How to Compute Σ1/n^2 by Solving Triangles

Mikael Passare

4. J. Borwein and P. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, John Wiley, New York, 1987.

5. J. B. Conrey, The Riemann hypothesis, *Notices of the AMS* **50** (2003) 341–353.

6. W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. II*, 2nd ed., John Wiley, New York, 1971.

7. G. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., Pure and Applied Mathematics, Wiley-Interscience, New York, 1999.

8. W. Gosset, The probable error of a mean, *Biometrika* **6** (1908) 1–25.

9. R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 5th ed., Prentice Hall, New York, 1994.

10. S. Lang, *Complex Analysis*, Graduate Texts in Mathematics, vol. 103, Springer-Verlag, New York, 1999.

11. M. Mehta, *Random Matrices*, 2nd ed., Academic Press, Boston, 1991.

12. I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications*, 7th ed., Prentice Hall, New York, 2004.

13. S. J. Miller, M. Schiffman, and B. Wieland, Irrationality measure and lower bounds for $\pi(x)$, available at `http://arxiv.org/pdf/0709.2184`.

14. S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.

15. J. F. Scott, *Mathematical Works of John Wallis*, Chelsea, New York, 1981.

16. J. Sondow, A faster product for π and a new integral for $\ln \pi/2$, this MONTHLY **112** (2005) 729–734.

17. G. K. Srinivasan, The Gamma function: an eclectic tour, this MONTHLY **114** (2007) 297–315.

18. E. Stein and R. Shakarchi, *Complex Analysis*, Princeton University Press, Princeton, NJ, 2003.

Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267
Steven.J.Miller@williams.edu

How to Compute $\sum 1/n^2$ by Solving Triangles

Mikael Passare

1. INTRODUCTION. The harmonic series $1 + 1/2 + 1/3 + 1/4 + \cdots$ is infinitely large, whereas the series of squared terms $1 + 1/4 + 1/9 + 1/16 + \cdots$ sums to a finite value. Indeed, the divergence of the harmonic series is easily verified by grouping its terms in packages whose lengths are increasing powers of two:

$$1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots > 1 + \frac{1}{2} + \frac{2}{4} + \frac{4}{8} + \cdots$$

and the boundedness of the squared series is equally simple to check:

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots < 1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots$$
$$= 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots = 2.$$

Alternatively, one can of course also establish these well-known facts by comparing with integrals:

$$\sum_{n=1}^{N} \frac{1}{n} > \int_0^N \frac{dx}{1+x} = \log(1+N) \to \infty,$$

$$\sum_{n=1}^{N} \frac{1}{n^2} < 1 + \int_1^N \frac{dx}{x^2} = 2 - \frac{1}{N} \to 2.$$

The problem of computing the exact value of $\sum 1/n^2$ is so classical (dating back to 1644) that it even has a name of its own: the Basel problem. The city of Basel was the hometown both of the famous brothers Jakob and Johann Bernoulli, who made serious but unsuccessful attempts to solve the problem, and to the prodigious mathematician Leonhard Euler, who found the first solution in 1735, see [3]. Since then a great many different ways of evaluating the sum have been discovered. Fourteen proofs are collected in [2], and two of them have been included in [1].

In this note we shall present a new simple geometric method for finding the exact value of $\sum_{n=1}^{\infty} 1/n^2$ using essentially just basic trigonometry: the sine rule for triangles; combined with some elementary undergraduate analysis: area-preserving maps, see for instance [4].

2. BIPOLAR COORDINATES AND THE (CO)SINE RULE.
Consider an arbitrary triangle with base of length 1, such as the one in Figure 1. The position of the top of the triangle can then be determined either by the two remaining side lengths A and B or, equivalently, by the two interior base angles α and β. One can view these data as representing two alternative sets of coordinates for the point at the top of the triangle. Thinking of the two base vertices as a pair of reference points (much like your own two eyes) it is natural to refer to (A, B) and (α, β) as *bipolar coordinates*, radial and angular respectively.

