Lattice Polygons and the Number 12

Bjorn Poonen; Fernando Rodriguez-Villegas

The American Mathematical Monthly, Vol. 107, No. 3. (Mar., 2000), pp. 238-250.

Stable URL:
http://links.jstor.org/sici?sici=0002-9890%28200003%29107%3A3%3C238%3ALPATN1%3E2.0.CO%3B2-P

The American Mathematical Monthly is currently published by Mathematical Association of America.

Lattice Polygons and the Number 12

Bjorn Poonen and Fernando Rodriguez-Villegas

1. PROLOGUE. In this article, we discuss a theorem about polygons in the plane, which involves in an intriguing manner the number 12. The statement of the theorem is completely elementary, but its proofs display a surprisingly rich variety of methods, and at least some of them suggest connections between branches of mathematics that on the surface appear to have little to do with one another.

We describe four proofs of the main theorem, but we give full details only for proof 4, which uses modular forms. Proofs 2 and 3, and implicitly the theorem, appear in [**4**].

2. THE THEOREM. A *lattice polygon* is a polygon $\mathscr{P}$ in the plane $\mathbf{R}^2$ all of whose vertices lie in the lattice $\mathbf{Z}^2$ of points with integer coordinates. It is *convex* if for any two points P and Q in the polygon, the segment $\overline{PQ}$ is contained in the polygon. Let $l(\mathscr{P})$ be the total number of lattice points on the boundary of $\mathscr{P}$. If we define the *discrete length* of a line segment connecting two lattice points to be the number of lattice points on the segment (including the endpoints) minus 1, then $l(\mathscr{P})$ equals also the sum of the discrete lengths of the sides of $\mathscr{P}$.

We are interested in convex lattice polygons such that $(0, 0)$ is the only lattice point in the interior of $\mathscr{P}$. For such $\mathscr{P}$, we can define a *dual polygon* $\mathscr{P}^\vee$ as follows. Let $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ be the vectors representing the lattice points along the boundary of $\mathscr{P}$, in counterclockwise order. For convenience, indices are considered modulo n, so that $\mathbf{p}_{n+1} = \mathbf{p}_1$, etc. Define $\mathbf{q}_i$ to be the vector difference $\mathbf{p}_{i+1} - \mathbf{p}_i$. We will soon see that the kindergarten process of *connecting the dots* with straight lines from $\mathbf{q}_1$ to $\mathbf{q}_2$ to $\ldots$ to $\mathbf{q}_n$ and back to $\mathbf{q}_1$ traces out counterclockwise the boundary of a new convex lattice polygon whose only interior lattice point is $(0, 0)$; some of the $\mathbf{q}_i$ may coincide. For now, however, we define $\mathscr{P}^\vee$ simply as the convex hull of $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$; the *convex hull* of a set S is the smallest convex set containing S. One can show that $\mathscr{P}^{\vee\vee}$ is the 180° rotation of $\mathscr{P}$.

We are now ready to state the theorem.

Theorem 1. *Let $\mathscr{P}$ be a convex lattice polygon whose only interior lattice point is $(0, 0)$, and let $\mathscr{P}^\vee$ be its dual. Then $l(\mathscr{P}) + l(\mathscr{P}^\vee) = 12$.*

An instance of the theorem is illustrated in Figure 1.

3. OTHER MANIFESTATIONS OF 12. How can we "explain" the 12? One way would be to relate it to other appearances of 12 in mathematics. People in different fields brainstorming for an answer to the question "What is 12?" would likely produce widely varying results:

(A) To one who specializes in algebraic geometry, 12 might be the number appearing in *Noether's formula* $12(1 + p_a) = K^2 + c_2$, which relates certain integer invariants of an algebraic surface. This formula is a special case of

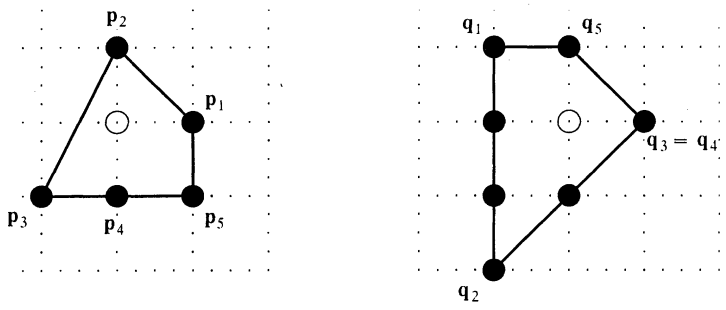

Figure 1. A polygon $\mathscr{P}$ and its dual. Note that 5 + 7 = 12.

the Hirzebruch-Riemann-Roch theorem [**6**, pp. 363, 432] and the 12 here comes from the coefficient of x^2 in the Taylor series

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} + \cdots,$$

whose n^{th} coefficient is $B_n/n!$, where B_n denotes the n^{th} Bernoulli number.

(B) To one who studies automorphic forms, 12 might be the weight of $\Delta(z)$, which is the cusp form of smallest weight for $\mathrm{SL}_2(\mathbf{Z})$.

(C) To one who dabbles in astrology, 12 might be the number of signs in the zodiac.

We relate the 12 in our theorem to (A) and (B) only!

4. THE PROOFS. We sketch four proofs, using the following:

1, Exhaustion
2. Stepping in the space of polygons
3. Toric varieties
4. Modular forms

The first two have the advantage of being completely elementary, but the last two do a better job of explaining the 12. Only the last proof is new, so it is the only one that we give in full.

We say that a polygon is *legal* if it is a convex lattice polygon in $\mathbf{R}^2$ and $\mathbf{o} := (0, 0)$ is its only interior lattice point.

5. PROOF 1: EXHAUSTION. Exhaustion means that we are going to list all legal polygons, and verify Theorem 1 for each, one at a time. If we take this literally, we will soon be truly exhausted, because there are infinitely many legal polygons.

To cut down the number of polygons we need to consider, we can define a notion of equivalence. Let $\mathrm{SL}_2(\mathbf{Z})$ (respectively, $\mathrm{GL}_2(\mathbf{Z})$) denote the group of 2-by-2 matrices $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $a, b, c, d \in \mathbf{Z}$ and $\det A = 1$ (respectively, $\det A \in \{1, -1\}$). Every matrix $A \in \mathrm{GL}_2(\mathbf{Z})$ determines a linear transformation of the plane $\mathbf{R}^2$ that maps $\mathbf{Z}^2$ bijectively onto itself, so A maps legal polygons to legal polygons.

