The combinatorics of Markoff numbers

(www.math.wisc.edu/~propp/markoff-slides.pdf)

Jim Propp
Department of Mathematics,
University of Wisconsin
(propp@math.wisc.edu)

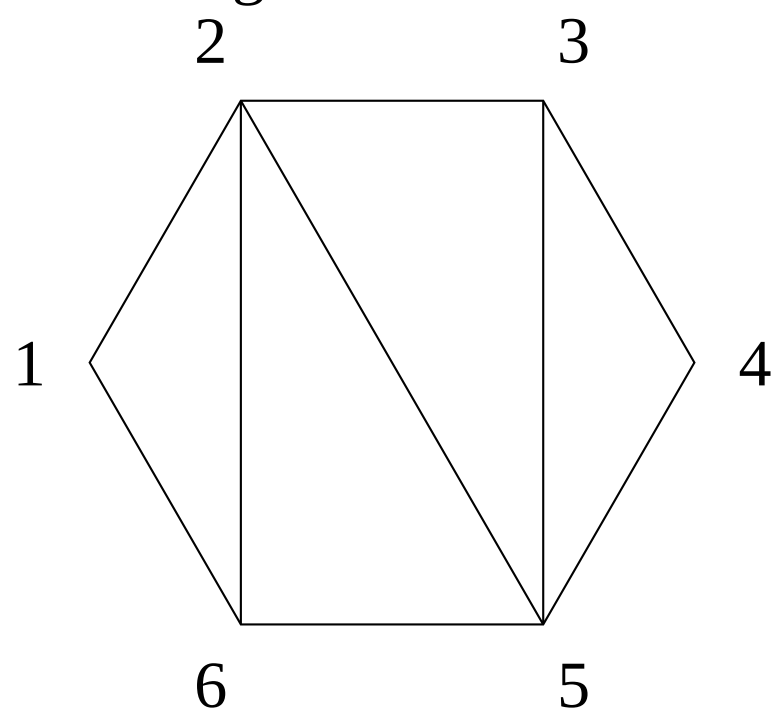
This talk describes joint work with Dylan Thurston and with (former or current) Boston-area undergraduates Gabriel Carroll, Andy Itsara, Ian Le, Gregg Musiker, Gregory Price, and Rui Viana, under the auspices of REACH (Research Experiences in Algebraic Combinatorics at Harvard). For details of proofs, see preprints on-line at www.math.wisc.edu/~propp/reach .

I. Triangulations and frieze patterns

To every triangulation T of an n-gon with vertices cyclically labelled 1 through n, Conway and Coxeter associate an $(n-1)$-rowed periodic array of numbers called a **frieze pattern** determined by the numbers $a_1, a_2, \ldots, a_n$, where a_k is the number of triangles in T incident with vertex k.

(See J. H. Conway and H. S. M. Coxeter, "Triangulated Polygons and Frieze Patterns," *Math. Gaz.* **57** (1973), 87–94 and J. H. Conway and R. K. Guy, in *The Book of Numbers*, New York : Springer-Verlag (1996), 75–76 and 96–97.)

E.g., the triangulation



of the 6-gon determines the 5-row frieze pattern

```
... 1 1 1 1 1 1 1 1 1   ...
...  1 3 2 1 3 2 1 3 2 ...
... 1 2 5 1 2 5 1 2 5   ...
...  1 3 2 1 3 2 1 3 2 ...
... 1 1 1 1 1 1 1 1 1   ...
```

Rules for constructing frieze patterns:

1. The top row is

$$\dots, 1, 1, 1, \dots$$

2. The second row (offset from the first) is

$$\dots, a_1, a_2, \dots, a_n, a_1, \dots$$

(with period n).

3. Each succeeding row (offset from the one before) is determined by the recurrence

```
 A
B C  :   D = (BC - 1) / A
 D
```

Facts:

- Every entry in rows 1 through $n-1$ is non-zero (so that the recurrence

```
D = (BC-1)/A
```

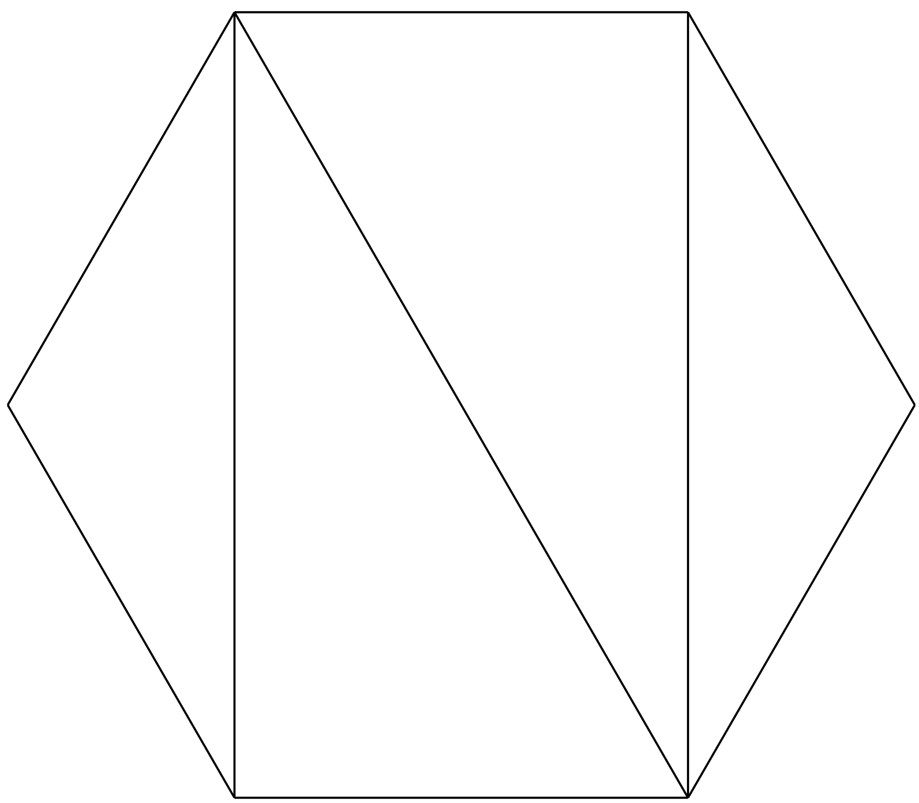
  never involves division by 0).

- Each of the entries in the array is a positive integer.

