X_E^S - \sin\theta Z_E^S, \\ Y_{ER}^S &= Y_E^S, \\ Z_{ER}^S &= \sin\theta X_E^S + \cos\theta Z_E^S. \end{aligned} \tag{3}$$

From (2) and (3), it follows that

$$\begin{aligned} X_{ER}^S &= R\cos\theta\cos\varphi, \\ Y_{ER}^S &= R\sin\varphi, \\ Z_{ER}^S &= R\sin\theta\cos\varphi. \end{aligned} \tag{4}$$

Having established the Sun's coordinates (see (4)) and those of the point P (see (1)), both in the (X_{ER}, Y_{ER}, Z_{ER}) system, we can write down the vector components of SP and EP in this system as follows:

$$\begin{aligned} \boldsymbol{SP} &= [X_{ER}^P - X_{ER}^S, Y_{ER}^P - Y_{ER}^S, Z_{ER}^P - Z_{ER}^S], \\ \boldsymbol{EP} &= [X_{ER}^P, Y_{ER}^P, Z_{ER}^P]. \end{aligned}$$

The X, Y, and Z type coordinates in the above are as given by (1) and (4). A square bracket notation has been employed in the sense that $[a, b, c]$ represents a vector with components a, b, and c along the respective coordinate axes.

At a point of tangency on the Earth the vectors $\boldsymbol{SP}$ and $\boldsymbol{EP}$ are at right angles and so their dot product is zero, leading to the first result that

$$(X_{ER}^P)^2 + (Y_{ER}^P)^2 + (Z_{ER}^P)^2 = X_{ER}^P X_{ER}^S + Y_{ER}^P Y_{ER}^S + Z_{ER}^P Z_{ER}^S.$$

Since the left-hand side of this equation is simply the square of the radius of the Earth, the above may be written, after substitution from above for the terms on the right-hand side of the equation,

$$\frac{r}{R} = \cos\lambda\cos\gamma\cos\theta\cos\varphi + \cos\lambda\sin\gamma\sin\varphi + \sin\lambda\sin\theta\cos\varphi. \tag{5}$$

However, the ratio r/R on the left-hand side of (5) is small enough to be neglected (see (i) above) and so this result may be approximated as follows:

$$\cos\lambda\cos\varphi[\cos\gamma\cos\theta + \sin\gamma\tan\varphi + \tan\lambda\sin\theta] = 0. \tag{6}$$

This is a sunray to Earth tangency condition that determines the azimuth parameter γ when all other parameters are specified. This equation may be solved in a number of different ways. Quite simply, the usual half-angle formulae for $\tan\gamma$ may be employed to reduce the above to a quadratic in $\tan(\gamma/2)$; the solution to the ensuing quadratic may then be shown to be

$$\tan\frac{\gamma}{2} = \frac{-\tan\varphi \pm \sqrt{\tan^2\varphi - \tan^2\lambda\sin^2\theta + \cos^2\theta}}{\tan\lambda\sin\theta - \cos\theta}. \tag{7}$$

For a given latitude value λ and for a given Earth-orbital position value φ, this equation yields two values for γ, one for a tangent point at sunrise and the other for a tangent point at sunset.

The value obtained for γ from the above formulae may be employed in a result, obtained below, to determine the angle between the tangent ray (from the Sun to the point P) and the tangent to the west/east latitude at that point. (This is the angle of interest.) To determine this it is necessary to write down the direction cosines of the vector $\boldsymbol{SP}$ and those of a unit vector along the latitude line. In the usual notation, denote the former by (l_1, m_1, n_1) and the latter by

(l_2, m_2, n_2). Then, since r is small in relation to both ρ and R, and, since ρ/R is therefore close to unity, we have

$$\begin{aligned} l_1 &= \frac{X_{ER}^P - X_{ER}^S}{\rho} \\ &= \frac{r\cos\lambda\cos\gamma - R\cos\theta\cos\varphi}{\rho} \\ &\approx -\cos\theta\cos\varphi, \\ m_1 &= \frac{Y_{ER}^P - Y_{ER}^S}{\rho} \\ &= \frac{r\cos\lambda\sin\gamma - R\sin\varphi}{\rho} \\ &\approx -\sin\varphi, \\ n_1 &= \frac{Z_{ER}^P - Z_{ER}^S}{\rho} \\ &= \frac{r\sin\lambda - R\sin\theta\cos\varphi}{\rho} \\ &\approx -\sin\theta\cos\varphi. \end{aligned}$$

For the latter, direction cosines of the tangent along the latitude line at the point of interest are

$$\begin{aligned} l_2 &= -\sin\gamma, \\ m_2 &= \cos\gamma, \\ n_2 &= 0. \end{aligned}$$

Thus, the angle (Ψ, say) between the two lines is determined in the usual way from $\cos\Psi = l_1 l_2 + m_1 m_2 + n_1 n_2$, where the direction cosines are as given in the above equation sets. Thus,

$$\cos\Psi = \cos\theta\cos\varphi\sin\gamma - \sin\varphi\cos\gamma. \tag{8}$$

Therefore, for specified inputs of θ (the tilt angle of the Earth's axis), φ (the orbital angle measured from a solstice position), and λ (the latitude of interest), the problem is now essentially one of determining first the value for γ from (7), say, to be used in (8) for the determination of Ψ.

Alternatively, it is also possible to eliminate γ completely from (6) and (8) to determine Ψ directly in terms of λ, θ, and φ. Specifically,

$$\sin\Psi = \frac{\sin\theta\cos\varphi}{\cos\lambda}. \tag{9}$$

However, this approach does not allow for an estimation of daylight hours, since specific values for γ are required for that purpose.

Discussion

For a specified latitude angle λ, the equations of interest above ((7) and (8)) may be programmed on a spreadsheet to show sunrise angles for various φ angles covering 360° (i.e. one year of

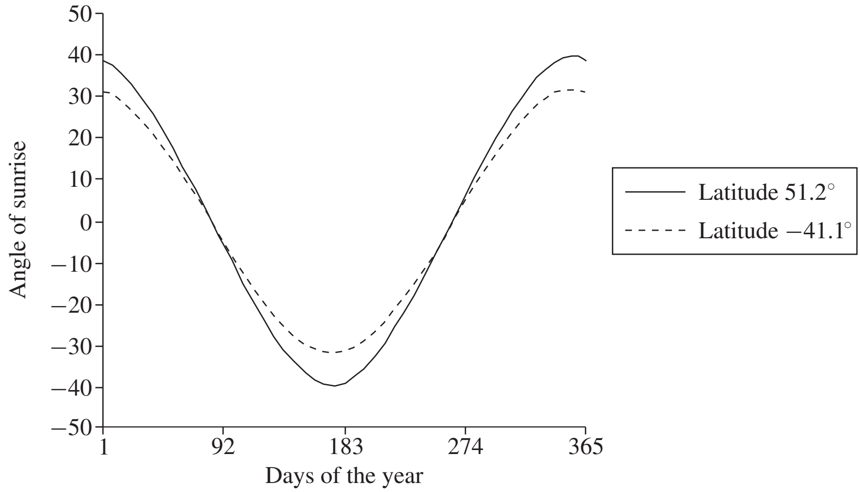


Figure 5

365 days), starting on day 1 (1st January) which is nine days after the winter solstice. The results of such an exercise are shown in figure 5 for two different latitude angles.

The first latitude angle is that for Stonehenge (circa $51.2°$) in the northern hemisphere and it can be seen from figure 5 that apparent sunrise swings through an angle whose magnitude is close to $80°$, from the summer solstice (circa day 173) to the winter solstice (circa day 355). This result would appear to be in agreement with that simply quoted in reference 1. Care has to be taken to ensure that the correct branches of the roots for γ are chosen on passing through the equinoxes – this will ensure a smoothly continuous curve in figure 5.

The second latitude angle is that for the recently constructed Stonehenge Aotearoa in the southern hemisphere. It shows an apparent sunrise swing between the solstices of about $63.5°$ and this would appear to be in reasonable agreement with what can be gleaned from reference 2. In reality, the positioning of the solstice heel stones at this henge was obtained by measurement, with reference to distant star-sets.

At the solstices, the situation simplifies. For example, to determine the magnitude of the swing between the solstices it will suffice to perform the calculations in the above when $\varphi = 0$ to obtain, from (9), values for Ψ, and hence 2Ψ. Specifically, Ψ is then given by

$$\Psi = \arcsin \frac{\sin\theta}{\cos\lambda}.$$

Using this formula in the case of the two latitudes considered above, it is a simple matter to verify the magnitude, 2Ψ, of the swing of the heel stone axis. At the northerly latitude of $51.2°$ the swing is $78.8°$ and at the southerly latitude of $41.1°$ the swing is $63.7°$, both results being in reasonable agreement with known facts.

Daylight hours

It is possible to determine, for a given latitude, the approximate length of daylight hours by first obtaining the difference between the γ-values of (7) in degrees, normalising by $360°$, and

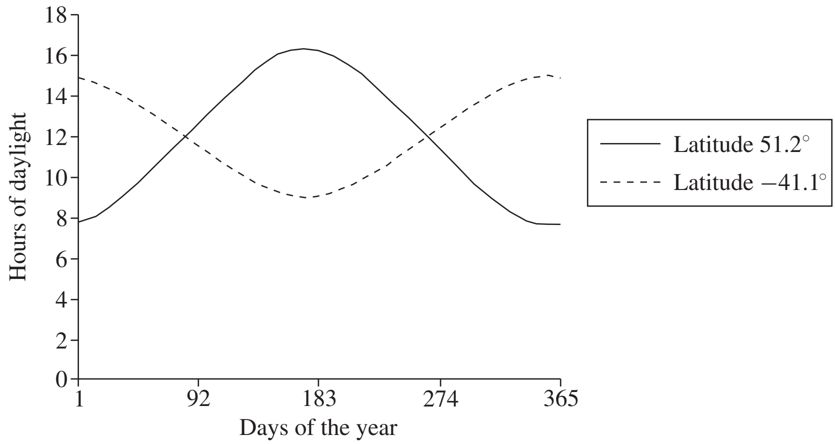


Figure 6

then multiplying by a factor of 24 hours; depending on the orbital position, it might also be necessary to subtract the answer from 24 hours. In the case of the examples given above it can be shown, again using spreadsheet facilities, that approximate daylight hours at those latitudes are as shown in figure 6.

From the data supporting figure 6 it could be seen that, for the Stonehenge, UK, latitude, the length of daylight on the shortest day (circa day 355) is about 7 hours 39 minutes, and this compares not unfavourably with known data. For example, in the year 2005 the expected shortest day length at the London latitude (not dissimilar to the latitude at Stonehenge) is about 7 hours 50 minutes. Conversely, the longest day-length at the London latitude in 2005 is about 16 hours 43 minutes, and the data supporting figure 6 show that this occurred at circa day 173 and was of a magnitude of about 16 hours 21 minutes, again making for a favourable comparison, given the initial assumptions. Not surprisingly, the story at the specified latitude for Stonehenge Aotearoa shows that the excursion of daylight hours across the year is smaller there than that at Stonehenge, UK, because the southerly latitude points for this example are closer to the equator.