We say that two legal polygons $\mathscr{P}$ and $\mathscr{Q}$ are *equivalent* if there exists an $A \in \mathrm{GL}_2(\mathbf{Z})$ that transforms $\mathscr{P}$ into $\mathscr{Q}$. In that case, $l(\mathscr{P}) = l(\mathscr{Q})$, and A transforms $\mathscr{P}^\vee$ into $\mathscr{Q}^\vee$ or the 180° rotation of $\mathscr{Q}^\vee$, depending on the sign of $\det A$.

Hence proving Theorem 1 for $\mathscr{P}$ is the same as proving it for $\mathscr{E}$. If we knew that there were only finitely many equivalence classes of legal polygons, and if we could find a list of legal polygons representing these classes, then we could prove Theorem 1 by checking the polygons on this list.

The desired finiteness does hold, and in fact, much more is true. For $d \geq 2$, a *convex lattice polytope* in $\mathbf{R}^d$ is the convex hull $\mathscr{P}$ of a finite set of points with integer coordinates, such that the points are not all contained in a hyperplane; this ensures that $\mathscr{P}$ is d-dimensional. Hensley [7] bounded the volume and the number of boundary lattice points in terms of d and the number k of interior lattice points when $k \geq 1$; this result for $d = 2$ was proved earlier by Scott [14], and Hensley's bounds have been improved in [11]. These results easily imply the following:

Theorem 2. *Fix integers $d \geq 2$ and $k \geq 1$. Up to the action of $\mathrm{GL}_d(\mathbf{Z})$ and translation by lattice points, there are only finitely many convex lattice polytopes in $\mathbf{R}^n$ having exactly k interior lattice points.*

Remark. Theorem 2 would be false if we allowed $k = 0$.

In the case of interest ($d = 2$ and $k = 1$), there are exactly 16 equivalence classes. Polygons representing these equivalence classes are listed in Figure 2. We leave it to the reader to pair them with their duals (up to equivalence), and to verify that Theorem 1 holds. Some polygons are self-dual.

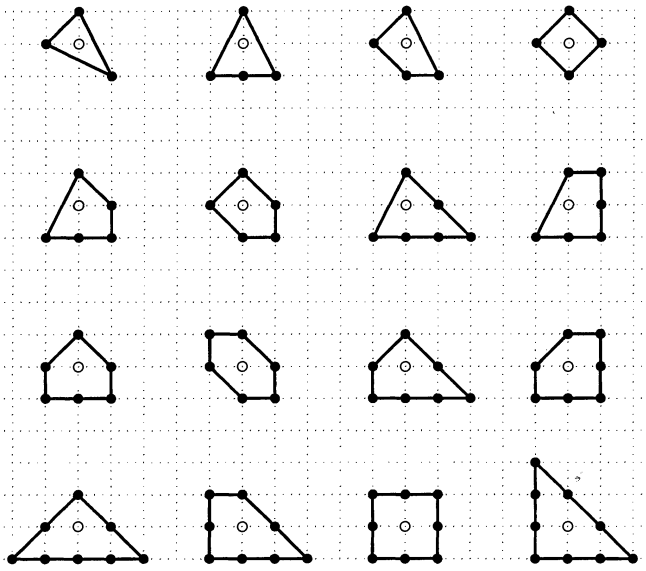

Figure 2. The 16 equivalence classes of legal polygons.

Although this method does prove Theorem 1, it is not very satisfying, because it does not really *explain* anything.

6. PROOF 2: STEPPING IN THE SPACE OF POLYGONS. We say that two legal polygons are *neighbors* if there is a counterclockwise labeling $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \ldots, \mathbf{p}_n$ of the boundary lattice points of one of them, such that the boundary lattice points of the other are $\mathbf{p}_1, \mathbf{p}_1 + \mathbf{p}_2, \mathbf{p}_2, \mathbf{p}_3, \ldots, \mathbf{p}_n$.

Warning. Suppose that $\mathscr{P}$ is a legal polygon with a counterclockwise labeling $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ of its boundary lattice points. Then the polygon with boundary lattice points $\mathbf{p}_1, \mathbf{p}_1 + \mathbf{p}_2, \mathbf{p}_2, \mathbf{p}_3, \ldots, \mathbf{p}_n$ need not be legal: it may fail to be convex.

We can prove Theorem 1 by

1. verifying it for a single legal polygon $\mathscr{P}_0$;
2. proving that for any $\mathscr{P}$, there is a sequence of legal polygons starting from $\mathscr{P}_0$ and ending with $\mathscr{P}$, such that each polygon is a neighbor of its successor; and
3. checking that the value of $l(\mathscr{P}) + l(\mathscr{P}^\vee)$ is unchanged when $\mathscr{P}$ is replaced by a neighboring legal polygon.

This method of proof is discussed in a series of exercises in [**4**, Section 2.5].

Most people would agree that such a proof is more satisfying than the one in Section 5: it explains why $l(\mathscr{P}) + l(\mathscr{P}^\vee)$ must be constant. But it does *not* explain why that constant should be 12.

7. PROOF 3: TORIC VARIETIES. We do not want to go too deeply into the geometry of toric varieties here; readers with a basic knowledge of algebraic geometry who want more details are encouraged to consult [**4**], especially Sections 2.5 and 4.3. Other readers may choose to skip this section; it is not needed in the rest of the paper.

To any legal polygon $\mathscr{P}$, one can associate a 2-dimensional toric variety $T_{\mathscr{P}}$, which is a kind of algebraic variety. The condition that $\mathbf{o}$ be the only interior lattice point of $\mathscr{P}$ is exactly what is required to make the surface $T_{\mathscr{P}}$ nonsingular. The arithmetic genus p_a, the self-intersection K^2 of the canonical bundle, and the second Chern class c_2 of the tangent bundle are geometric invariants of the surface $T_{\mathscr{P}}$ that can be expressed in terms of the combinatorics of $\mathscr{P}$; in fact it turns out that they equal 0, $l(\mathscr{P})$, and $l(\mathscr{P}^\vee)$, respectively. Hence Theorem 1 is simply a restatement of Noether's formula $12(1 + p_a) = K^2 + c_2$ for 2-dimensional nonsingular toric varieties. If one proves Noether's formula for surfaces in general using algebraic-geometric methods, as done in [**9**, p. 154], one obtains a new proof of Theorem 1. The point of view taken in [**4**] is the reverse; the combinatorial proof of Section 6 can be used to give an independent proof of Noether's formula in the special case of toric surfaces.