- For $1 \leq m \leq n-1$, the $n-m$th row is the same as the mth row, shifted. (That is, the array as a whole is invariant under a glide reflection.)

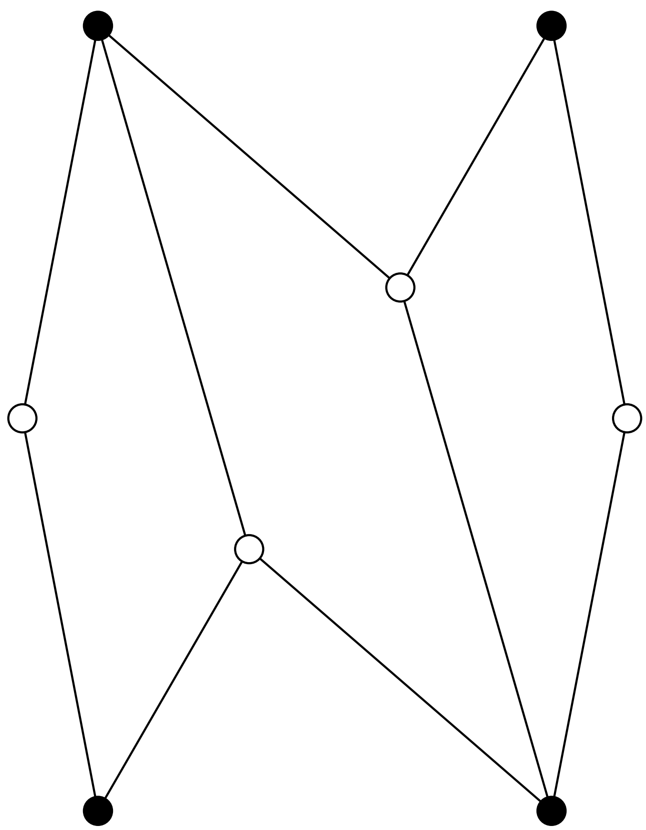
Question: What do these positive integers count? (And why does the array possess this symmetry?)

E.g., in the picture



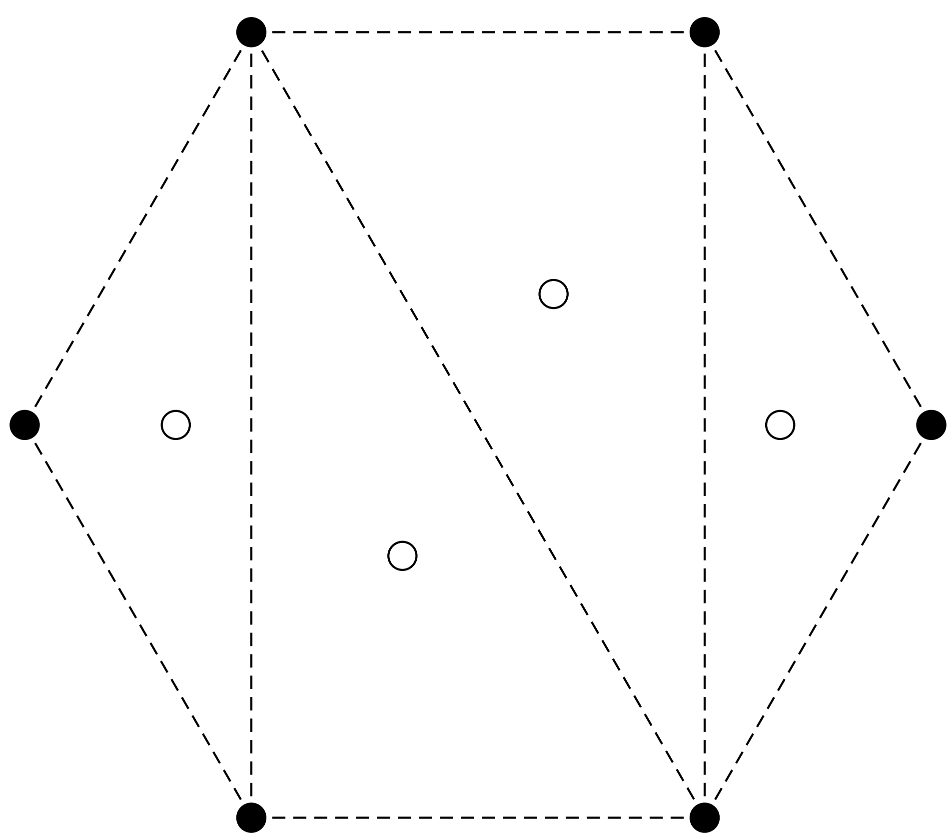
what are there 5 of?

Answer: Perfect matchings of the graph

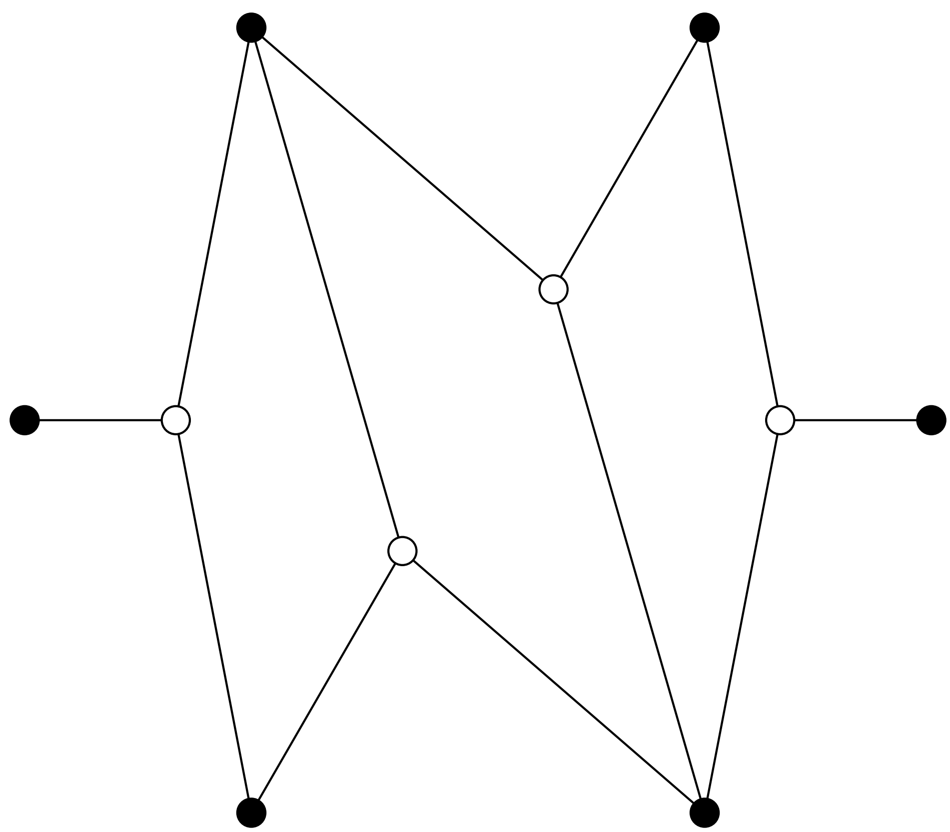


General construction:

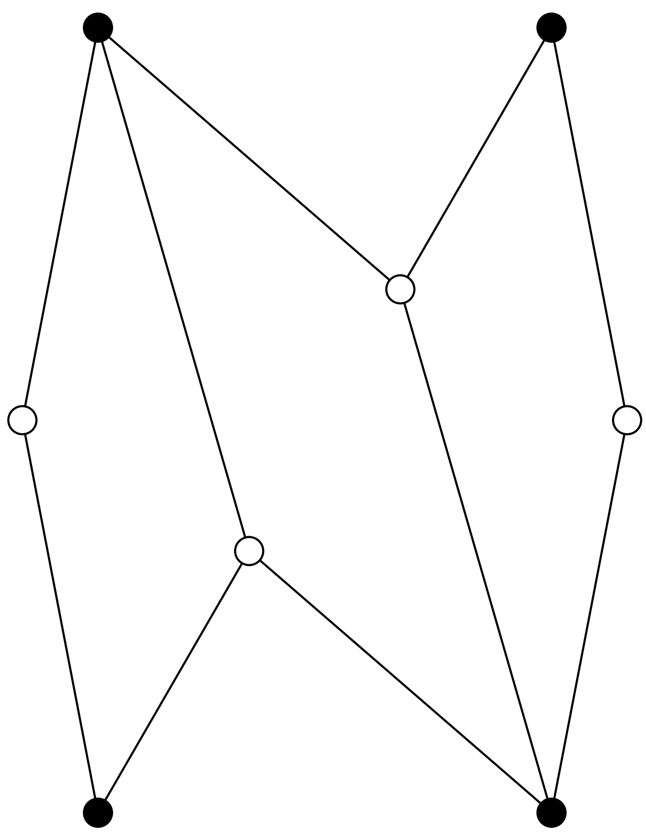
Put a black vertex at each of the n vertices of the n-gon.
Put a white vertex in the interior of each of the $n - 2$ triangles in the triangulation T.



For each of the $n - 2$ triangles, connect the black vertices of the triangle to the white vertex inside the triangle. This gives a connected planar bipartite graph with n black vertices and $n - 2$ white vertices.



If we remove 2 of the black vertices (say vertices i and j), we get a graph with equally many black and white vertices. Let $C_{i,j}$ be the number of perfect matchings of this graph.



Theorem (Gabriel Carroll and Gregory Price): The Conway-Coxeter frieze pattern is just

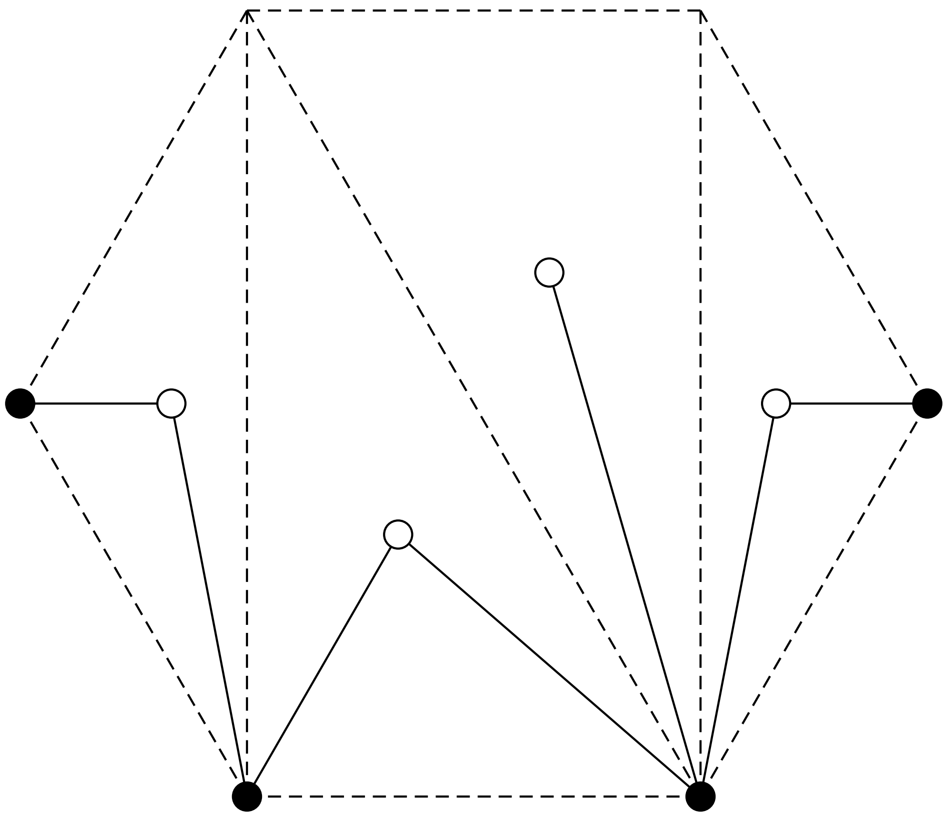
$$
\begin{array}{ccccccccccc}
\ldots & C_{1,2} & & C_{2,3} & & C_{3,4} & & C_{4,5} & \ldots \\
\ldots & & C_{1,3} & & C_{2,4} & & C_{3,5} & & \ldots \\
\ldots & C_{n,3} & & C_{1,4} & & C_{2,5} & & C_{3,6} & \ldots \\
\ldots & & C_{n,4} & & C_{1,5} & & C_{2,6} & & \ldots \\
 & \vdots & & \vdots & & \vdots & & \vdots &
\end{array}
$$

(interpret all subscripts mod n).