For daylight hours at the solstices only (put $\varphi = 0$, in (7)), the calculations simplify. For example, it can be shown that γ is then readily secured from

$$\cos\gamma = -\tan\theta\tan\lambda.$$

Root-pairs of interest in this equation are $\pm\gamma$ and $\pm(\pi - \gamma)$. Accordingly, the daylight quantities of interest are $[2\gamma/2\pi] * 24$ hours and $[(2\pi - 2\gamma)/2\pi] * 24$ hours for the shortest and longest days respectively. Using input data, say, for the Stonehenge, UK, latitude, these reduce to 7 hours 39 minutes and 16 hours 21 minutes respectively, as cited above.

The analemma

A detailed technical discussion on the sums associated with the analemma is beyond the scope of the present text. However, it is appropriate to dwell briefly on the topic, since the analemma falls within the lexicon of sundials and henges, as follows. When the Sun illuminates an object

Figure 7 The obelisk near the centre of Stonehenge Aotearoa, New Zealand, serves to cast a noon shadow that provides the analemma shape on the ground.

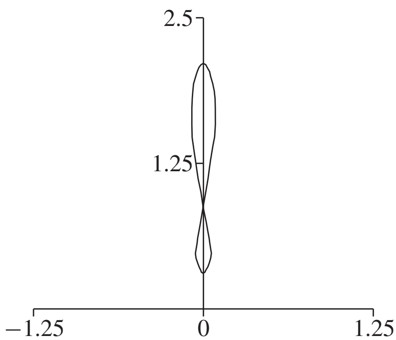


Figure 8 The predicted analemma shape at Stonehenge Aotearoa, New Zealand.

on the Earth, such as a pole (or obelisk) stuck vertically in the ground, a shadow is cast and a mark on the ground can be made at the tip of the shadow point at noon local mean time (LMT), rather than at noon local apparent time (LAT) which is when the Sun lies in the local meridian and when the shadow will be at its shortest length. Using a regular clock, such marks can be recorded every 24 hours thereafter throughout the year and the subsequently accrued locus of such shadow points is called the *analemma*. It appears roughly as an elongated hourglass or figure eight and its extent is latitude dependent. A typical, predicted shape relevant to the latitude at Stonehenge Aotearoa for an obelisk of unit height is shown in figure 8.

In figure 8, the major lobe points towards the south pole and effects due only to the 'obliquity of the ecliptic' (the tilt of the Earth's axis) have been taken into account in the calculations; there are other 'eccentricity' effects due to the offset position of the Sun within an otherwise circular orbit that have not been taken into consideration. For completeness, there is one final concept worthy of mention that is inextricably linked to that of the analemma. This is 'the equation of time'. The interested reader can find more about this equation in the references cited and it will suffice to simply reiterate here that it is the difference between LAT and LMT, or for the more practically minded, the amount by which a regular clock and a sundial differ at any instant.

Conclusion

The motion of the Earth around the Sun has been examined briefly to determine quantities of interest in matters pertaining to henges, heel stones, and analemmas. Simple formulae have been derived from which it is possible to determine, for example, from any latitude point at the solstices, the number of daylight hours and the relative position of the Sun above/below an east–west axis on the Earth and the shape that is to be expected for an analemma at any given latitude.

References

1 F. Hoyle, *On Stonehenge* (Heinemann, London, 1977).
2 *Stonhenge Aotearoa: The Complete Guide* (2005).
3 http://www.analemma.com.
4 A. E. Waugh, *Sundials: Their Theory and Construction* (Dover, New York, 1973).

***John D. Mahony** has worked in academia (teaching and research) and industry (antennas in aerospace and defence) since gaining a PhD from the University of Surrey in 1973. He has published articles of a mathematical persuasion in various engineering journals and magazines and is now planning a relocation to live in New Zealand and continue his interest in mathematical matters. He is a CMath (Chartered Mathematician), an FIMA (Fellow of the Institute of Mathematics and its Applications), and a CSc (Chartered Scientist).*

Summing divisors

The sum of the positive divisors of 276 is 672, i.e. 276 reversed. Are there any other numbers with this property?

57 Conduit Road, Stamford,
Lincolnshire PE9 1QL, UK

Alastair Summers

Circles, Chords, and Difference Equations

STUART SIMONS

Consider a set of circles $C_1, C_2, \ldots, C_n, \ldots$ whose centres are collinear with the circles touching (see figure 1) and let their radii be $a_1, a_2, \ldots, a_n, \ldots$ respectively. Let P_1Q be a straight line making an angle θ with the line of centres and suppose that it intersects all the circles, defining chords P_1P_1', P_2P_2', $\ldots$, P_nP_n', $\ldots$. If these chords all have the same length (i.e. $P_1P_1' = P_2P_2' = \cdots = P_nP_n' = \cdots$), then what will the value of a_n/a_1 be in terms of n and θ?

Without loss of generality, we take $a_1 = 1$ and our task is then to calculate a_n in terms of θ for all integer values of n ($n \geq 2$). We let R_n be the mid-point of the chord P_nP_n' and begin by considering the case of $n = 2$. Since $P_1R_1 = R_2P_2$, we have $\cos^2\theta = a_2^2(2 + a_2)^2 \sin^2\theta$ (as $a_1 = 1$); thus, we readily obtain

$$a_2 = 1 + 4\tan^2\theta. \tag{1}$$

For general values of n, the fact that R_nP_n' is independent of n tells us that the expression

$$a_n^2 - (2 + 2a_2 + \cdots + 2a_{n-1} + a_n)^2 \sin^2\theta \tag{2}$$

is independent of n, and equating this to $\cos^2\theta$ then yields a difference equation which essentially defines $a_3, a_4, \ldots$ when n takes successively the values $3, 4, \ldots$. However, solving this equation to yield an explicit solution for a_n is more difficult than solving the difference equations usually encountered in elementary work. This is because such elementary difference equations are generally linear with a fixed number of terms, while our difference equation is nonlinear and additionally the number of terms it contains increases as n increases. We therefore begin our treatment by showing how our difference equation can be reduced to a

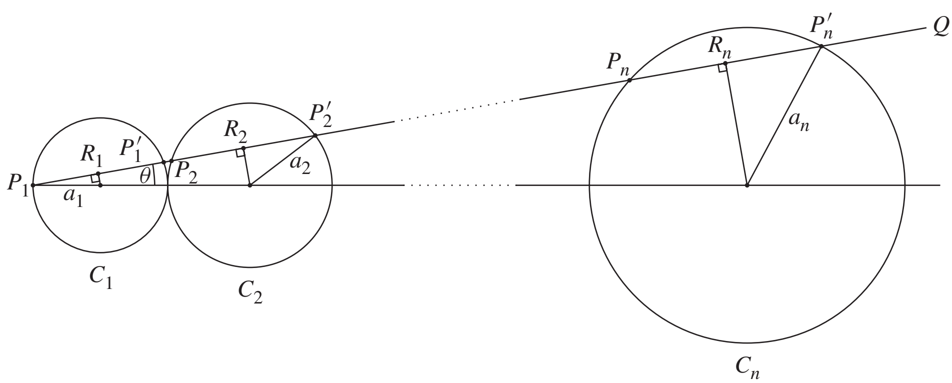


Figure 1

linear form. We do this by equating (2) to the same expression with n replaced by $n+1$. This equality then yields

$$\begin{aligned} a_{n+1}^2 - a_n^2 &= \sin^2\theta[(2a_1 + 2a_2 + \cdots + 2a_n + a_{n+1})^2 - (2a_1 + 2a_2 + \cdots + 2a_{n-1} + a_n)^2] \\ &= \sin^2\theta[(4a_1 + 4a_2 + \cdots + 4a_{n-1} + 3a_n + a_{n+1})(a_{n+1} + a_n)], \end{aligned}$$

and dividing both sides by the common factor $a_{n+1} + a_n$ ($\neq 0$) then gives the linear relationship

$$a_{n+2} - a_n = \sin^2\theta(4a_1 + 4a_2 + \cdots + 4a_{n-1} + 3a_n + a_{n+1}). \tag{3}$$

We now reduce the number of terms appearing on the right-hand side of (3) by subtracting (3) from the equation obtained from it when n is replaced by $n+1$. This yields

$$a_{n+2} - 2a_{n+1} + a_n = \sin^2\theta(a_n - 2a_{n+1} + a_{n+2}),$$

leading to

$$a_{n+2} - 2(\sec^2\theta + \tan^2\theta)a_{n+1} + a_n = 0. \tag{4}$$

We have thus succeeded in reducing our nonlinear difference equation to a second-order linear equation which determines $a_3, a_4, \ldots$ uniquely when a_1 and a_2 are specified. The general solution to (4) is now found by the standard procedure of assuming a solution of the form $a_n = x^n$. On substituting this into (4) we obtain a quadratic equation for x with solutions

$$x_1 = (\sec\theta + \tan\theta)^2 \quad \text{and} \quad x_2 = (\sec\theta - \tan\theta)^2.$$

These in turn yield the general solution to (4) in the form

$$a_n = A(\sec\theta + \tan\theta)^{2n-2} + B(\sec\theta - \tan\theta)^{2n-2}, \tag{5}$$

where A and B are arbitrary constants which can be determined from the 'boundary conditions' $a_1 = 1$ and $a_2 = 1 + 4\tan^2\theta$ (see (1)). Thus,

$$A + B = 1, \qquad A(\sec\theta + \tan\theta)^2 + B(\sec\theta - \tan\theta)^2 = \sec^2\theta + 3\tan^2\theta,$$

with solutions

$$A = \tfrac{1}{2}\cos\theta(\sec\theta + \tan\theta) \quad \text{and} \quad B = \tfrac{1}{2}\cos\theta(\sec\theta - \tan\theta).$$

On substituting these values back into the general solution (5), we thus finally obtain

$$a_n = \tfrac{1}{2}\cos\theta[(\sec\theta + \tan\theta)^{2n-1} + (\sec\theta - \tan\theta)^{2n-1}], \tag{6}$$

with $a_n \approx \frac{1}{2}\cos\theta(\sec\theta + \tan\theta)^{2n-1}$ for $n >> 1$. As an illustration of (6) we take $\theta = 30°$, when $a_n = \frac{1}{4}[3^n + (1/3^{n-1})]$, giving $a_1 = 1, a_2 = \frac{7}{3}, a_3 = \frac{61}{9}, a_4 = \frac{547}{27}, \ldots$.

Stuart Simons *was Reader in Applied Mathematics at the University of London before his retirement, with main interests in transport theory and the mathematical theory of aerosols. His current interests include developing novel approaches to relatively elementary problems.*

Triangles and Parallelograms of Equal Area in an Ellipse

ROBERT BUONPASTORE and THOMAS J. OSLER

Introduction

In reference 1, Euler looked at certain properties of the conic sections and tried to find other curves that shared these properties. In the eighteenth century, mathematicians were familiar with many properties of the parabola, ellipse, and hyperbola that have been neglected in our modern education. This article is about one such ignored property of the ellipse which we rediscovered in order to understand Euler's work. We will study an interesting family of parallelograms inscribed in the ellipse, all of which have the same area.

We begin by defining a few new terms, *diameters*, *reciprocal diameters,* and *reciprocal points* in an ellipse.

A *diameter* of an ellipse is any chord that passes through the center. In figure 1 MM' and mm' are diameters. Now, start with any diameter MM'. We say that the diameter mm' is *reciprocal* to the diameter MM' if it is parallel to the tangent line to the ellipse at M. If we started with diameter mm', then MM' would be the reciprocal diameter. We say that the points m and m' are *reciprocal* to the point M. Euler assumed that his readers were familiar with reciprocal diameters and points. He also assumed that his readers would be aware that the area of the parallelogram $MmM'm'$ is constant, regardless of the choice of the initial diameter, and is equal to $2ab$. The areas of the triangles CMm and CMm' are also constant and equal to $ab/2$.