This proof is just one instance of a great exchange that has been taking place between two disciplines. Results in the combinatorics of lattices are being used to prove results about toric varieties, and vice versa, to the benefit of both sides.

Remark. The proof here and the proof of the previous section are related, as discussed in [**4**, Section 2.5]. If $\mathscr{P}$ and $\mathscr{Q}$ are neighboring legal polygons and $\mathscr{Q}$ is the one with more boundary lattice points, then $T_{\mathscr{Q}}$ can be constructed geometrically from $T_{\mathscr{P}}$ by "blowing up a point" [**6**, p. 28]. The relation $p_a = 0$ for toric varieties is implied by the classical fact that the arithmetic genus p_a of a nonsingular surface is a birational invariant.

8. PROOF 4: MODULAR FORMS. We now give the final proof, using transformation properties of the logarithm of the modular form $\Delta(z)$. No prior knowledge of modular forms is required (we state a few facts without proof), but we use some undergraduate topology and complex analysis.

8.1. Interpretation via matrices. Let $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ be as in Section 2.

Lemma 3. *The vectors* $\mathbf{p}_1$ *and* $\mathbf{p}_2$ *form a basis for the lattice* $\mathbf{Z}^2$ *that has the same orientation as the standard basis* $(1, 0)$, $(0, 1)$.

Proof: Since the $\mathbf{p}_i$ were chosen in counterclockwise order, it suffices to show that $\mathbf{p}_1$ and $\mathbf{p}_2$ span $\mathbf{Z}^2$. If v is a lattice point not of the form $m \cdot \mathbf{p}_1 + n \cdot \mathbf{p}_2$, then translating v by such a combination yields a new lattice point w in the closed parallelogram P that has the vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ as sides, such that w is not a vertex of P. Replacing w by $\mathbf{p}_1 + \mathbf{p}_2 - w$ if necessary, we can assume that w is in or on the triangle with vertices $\mathbf{o}, \mathbf{p}_1, \mathbf{p}_2$ but is not equal to any of these vertices. Since $\mathscr{P}$ has no interior lattice point other than $\mathbf{o}$, the point w cannot be in the interior of the triangle. For the same reason it cannot be in the interior of either side with endpoint $\mathbf{o}$. Hence w is in the interior of the segment joining $\mathbf{p}_1$ to $\mathbf{p}_2$. This also is a contradiction, since then w should have been listed as a boundary lattice point between $\mathbf{p}_1$ and $\mathbf{p}_2$. ■

Remark. Lemma 3 could also be derived by applying Pick's formula [**4**, p. 113] to the triangle with vertices $\mathbf{o}, \mathbf{p}_1, \mathbf{p}_2$. Pick's formula states that the area of a lattice polygon equals $I + B/2 - 1$, where I is the number of interior lattice points, and B is the number of boundary lattice points. See [**3**] and [**5**] for further discussion of this formula.

Lemma 3 implies that the 2×2 matrix $\begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}$ whose rows are $\mathbf{p}_1$ and $\mathbf{p}_2$ written as row vectors belongs to $\mathrm{SL}_2(\mathbf{Z})$. Similarly, $\begin{bmatrix} \mathbf{p}_2 \\ \mathbf{p}_3 \end{bmatrix} \in \mathrm{SL}_2(\mathbf{Z})$. The matrix M such that

$$M \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{p}_2 \\ \mathbf{p}_3 \end{bmatrix}$$

has the form $M = \begin{bmatrix} 0 & 1 \\ c & d \end{bmatrix}$ Since $M \in \mathrm{SL}_2(\mathbf{Z})$ also, $c = -1$. In general, for each i, we have a matrix $M_i = \begin{bmatrix} 0 & 1 \\ -1 & d_i \end{bmatrix}$ such that

$$M_i \begin{bmatrix} \mathbf{p}_{i-1} \\ \mathbf{p}_i \end{bmatrix} = \begin{bmatrix} \mathbf{p}_i \\ \mathbf{p}_{i+1} \end{bmatrix}$$

With respect to the basis $\{\mathbf{p}_{i-1}, \mathbf{p}_i\}$ we then have

$$\begin{aligned} &\mathbf{p}_{i-1} = (1, 0), \quad \mathbf{p}_i = (0, 1), \quad \mathbf{p}_{i+1} = (-1, d_i), \\ &\mathbf{q}_{i-1} = (-1, 1), \quad \mathbf{q}_i = (-1, d_i - 1), \end{aligned} \tag{1}$$

and $\mathbf{q}_i - \mathbf{q}_{i-1} = n_i \mathbf{p}_i$ where $n_i = d_i - 2$. Since $\mathscr{P}$ is convex, the point $\mathbf{p}_{i+1}$ of $\mathscr{P}$ must lie in the half plane $x + y \leq 1$ below the line through $\mathbf{p}_i$ and $\mathbf{p}_{i-1}$; this forces $n_i \leq 0$. If $\mathbf{p}_i$ is a vertex (as opposed to being in the interior of one of the sides), then $n_i < 0$. At this point, we can explain why the dual of a legal polygon is legal.

Proposition 4. *If* $\mathscr{P}$ *is a legal polygon, then drawing segments from* $\mathbf{q}_1$ *to* $\mathbf{q}_2$ *to* ... *to* $\mathbf{q}_n$ *and back to* $\mathbf{q}_1$ *results in a new legal polygon. In particular,* $\mathscr{P}^\vee$ *is legal.*

Proof: As we traverse the boundary of $\mathscr{P}$ once in a counterclockwise direction, the direction we face also rotates 360° counterclockwise. Since the vectors $\mathbf{q}_i$ are

translates of the vectors forming the sides of the polygon (the vectors that show the direction of our motion), the segments between them trace out some lattice polygon $\mathscr{P}'$ that contains $\mathbf{o}$.