Note: This claim explains the glide-reflection symmetry.

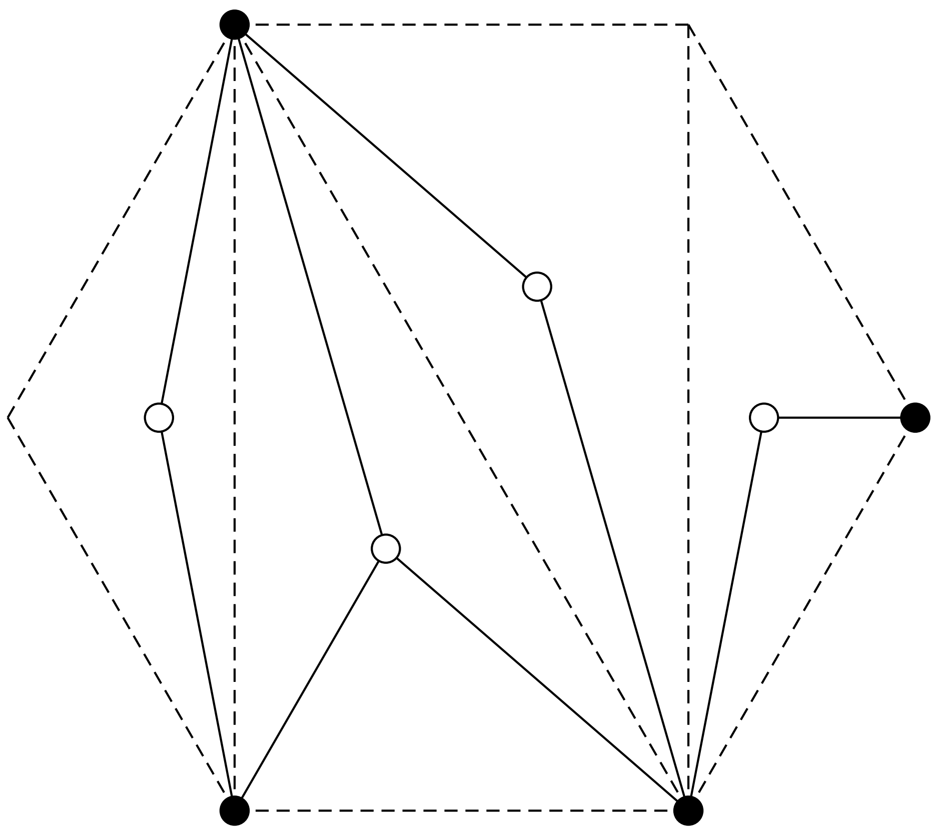
Proof of theorem:

1. $C_{i,i+1} = 1$.



(proof of theorem, continued)

2. $C_{i-1,i+1} = a_i$.

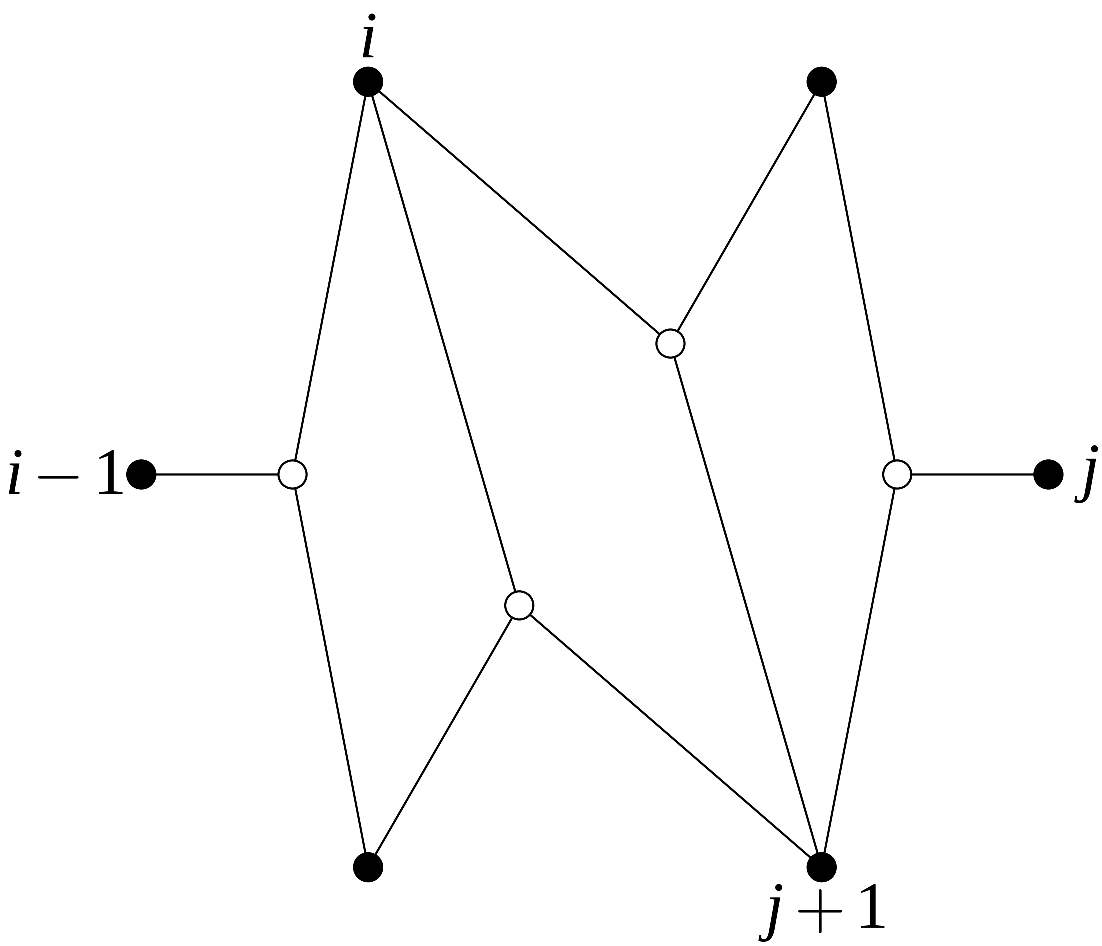


(proof of theorem, continued)

3. $C_{i,j}C_{i-1,j+1} = C_{i-1,j}C_{i,j+1} - 1.$

Move the 1 to the left-hand side, and write the equation in the form

$$C_{i,j}C_{i-1,j+1} + C_{i-1,i}C_{j,j+1} = C_{i-1,j}C_{i,j+1}$$



(proof of theorem, concluded)

This is a consequence of a lemma due to Eric Kuo (see Theorem 2.5 in "Applications of graphical condensation for enumerating matchings and tilings," `math.CO/0304090`):

If a bipartite planar graph G has 2 more black vertices than white vertices, and black vertices a, b, c, d lie in cyclic order on some face of G, then

$$M(a,c)M(b,d) =$$

$$M(a,b)M(c,d) + M(a,d)M(b,c),$$

where $M(x,y)$ denotes the number of perfect matchings of the graph obtained from G by deleting vertices x and y and all incident edges.