Before we derive our main result, we review parametric equations for the ellipse and their geometric consequences.

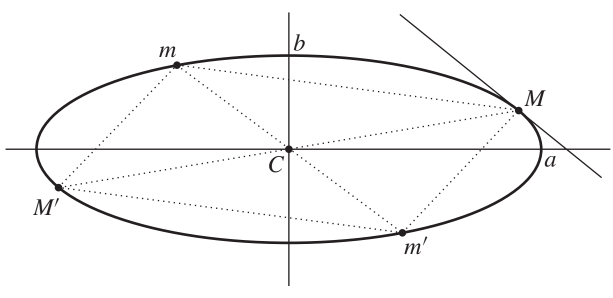


Figure 1

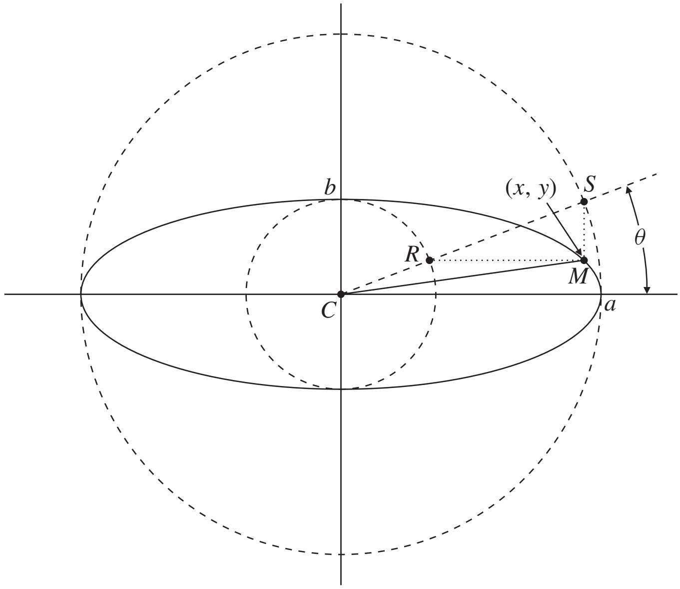


Figure 2

Parametric equations for the ellipse

These reciprocal diameters have an interesting relation to the parametric form of the equation of the ellipse given by

$$x = a\cos\theta \quad \text{and} \quad y = b\sin\theta. \tag{1}$$

In figure 2 we see two circles centered at point C with radii b and a. The ray CRS makes an angle θ with the x-axis and intersects the smaller circle at R and the larger circle at S. From R extend a horizontal line and from S drop a vertical line. These two lines intersect at the point M. This point M is on the ellipse given by the parametric equations (1). As the angle θ varies between 0 and 2π, the point M generates the entire ellipse. Notice that the ray CM which identifies the point (x, y) on the perimeter of the ellipse differs from the ray CS that is made by the parameter θ. The angle θ is known historically as the *eccentric anomaly.* It is important to note that this is not the usual polar angle associated with the point (x, y).

Reciprocal diameters and the eccentric anomaly

To see the relation between the reciprocal diameters and the eccentric anomaly, consider figure 3. Start with the radius CR making eccentric anomaly θ to identify the point on the ellipse M. Now increase the eccentric anomaly by $\pi/2$ to identify the radius ray CS and corresponding point m on the ellipse. We will show that this point m is reciprocal to M. Thus, reciprocal points on the ellipse have their related eccentric anomalies separated by the angle $\pi/2$.

To see that this is true, we use (1) to calculate the slope of the tangent at M. We get

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b\cos\theta\,\mathrm{d}\theta}{-a\sin\theta\,\mathrm{d}\theta} = -\frac{b}{a}\cot\theta.$$

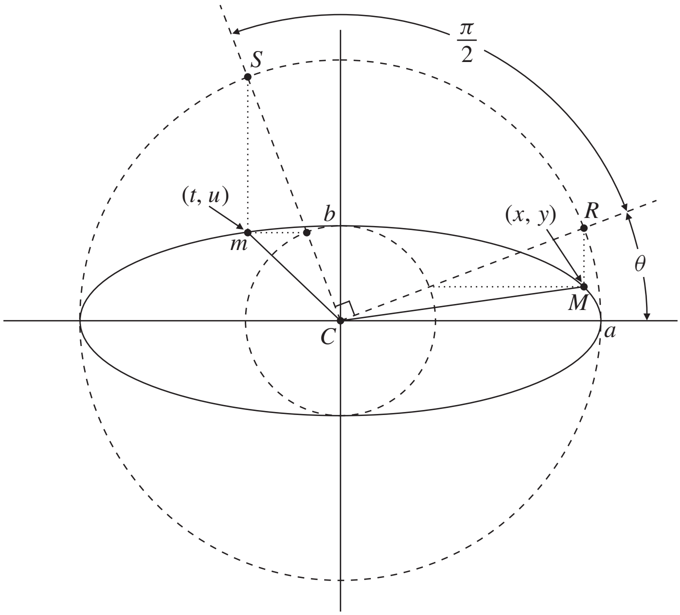


Figure 3

Therefore, the slope of the reciprocal ray Cm is given by $-(b/a)\cot\theta$. (Notice that when the slope of the ray CR defined by the eccentric anomaly is $\tan\theta$, then the slope of the corresponding ray CM defined by the point on the ellipse is always given by $(b/a)\tan\theta$.) Thus, the slope of the ray CS is given by $-\cot\theta$, but the identity $-\cot\theta = \tan(\theta+\pi/2)$ demonstrates the truth of the relation between M and m just stated. Thus, the coordinates (t, u) of the point m reciprocal to M are given by

$$t = a\cos\left(\theta + \frac{\pi}{2}\right) = -a\sin\theta \quad \text{and} \quad u = b\sin\left(\theta + \frac{\pi}{2}\right) = b\cos\theta. \tag{2}$$

The area of the triangle *CMm* is constant

The area of the quadrilateral $ABMm$ shown in figure 4 is $((y+u)/2)(x-t)$. Subtracting the areas of triangles ACm and CBM we get the area of the triangle CMm:

$$\text{Area }\Delta CMm = \frac{y+u}{2}(x-t) - \frac{(-t)u}{2} - \frac{xy}{2}.$$

This simplifies to $(xu - ty)/2$. Substituting the values of these variables in terms of θ from (1) and (2), we obtain

$$\text{Area }\Delta CMm = \frac{ab\cos^2\theta + ab\sin^2\theta}{2} = \frac{ab}{2}.$$

This proves that the area of the triangle CMm is constant. In the same way we can show that the area of triangle CMm' is $ab/2$, and thus the area of the parallelogram $MmM'm'$ is $2ab$.

This completes our study of the triangles and parallelograms of equal area in the ellipse.

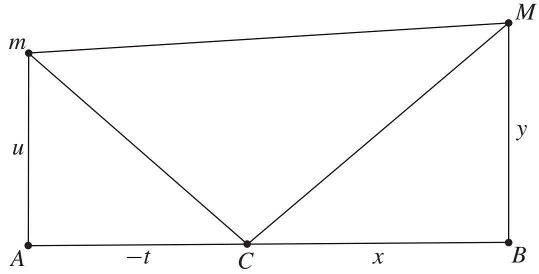


Figure 4

Reference

1 E. Greve and T. J. Osler, Translation with notes of Euler's paper E83: On some properties of conic sections that are shared with infinitely many other curved lines. Available at http://www.math.dartmouth.edu/∼euler/.

Robert Buonpastore *is graduating from Rowan University with dual degrees in physics and mathematics. His research thus far includes an assortment of projects in number theory and various theoretical projects in quantum optics and general relativity. In his free time, Rob enjoys camping, concerts, and the performing arts, namely theatre.*

Tom Osler *is professor of mathematics at Rowan University and is the author of over 100 mathematical papers. In addition to teaching university mathematics for the past 46 years, Tom has a passion for long distance running. He has been competing for the past 53 consecutive years. Included in his over 2000 races are wins in three national championships in the late 1960s at distances from 25 kilometers to 50 miles. He is the author of two running books and is in the Road Runners Club of the American Hall of Fame.*

Mathematics for mice

A $3 \times 3 \times 3$ cube of cheese is cut into 27 $1 \times 1 \times 1$ cubes. A mouse eats one exterior cube on the first day and on each succeeding day it eats a cube adjacent to the one just eaten. Is it possible to have only the central cube left on the last day?

Department of Mathematics and Statistics
San Diego State University, USA

Farshid Arjomandi

Counting the Perfect Powers

M. A. NYBLOM

1. Introduction

A number of the form a^b, where a and b are positive integers with $b > 1$, is called a *perfect power*. The first few terms of the sequence of perfect powers arranged in ascending order are

$$1,\ 4,\ 8,\ 9,\ 16,\ 25,\ 27,\ 32,\ 36,\ 49, \ldots,$$

and is listed in the On-Line Encyclopedia of Integer Sequences under A001597 (see reference 1). It is interesting to note that the subject of perfect powers arose recently in connection with the settling of the famous Catalan conjecture (see reference 2) which states that the only perfect powers in the above sequence that differ by one are the integers 8 and 9. In this article, we return to the theme of perfect powers by examining the following question. Suppose that, for a positive real number x, we define $N(x)$ as the number of perfect powers less than or equal to x; can an expression for $N(x)$ then be found in terms of x, so allowing the calculation of $N(x)$ to occur without having to find and count all the perfect powers less than or equal to x? This question will be answered in the affirmative by showing that if $p_1, p_2, \ldots, p_m$ are the prime numbers less than or equal to $\lfloor \log_2(x) \rfloor$, then

$$N(x) = \sum_{k=1}^{m} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le m} \lfloor x^{(p_{i_1} \cdots p_{i_k})^{-1}} \rfloor, \tag{1}$$

where the expression $1 \le i_1 < \cdots < i_k \le m$ indicates that the sum is taken over all ordered k-element subsets $\{i_1, \ldots, i_k\}$ of the set $\{1, 2, \ldots, m\}$ and $\lfloor \cdot \rfloor$ denotes the integer-part function. For example, consider $x = 16 = 2^4$. As 2 and 3 are the only prime numbers less than or equal to $\lfloor \log_2(16) \rfloor = 4$ then (1) evaluates to $N(16) = \lfloor \sqrt{16} \rfloor + \lfloor \sqrt[3]{16} \rfloor - \lfloor \sqrt[6]{16} \rfloor = 4 + 2 - 1 = 5$, which from the above sequence we can verify is correct. As will be seen, (1) will follow from an application of the inclusion–exclusion principle, which is an elementary counting technique used in determining the number of elements in a finite union of sets, not necessarily all disjoint. In spite of the somewhat convoluted appearance of the expression in (1), we shall demonstrate its practicability in computing $N(x)$ for a large value of x by providing a numerical example in Section 4. We begin with a technical result concerning the intersection of certain sets which will be needed later in proving our main result.

2. A preliminary result

In determining $N(x)$, we must first know how many perfect powers of a certain fixed exponent are less than or equal to x, and in addition how many perfect powers less than or equal to x can be simultaneously expressed as a power of an integer with two or more distinct exponents. The first of these questions will be answered in the course of the proof of our main result, but to help answer the second question we shall need to consider the family of sets $A_n(x) = \{k^n : k \in \mathbb{N}, k^n \le x\}$, where $n \in \mathbb{N} \setminus \{1\}$, and establish the following result.