Using (1), we see that if $\mathbf{q}_{i-1} \neq \mathbf{q}_i$, then the only lattice points inside or on the solid triangle with vertices $\mathbf{o}$, $\mathbf{q}_{i-1}$, $\mathbf{q}_i$ are $\mathbf{o}$ and those on the side joining $\mathbf{q}_{i-1}$, $\mathbf{q}_i$. This holds for all i, so $\mathbf{o}$ is the only interior lattice point of $\mathscr{P}'$.

It remains to show that $\mathscr{P}'$ is convex. If $\mathbf{p}_i$ is a vertex of $\mathscr{P}$ (or equivalently, $n_i \neq 0$), then we may characterize $\mathbf{q}_{i-1}, \mathbf{q}_i$ as the primitive vectors along sides of $\mathscr{P}$ coming in and out of $\mathbf{p}_i$, respectively, when $\mathscr{P}$ is given the counterclockwise orientation. A vector (a, b) is *primitive* if $\gcd(a, b) = 1$.

Now let $\mathbf{p}_i, \mathbf{p}_j$, with $i < j$, be two consecutive vertices of $\mathscr{P}$ that determine a side of $\mathscr{P}$ of length m. Note that $\mathbf{q}_i = \mathbf{q}_{j-1}$. By (1), $\mathbf{q}_i - \mathbf{q}_{i-1} = n_i\mathbf{p}_i$ and $\mathbf{q}_j - \mathbf{q}_{j-1} = n_j\mathbf{p}_j$. Then, since $\mathbf{p}_i$ and $\mathbf{p}_j$ must be linearly independent, $\mathbf{q}_i$ is a vertex of $\mathscr{P}'$. Since $\mathbf{p}_i$ and $\mathbf{p}_j$ are vertices, we have $n_i < 0$ and $n_j < 0$, so $-\mathbf{p}_i$ and $-\mathbf{p}_j$ are the primitive vectors along sides of $\mathscr{P}'$ coming in and out of $\mathbf{q}_i$, respectively. Finally, $\mathbf{p}_j - \mathbf{p}_i = m\mathbf{q}_i$, so $\mathbf{o} = \mathbf{q}_i + (1/m)\mathbf{p}_i + (1/m)(-\mathbf{p}_j)$ is in the cone $\{\mathbf{q}_i + \alpha\mathbf{p}_i + \beta(-\mathbf{p}_j) : \alpha, \beta \in \mathbf{R}_{\geq 0}\}$ determined by the sides of $\mathscr{P}'$ at $\mathbf{q}_i$. It follows that $\mathscr{P}'$ is convex. ■

Since

$$M_nM_{n-1} \cdots M_1\begin{bmatrix}\mathbf{p}_0\\ \mathbf{p}_1\end{bmatrix} = \begin{bmatrix}\mathbf{p}_n\\ \mathbf{p}_{n+1}\end{bmatrix} = \begin{bmatrix}\mathbf{p}_0\\ \mathbf{p}_1\end{bmatrix},$$

we have

$$\begin{bmatrix}0 & 1\\ -1 & d_n\end{bmatrix}\begin{bmatrix}0 & 1\\ -1 & d_{n-1}\end{bmatrix}\cdots\begin{bmatrix}0 & 1\\ -1 & d_1\end{bmatrix} = I, \tag{2}$$

the identity matrix.

Using (1), we find that the discrete length of the segment from $\mathbf{q}_{i-1}$ to $\mathbf{q}_i$ is $|(d_i - 1) - 1| = 2 - d_i$, since $d_i - 2 \leq 0$. Hence $l(\mathscr{P}^\vee) = \sum_{i=1}^{n}(2 - d_i)$. On the other hand, $l(\mathscr{P}) = n$, so Theorem 1 is equivalent to $\sum_{i=1}^{n}(3 - d_i) = 12$. But this equality cannot be a consequence of (2) alone: if we took the sequence of matrices M_i corresponding to $\mathscr{P}$ and repeated it twice, the resulting sequence would still multiply to give the identity, but now the sum of the $(3 - d)$'s would be 24. We need somehow to incorporate the information that our polygon $\mathscr{P}$ winds exactly once around the origin. To do this we lift (2) to an equation in an extension of the group $\mathrm{SL}_2(\mathbf{Z})$ by $\mathbf{Z}$, in which the $\mathbf{Z}$ keeps track of the winding number.

8.2. The universal cover $\widetilde{\mathrm{SL}_2(\mathbf{R})}$ of $\mathrm{SL}_2(\mathbf{R})$. Let $\mathrm{SL}_2(\mathbf{R})$ denote the group of 2-by-2 matrices $A = \left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right]$ such that $a, b, c, d \in \mathbf{R}$ and $\det A = 1$. As a topological space, $\mathrm{SL}_2(\mathbf{R})$ is the set of oriented bases of $\mathbf{R}^2$ of determinant one. Geometric intuition suggests that the only homotopy invariant of a loop in this space is "the number of times the basis gets rotated around the origin", so that the fundamental group $\pi_1(\mathrm{SL}_2(\mathbf{R}))$ should be $\mathbf{Z}$; we recommend [**12**] as an introduction to fundamental groups and universal covers.

This is not hard to prove rigorously; for example, one could use the *Iwasawa decomposition*, which for $\mathrm{SL}_2(\mathbf{R})$ says that each $M \in \mathrm{SL}_2(\mathbf{R})$ can be factored uniquely as

$$M = \begin{bmatrix}1 & u\\ 0 & 1\end{bmatrix}\begin{bmatrix}a & 0\\ 0 & a^{-1}\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$$

with $u \in \mathbf{R}$, $a \in \mathbf{R}^+$, and $0 \le \theta < 2\pi$. Thus there is a homeomorphism

$$\mathrm{SL}_2(\mathbf{R}) \approx \mathbf{R} \times \mathbf{R}^+ \times \mathbf{R}/2\pi\mathbf{Z}$$

(not a homomorphism). Since $\mathbf{R}$ and $\mathbf{R}^+$ are simply connected, and since the last factor $\mathbf{R}/2\pi\mathbf{Z}$ has simply connected cover $\mathbf{R}$ with covering group $2\pi\mathbf{Z} \cong \mathbf{Z}$, we find that $\pi_1(\mathrm{SL}_2(\mathbf{R})) = \mathbf{Z}$.