Note that if we replace $C_{i,j}$ by the distance $D_{i,j}$ between points i and j, and all the points on the n-gon lie on a circle, we get the three-term quadratic relation

$$D_{i,j}D_{i-1,j+1}+D_{i-1,i}D_{j-1,j}=D_{i-1,j}D_{i,j+1}$$

which is a consequence of Ptolemy's theorem on the lengths of the sides and diagonals of an inscriptible quadrilateral.

In fact, the Carroll-Price theorem does have geometric content, but not for Euclidean geometry.

Dylan Thurston pointed out that this relation can be understood in terms of the topology and geometry of the hyperbolic manifold with boundary obtained from the closed disk by removing n points on the boundary (where we require the n boundary components to be geodesics, and we require the metric in the interior to have constant curvature -1).

A version of this construction that includes edge-weights gives the cluster algebras of type A introduced by Sergey Fomin and Andrei Zelevinsky. (See section 3.5 of Fomin and Zelevinsky, "Y-systems and generalized associahedra", `hep-th/0111053`.)

II. The Stern-Brocot tree and superbases of $\mathbf{Z}^2$, or, The topography of Fareyland

The **mediant** of two fractions $\frac{a}{b}$, $\frac{c}{d}$, each expressed in lowest terms, is the fraction $\frac{a+c}{b+d}$.

Aside from $\frac{0}{1} = 0$ and $\frac{1}{0} = \infty$ (included by special allowance), we require numerators and denominators to be positive.

In the **Stern-Brocot process**, we start with the row

$$\frac{0}{1} < \frac{1}{0}$$

and repeatedly insert mediants between every pair of adjacent fractions in the current row, to get the next row:

$$\frac{0}{1} < \frac{1}{0}$$

$$\frac{0}{1} < \frac{1}{1} < \frac{1}{0}$$

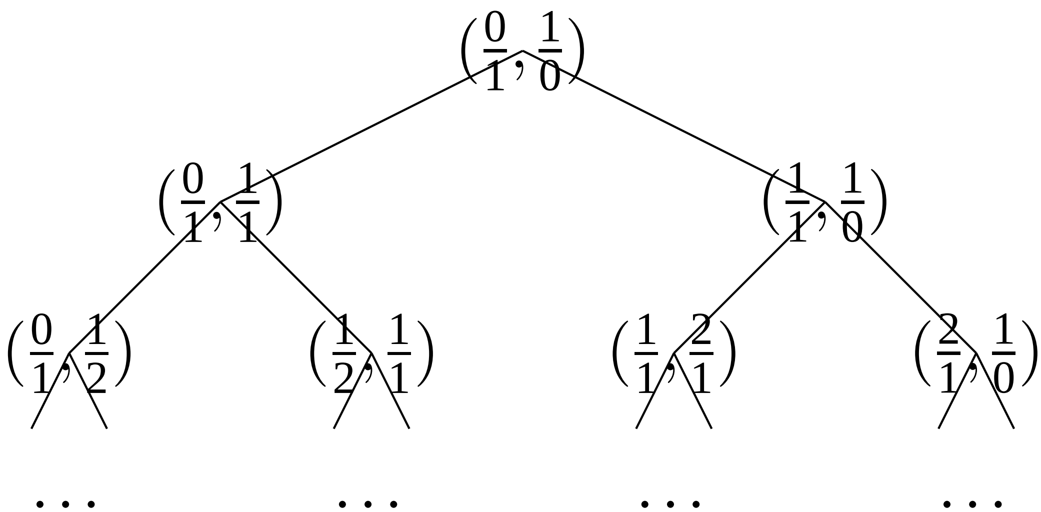
$$\frac{0}{1} < \frac{1}{2} < \frac{1}{1} < \frac{2}{1} < \frac{1}{0}$$

$$\frac{0}{1} < \frac{1}{3} < \frac{1}{2} < \frac{2}{3} < \frac{1}{1} < \dots$$

$$\frac{0}{1} < \frac{1}{4} < \frac{1}{3} < \frac{2}{5} < \frac{1}{2} < \frac{3}{5} < \frac{2}{3} < \frac{3}{4} < \frac{1}{1} < \dots$$

It's natural to write these numbers in a "tree" with two roots.

But it's even more natural to put pairs of fractions at the nodes and just have one root $(\frac{0}{1}, \frac{1}{0})$, where the two children of $(\frac{a}{b}, \frac{c}{d})$ are $(\frac{a}{b}, \frac{a+c}{b+d})$ and $(\frac{a+c}{b+d}, \frac{c}{d})$.



We can represent each pair $(\frac{a}{b}, \frac{c}{d})$ by the two-by-two matrix

$$\begin{pmatrix} c & a \\ d & b \end{pmatrix}$$

(note the switch!) whose two descendants are then

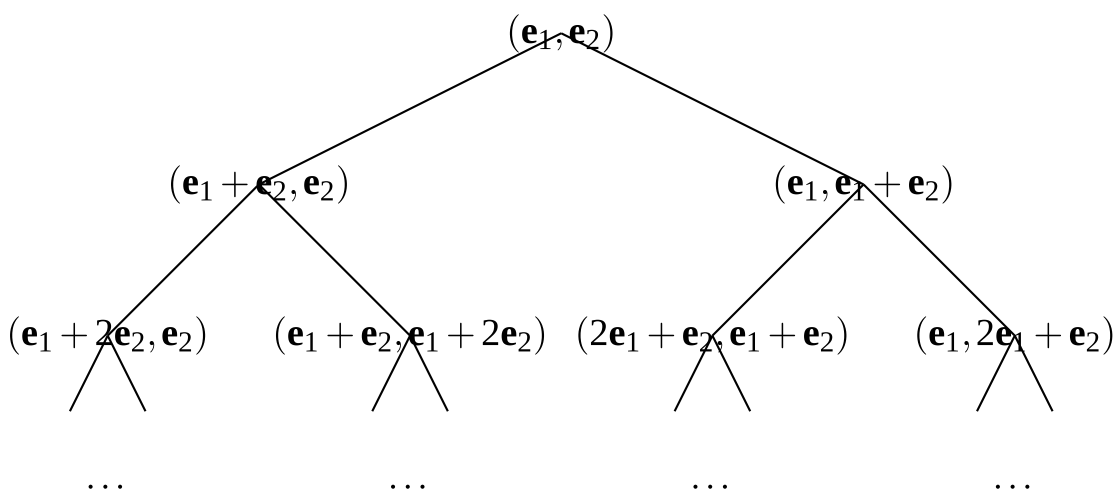
$$\begin{pmatrix} c & a \\ d & b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} a+c & a \\ b+d & b \end{pmatrix}$$