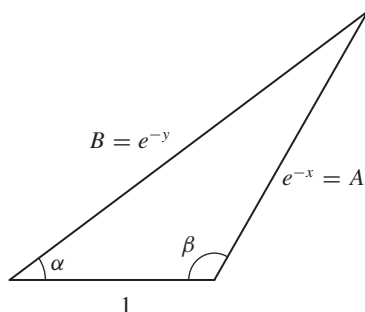**Figure 1.** The side lengths and interior angles in a triangle.

Let us check what the possible values of the two sets (A, B) and (α, β) of bipolar coordinates are. For the radial coordinates (A, B) the triangle inequality amounts to the obvious restriction that each side length must be less than the sum of the remaining two. That is, one has the following three inequalities:

$$1 < A + B, \qquad A < 1 + B, \qquad B < 1 + A.$$

These are in fact the only conditions imposed on the radial bipolar coordinates, so the collection of all possible (A, B) is given by the infinite polygon S depicted on the left in Figure 2. For the angular coordinates (α, β), which we always measure in

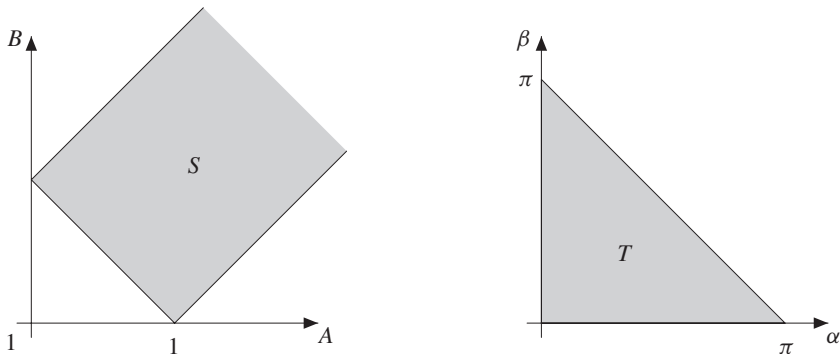© THE MATHEMATICAL ASSOCIATION OF AMERICA [Monthly 115

Figure 2. The possible values for radial and angular bipolar coordinates.

radians, the corresponding restriction comes from the fact that the sum of all three angles in a triangle is equal to π. The possible values of the angular coordinates (α, β) are therefore given by the half square T shown on the right of Figure 2.

To solve a triangle means to find all its side lengths and all its interior angles, with only some of these data being known beforehand. This is precisely what one needs to do in order to pass from one set of bipolar coordinates to another. The classical tool for this is the *sine rule*

$$\frac{\sin \alpha}{A} = \frac{\sin \beta}{B} = \frac{\sin(\pi - \alpha - \beta)}{1} = \sin(\alpha + \beta), \tag{1}$$

which permits one to express the side lengths (A, B) in terms of the angles (α, β), and the *cosine rule*

$$A^2 = 1 + B^2 - 2B \cos \alpha, \quad B^2 = 1 + A^2 - 2A \cos \beta, \tag{2}$$

which conversely expresses (α, β) in terms of (A, B). More precisely, the sine rule provides us with the explicit bijective map $F \colon T \to S$ given by

$$F(\alpha, \beta) = \left[\frac{\sin \alpha}{\sin(\alpha + \beta)}, \frac{\sin \beta}{\sin(\alpha + \beta)} \right],$$

whereas the cosine rule allows us to write down the inverse map

$$F^{-1}(A, B) = \left[\arccos\left(\frac{1 - A^2 + B^2}{2B} \right), \arccos\left(\frac{1 + A^2 - B^2}{2A} \right) \right].$$

It will now be useful to make a logarithmic change of scale for the side lengths in Figure 1, and thus to replace the radial bipolar coordinates (A, B) by the new coordinates (x, y) given by $x = -\log A$, $y = -\log B$. The image of the set S under this change of coordinates, that is, the result of drawing it on log-log graph paper, is the set U shown in Figure 3. This set U, with its "tentacles" going off to infinity, is an example of what is nowadays known as an *amoeba* in complex geometry, see [**5**] and [**6**]. In the first quadrant its boundary is given by the curve $e^{-x} + e^{-y} = 1$, which can also be represented as $y = -\log(1 - e^{-x})$.