Lemma 1 *For any set consisting of $m \geq 2$ positive integers $\{n_1, \ldots, n_m\}$ all greater than unity, we have the set equality*

$$\bigcap_{i=1}^{m} A_{n_i}(x) = A_{[n_1,\ldots,n_m]}(x), \tag{2}$$

where $[n_1, \ldots, n_m]$ denotes the least common multiple of the m integers $n_1, \ldots, n_m$.

It is of interest to note that the following proof is an example of an inductive argument in which the base step, namely $m = 2$, is more difficult to establish in comparison with the inductive step.

Proof of Lemma 1 We begin by demonstrating that $A_n(x) \cap A_m(x) = A_{[n,m]}(x)$ for any $n, m \in \mathbb{N} \setminus \{1\}$, which is the base step of our inductive argument. Now, since $n \mid [n, m]$ and $m \mid [n, m]$, any number of the form $k^{[n,m]}$, where $k \in \mathbb{N}$, can be rewritten as a perfect power having an exponent n and as a perfect power having exponent m. Thus, $A_{[n,m]}(x) \subseteq A_n(x) \cap A_m(x)$. Let $s \in A_n(x) \cap A_m(x)$ with $s \neq 1$. Then $s = k_1^n = k_2^m$, for some $k_1, k_2 \in \mathbb{N} \setminus \{1\}$. We have to produce a $k \in \mathbb{N} \setminus \{1\}$ such that $s = k^{[n,m]}$. As $k_1^n = k_2^m$, both k_1 and k_2 must have the same prime divisors. Writing $k_1 = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$ and $k_2 = p_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r}$, we deduce from the equality $k_1^n = k_2^m$ that $n\alpha_i = m\beta_i$ for each $i = 1, 2, \ldots, r$. Consequently, $n \mid n\alpha_i$ and $m \mid n\alpha_i$ and so $n\alpha_i = [n, m]\gamma_i$ for some $\gamma_i \in \mathbb{N}$. Thus, $s = k^{[n,m]}$ where $k = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_r^{\gamma_r}$, which establishes that $A_n(x) \cap A_m(x) \subseteq A_{[n,m]}(x)$.

Now suppose that, for $m > 2$, the set identity in (2) holds for an arbitrary set of $m - 1$ positive integers $\{n_1, \ldots, n_{m-1}\}$ all greater than unity. Then, as $[[n_1, n_2, \ldots, n_{m-1}], n_m] = [n_1, \ldots, n_m]$, we observe from the inductive assumption and the base step that

$$\begin{aligned} \bigcap_{i=1}^{m} A_{n_i}(x) &= \left(\bigcap_{i=1}^{m-1} A_{n_i}(x) \right) \cap A_{n_m}(x) \\ &= A_{[n_1,\ldots,n_{m-1}]}(x) \cap A_{n_m}(x) \\ &= A_{[[n_1,\ldots,n_{m-1}],n_m]}(x) \\ &= A_{[n_1,\ldots,n_m]}(x). \end{aligned}$$

Hence, (2) holds for m arbitrary positive integers greater than unity and so the result is established by the principle of mathematical induction.

3. Main result

Before establishing the expression for $N(x)$ in (1) we shall, for the benefit of the reader, give a brief description of the inclusion–exclusion principle, which will be used shortly in proving our main result.

For a given finite collection of sets $A_1, A_2, \ldots, A_n$, all of which are mutually disjoint (i.e. $A_i \cap A_j = \varnothing$ for $i \neq j$), the total number of elements in $A = \cup_{i=1}^{n} A_i$ is clearly given by $|A_1| + |A_2| + \cdots + |A_n|$, where $|A_i|$ denotes the number of elements in A_i. If, on the other hand, some of the sets in the collection are not disjoint, then the previous formula will count at least twice those elements contained in more than one of the sets in the above collection. In this instance, to obtain the correct number of elements in $A = \cup_{i=1}^{n} A_i$ the inclusion–exclusion principle states that the exact number of elements in A is given by

$$\sum_{k=1}^{n}(-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le n} |A_{i_1} \cap \cdots \cap A_{i_k}|, \tag{3}$$

where the expression $1 \le i_1 < \cdots < i_k \le n$ indicates that the sum is taken over all the k-element subsets $\{i_1, \ldots, i_k\}$ of the set $\{1, 2, \ldots, n\}$. We can view the alternating sign in (3) as reflecting the process of including and excluding those elements belonging to more than one of the sets A_i, thereby counting each element in A exactly once. We now apply this principle to the sets of Section 2 to formulate the desired expression for $N(x)$.

Theorem 1 *If $x \ge 4$ and $p_1, p_2, \ldots, p_m$ denote the prime numbers less than or equal to $\lfloor \log_2(x) \rfloor$, then the number of perfect powers less than or equal to x is given by*

$$N(x) = \sum_{k=1}^{m}(-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le m} \lfloor x^{(p_{i_1}\cdots p_{i_k})^{-1}} \rfloor.$$

Proof Recalling for each $n \in \mathbb{N}\setminus\{1\}$ that $A_n(x) = \{k^n : k \in \mathbb{N}, k^n \le x\}$, observe since $x > 1$ that 1 is always an element of $A_n(x)$ but, for n sufficiently large, $A_n(x) \setminus \{1\} = \varnothing$. Thus, by defining the auxiliary function $M(x) = \max\{n \in \mathbb{N} \setminus \{1\} \colon A_n(x) \setminus \{1\} \ne \varnothing\}$ we clearly see that $M(x) \ge 2$ for $x \ge 4$, and that $N(x)$ must equal the number of distinct elements in the set $A = \cup_{n=2}^{M(x)} A_n(x)$. Furthermore, we observe from the inequality $2^{\lfloor \log_2 x \rfloor} \le x < 2^{\lfloor \log_2 x \rfloor + 1}$ that $M(x) = \lfloor \log_2 x \rfloor$. As the sets $A_n(x)$ are not necessarily all disjoint, we will need to apply the inclusion–exclusion principle to the set A in order to derive the expression in (1) for $N(x)$. However, before doing this it will be necessary to show that if $p_1, \ldots, p_m$ are the prime numbers less than or equal to $\lfloor \log_2 x \rfloor$, then in fact $A = B$ where

$$B = \bigcup_{r=1}^{m} A_{p_r}(x).$$

The inclusion $B \subseteq A$ follows automatically by definition as each set $A_{p_k}(x)$ is included in the union of sets which form A. To establish the reverse inclusion $A \subseteq B$, first observe that, as $p_1, \ldots, p_m$ represents the complete list of prime numbers less than or equal to $\lfloor \log_2 x \rfloor$, every integer $n \in \{2, 3, \ldots, \lfloor \log_2 x \rfloor\}$ must be divisible by at least one of these prime numbers since otherwise, by the fundamental theorem of arithmetic, n would be divisible by a prime $p' > \lfloor \log_2 x \rfloor$ and so $n > \lfloor \log_2 x \rfloor$, i.e. a contradiction. Consequently, if given any $s \in A_n(x)$, then $s = k^n$ and we may write $n = p_r \gamma$ for some $r \in \{1, 2, \ldots, m\}$ and $\gamma \in \mathbb{N}$. Thus, $s = (k^\gamma)^{p_r} \in A_{p_r}(x)$, and so every element of A is contained in the set B.

Now, applying the inclusion–exclusion principle to the set $A = B$, we deduce from (3) that

$$N(x) = |B| = \sum_{k=1}^{m}(-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le m} |A_{p_{i_1}}(x) \cap \cdots \cap A_{p_{i_k}}(x)|. \tag{4}$$

As the least common multiple of the k prime numbers $p_{i_1}, \ldots, p_{i_k}$ is clearly the product $(p_{i_1} \cdots p_{i_k})$, we observe from Lemma 1 that $|A_{p_{i_1}}(x) \cap \cdots \cap A_{p_{i_k}}(x)| = |A_{(p_{i_1}\cdots p_{i_k})}(x)|$. Thus, to complete the argument all that is required is to determine the number of elements of a typical set $A_n(x)$. As $A_n(x) \neq \varnothing$, there must exist a largest integer $m \geq 1$ such that $m^n \leq x < (m+1)^n$. By taking the nth root through the previous inequality we deduce that $m \leq \sqrt[n]{x} < m+1$ (i.e. $m = \lfloor \sqrt[n]{x} \rfloor$), and so $A_n(x)$ must contain $\lfloor \sqrt[n]{x} \rfloor$ elements. Hence, upon substituting $\lfloor x^{(p_{i_1} \cdots p_{i_k})^{-1}} \rfloor$ for $|A_{(p_{i_1} \cdots p_{i_k})}(x)|$ into (4), we obtain the required expression in (1).

4. Numerical example

We now examine how the explicit expression for $N(x)$ in (1) can be practically implemented to compute the number of perfect powers less than or equal to a given large positive real x. For notational convenience, let the inner summation of (1) be denoted by

$$S_k(x) = \sum_{1 \leq i_1 < \cdots < i_k \leq m} \lfloor x^{(p_{i_1} \cdots p_{i_k})^{-1}} \rfloor.$$

Observe that in order to evaluate each $S_k(x)$, we must sum the terms $\lfloor x^{(p_{i_1} \cdots p_{i_k})^{-1}} \rfloor$ over those subscripts $i_1 < \cdots < i_k$ whose values are chosen from the ordered k-element subsets of $\{1, 2, \ldots, m\}$; consequently, the number of summands is ${}^m\mathrm{C}_k$. Thus, on first acquaintance, it would appear that the calculation of $S_k(x)$ would involve having to determine, for each $1 \leq k \leq m$, all ${}^m\mathrm{C}_k$ combinations of prime numbers from the set $\{p_1, \ldots, p_m\}$. However, for sufficiently large x this may not be necessary since, for certain values of k, we can show that $S_k(x) = {}^m\mathrm{C}_k$ as follows.

To begin, consider for any $x > 2$ the arithmetic function

$$k(x) = \min\{k \in \mathbb{N} \colon p_1 p_2 \cdots p_k > x\},$$

where again p_i denotes the ith prime number. We wish to first show that if there are m prime numbers less than or equal to $\lfloor \log_2 x \rfloor$, then $k(\lfloor \log_2 x \rfloor)$ will be at most $m-2$ when $m > 5$. Recalling that, for any $n \geq 2$, there exists a prime number strictly between n and $2n$ (Bertrand's postulate), we observe, as each $p_i \geq 2$, that

$$p_{m-5} p_{m-4} (p_{m-3} p_{m-2}) > p_{m-5} (p_{m-4} p_{m-1}) > p_{m-5} p_m > p_{m+1} > \lfloor \log_2 x \rfloor.$$

Thus, when $m > 5$ we have $p_1 \cdots p_{m-2} > \lfloor \log_2 x \rfloor$, and so $k(\lfloor \log_2 x \rfloor) \leq m-2$. Now, for $\lfloor \log_2 x \rfloor > p_5 = 11$ and $k \geq k(\lfloor \log_2 x \rfloor)$, we note that in the summation $S_k(x)$ for all ${}^m\mathrm{C}_k$ combinations of products $p_{i_1} \cdots p_{i_k} \geq p_1 \cdots p_{k(\lfloor \log_2 x \rfloor)} > \lfloor \log_2 x \rfloor$. Consequently, from the inequality $2^{\lfloor \log_2 x \rfloor} \leq x < 2^{\lfloor \log_2 x \rfloor + 1}$ it is immediate that

$$1 < 2^{\lfloor \log_2 x \rfloor (p_{i_1} \cdots p_{i_k})^{-1}} \leq x^{(p_{i_1} \cdots p_{i_k})^{-1}} < 2^{(\lfloor \log_2 x \rfloor + 1)(p_{i_1} \cdots p_{i_k})^{-1}} \leq 2.$$

Thus, $\lfloor x^{(p_{i_1} \cdots p_{i_k})^{-1}} \rfloor = 1$, and so the summation $S_k(x)$ must consist of adding ${}^m\mathrm{C}_k$ terms all of which are identically 1, i.e. $S_k(x) = {}^m\mathrm{C}_k$. Hence, for $x > 2^{p_5} = 2^{11}$, the number of perfect powers less than or equal to x can be calculated by the alternate expression

$$N(x) = \sum_{k=1}^{k(\lfloor \log_2 x \rfloor) - 1} (-1)^{k+1} S_k(x) + \sum_{k=k(\lfloor \log_2 x \rfloor)}^{m} (-1)^{k+1}\, {}^m\mathrm{C}_k. \tag{5}$$

For $x > 2^{11}$, the value of the arithmetic function $k(\lfloor \log_2 x \rfloor)$ will in practice be much smaller than the number of prime numbers less than or equal to $\lfloor \log_2 x \rfloor$; consequently, in calculating $N(x)$, we shall only have to evaluate $S_k(x)$ for the few values of $1 \le k < k(\lfloor \log_2 x \rfloor)$. In what follows, the reader may wish to consult a table of perfect powers less than or equal to 10^9 (see reference 3).