and

$$\begin{pmatrix} c & a \\ d & b \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} c & a+c \\ d & b+d \end{pmatrix}.$$

Every 2-by-2 matrix with non-negative integer entries and determinant $+1$ arises in a unique fashion from this process.

We can think of the columns of each matrix as giving an ordered base (usually called a basis) for a two-dimensional lattice L. The vectors in the base are primitive vectors, where a non-zero vector $\mathbf{u}$ is called **primitive** if it cannot be written as $k\mathbf{v}$ for $k > 1$ and $v \in L$.



If we want to include negative numbers in this story (after all, $\frac{1}{2}$ is also $\frac{-1}{-2}$) and arbitrary bases for $\mathbf{Z}^2$, one natural way to do this, introduced by Conway in *The Sensual (Quadratic) Form*, is to replace vectors and bases by “lax vectors” and “lax bases”, and to use “super-bases” as well, and to use these as the faces, edges, and vertices of a picture called the “topograph” of $\mathbf{Z}^2$.

A **lax vector** is a primitive vector, only defined up to sign. If $\mathbf{u}$ is a primitive vector, the associated lax vector is written $\pm\mathbf{u}$. We call $\mathbf{u}$ (in contrast to $\pm\mathbf{u}$) a **strict** (primitive) **vector**.

A **strict base** is an ordered pair $(\mathbf{u}, \mathbf{v})$ of primitive vectors whose integral linear combinations are exactly the elements of L.

A **lax base** is a set $\{\pm\mathbf{u}, \pm\mathbf{v}\}$ obtained from a strict base.

A **strict superbase** is an ordered triple $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ for which $\mathbf{u} + \mathbf{v} + \mathbf{w} = \mathbf{0}$ and $(\mathbf{u}, \mathbf{v})$ is a strict base (implying that $(\mathbf{u}, \mathbf{w})$ and $(\mathbf{v}, \mathbf{w})$ are also strict bases for L).

A **lax superbase** is a set $\{\pm\mathbf{u}, \pm\mathbf{v}, \pm\mathbf{w}\}$ where $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is a strict superbase.

$\{\pm\mathbf{u}, \pm\mathbf{v}, \pm\mathbf{w}\}$ is a lax superbase if and only if $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are primitive vectors any two of which form a base, and

$$\pm\mathbf{u} \pm \mathbf{v} \pm \mathbf{w} = \mathbf{0}$$

for some choice of signs.

Each superbase

$$\{\pm\mathbf{u}, \pm\mathbf{v}, \pm\mathbf{w}\}$$

contains the three bases

$$\{\pm\mathbf{u}, \pm\mathbf{v}\},\ \{\pm\mathbf{u}, \pm\mathbf{w}\},\ \{\pm\mathbf{v}, \pm\mathbf{w}\}$$

and no others.

Each base

$$\{\pm\mathbf{u}, \pm\mathbf{v}\}$$

is in the two superbases

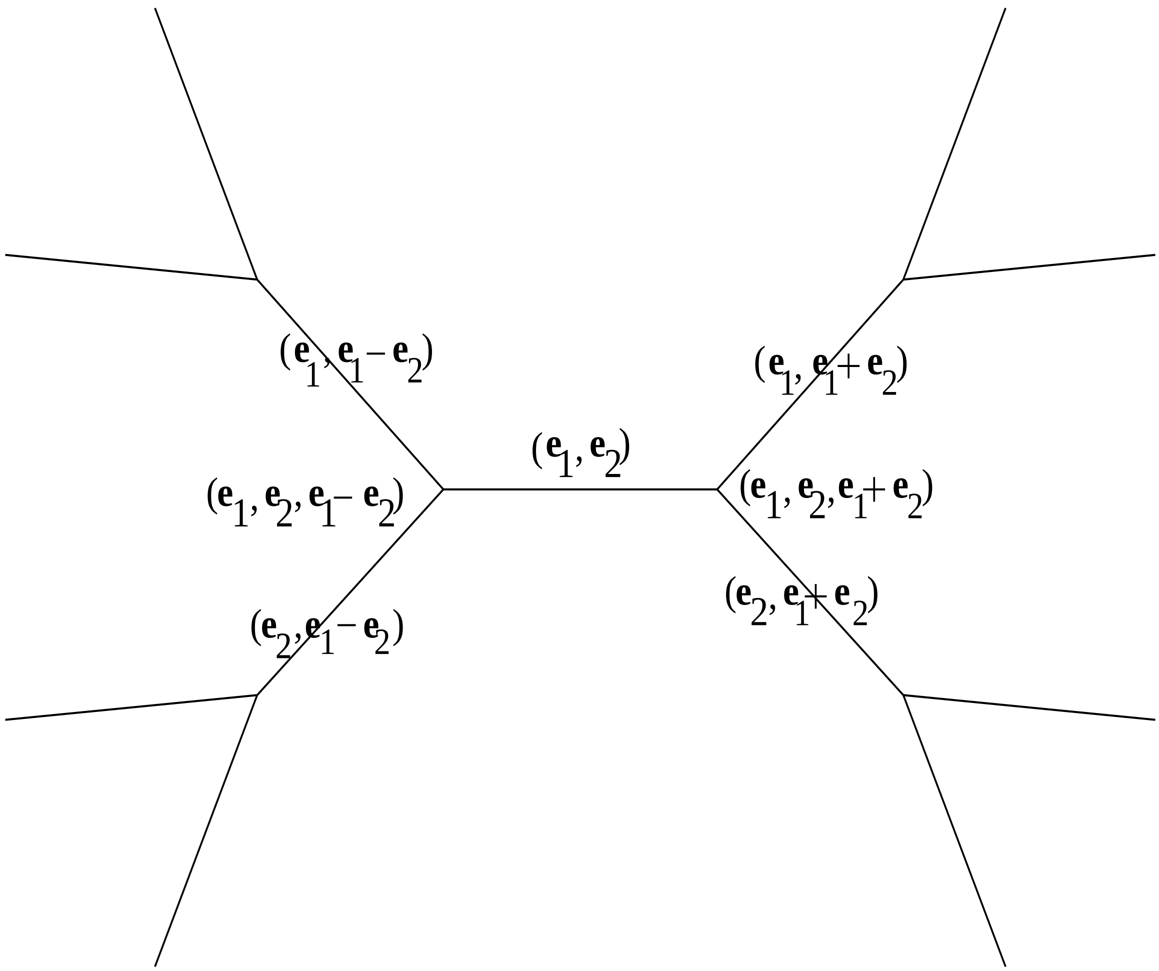
$$\{\pm\mathbf{u}, \pm\mathbf{v}, \pm(\mathbf{u}+\mathbf{v})\},\ \{\pm\mathbf{u}, \pm\mathbf{v}, \pm(\mathbf{u}-\mathbf{v})\}$$