We let $\widetilde{\mathrm{SL}_2(\mathbf{R})}$ denote the universal cover of $\mathrm{SL}_2(\mathbf{R})$, which is a connected topological group fitting into an exact sequence

$$0 \to \mathbf{Z} \to \widetilde{\mathrm{SL}_2(\mathbf{R})} \to \mathrm{SL}_2(\mathbf{R}) \to 0.$$

Although $\widetilde{\mathrm{SL}_2(\mathbf{R})}$ cannot be described as a subgroup of matrices satisfying algebraic conditions [**18**, Exercise 15(b), p. 137], we can give a fairly concrete description of it. Consider pairs $(M, [\gamma])$, where

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbf{R})$$

and $[\gamma]$ is the *path-homotopy class* [**12**, p. 319] of a path γ in $\mathbf{R}^2 \setminus \mathbf{o}$ from $\begin{bmatrix} 0 & 1 \end{bmatrix} I$ to $\begin{bmatrix} 0 & 1 \end{bmatrix} M$; i.e., from $(0, 1)$ to (c, d). Every $M \in \mathrm{SL}_2(\mathbf{R})$ acts on the right on $\mathbf{R}^2$ (whose elements we identify with row vectors), and hence also acts on paths in $\mathbf{R}^2 \setminus \mathbf{o}$. We obtain a group structure on the set of pairs $(M, [\gamma])$ by letting

$$(M_1, [\gamma_1]) \cdot (M_2, [\gamma_2]) = \left(M_1 M_2, \left[\gamma_2 + \gamma_1^{M_2}\right]\right),$$

where $\gamma_1^{M_2}$ denotes the action described in the previous sentence, and $+$ denotes the join of two paths sharing an endpoint. The matrix M_2 transforms γ_1 into a path that begins where γ_2 ends. This group is connected, and it covers $\mathrm{SL}_2(\mathbf{R})$ with covering group $\mathbf{Z}$, so it is isomorphic to $\widetilde{\mathrm{SL}_2(\mathbf{R})}$.

8.3. The extension $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$ of $\mathrm{SL}_2(\mathbf{Z})$. Our desired extension of $\mathrm{SL}_2(\mathbf{Z})$ by $\mathbf{Z}$ is the preimage of $\mathrm{SL}_2(\mathbf{Z})$ under the covering map $\widetilde{\mathrm{SL}_2(\mathbf{R})} \to \mathrm{SL}_2(\mathbf{R})$. We call this group $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$, even though $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$ is, of course, not the universal cover of $\mathrm{SL}_2(\mathbf{Z})$, since $\mathrm{SL}_2(\mathbf{Z})$ is a discrete group.

Now let M_i and d_i be as in Section 8.1 and furthermore assume without loss of generality that $\mathbf{p}_0 = (1, 0)$ and $\mathbf{p}_1 = (0, 1)$. Let γ_i be the straight-line path from $(0, 1)$ to $(-1, d_i)$. Then by induction on j, we have

$$M_j M_{j-1} \cdots M_1 = \begin{bmatrix} \mathbf{p}_j \\ \mathbf{p}_{j+1} \end{bmatrix}$$

and $M_j M_{j-1} \cdots M_1$ transforms γ_{j+1} into the straight-line path from $\mathbf{p}_{j+1}$ to $\mathbf{p}_{j+2}$. Hence

$$(M_j, [\gamma_j]) \cdot (M_{j-1}, [\gamma_{j-1}]) \cdots (M_1, [\gamma_1]) = \left(M_j M_{j-1} \cdots M_1, [\Gamma_j]\right)$$

where Γ_j is the polygonal path from $\mathbf{p}_1$ to $\mathbf{p}_2$ to $\ldots$ to $\mathbf{p}_{j+1}$. Taking $j = n$, we obtain

$$(M_n, [\gamma_n]) \cdot (M_{n-1}, [\gamma_{n-1}]) \cdots (M_1, [\gamma_1]) = (I, \mathrm{loop}), \tag{3}$$

where *loop* denotes the path-homotopy class of a counterclockwise loop around $\mathbf{o}$.

Remark. The group we are calling $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$ can be presented in terms of generators and relations either as $\langle a, b : aba = bab \rangle$ or as $\langle x, y : x^2 = y^3 \rangle$, where a, b, x, y

equal (respectively)

$$\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\ \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix},\ \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix},\ \begin{bmatrix}0 & -1\\ 1 & 1\end{bmatrix}$$

equipped with the straight-line paths. It is isomorphic also to groups occurring naturally in several other contexts:

- the fundamental group of $\mathbf{R}^3 \setminus T$, where T is a trefoil knot,
- the braid group B_3 (an isomorphism being given, for example, by sending a, b to standard generators of B_3 [**2**, p. 18]), and
- the local fundamental group of the ordinary cusp singularity, which also equals the fundamental group of $\mathbf{C}^2 \setminus \{(x, y) \in \mathbf{C}^2 : y^2 = x^3\}$, where $\mathbf{C}$ denotes the field of complex numbers.

See [**16**, p. 11]

8.4. The modular form $\Delta(z)$ and its logarithm. Let $\mathscr{H} = \{z \in \mathbf{C} : \mathrm{Im}(z) > 0\}$ denote the upper half plane. Matrices

$$M = \begin{bmatrix}a & b\\ c & d\end{bmatrix} \in \mathrm{SL}_2(\mathbf{R})$$

act on points $z \in \mathscr{H}$ according to the rule: $Mz := (az + b)/(cz + d)$.

An excellent introduction to the theory of modular forms is given in [**15**]. In fact, the only external fact we need is that there exists a holomorphic function $\Delta(z)$ on $\mathscr{H}$ such that

(a) $\Delta(z) \neq 0$ for all $z \in \mathscr{H}$; and
(b) $\Delta(Mz) = (cz + d)^{12}\Delta(z)$ for all $M = \begin{bmatrix}a & b\\ c & d\end{bmatrix} \in \mathrm{SL}_2(\mathbf{Z})$ and $z \in \mathscr{H}$.

Property (b) is part of what it means for a function to be a modular form. To be precise, if k is an integer, a *modular form of weight k* for the group $\mathrm{SL}_2(\mathbf{Z})$ is a holomorphic function f on $\mathscr{H}$ such that

(i) $f(Mz) = (cz + d)^k f(z)$ for all $M = \begin{bmatrix}a & b\\ c & d\end{bmatrix} \in \mathrm{SL}_2(\mathbf{Z})$ and $z \in \mathscr{H}$; and
(ii) f is "holomorphic at infinity;" i.e., it has a Fourier expansion of the form $f(z) = \sum_{n=0}^{\infty} a_n e^{2\pi i n z}$.