Example 1 Consider $x = 2^{18} = 262\,144$. From reference 3 we can, by inspection, deduce that $N(x) = 583$. To demonstrate the use of (1) we shall apply the alternate expression (5) to verify that the number of perfect powers less than or equal to x is 583. Now, $\lfloor \log_2 x \rfloor = 18$ and so there are $m = 7$ prime numbers, less than $\lfloor \log_2 x \rfloor$, namely $2, 3, 5, 7, 11, 13, 17$. As $2 \cdot 3 \cdot 5 > 18 > 2 \cdot 3$, we have that $k(\lfloor \log_2 x \rfloor) = 3$. So, from (5), we obtain

$$N(x) = S_1(x) - S_2(x) + \sum_{k=3}^{7} (-1)^{k+1}\, {}^7\mathrm{C}_k. \tag{6}$$

Using a calculator we find in this instance that

$$\begin{aligned} S_1(x) &= \lfloor \sqrt{2^{18}} \rfloor + \lfloor \sqrt[3]{2^{18}} \rfloor + \lfloor \sqrt[5]{2^{18}} \rfloor + \lfloor \sqrt[7]{2^{18}} \rfloor + \lfloor \sqrt[11]{2^{18}} \rfloor + \lfloor \sqrt[13]{2^{18}} \rfloor + \lfloor \sqrt[17]{2^{18}} \rfloor \\ &= 512 + 64 + 12 + 5 + 3 + 2 + 2 \\ &= 600. \end{aligned}$$

To evaluate $S_2(x)$, first recall from the definition that

$$S_2(x) = \sum_{1 \le i_1 < i_2 \le 7} \lfloor 2^{18(p_{i_1} p_{i_2})^{-1}} \rfloor.$$

Now, if $p_{i_1} p_{i_2} > 18$ then $\lfloor 2^{18(p_{i_1} p_{i_2})^{-1}} \rfloor = 1$. However, of the ${}^7\mathrm{C}_2 = 21$ combinations of products $p_{i_1} p_{i_2}$ with $1 \le i_1 < i_2 \le 7$, the only products less than 18 are $2 \cdot 3$, $2 \cdot 5$, $2 \cdot 7$, and $3 \cdot 5$. Thus, the summation $S_2(x)$ will consist of adding $21 - 4 = 17$ terms all of which are identically 1, together with the sum of the terms $\lfloor \sqrt[6]{2^{18}} \rfloor$, $\lfloor \sqrt[10]{2^{18}} \rfloor$, $\lfloor \sqrt[14]{2^{18}} \rfloor$, and $\lfloor \sqrt[15]{2^{18}} \rfloor$, which are 8, 3, 2, and 2 respectively. Consequently, $S_2(x) = 17 + 8 + 3 + 2 + 2 = 32$ and so finally adding in the alternating sum of binomial coefficients in (6) yields

$$N(x) = 600 - 32 + 35 - 35 + 21 - 7 + 1 = 583,$$

as required.

References

1 http://www.research.att.com/∼njas/sequences/

2 P. Mihailescu, A class number free criterion for Catalan's conjecture, *J. Number Theory* **99** (2003), pp. 225–231.

3 http://www.ug.bcc.bilkent.edu.tr/∼sevki/powers_up_to_1e9.html

Michael Nyblom *is a lecturer of mathematics at RMIT University in Melbourne, Australia. His research interests include Combinatorics, Elementary Number Theory, and Mathematical Analysis. When not doing research he likes to solve the various problems that appear in publications such as Mathematical Spectrum.*

Curious Properties of the Circumcircle and Incircle of an Equilateral Triangle

PRITHWIJIT DE

In this article we consider a problem in plane geometry involving a plane figure and integral distances. Such problems are well known in mathematical literature and the reader may find copious examples in references 1 and 2. The plane figure that we will work with is an equilateral triangle ABC with integer sides of length s. Before we focus on the central theme of the article we will state and prove an interesting result that relates the distances of a point P in the plane of the triangle from its vertices A, B, and C respectively with the side-length s.

Theorem 1 *Let ABC be an equilateral triangle with $|AB| = s$ and let P be any point in the plane of the triangle such that $|PA| = p$, $|PB| = q$, and $|PC| = r$. Then we obtain*

$$3(p^4 + q^4 + r^4 + s^4) = (p^2 + q^2 + r^2 + s^2)^2. \tag{1}$$

Proof To prove theorem 1 we may assume, without loss of generality, that the point P is located between the sides AB and AC. Also, it does not matter if P is inside or outside ABC. Let $\angle BAP = \alpha$ so that $\angle PAC = 60^\circ - \alpha$. The cosine rule on the triangles PBA and PCA yields

$$\cos\alpha = \frac{s^2 + p^2 - q^2}{2sp} \quad \text{and} \quad \cos(60^\circ - \alpha) = \frac{s^2 + p^2 - r^2}{2sp}$$

respectively. Expanding $\cos(60^\circ - \alpha)$ and substituting the expression for $\cos\alpha$ into it we get

$$\sin\alpha = \frac{s^2 + p^2 + q^2 - 2r^2}{2\sqrt{3}sp}.$$

Now, using the familiar trigonometric identity $\cos^2\alpha + \sin^2\alpha = 1$ and the expressions for $\cos\alpha$ and $\sin\alpha$ obtained above, after simplification we obtain

$$p^4 + q^4 + r^4 + s^4 = p^2q^2 + p^2r^2 + p^2s^2 + q^2r^2 + q^2s^2 + r^2s^2. \tag{2}$$

Multiplying both sides by 2, adding $p^4 + q^4 + r^4 + s^4$ to both sides, and simplifying yields

$$3(p^4 + q^4 + r^4 + s^4) = (p^2 + q^2 + r^2 + s^2)^2.$$

We shall now see how the result proved in theorem 1 can be used to prove two interesting properties concerning the circumcircle and incircle of ABC.

Theorem 2 *If P lies on the circumcircle of ABC then (1) has infinitely many solutions in the positive integers.*

To prove this result we need the following theorem due to Van Schooten.

Theorem 3 (Van Schooten) *Let ABC be an equilateral triangle and P be a point on the minor arc BC of the circumcircle of the triangle, then $|PA| = |PB| + |PC|$.*

Proof Observe that $ABPC$ is a cyclic quadrilateral with PA and BC as the diagonals. From Ptolemy's theorem we obtain

$$|PA|\,|BC| = |PB|\,|CA| + |PC|\,|AB|.$$

Now, $|AB| = |AC| = |BC|$. Hence, $|PA| = |PB| + |PC|$, as required.

Now we are in a position to prove theorem 2.

Proof of theorem 2 Note that (1) can be expressed as (2) and can be written as a quadratic in s^2 as follows:

$$s^4 - (p^2 + q^2 + r^2)s^2 + (p^4 + q^4 + r^4 - p^2q^2 - p^2r^2 - q^2r^2) = 0.$$

This can be solved to give

$$s^2 = \tfrac{1}{2}[(p^2 + q^2 + r^2) \pm \sqrt{3(p+q+r)(-p+q+r)(p-q+r)(p+q-r)}].$$

By theorem 3, $p = q + r$. Hence,

$$p^2 + q^2 + r^2 = 2s^2. \tag{3}$$

By substituting $p = q + r$ into (3) and simplifying, we get

$$q^2 + qr + r^2 = s^2. \tag{4}$$

An important observation in this case is that the integers q, r, and s represent the side-lengths of a triangle in which the size of the angle opposite s is $120°$. The positive integral solutions to (4) are well known (see reference 1) or can be determined by employing the factorization technique as follows:

$$(q, r, s) = (m^2 - n^2, 2mn + n^2, m^2 + mn + n^2)$$

(where m and n are positive integers). Hence, we have

$$(p, q, r, s) = (m^2 + 2mn, m^2 - n^2, 2mn + n^2, m^2 + mn + n^2),$$

and this establishes the claim that there are infinitely many solutions in the positive integers of (1) when the point P lies on the circumcircle.

Having observed this property we may be inclined to look for a similar property of the incircle of ABC. Thus, the question that we are now going to address is the following.

> Does (1) have infinitely many solutions in the positive integers when P is restricted to lie on the incircle of triangle ABC?

To answer this question we need another result, apart from (1), from plane geometry which relates p, q, r, and s when P lies on the incircle of ABC.

Theorem 4 *If P is a point on the incircle of an equilateral triangle ABC then*

$$4(p^2 + q^2 + r^2) = 5s^2, \tag{5}$$

where $|PA| = p$, $|PB| = q$, $|PC| = r$, and $|AB| = s$.

Proof Probably the easiest way to prove theorem 4 is by using coordinate geometry. Without loss of generality we may choose the coordinate axes so that the coordinates of B and C are $(-s/2, 0)$ and $(s/2, 0)$ respectively. Hence, the coordinates of A are $(0, s\sqrt{3}/2)$. The incentre, I, lies on the line joining A and the mid-point of BC and is one third of the distance away from BC towards A. Therefore, its coordinates are $(0, s/2\sqrt{3})$. Thus, the equation of the incircle is

$$x^2 + \left(y - \frac{s}{2\sqrt{3}}\right)^2 = \frac{s^2}{12}. \tag{6}$$

If the coordinates of P are (h, k) then

$$p^2 + q^2 + r^2 = 3\left(h^2 + k^2 - \frac{ks}{\sqrt{3}}\right) + \frac{5s^2}{4}. \tag{7}$$

The reader may verify that (6) and (7) together yield the desired result.

Now we are ready to answer the question that we raised above. Observe that (5) forces s to be even. Replacing s by $2t$ in (1) and (5), and using (5) in (1), leads us to

$$p^4 + q^4 + r^4 = 11t^4. \tag{8}$$

Thus, the problem boils down to finding positive integral solutions to (8). In order to do so we use some number theory. In general it is not easy to determine all integral solutions to an equation, but often it is easy to show that such an equation does not have any integral solutions. This can be done using congruence relations with respect to an appropriate base. The rationale for such a line of argument is simple in that, if an equation has a solution at all in positive integers, then both sides of the equation must leave the same remainder when divided by any positive integer. The task of choosing a base requires some experience, cleverness, and luck.