and no others.

The **topograph** is the graph whose vertices are lax superbases and whose edges are lax bases, where each superbase is incident with the three bases in it.

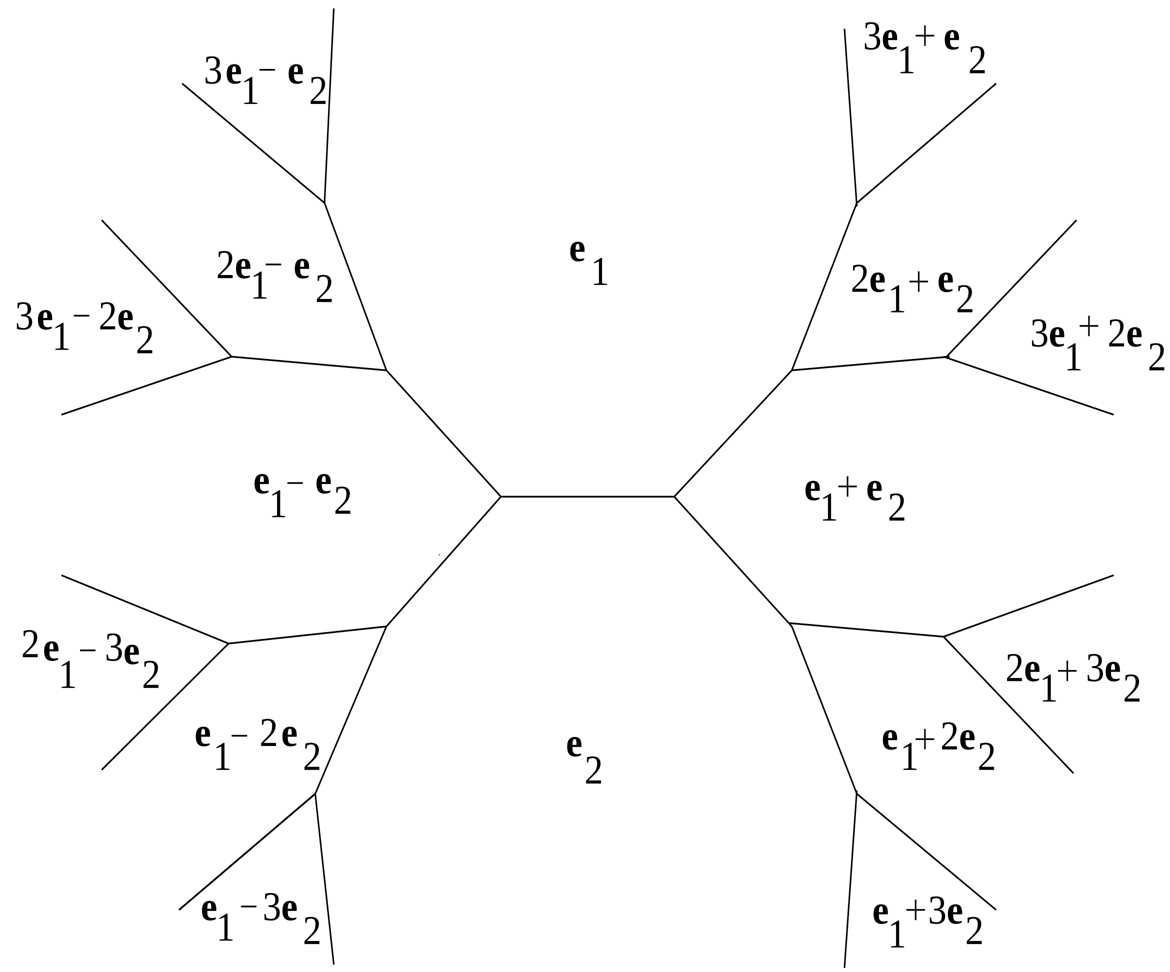
This gives a 3-valent tree whose vertices correspond to the lax superbases of L, whose edges correspond to the lax bases of L, and whose "faces" correspond to the lax vectors in L.

(Highbrows may wish to call this tree the dual of the triangulation of the hyperbolic plane by images of the modular domain under the action of the modular group.)



(from *The Sensual (Quadratic) Form*)

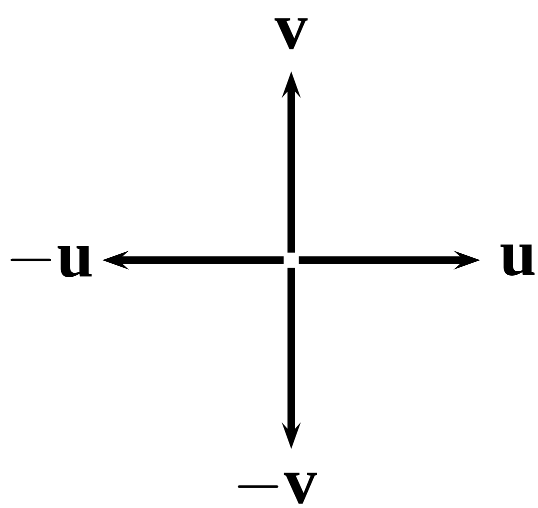
[Note that the distinction between lax and strict is ignored here, for notational simplicity.]