The advantage of having passed to logarithmic coordinates becomes apparent when we consider the composed bijective map $G \colon T \to U$ given by

$$G(\alpha, \beta) = \left[\log\left(\frac{\sin(\alpha + \beta)}{\sin \alpha} \right), \log\left(\frac{\sin(\alpha + \beta)}{\sin \beta} \right) \right].$$
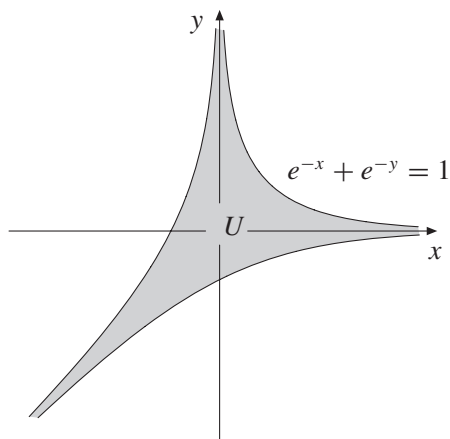
Figure 3. The set of values of the logarithmic side lengths: an amoeba.

Theorem. *The Jacobian determinant of the map G is identically equal to* 1, *that is, G is an area-preserving map.*

Proof. Componentwise differentiation of G yields the Jacobian determinant

$$\begin{vmatrix} \cot(\alpha + \beta) - \cot\alpha & \cot(\alpha + \beta) \\ \cot(\alpha + \beta) & \cot(\alpha + \beta) - \cot\beta \end{vmatrix}$$

$$= -\cot(\alpha + \beta)\big(\cot\alpha + \cot\beta\big) + \cot\alpha\cot\beta,$$

and this expression is indeed identically equal to 1, in view of the addition formula $\cot(\alpha + \beta) = (\cot\alpha\cot\beta - 1)/(\cot\alpha + \cot\beta)$. ∎

Since G is a bijection, that is, a one-to-one map, the theorem has the following immediate and remarkable consequence.

Corollary. $\text{Area}(U) = \text{Area}(T) = \pi^2/2.$

The threefold shape of the amoeba U suggests that one should let its asymptotes divide it into three parts U_0, U_1, and U_2, as indicated on the left in Figure 4. The points (x, y) on the asymptotes correspond precisely to the side lengths $(A, B) = (e^{-x}, e^{-y})$ of the isosceles triangles for which the interior angles (α, β) satisfy one of the conditions $\alpha = \beta$, $\alpha = \pi - \alpha - \beta$, or $\beta = \pi - \alpha - \beta$, and these are just the equations of the three medians of the triangle T. (In particular, the origin $(x, y) = (0, 0)$ represents the equilateral triangle, with $\alpha = \beta = \pi/3$.) In other words, the map G sends the medians of T to the asymptotes of U, and one obtains a corresponding polygonal subdivision of T into three parts T_0, T_1, and T_2, shown on the right in Figure 4.

Theorem. *The area-preserving bijective map* $(x, y) \mapsto (-y, x - y)$ *permutes the amoeba subsets cyclically:* $U_0 \mapsto U_1 \mapsto U_2 \mapsto U_0.$

Proof. To check that $U_0 \mapsto U_1$ one can simply write down the three defining inequalities for U_0 in the new coordinates $(\xi, \eta) = (-y, x - y)$:

$$x > 0 \Leftrightarrow \xi < \eta, \qquad y > 0 \Leftrightarrow \xi < 0, \qquad 1 < e^{-x} + e^{-y} \Leftrightarrow e^{-\xi} < e^{-\eta} + 1,$$
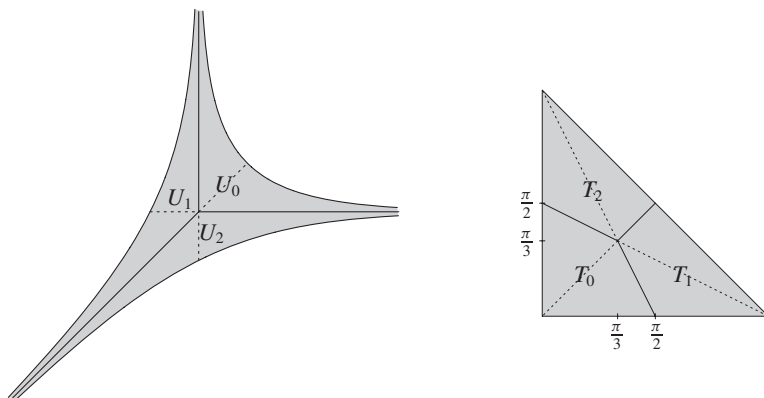
© THE MATHEMATICAL ASSOCIATION OF AMERICA [Monthly 115