Suppose for argument sake that (8) has a solution in the positive integers. Let (p, q, r, t) be a solution to (8) such that t is minimal. Now observe that

$$t^4 \equiv \begin{cases} 0 \pmod{16} & \text{if } t \text{ is even,} \\ 1 \pmod{16} & \text{if } t \text{ is odd.} \end{cases}$$

To help the understanding of the reader we provide a brief sketch of the proof of the congruence relation when t is odd. Let $t = 2k + 1$. Then we have

$$t^4 - 1 = (2k+1)^4 - 1 = 8k(k+1)(2k^2 + 2k + 1)$$

and we observe that $k(k+1)$ is even for any positive integer k. If t is odd then, modulo 16, the right-hand side of (8) is 11 and the left-hand side is at most 3. Therefore, t cannot be odd and has to be even. This implies that $16 \mid (p^4 + q^4 + r^4)$, and consequently forces p, q, and r to be even too. Thus, we see that $(p/2, q/2, r/2, t/2)$ satisfies (8), contradicting the minimality of t. Hence, (8) does not have any solution in positive integers.

To summarize, in this article we have presented interesting properties of the circumcircle and incircle of an equilateral triangle with integer side length. On one hand we have shown that we may find infinitely many points on the circumcircle of the triangle whose distances from the vertices of the triangle are integers and on the other hand we have proved that no such point exists on the incircle of the triangle. An interesting exercise would be to find other subsets of the plane of the triangle which produce infinitely many integral solutions to (1). We also note in passing that the quadruple $(2, 4, 5, 6)$ is a solution to (5) but not to (1). Note that $p = 2$, $q = 4$, and $s = 6$ give $p + q = s$, implying that the point P is located on side AB of the triangle. By the cosine rule, we have $|PC| = r = 2\sqrt{7}$ $(\neq 5)$. It will be worth the effort to find the general solutions to (5) in natural numbers.

References

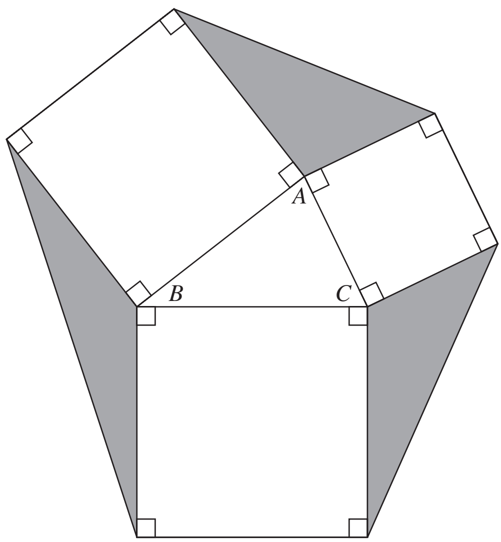
1 L. E. Dickson, *History of the Theory of Numbers*, Vol. II (Chelsea, New York, 1971).

2 C. J. Bradley, *Challenges In Geometry: For Mathematical Olympians Past and Present* (Oxford University Press, 2005).

Prithwijit De *is a PhD student in Statistics at the National University of Ireland, Cork. He dabbles in mathematical problem solving and recreational mathematics. His other interests are reading, music, and cricket.*

The bride's chair

How do the areas of the shaded triangles compare with the area of the triangle ABC?



Reference

1 R. Webster, Bride's chair revisited, *Math. Gazette* **78** (1994), pp. 345–346.

Mathematics and Music: Relating Science to Arts?

MICHAEL BEER

1. Introduction

For many people, mathematics is an enigma. Characterised by the impression of numbers and calculations taught at school, it is often accompanied by feelings of rejection and disinterest, and it is believed to be strictly rational, abstract, cold, and soulless.

Music, on the other hand, has something to do with emotion, with feelings, and with life. It is present in all daily routines. Everyone has sung a song, pressed a key on a piano, blown into a flute, and therefore, in some sense, made music. It is something people can interact with, it is a way of expression and a part of everyone's existence.

The motivation for investigating the connections between these two apparent opposites therefore is not very obvious, and it is unclear in what aspects of both topics such a relationship could be sought. Moreover, even if some mathematical aspects in music such as rhythm and pitch are accepted, it is far more difficult to imagine any musicality in mathematics. The countability and the strong order of mathematics do not seem to coincide with an artistic pattern.

However, there are different aspects which indicate this sort of relationship. Firstly, research has proved that children playing the piano often show improved reasoning skills like those applied in solving jigsaw puzzles, playing chess, or conducting mathematical deductions (see reference 1, p. 17). Secondly, it was noticed in a particular investigation that the percentage of undergraduate students having taken a music course was about eleven percent above average amongst mathematics majors (see reference 2, p. 18). This affinity of mathematicians for music is not only a recent phenomenon, but has been mentioned previously by Bloch in 1925 (see reference 3, p. 183).

This article examines the relationship between mathematics and music from three different points of view. The first describes some ideas about harmony, tones, and tunings generated by the ancient Greeks, the second shows examples of mathematical patterns in musical compositions, and the last illuminates artistic attributes of mathematics.

It is not the intention of this article to provide a complete overview of the complex connections between these two subjects. Neither is it to give detailed explanations and reasons for the cited aspects. However, this assignment will show that mathematics and music do not form such strong opposites as they are commonly considered to do, but that there are connections and similarities between them, which may explain why some musicians like mathematics and why mathematicians frequently love music.

2. Tone and tuning: the Pythagorean perception of music

In the time of the ancient Greeks, mathematics and music were strongly connected. Music was considered as a strictly mathematical discipline, handling number relationships, ratios, and proportions. In the quadrivium (the curriculum of the Pythagorean School) music was placed

Mathematics *(the study of the unchangeable)*			
quantity *(the discreet)*		**magnitude** *(the continued)*	
alone *(the absolute)*	**in relation** *(the relative)*	**at rest** *(the stable)*	**in motion** *(the moving)*
Arithmetic	Music	Geometry	Astronomy

Figure 1 The quadrivium (see reference 4, p. 64).

on the same level as arithmetic, geometry, and astronomy (see figure 1). This interpretation totally neglected the creative aspects of musical performance. Music was the science of sound and harmony.

The basic notions in this context were those of consonance and dissonance. People had noticed very early on that two different notes do not always sound pleasant (consonant) when played together. Moreover, the ancient Greeks discovered that to a note with a given frequency only those other notes whose frequencies were integer multiples of the first could be properly combined. If, for example, a note of the frequency 220 Hz was played, the notes of frequencies 440 Hz, 660 Hz, 880 Hz, 1100 Hz, and so on, sounded best when played together with the first.

Furthermore, examinations of different sounds showed that these integer multiples of the base frequency always appear in a weak intensity when the basic note is played. If a string whose length defines a frequency of 220 Hz is vibrating, the generated sound also contains components of the frequencies 440 Hz, 660 Hz, 880 Hz, 1100 Hz, and so on. Whereas the listeners perceive mainly the basic note, the intensities of these so-called overtones define the character of an instrument. It is primarily due to this phenomenon that a violin and a trumpet do not sound similar even if they play the same note. (The respective intensities of the overtones are expressed by the Fourier coefficients when analysing a single note played. This concept, however, will not be explained within the scope of this article.)

The most important frequency ratio is $1:2$, which is called an octave in the Western system of music notation. Two different notes in such a relation are often considered as principally the same (and are therefore given the same name), only varying in their pitch but not in their character. The Greeks saw in the octave a 'cyclic identity'. The following ratios build the musical fifth $(2:3)$, fourth $(3:4)$, major third $(4:5)$, and minor third $(5:6)$, which all have their importance in the creation of chords. The difference between a fifth and a fourth was defined as a 'whole' tone, which results in a ratio of $8:9$. These ratios correspond not only to the sounding frequencies but also to the relative string lengths, which made it easy to find consonant notes starting from a base frequency. Shortening a string to two thirds of its length creates the musical interval of a fifth for example.

All these studies of 'harmonic' ratios and proportions were the essence of music during Pythagorean times. This perception, however, lost its importance at the end of the Middle Ages, when more complex music was developed. Despite the 'perfect' ratios, there occurred new dissonances when particular chords, different keys, or a greater scale of notes were used. The explanation for this phenomenon was the incommensurability of thirds, fifths, and octaves when defined by integer ratios. By adding several intervals of these types to a base note, we never reach an octave of the base note again. In other words, an octave $(1:2)$ cannot be subdivided into a finite number of equal intervals of this Pythagorean type $(x : x+1 \mid x$ being an integer$)$. Adding whole tones defined by the ratio $9:8$ to a base note with the frequency f, for example, never creates a new note with the frequency $2f$, $3f$, $4f$, or similar. However, adding six whole tones to a note almost creates its first octave defined by the following double frequency:

$$\left(\tfrac{9}{8}\right)^6 f \approx 2.0273 f > 2f.$$

The amount six whole tones overpass an octave is called the 'Pythagorean comma':

$$\frac{\left(\frac{9}{8}\right)^6}{2} = 1.013\,643\,2\ldots.$$

Considering these characteristics of the Pythagorean intervals, the need for another tuning system developed. Several attempts were made, but only one has survived until nowadays: the system of dividing an octave into twelve equal ('even-tempered') semi-tones introduced by Johann Sebastian Bach. Founding on the ratio $1:2$ for octaves, all the other Pythagorean intervals were slightly tempered (adjusted) in order to fit into this new pattern. A whole tone no longer was defined by the ratio $9/8 = 1.125$, but by two semi-tones (each expressed by $\sqrt[12]{2}$) obtaining the numerical value $\sqrt[12]{2}\,\sqrt[12]{2} = \sqrt[6]{2} \approx 1.1225$. The even-tempered fifth then was defined by seven semi-tones and therefore slightly smaller than the Pythagorean fifth, the fourth by five semi-tones and therefore slightly bigger than the Pythagorean fourth.

The controversy within this tempering process is that the human ear still prefers the 'pure' Pythagorean intervals, whereas a tempered scale is necessary for complex chordal music. Musicians nowadays have to cope with these slight dissonances in order to tune an instrument so that it fits into this even-tempered pattern.

With the evolution of this more complicated mathematical model for tuning an instrument, and with the increased importance of musicality and performance, music and mathematics in this aspect have lost the close relationship known in ancient Greek times. As an even-tempered interval could no longer be expressed as a ratio ($\sqrt[12]{2}$ is an irrational number), the musicians learnt to tune an instrument by training their ear rather than by applying mathematical principles. Music from this point of view released itself from mathematical domination; see references 4 (pp. 36–67), 5, 6 (pp. 3–5), and 7 (Chapter 1, pp. 13–27).

3. Mathematical music: Fibonacci numbers and the golden section in musical compositions

The questions of tone and tuning are one aspect in which mathematical thoughts enter the world of music. However, music – at least in a modern perception – does not only consist of notes and harmony. More important are the changes of notes in relation to time, that is the aspect of rhythm and melody. Here again mathematical concepts are omnipresent. Not only is the symbolic musical notation in all its aspects very mathematical, but also particular arithmetic and geometric reflections can be found in musical compositions, as will be seen in the following paragraphs. (This article is not going to deal with highly sophisticated mathematical music theories as established by Xenakis (see reference 8) or Mazzola (see references 9 and 10) for example, based on an algebraic composition model or on group and topos theories respectively. These two concepts would exceed this overview of the relations between mathematics and music.)