Note that condition (i) applied with $M = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ yields $f(z + 1) = f(z)$, which implies that f has a Fourier expansion of the form $f(z) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n z}$; the point of condition (ii) is to require that $a_n = 0$ *for* $n < 0$. If in addition the Fourier coefficient a_0 is zero, then f is said to "vanish at the cusp ∞", and f is called a *cusp form*.

The function we need is given by the following striking result, which is proved, in [**15**].

Theorem 5. *The set of cusp forms of weight* 12 *for* $\mathrm{SL}_2(\mathbf{Z})$ *is a* 1*-dimensional vector space over* $\mathbf{C}$, *spanned by*

$$\Delta(z) := (2\pi)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24} \quad \text{where } q := e^{2\pi i z}.$$

The fact that $\Delta(z) \neq 0$ for $z \in \mathscr{H}$ follows from the convergence of the product for $|q| < 1$. Also, $\mathscr{H}$ is simply connected, so we may fix once and for all a branch of

$\log \Delta(z)$ on $\mathscr{H}$. Then

$$\log \Delta(Mz) - \log \Delta(z) = 12 \log(cz + d) + 2\pi i m \tag{4}$$

for some integer m depending on the choice of branch of $\log(cz + d)$. If when $c \neq 0$ one chooses the branch of $\log(cz + d)$ so that its range is the same as that of $\log(z)$ in $\mathscr{H}$, then the integer m is related to Dedekind sums [**13**, p. 47]. The complications arising from Dedekind sums and their transformation formulas can be avoided, however, if we equip M with a path as in Section 8.2.

If $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbf{R})$, then having a path γ from $(0, 1)$ to (c, d) in $\mathbf{R}^2 \setminus \mathbf{o}$ lets us make a *canonical* choice of branch of $\log(cz + d)$: for fixed $z \in \mathscr{H}$, we set $\log(0 \cdot z + 1) = 0$ and then make $\log(c'z + d')$ a continuous function of the path parameter, as (c', d') moves from $(0, 1)$ to (c, d). Moreover this choice of branch depends only on the path-homotopy class of γ; we call it $L(M, [\gamma]; z)$. For $(M, [\gamma]) \in \widetilde{SL_2(\mathbf{Z})}$,

$$\log \Delta(Mz) - \log \Delta(z) = 12L(M, [\gamma]; z) + 2\pi i \Phi(M, [\gamma]) \tag{5}$$

now defines a function $\Phi : \widetilde{\mathrm{SL}_2(\mathbf{Z})} \to \mathbf{Z}$.

If $(M_1, [\gamma_1]) \cdot (M_2, [\gamma_2]) = (M_3, [\gamma_3])$ in $\widetilde{\mathrm{SL}_2(\mathbf{R})}$, and if $[c_3 \quad d_3]$ is the bottom row of M_3, then a computation shows that

$$L(M_1, [\gamma_1]; M_2 z) + L(M_2, [\gamma_2]; z) \quad \text{and} \quad L(M_3, [\gamma_3]; z)$$

are both branches of $\log(c_3 z + d_3)$ on $\mathscr{H}$, so they differ by $2\pi i N((M_1, [\gamma_1]), (M_2, [\gamma_2]))$ for some integer-valued function $N : \widetilde{\mathrm{SL}_2(\mathbf{R})} \times \widetilde{\mathrm{SL}_2(\mathbf{R})} \to \mathbf{Z}$. But N is continuous, $\widetilde{\mathrm{SL}_2(\mathbf{R})}$ is connected, and $\mathbf{Z}$ is discrete, so the image of N is constant. Evaluating N when both arguments are the identity in $\widetilde{\mathrm{SL}_2(\mathbf{R})}$ shows that N is identically zero. Now adding (5) for $(M_2, [\gamma_2])$ to the corresponding equation for $(M_1, [\gamma_1])$ with z replaced by $M_2 z$, and comparing with (5) for $(M_1, [\gamma_1])(M_2, [\gamma_2])$, we find that $\Phi : \widetilde{\mathrm{SL}_2(\mathbf{Z})} \to \mathbf{Z}$ is a homomorphism.

Remark. The modular form Δ was used only to construct Φ. There are other, more elementary means to construct Φ, but these are also more *ad hoc*. For instance, we could have used one of the explicit presentations of $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$ mentioned in Section 8.3.

8.5. Values of the homomorphism Φ. Fix $z \in \mathscr{H}$. As (c, d) winds around $\mathbf{o}$ once in the counterclockwise direction, $cz + d$ winds around $0 \in \mathbf{C}$ once in the *clockwise* direction. Hence by definition $L(I, \text{loop}; z) = -2\pi i$ and $\Phi(I, \text{loop}) = 12$.

Let

$$S := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad T := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Let $\tilde{S}$ and $\tilde{T}$ be the elements of $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$ obtained by equipping S with the straight-line path from $(0, 1)$ to $(1, 0)$, and T with the trivial path. It is known that S and T generate $\mathrm{SL}_2(\mathbf{Z})$ (and in fact $\tilde{S}$ and $\tilde{T}$ generate $\widetilde{\mathrm{SL}_2(\mathbf{Z})}$), but we will not need to use this. A short calculation shows that

$$\tilde{S}^{-1} 1 \cdot \tilde{T}^{-d} = \left(\begin{bmatrix} 0 & 1 \\ -1 & d \end{bmatrix}, [\gamma] \right),$$

where γ is the straight-line path from $(0, 1)$ to $(-1, d)$. Applying (3) to the diamond (the 4th polygon in Figure 2) shows that

$$(\tilde{S}^{-1})^4 = (I, \text{loop})$$

and applying Φ to both sides shows that $4\Phi(\tilde{S}^{-1}) = 12$, so $\Phi(\tilde{S}) = -3$. Similarly, applying (3) to the hexagon (the 10^{th} polygon in Figure 2) shows that

$$\left(\tilde{S}^{-1} \cdot \tilde{T}^{-1}\right)^6 = (I, \text{loop})$$

so $\Phi(S^{-1} \cdot \tilde{T}^{-1}) = 2$, and $\Phi(\tilde{T}) = 1$. (Alternatively, one could calculate $\Phi(\tilde{S}) = -3$ and $\Phi(\tilde{T}) = 1$ directly from (5), using the fixed point $z = i$ of S for the former.) Hence

$$\Phi\left(\begin{bmatrix} 0 & 1 \\ -1 & d \end{bmatrix}, [\gamma]\right) = \Phi\left(\tilde{S}^{-1} \cdot \tilde{T}^{-d}\right) = 3 - d.$$

Applying Φ to (3) shows that $\sum_{i=1}^{n}(3 - d_i) = 12$, which, as we saw in Section 8.1, is equivalent to Theorem 1.