Figure 4. All six subsets have equal area.

and observe that the new inequalities are precisely the ones that define U_1. The rest of the theorem is proved analogously. ∎

It follows immediately from this theorem that U_0, U_1, and U_2 all have the same area, which must then be equal to $\pi^2/6$, the total area of U being $\pi^2/2$. So, either using again that G is area-preserving and that $G(T_0) = U_0$, or directly calculating the area of the simple polygon T_0, one obtains the following conclusion.

Corollary. $\text{Area}(U_0) = \text{Area}(T_0) = \pi^2/6$.

Notice that the points in T_0 and U_0 correspond to triangles in which the longest side is the base of length 1.

3. SPREADING AND PILING OF SQUARES. Geometrically, the Basel problem amounts to computing the total area of a collection of squares with decreasing side lengths $1, 1/2, 1/3, 1/4, \ldots$, and in order to achieve this, it will be advantageous to spread out each such square by means of an exponential function. More precisely, one observes that a square with side length $1/n$ has the same area as the part in the first quadrant lying under the exponential curve $y = e^{-nx}/n$, see Figure 5.
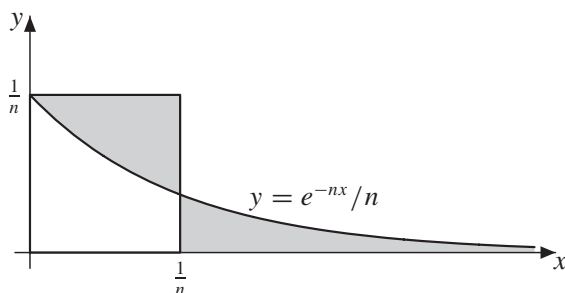

Figure 5. The two areas are the same.

Indeed, the area under the curve is readily calculated by means of the integral

$$\frac{1}{n} \int_0^\infty e^{-nx}\, dx = \frac{1}{n} \left[-\frac{e^{-nx}}{n} \right]_0^\infty = \frac{1}{n^2}.$$

(Incidentally, this elementary fact can also be deduced by observing that the area-preserving bijection $(x, y) \mapsto ((1 - e^{-nx})/n, ye^{nx})$ maps the area under the curve $y = e^{-nx}/n$ to the square $0 < x, y < 1/n$.)

The next step will be to pile these spread out squares on top of each other. In order to understand what such a process will yield, let us take a new look at the harmonic series. Even though the harmonic series itself is not convergent, it can be made to converge by introducing powers in the numerators.

Theorem. *For any positive real number $t < 1$ the power series $t + t^2/2 + t^3/3 + t^4/4 + \cdots$ is convergent and its sum is equal to $-\log(1 - t)$.*

Proof. Differentiation gives

$$D\left(t + \frac{t^2}{2} + \frac{t^3}{3} + \ldots + \frac{t^N}{N}\right) = 1 + t + t^2 + \cdots + t^{N-1} = \frac{1 - t^N}{1 - t},$$

which means that

$$t + \frac{t^2}{2} + \frac{t^3}{3} + \cdots + \frac{t^N}{N} = \int_0^t \frac{1 - \tau^N}{1 - \tau} \, d\tau = \int_0^t \frac{d\tau}{1 - \tau} - \int_0^t \frac{\tau^N \, d\tau}{1 - \tau}.$$

Here the first integral on the right has the desired value $-\log(1 - t)$ while the second integral is less than

$$t^N \int_0^t \frac{d\tau}{1 - \tau} = -t^N \log(1 - t),$$

which tends to zero as $N \to \infty$. ∎

Writing $t = e^{-x}$, so that $t^n = e^{-nx}$, we thus see that for every $x > 0$ the series $\sum e^{-nx}/n$ has the sum $-\log(1 - e^{-x})$. But piling the spread out squares on top of each other, as indicated on the right in Figure 6, precisely amounts to summing the functions e^{-nx}/n, so from the theorem one deduces the following fact.