A very interesting aspect of mathematical concepts in musical compositions is the appearance of Fibonacci numbers and the theory of the golden section. The former is an infinite sequence of integers named after Leonardo de Pisa (alias Fibonacci), a medieval mathematician. Its first two members are both 1, whereas every new member of the sequence is formed by the addition of the two preceding numbers, giving $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89 \ldots$. However, their most important feature in this context is that the sequence of Fibonacci ratios (that is the ratio of a Fibonacci number with its larger adjacent number) converges to a constant limit, called the golden ratio, golden proportion, or golden section, i.e. 0.618 033 98 . . . .

More common is the geometric interpretation of the golden section: a division of a line into two unequal parts is called 'golden' if the relation of the length of the whole line to the length of the bigger part is the same as the relation of the length of the bigger part to the length of the smaller part. This proportion cannot only be found in geometric forms (for example the length of a diagonal related to the length of an edge in a regular pentagon), but also in nature (for example the length of the trunk in relation to the diameter of the tree for some particular trees, such as the Norway spruce; see reference 4, p. 113).

Due to its consideration as well-balanced, beautiful, and dynamic, the golden section has found various applications in the arts, especially in painting and photography, where important elements often divide a picture's length or width (or both) following the golden proportion. However, such a division is not necessarily undertaken consciously, but results from an impression of beauty and harmony.

Diverse studies have discovered that this same concept is also very common in musical compositions. The golden section – expressed by Fibonacci ratios – is either used to generate rhythmic changes or to develop a melody line (see reference 4, p. 116). Examples of deliberate applications can be found in the widely used 'Schillinger System of Musical Composition' or concretely in the first movement of Béla Bartók's piece 'Music for Strings, Percussion and Celeste', where, for instance, the climax is situated at bar 55 of 89 (see reference 11).

Furthermore, Rothwell's study (see reference 12) has revealed examples of the golden proportion in various musical periods. While the characteristics of the examined compositions varied greatly, the importance of proportional organisation was mostly similar. Important structural locations, marked by melodic, rhythmic, or dynamic events, were often discovered to divide the composition in two parts, either symmetrically or in the golden proportion.

A well-known example is the 'Hallelujah' chorus in Handel's Messiah. Whereas the whole consists of 94 measures, one of the most important events (entrance of solo trumpets: 'King of Kings') happens in measures 57 to 58, after about $\frac{8}{13}$ (!) of the whole piece. In addition to that, we can find a similar structure in both of the divisions of the whole piece. After $\frac{8}{13}$ of the first 57 measures, that is in measure 34, the entrance of the theme 'The kingdom of glory . . . ' marks another essential point; and after $\frac{8}{13}$ of the second 37 measures, in measure 79 ('And he shall reign . . . '), again, the importance of the location is enforced by the appearance of solo trumpets (see reference 12, p. 89). It is hard to say whether Handel chose these locations deliberately, but at least this phenomenon outlines the importance of the golden section not only in visual but also in performing arts.

Another study (see reference 13, pp. 118–119) has shown that in almost all of Mozart's piano sonatas, the relation between the exposition and the development and recapitulation conforms to the golden proportion. Here, again, we cannot ascertain whether Mozart was conscious of his application of the golden section, even though some evidence suggests his attraction to mathematics.

It is probably less important to evaluate whether people consider mathematics when they apply or perceive a golden proportion than to notice that harmony and beauty – at least in this aspect – can be expressed by mathematical means. Fibonacci ratios in relation to the division of a composition, as well as integer ratios in relation to Pythagorean intervals, are examples of the fact that harmony can sometimes be described by numbers (even by integers) and therefore has a very mathematical aspect. This could be one way to introduce an additional idea: that beauty is inherent in mathematics.

4. Musical mathematics: reflections on an artistic aspect of mathematics

All these aspects of mathematical patterns in sound, harmony, and composition do not convincingly explain the outstanding affinity of mathematicians for music. Being a mathematician does not mean discovering numbers everywhere and enjoying only issues with strong mathematical connotations. The essential relation is therefore presumed to be found on another level.

It is noticeable that the above-mentioned affinity is not reciprocated. Musicians do not usually show the same interest for mathematics as mathematicians for music. We therefore must suppose that the decisive aspect cannot lie in arithmetic, the part of mathematics people sometimes consider to be in fact the whole subject. It is probably more the area of mathematical thinking, mind-setting, and problem-solving which creates these connections.

An example given by both Henle (see reference 2, p. 19) and Reid (see reference 5) is the omnipresence of words such as beauty, harmony, and elegance in mathematical research. Whereas musicians sometimes develop a particularly well-formed melody or apply an outstanding harmony, mathematicians often seek 'simple' and elegant proofs. Moreover, the sensations in solving a mathematical problem seem to be similar to those appearing when performing a musical work. Most important is the creative aspect, which lies within both of these disciplines.

Interesting evidence for this idea has been presented by Henle (see reference 2, p. 19), who compared the history of music with the history of mathematics based on the following three main arguments.

1. Mathematics has many of the characteristics of an art.
2. Viewed as an art, it is possible to identify artistic periods in mathematics: Renaissance, Baroque, Classical, and Romantic.
3. These periods coincide nicely and share many characteristics with the corresponding musical epochs, but *are significantly different* from those of painting and literature.

Relating to concepts such as dualism (Baroque), universality (Classical), and eternity (Romantic), Henle drew out surprising similarities between the evolution of mathematics and music.

Moreover, Henle outlined the necessity of a change in mathematical education towards a more musical style (see reference 2, p. 28).

> Students should make mathematics *together* (as in fact professional mathematicians do), not alone. [. . .] And finally, students should perform mathematics; they should *sing* mathematics and *dance* mathematics.

This would probably help people understand what mathematics really is, namely not divine, but mortal, and not law, but taste.

In spite of the highly speculative aspect within such ideas, this is probably the fundamental point of view when seeking connections between mathematics and music. It is the musicality in the mathematical way of thinking that attracts mathematicians to music. This, however, is difficult for people who are not familiar with this particular pattern of mind to comprehend. It is therefore probable – as has been stated by Reid (see reference 5) – that the degree of understanding such relationships is proportional to the observer's understanding of both mathematics and music. (In this context, we should mention some ideas of Hofstadter (see reference 14), who linked the music of J. S. Bach, the graphic art of Escher, and the mathematical theorems of Gödel in order to illuminate the nature of human thought processes. Once more, however, this would go beyond the framework of this article.)

5. Conclusion

This article has outlined three different approaches to the question of how mathematics and music relate to each other. The first showed the particular perception of music by the ancient Greeks, putting less importance on melody and movement than on tone, tuning, and static harmony. In the second, the concept of the golden section was brought into relation with number ratios and their occurrence in diverse compositions. The most fundamental approach, however, was the third, in which connections were revealed concerning the artistic aspect of the mathematical way of thinking.

It is obvious that these are only examples for investigating such a relationship and that other comparisons could be attempted (apart from those already mentioned). However, these three approaches represent probably the most often discussed concepts and ideas and are particularly suitable for providing a first impression of this topic.

Whatever links between music and mathematics exist, both of them are obviously still very different disciplines, and we should not try to impose one on the other. It would be wrong to attempt to explain all the shapes of music by mathematical means, just as there would be no sense in studying mathematics from a musicological point of view. However, it would be enriching if these relationships were introduced into mathematical education in order to release mathematics from its often too serious connotations.

It is important to show people that mathematics, in one way, is as much an art as it is a science. This probably would alter its common perception, and people would understand better its essence and its universality.

References

1 A. Motluk, Can Mozart make maths add up?, *New Scientist* **2073** (15 March 1997).

2 J. Henle, Classical mathematics, *Amer. Math. Monthly* **103** (1996), pp. 18–29.

3 E. Bloch, *Essays on the Philosophy of Music* (Cambridge University Press, 1985).

4 T. H. Garland and C. V. Kahn, *Math and Music: Harmonious Connections* (Dale Seymour Publications, Palo Alto, CA, 1995).

5 H. Reid, On mathematics and music, available at http://www.woodpecker.com/writing/essays/math+music.html.

6 E. G. McClain, *The Pythagorean Plato* (Hays, Stony Brook, NY, 1978).

7 N. Bibby, Tuning and temperament: closing the spiral. In *Music and Mathematics: From Pythagoras to Fractals* (Oxford University Press, 2003).

8 I. Xenakis, *Formalized Music: Thought and Mathematics in Composition*, (Indiana University Press, 1971).

9 G. Mazzola, *Gruppen und Kategorien in der Musik: Entwurf einer mathematischen Musiktheorie* (Heldermann, Berlin, 1985).

10 G. Mazzola, *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance* (Birkhäuser, Basel, 2002).
11 D. Rusin, Mathematics and music, available at http://www.math.niu.edu/~rusin/uses-math/music/.
12 J. A. Rothwell, *The Phi Factor: Mathematical Proportions in Musical Forms* (University of Missouri, Kansas City, 1977).
13 M. May, Did Mozart use the golden section?, *Amer. Scientist* **84** (1996), pp. 118–119.
14 D. R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid*, (Penguin Books, Harmondsworth, 1979).

Michael Beer *is a PhD student and research assistant at the Department of Quantitative Economics, University of Fribourg, Switzerland. This article is a slightly amended version of his essay entitled 'How do mathematics and music relate to each other?' written in October 1998 as an assignment at the East Coast College of English, Brisbane, Australia.*

Letters to the Editor

Dear Editor,

Regular polygons

I enjoyed reading about Daniel Schultz's interesting discovery in Volume 40, Number 2, p. 84. A short, albeit rather mechanical, proof of his general result can be given using complex numbers.

Put O, the centre of the regular n-sided polygon, at the origin and take its vertices to be (P_k), represented by the complex numbers $OP_k = z_k = R\omega^k$, where $\omega = \cos(2\pi/n) + \mathrm{i}\sin(2\pi/n)$ and, as usual, $P_{n+1} = P_1$. If the fixed point in the plane of the polygon is denoted by P with $OP = z$, then A'_k, the foot of the perpendicular from P to side P_kP_{k+1}, is of the form $OA'_k = z_k + \lambda(z_{k+1} - z_k)$ with the real number λ characterised by the fact that

$$\frac{z_k + \lambda(z_{k+1} - z_k) - z}{z_k - z_{k+1}}$$

is purely imaginary. Thus,

$$\frac{z_k + \lambda(z_{k+1} - z_k) - z}{z_k - z_{k+1}} + \frac{\bar{z}_k + \lambda(\bar{z}_{k+1} - \bar{z}_k) - \bar{z}}{\bar{z}_k - \bar{z}_{k+1}} = 0$$

which, on solving for λ and substituting $z_k = R\omega^k$, leads eventually to the expression

$$OA'_k = \tfrac{1}{2}z - \tfrac{1}{2}\bar{z}\omega^{2k+1} + \tfrac{1}{2}R\omega^k + \tfrac{1}{2}R\omega^{k+1}.$$

Since $\sum_{k=1}^{n} \omega^k = 0 = \sum_{k=1}^{n} \omega^{2k+1}$, it follows that

$$\frac{1}{n}\sum_{k=1}^{n} OA'_k = \frac{1}{2}z,$$

which proves David Wells' observation in Daniel's letter.