9. GENERALIZATIONS. Theorem 1 can be generalized in various ways. We do not know, however, how to generalize it to polygons with more than one interior lattice point.

9.1. Legal loops. These are generalizations of legal polygons. A *legal loop* $\mathscr{L}$ is a closed path in the plane formed by consecutive legal moves. A *legal move* is an oriented segment joining two lattice points such that its initial and end point together with the origin form a non-degenerate triangle with no other lattice point (in the interior or on its boundary) except those three; equivalently, by Pick's formula, we could require the triangle to have area $1/2$. Clearly, a legal polygon is a legal loop. Notice that we do not require a legal loop to have acute angles at its corners, and hence its winding number w with respect to $\mathbf{o}$ (in the sense of algebraic topology) can be an arbitrary integer. Boundary lattice points now have to be counted with a sign; precisely, if s is an oriented segment joining two lattice points $\mathbf{p}$ and $\mathbf{p}'$, and if there are k lattice points on s, then define

$$l(s) = (k - 1) \cdot \det\begin{bmatrix} \mathbf{p} \\ \mathbf{p}' \end{bmatrix}.$$

The determinant is ± 1. We then define $l(\mathscr{L})$ as the sum of $l(s)$ over the oriented segments s forming the loop.

Let

$$\mathbf{q}_i = \left(\det\begin{bmatrix} \mathbf{p}_i \\ \mathbf{p}_{i+1} \end{bmatrix}\right) \cdot (\mathbf{p}_{i+1} - \mathbf{p}_i),$$

and define the dual $\mathscr{L}^\vee$ to be the loop obtained by "connecting the dots" from $\mathbf{q}_1$ to $\mathbf{q}_2$ to ... to $\mathbf{q}_n$ and back to $\mathbf{q}_1$, listing also the lattice points *along* the segments drawn. It is not hard to see that $\mathscr{L}^\vee$ is again a legal loop with the same winding number w. See Figure 3 for an example. The statement of Theorem 1 becomes

$$l(\mathscr{L}) + l(\mathscr{L}^\vee) = 12 \cdot w.$$

Of all the proofs of Theorem 1 that we have discussed, proof 4 seems to be the best suited for this generalization. Proof 3 is difficult to generalize, because the natural object associated to a legal loop by the construction of Section 7 need not be an algebraic variety; it could also be a "non-separated scheme", which is the algebraic analogue of a non-Hausdorff topological space. Hence Noether's formula as stated does not apply.

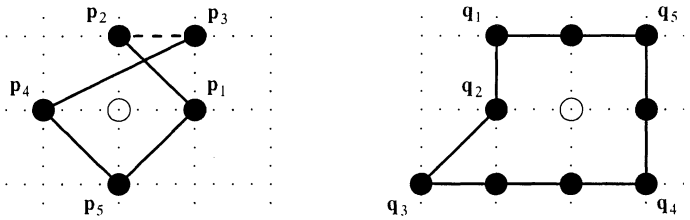

Figure 3. A legal loop $\mathscr{L}$ of winding number 1 and its dual. Because the dotted line counts as having length -1, $l(\mathscr{L}) = 3$. Note that $3 + 9 = 12 \cdot 1$.

9.2. Higher dimensions. Theorem 1 has generalizations to higher dimensions, but the statements are not as simple, and in fact we do not give any explicitly here. The appropriate notion of legal polytope would have to be that of a *reflexive polytope*; i.e., a convex lattice polytope that can be described as the solution set of a system of linear inequalities of the form $a_1x_1 + \cdots + a_dx_d \leq 1$ with $a_1, \ldots, a_d$ integers with gcd 1. The dual polytope (usually called the *polar* polytope) is defined as the convex hull of the points $(a_1, \ldots, a_d)$. Our original definition of dual polygon differs from this one when $d = 2$ by a 90° rotation. A reflexive polytope has $\mathbf{o}$ as its only interior lattice point, but in any dimension $d > 2$ there exist *non-reflexive* convex lattice polytopes having $\mathbf{o}$ as the only interior lattice point [**8**]. Applying the Hirzebruch-Riemann-Roch theorem to the toric variety associated to a reflexive polytope gives information about the combinatorics of the polytope.

Physicists are interested in reflexive polytopes because of their relation to mirror symmetry [**1**]; in fact, they now have a complete classification in dimension 3 [**10**]: there are 4,319 of them! The complete list and other interesting information may be obtained from the website `http://tph16.tuwien.ac.at/~kreuzer/CY.html`.

10. GAUSS-BONNET?. There is a potential connection of Theorem 1 to the Gauss-Bonnet theorem. For a geodesic polygon $\mathscr{P}$ on a surface of constant curvature c, the classical Gauss-Bonnet theorem [**17**, pp. 247–250] states that

$$c \cdot \text{Area}(\mathscr{P}) + \sum(\text{exterior angles}) = 2\pi. \tag{6}$$

For example, on a sphere of a radius 1, if a "triangle" bounded by arcs of great circles has interior angles α, β, γ measured in radians, then its area is $\alpha + \beta + \gamma - \pi$.

Now suppose instead that $\mathscr{P}$ is one of our legal lattice polygons. If $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ are two adjacent boundary lattice points, as in Section 8.1, then the area of the triangle with vertices $\mathbf{o}, \mathbf{p}_i, \mathbf{p}_{i+1}$ equals

$$\frac{1}{2}\det\begin{bmatrix}\mathbf{p}_i \\ \mathbf{p}_{i+1}\end{bmatrix} = \frac{1}{2}.$$

Summing over i shows that $\text{Area}(\mathscr{P}) = n/2$, where $n = l(\mathscr{P})$; alternatively, this follows from Pick's formula.