Corollary. *The infinite pile of spread out squares exactly covers the set U_0 defined by the inequalities $x > 0$, $y > 0$, $y < -\log(1 - e^{-x})$.*
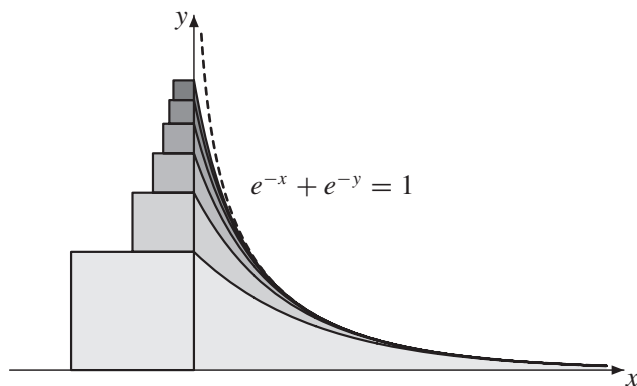

$$e^{-x} + e^{-y} = 1$$

Figure 6. Building two infinite piles. Equally shaded areas are the same.

Since the convergence is not uniform in x a little argument may be in order here: One easily verifies that for any compact subset K of U_0 there is some finite pile $0 < y < e^{-x} + e^{-2x}/2 + \cdots + e^{-Nx}/N$ which contains K. On the other hand, it is clear that every finite pile is contained in U_0.

Conclusion. *The series $1 + 1/4 + 1/9 + \cdots$ of squared reciprocals of positive integers has the exact sum $\pi^2/6$.*

In order to emphasize the conciseness of the solution to the Basel problem that we have just presented, let us recapitulate it again in formulas:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=1}^{\infty} \int_0^{\infty} \frac{e^{-nx}}{n}\,dx = -\int_0^{\infty} \log(1 - e^{-x})\,dx$$

$$= \int_{U_0} dx\,dy = \int_{T_0} d\alpha\,d\beta = \frac{\pi^2}{6}.$$

Here the notations U_0 and T_0 again refer to the sets appearing earlier, see Figure 4.

4. EPILOGUE: THE COMPLEX LOGARITHM EXPLAINS THE MYSTERY.
The most surprising part of the arguments in the preceding sections is probably the fact that the map G, connecting the logarithmic side lengths (x, y) and the angles (α, β), turned out to be area-preserving. Why on earth should the Jacobian of G be identically equal to 1?

In order to shed some light on this enigma, we shall take a look at logarithms of complex numbers. Vaguely put, it is their marvelous property

$$\log z = \log |z| + i \arg z,$$

combining (the logarithm of) the modulus $|z|$ with the argument $\arg z$, that explains the presence of a simple relation between the logarithmic side lengths (x, y) and the angles (α, β).

Let us now be more precise, and consider the linear equation $1 + z + w = 0$, with the unknowns z and w being complex numbers. Depicting 1, z, and w as vectors in the plane, one can interpret the fact that they sum to zero as saying the three vectors should form a closed triangle, quite similar to the one in Figure 1.

We now get $1 + z + w = 1 + Ae^{i(\pi - \beta)} + Be^{i(\alpha + \pi)} = 0$, and, taking real and imaginary parts of this complex equation, we see that it amounts to the two real equations

$$\begin{cases} A \sin \beta - B \sin \alpha = 0, & (3) \\ 1 - A \cos \beta - B \cos \alpha = 0. & (4) \end{cases}$$

It is an amusing exercise to verify that this system of equations is in fact equivalent to the sine and cosine rules (1) and (2). For instance, to deduce the cosine rule from (3) and (4), one can first rewrite (4) as $A \cos \beta = 1 - B \cos \alpha$. Squaring both sides, and using (3) to replace the squared left-hand side $A^2 \cos^2 \beta = A^2(1 - \sin^2 \beta)$ by $A^2 - B^2 \sin^2 \alpha$, one obtains the cosine rule $A^2 = 1 + B^2 - 2B \cos \alpha$.