Finally, if (A_k) are the intersection points constructed by Daniel such that PA_k is the common chord of the two circles whose centres are joined by $P_k P_{k+1}$, then A'_k is the midpoint of PA_k so that $OA'_k = \frac{1}{2}(OP + OA_k)$ or $OA_k = 2OA'_k - z$. Then we have

$$\frac{1}{n}\sum_{k=1}^{n} OA_k = 2\frac{1}{n}\sum_{k=1}^{n} OA'_k - z = 2\left(\frac{1}{2}z\right) - z = 0,$$

which proves Daniel's result.

I have not seen Daniel's result before, but the great 19th century geometer Jakob Steiner did prove some results about the 'pedal polygon' $A'_1 A'_2 \cdots A'_n$. For example, the area of the pedal polygons constructed from points which are the same distance from O have the same area.

Yours sincerely,

Nick Lord
(Tonbridge School
High Street
Tonbridge
Kent TN9 1JP
UK)

Problems and Solutions

Students are invited to submit solutions to some or all of the problems below. The most attractive solutions received by 1st March will be published in a subsequent issue and are eligible for annual prizes. When writing to the Editorial Office, please state your full name and also the postal address of your school, college, or university.

Problems

41.1 Is there a number greater than 1 which leaves the same remainder when it divides into 1716, 2154, and 4271?

(Submitted by Bob Bertuello, Midsomer Norton, Bath, UK)

41.2 (i) Given positive integers m, n, t, show that the equation

$$\frac{1}{x^m} + \frac{1}{y^n} = \frac{1}{z^t}$$

has no solution in positive integers x, y, z with x, y coprime and y, z coprime.

(ii) Fermat's last theorem states that, for a given integer $n \geq 3$, the equation

$$x^n + y^n = z^n$$

has no solution in positive integers x, y, z. Prove that, when $n \geq 3$, the equation

$$\frac{1}{x^n} + \frac{1}{y^n} = \frac{1}{z^n}$$

has no solution in positive integers x, y, z.

(Submitted by H. Sazegar, Mashhad, Iran)

41.3 Let c_1, c_2 be two given unequal numbers. The function f is differentiable everywhere and satisfies the condition

$$c_1 f(x-y) + c_2 f(x+y) \leq (c_1 + c_2) f(x)$$

for all real numbers x and y. Determine all such functions f.

(Submitted by P. Andriopoulos, 3rd High School of Amaliada, Greece)

41.4 The regular n-sided polygon $A_1 A_2 \cdots A_n$ is inscribed in the unit circle. What is the product of the distances $A_1 A_n, A_2 A_n, \ldots, A_{n-1} A_n$?

(Submitted by Theodoros Valahas and Andreas Boukas, The American College of Greece)

Solutions to Problems in Volume 40 Number 2

40.5 Let $P_1 P_2 P_3 P_4 P_5$ be a regular pentagon, centre O, and let P be any point in the plane of the pentagon. For $i = 1, 2, 3, 4, 5$, denote by C_i the circle with centre P_i which passes through P, and denote the point of intersection of circles C_i and C_{i+1} other than P by A_i ($C_6 = C_1$). Prove that the centre of gravity of A_1, A_2, A_3, A_4, A_5 is O.

Let A_i' be the foot of the perpendicular from P to $P_i P_{i+1}$ ($P_6 = P_1$). Show that A_i' is the midpoint of $A_i P$ and that the centre of gravity of A_1', A_2', A_3', A_4', A_5' is the midpoint of OP.

Solution see the letter from Nick Lord on pages 42–43

40.6 Let a and b be real numbers such that $a \geq b > 0$ and let $g(x)$ be a function such that $\lim_{x \to +\infty} g(x) = \lim_{x \to +\infty} g'(x) = +\infty$. Compute the limit

$$\lim_{x \to +\infty} \frac{\ln(a\sqrt{e^{2g(x)} + 1} - b e^{g(x)})}{x}.$$

Solution by Spiros P. Andriopoulos, who proposed the problem

When $a \neq b$,

$$\begin{aligned} a\sqrt{e^{2g(x)} + 1} - b e^{g(x)} &= e^{g(x)}(a\sqrt{1 + e^{-2g(x)}} - b) \\ &\to +\infty \quad \text{as } x \to +\infty. \end{aligned}$$

When $a = b$,

$$\begin{aligned} a\sqrt{\mathrm{e}^{2g(x)}+1} - a\mathrm{e}^{g(x)} &= a\mathrm{e}^{g(x)}(\sqrt{1+\mathrm{e}^{-2g(x)}} - 1) \\ &= \frac{a}{\mathrm{e}^{g(x)}(\sqrt{1+\mathrm{e}^{-2g(x)}} + 1)} \\ &\to 0 \quad \text{as } x \to +\infty \end{aligned}$$

and

$$\ln(a\sqrt{\mathrm{e}^{2g(x)}+1} - a\mathrm{e}^{g(x)}) \to -\infty \quad \text{as } x \to +\infty.$$

In both cases, we use l'Hôpital's rule to obtain

$$\begin{aligned} &\lim_{x\to+\infty} \frac{\ln(a\sqrt{\mathrm{e}^{2g(x)}+1} - b\mathrm{e}^{g(x)})}{x} \\ &\quad = \lim_{x\to+\infty} \frac{(a\mathrm{e}^{g(x)}/\sqrt{\mathrm{e}^{2g(x)}+1} - b)\mathrm{e}^{g(x)}g'(x)}{a\sqrt{\mathrm{e}^{2g(x)}+1} - b\mathrm{e}^{g(x)}} \\ &\quad = \lim_{x\to+\infty} \left(\frac{a/\sqrt{1+\mathrm{e}^{-2g(x)}} - b}{a\sqrt{1+\mathrm{e}^{-2g(x)}} - b}\right) g'(x) \qquad (1) \\ &\quad = \frac{a-b}{a-b} \lim_{x\to+\infty} g'(x) \\ &\quad = +\infty \quad \text{when } a \neq b. \end{aligned}$$

When $a = b$, (1) becomes

$$\lim_{x\to+\infty} \frac{-g'(x)}{\sqrt{1+\mathrm{e}^{-2g(x)}}} = -\infty.$$

40.7 Let $x_1, \ldots, x_n$ be positive real numbers and let α be a positive integer. Prove that

$$\sum_{1\le i<j\le n} \frac{x_i^{2\alpha} + x_j^{2\alpha}}{x_i^{2\alpha+2} + x_j^{2\alpha+2}} \le \frac{n-1}{2} \sum_{k=1}^{n} \frac{1}{x_k^2}.$$

Solution by José Luis Díaz-Barrero, who proposed the problem

From

$$(x_i - x_j)(x_i^{2\alpha+1} - x_j^{2\alpha+1}) \ge 0,$$

we see that

$$x_i^{2\alpha+2} + x_j^{2\alpha+2} \ge x_i x_j (x_i^{2\alpha} + x_j^{2\alpha}),$$

so that

$$\frac{x_i^{2\alpha} + x_j^{2\alpha}}{x_i^{2\alpha+2} + x_j^{2\alpha+2}} \le \frac{1}{x_i x_j} \le \frac{1}{2}\left(\frac{1}{x_i^2} + \frac{1}{x_j^2}\right).$$

Hence,

$$\sum_{1\le i<j\le n} \frac{x_i^{2\alpha} + x_j^{2\alpha}}{x_i^{2\alpha+2} + x_j^{2\alpha+2}} \le \frac{n+1}{2} \sum_{k=1}^{n} \frac{1}{x_k^2}.$$

40.8 For a positive integer n, let

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \quad \text{and} \quad \delta_n = \sum_{k=1}^{n} \frac{2H_k}{k} - H_n^2.$$

Prove that

(a)

$$\lim_{n\to\infty} \delta_n = \zeta(2),$$

(b)

$$\sum_{n=1}^{\infty} \frac{\delta_n}{n(n+1)} = \zeta(3),$$

where $\zeta(s) = \sum_{n=1}^{\infty}(1/n^s)$.

Solution by Henry Ricardo, Medgar Evers College, New York, USA

(a) First we note that, for $n > 1$,

$$\begin{aligned}
\delta_n - \delta_{n-1} &= \frac{2H_n}{n} - (H_n^2 - H_{n-1}^2) \\
&= \frac{2H_n}{n} - (H_n - H_{n-1})(H_n + H_{n-1}) \\
&= \frac{2H_n}{n} - \frac{1}{n}\left(2H_{n-1} + \frac{1}{n}\right) \\
&= \frac{2}{n^2} - \frac{1}{n^2} \\
&= \frac{1}{n^2}.
\end{aligned}$$

Hence,

$$\sum_{n=2}^{N}(\delta_n - \delta_{n-1}) = \sum_{n=2}^{N} \frac{1}{n^2},$$

so that

$$\delta_N - 1 = \sum_{n=2}^{N} \frac{1}{n^2}, \qquad \delta_N = \sum_{n=1}^{N} \frac{1}{n^2},$$

and

$$\lim_{N\to\infty} \delta_N = \sum_{n=1}^{\infty} \frac{1}{n^2} = \zeta(2) \qquad \left(= \frac{\pi^2}{6}\right).$$

(b) For a positive integer N,

$$\sum_{N=1}^{N}\frac{\delta_n}{n(n+1)} = \sum_{N=1}^{N}\left(\frac{\delta_n}{n} - \frac{\delta_n}{n+1}\right)$$
$$= \delta_1 + \sum_{n=2}^{N}\frac{\delta_n - \delta_{n-1}}{n} - \frac{\delta_N}{N+1}$$
$$= 1 + \sum_{n=2}^{N}\frac{1}{n^3} - \frac{\delta_N}{N+1}$$
$$= \sum_{n=1}^{N}\frac{1}{n^3} - \frac{\delta_N}{N+1}.$$

Hence,

$$\sum_{n=1}^{\infty}\frac{\delta_n}{n(n+1)} = \sum_{n=1}^{\infty}\frac{1}{n^3} = \zeta(3),$$

called *Apéry's constant*.

Reviews

Game Theory: A Very Short Introduction. By Ken Binmore. Oxford University Press, 2007. Paperback, 184 pages, £6.99 (ISBN 978-0-19-921846-2).

This book is written by the eminent economist Professor Ken Binmore, who used Game Theory in organizing the British Telecom auction; he proudly reports that this auction made a total of 35 billion dollars. The book consists of ten chapters:

1. The name of the game,
2. Chance,
3. Time,
4. Conventions,
5. Reciprocity,
6. Information,
7. Auctions,
8. Evolutionary Biology,
9. Bargaining and Coalitions,
10. Puzzles and Paradoxes.

These are followed by $5\frac{1}{4}$ pages of References and Further Reading, and a 4-page Index. There are 37 illustrations spread throughout the text.

The author gives a concise account of Game Theory; he comments 'In brief, a game is being played whenever human beings interact'. His book contains no algebra and a minimum of technical jargon, but provides a readable summary of what Game Theory is about. David Hume, John Von Neumann, John Nash, John Maynard Smith, the Prisoner's Dilemma, and the Monty Hall game all figure in this account, which provides an informative introduction to the subject. I particularly enjoyed Chapter 8 which describes the Hawk-Dove game, Hamilton's rule, and the evolution of cooperation among vampire bats. The book should prove a valuable introductory text for students wishing to learn about Game Theory.

Australian National University **Joe Gani**

Mathematical Spectrum

Volume 41 2008/2009 Number 1

ISSN 0025-5653

http://www.appliedprobability.org

Published by the Applied Probability Trust

Printed by MFK Pear Tree Press Ltd, Stevenage, Hertfordshire, UK