Recall that the nonnegative integer $2 - d_i$ had an interpretation as the discrete length of a segment in the dual polygon. We now explain why it can also be interpreted as a combinatorial analogue of an exterior angle. With respect to the basis $\{\mathbf{p}_{i-1}, \mathbf{p}_i\}$ of $\mathbf{Z}^2$, we have $\mathbf{p}_{i-1} = (1, 0)$ and $\mathbf{p}_i = (0, 1)$, and $\mathbf{p}_{i+1} = (-1, d_i)$. If $d_i = 2$, then at $\mathbf{p}_i$ there is a "straight angle"; as d_i decreases, the exterior angle (with respect to our basis) at $\mathbf{p}_i$ *increases*; see Figure 4. Therefore we define the

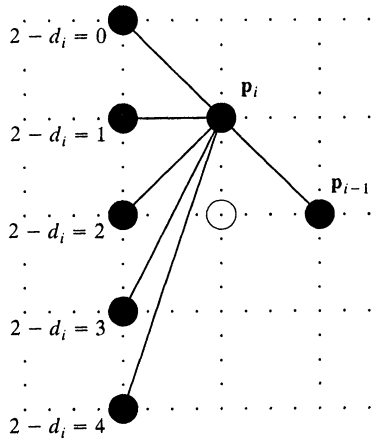

Figure 4. The dependence of the exterior angle at $\mathbf{p}_i$ on $2 - d_i$.

discrete exterior angle of $\mathscr{P}$ at $\mathbf{p}_i$ to be $1 - d_i/2$: there is no intrinsic reason *not* to multiply by $1/2$, and doing so makes the analogy to Gauss-Bonnet clearer.

Theorem 1, reinterpreted as in Section 8.1, says that $n + \sum_{i=1}^{n}(2 - d_i) = 12$. Dividing by 2, we obtain

$$\text{Area}(\mathscr{P}) + \sum(\text{discrete exterior angles}) = 2\lfloor \pi \rfloor, \tag{7}$$

where $\lfloor \pi \rfloor = 3$ is the discrete analogue of π!

We leave it to the reader to mull over whether there exists an explanation for the similarities between (6) and (7). We do not know one.

ACKNOWLEDGMENTS. The first author thanks C. Kenneth Fan and Jeff Lagarias for comments on a draft of this paper, and Bill Casselman for some comments about graphics in mathematical papers. The second author thanks Sean Keel, Harald Skarke, and John Tate for several helpful conversations.

REFERENCES

1. V. Batyrev, Dual polyhedra and mirror symmetry for Calabi-Yau hypersurfaces in toric varieties, *J. Algebraic Geom.* **3** (1994) 493–535.
2. J. Birman, *Braids, links, and mapping class groups*, Annals of Mathematics Studies, No. 82, Princeton Univ. Press, Princeton, N.J.; Univ. of Tokyo Press, Tokyo, 1974.
3. R. Diaz and S. Robins, Pick's formula via the Weierstrass $\wp$-function, *Amer. Math. Monthly* **102** (1995) 431–437.
4. W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies, No. 131, Princeton Univ. Press, Princeton, NJ, 1993.
5. B. Grünbaum and G. C. Shephard, Pick's theorem, *Amer. Math. Monthly* **100** (1993) 150–161.
6. R. Hartshorne, *Algebraic geometry*, Springer, New York, 1977.
7. D. Hensley, Lattice vertex polytopes with interior lattice points, *Pacific J. Math.* **105** (1983) 183–191.
8. T. Hibi, Some results on Ehrhart polynomials of convex polytopes. *Discrete Math.* **83** (1990) 119–121.
9. F. Hirzebruch, *Topological methods in algebraic geometry*, Springer-Verlag, Berlin, 1995.
10. M. Kreuzer and H. Skarke, Classification of reflexive polyhedra in three dimensions, preprint hep-th/9805190 at `http://xxx.lanl.gov/`.
11. J. C. Lagarias and G. M. Ziegler, Bounds for lattice polytopes containing a fixed number of interior points in a sublattice, *Canad. J. Math.* **43** (1991) 1022–1035.
12. J. Munkres, *Topology: a first course*, Prentice Hall, Englewood Cliffs, N.J., 1975.

13. H. Rademacher and E. Grosswald, *Dedekind sums*, The Mathematical Association of America, Washington, D.C., 1972.
14. P. Scott, On convex lattice polygons, *Bull. Austral. Math. Soc.* **15** (1976) 395–399.
15. J.-P. Serre, *A course in arithmetic*, Translated from the French, Springer, New York, 1973.
16. J.-P. Serre, *Trees*, Translated from the French by J. Stillwell, Springer, Berlin-New York, 1980.
17. J. Stillwell, *Mathematics and its history*, Springer, New York, 1989.
18. V. S. Varadarajan, *Lie groups, Lie algebras, and their representations*, Prentice--Hall, Inc., Englewood Cliffs, N.J., 1974.

BJOAN POONEN received an A.B. summa cum laude in mathematics and physics from Harvard in 1989, and a Ph. D. in mathematics from the University of California at Berkeley in 1994. A four-time winner of the Putnam Competition, he returned to Berkeley in 1997 as a faculty member, after spending one year at the Mathematical Sciences Research Institute and two years at Princeton University. His main research interests are in number theory and algebraic geometry, but he has also collaborated with researchers at AT & T and Lucent Technologies on several papers in combinatorics and optimization. He is an Alfred P. Sloan Research Fellow and a David and Lucile Packard Research Fellow.
University of California, Berkeley, CA 94720-3840
poonen@math.berkeley.edu

FERNANDO RODRIGUEZ-VILLEGAS obtained the degree of Licenciado en Ciencias Matemáticas from the Universidad de Buenos Aires, Argentina, in 1985 and received a Ph. D. in mathematics from The Ohio State University in 1990. Except for a year at the Institute for Advanced Study (90–91) and a year at the Max Planck Institut, Bonn (94–95) he was at Princeton University before joining the faculty at the University of Texas at Austin in 1998. His main research interests are in number theory, modular forms, and special values of L-functions. He is a current Alfred P. Sloan Research Fellow.
University of Texas at Austin, TX 78712-1082
villegas@math.utexas.edu