Letting the arguments $\arg z$ and $\arg w$ be quite general, it is natural to "lift" the complex line $1 + z + w = 0$ to the exponential complex curve

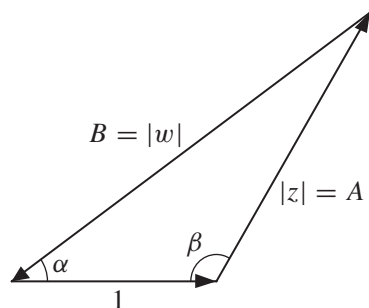$$X = \left\{ (s, t) \in \mathbf{C}^2 : 1 + e^s + e^t = 0 \right\}.$$

Figure 7. The triangle formed by the vectors 1, z, and w.

Denote by P and Q respectively the restrictions to X of the two linear projections

$$(s, t) \longmapsto (\operatorname{Re} s, \operatorname{Re} t) \quad \text{and} \quad (s, t) \longmapsto (\operatorname{Im} s, \operatorname{Im} t)$$

on the real and imaginary parts, and let $\widetilde{U} = P(X)$, $\widetilde{T} = Q(X)$ be the corresponding images in $\mathbf{R}^2$. Observe that Q is invertible on the interior of $\widetilde{T}$, so we can introduce a map $\widetilde{G} = P \circ Q^{-1}$ going from imaginary parts to real parts. This map $\widetilde{G}$ is closely related to our previous map G. In fact, one has the identity

$$G = -\widetilde{G} \circ L,$$

where L is the affine linear map $(\alpha, \beta) \mapsto (-\beta + \pi, \alpha + \pi)$. Since clearly both L and the sign-changing map $(x, y) \mapsto (-x, -y)$ are area-preserving, what remains to be explained is why $\widetilde{G}$ has its Jacobian identically equal to 1.

Fix any point $p \in X$ and let T_p denote the (real two-dimensional) tangent space to X at p. Since P and Q are (restrictions to X of) linear maps, their derivatives $P'(p)$ and $Q'(p)$ coincide with (restrictions to T_p of) the maps themselves. The Jacobian of $\widetilde{G}$ is therefore equal to the determinant of $P \circ Q^{-1}$, where we now use the same notation P, Q to mean the real and imaginary projections $T_p \to \mathbf{R}^2$.

Notice next that T_p is in fact a (one-dimensional) complex subspace of $\mathbf{C}^2$, so coordinate-wise multiplication by i defines a linear map J on T_p. Since J just amounts to a rotation (by an angle $\pi/2$) around the origin it is clear that J is area-preserving and thus has determinant 1. Moreover, one has $J = Q^{-1} \circ P$, so $P \circ Q^{-1} = Q \circ (Q^{-1} \circ P) \circ Q^{-1}$ also has determinant 1, which means that $\widetilde{G}$ indeed has its Jacobian identically equal to 1.

REFERENCES

1. M. Aigner and G. Ziegler, *Proofs from The Book*, Springer-Verlag, Berlin/Heidelberg, 2001.
2. R. Chapman, Evaluating $\zeta(2)$, Manuscript, 1999 (corrected 2003), available at `http://www.secamlocal.ex.ac.uk/~rjc/etc/zeta2.pdf`.
3. L. Euler, De summis serierum reciprocarum (On the sums of series of reciprocals), *Comm. Acad. Sci. Petrop.* **7** (1734/1735) 123–134. (Reprinted in *Opera Omnia I:14*, 1924, pp. 73–86.)
4. R. Osserman, *Two-Dimensional Calculus*, Harcourt, Brace & World, New York, 1968.
5. M. Passare and A. K. Tsikh, Amoebas: Their spines and their contours, *Contemp. Math.* **377** (2005) 275–288.
6. O. Viro, What is an amoeba? *Notices Amer. Math. Soc.* **49** (2002) 916–917.

Matematiska institutionen, Stockholms universitet, SE-10691 Stockholm, SWEDEN
passare@math.su.se