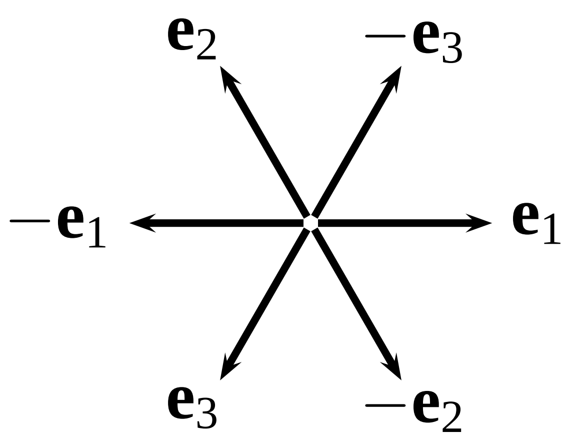
(from *The Sensual (Quadratic) Form*)

[Here too the distinction between lax and strict is ignored.]

If you want to use the square lattice $L = \mathbf{Z}^2$, it's most natural to center the topograph on the edge associated to the lax base $\{\pm\mathbf{u}, \pm\mathbf{v}\}$ where $\mathbf{u} = (1,0)$ and $\mathbf{v} = (0,1)$:



If you want to use the triangular lattice $L = \{(x,y,z) \in \mathbf{Z}^3 \ : \ x+y+z = 0\}$ (or $\mathbf{Z}^3/\mathbf{Z}\mathbf{v}$ where $\mathbf{v} = (1,1,1)$, if you prefer) it's most natural to center the topograph on the vertex associated to the lax superbase $\{\pm\mathbf{e}_1, \pm\mathbf{e}_2, \pm\mathbf{e}_3\}$ where $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ are shortest-length vectors in L summing to $\mathbf{0}$.



III. Markoff numbers

A **Markoff triple** is a triple (x,y,z) of positive integers satisfying $x^2+y^2+z^2 = 3xyz$; e.g., the triple (2,5,29).

A **Markoff number** is a positive integer that occurs in at least one such triple.

Writing the Markoff equation as

(*) $\qquad z^2-(3xy)z+(x^2+y^2) = 0,$

a quadratic equation in z, we see that if (x,y,z) is a Markoff triple, then so is (x,y,z'), where $z' = 3xy-z = (x^2+y^2)/z$, the other root of (*).

(z' is positive because $z' = (x^2+y^2)/z$, and is an integer because $z' = 3xy-z$.)

Likewise for x and y.

Claim: Every Markoff triple (x, y, z) can be obtained from the Markoff triple $(1, 1, 1)$ by a sequence of such exchange operations. E.g., $(1, 1, 1) \to (2, 1, 1) \to (2, 5, 1)$ $\to (2, 5, 29)$.

Proof idea: Use high-school algebra and some Olympiad-level cleverness to show that if (x, y, z) is a Markoff triple with $x \geq y \geq z$, and we take $x' = (y^2 + z^2)/x$, then $x' < x$ unless $x = y = z = 1$. See A. Baragar, “Integral solutions of the Markoff-Hurwitz equations,” (*Journal of Number Theory* **49** (1994), 27–44).

So in fact, each Markoff triple can be obtained from (1,1,1) by a sequence of moves that leaves two numbers alone and increases the third.

Create a graph whose vertices are the Markoff triples and whose edges correspond to the exchange operations

$$(x,y,z) \longrightarrow (x',y,z),$$
$$(x,y,z) \longrightarrow (x,y',z),$$
$$(x,y,z) \longrightarrow (x,y,z')$$

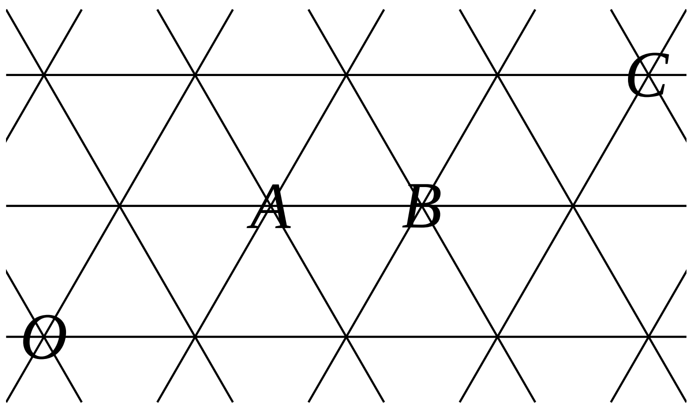
where

$$x' = \frac{y^2+z^2}{x},$$
$$y' = \frac{x^2+z^2}{y},$$
$$z' = \frac{x^2+y^2}{z}.$$

This 3-regular graph is connected (see the preceding claim), and it is not hard to show that it is acylic. Hence the graph is the 3-regular infinite tree.

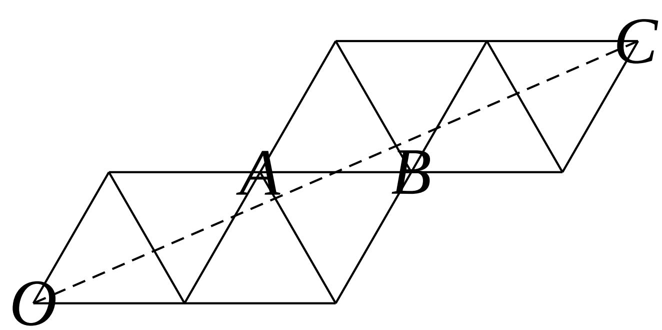
Unordered Markoff triples are associated with lax superbases of the triangular lattice, and Markoff numbers with lax vectors of the triangular lattice.

For example, the unordered Markoff triple $2, 5, 29$ corresponds to the lax superbase $\{\pm\mathbf{u}, \pm\mathbf{v}, \pm\mathbf{w}\}$ where $\mathbf{u} = \vec{OA}$, $\mathbf{v} = \vec{OB}$, and $\mathbf{w} = \vec{OC}$, with O, A, B, and C forming a fundamental parallelogram for the triangular lattice, as shown below.



The Markoff numbers 1, 2, 5, and 29 correspond to the primitive vectors $\vec{AB}$, $\vec{OA}$, $\vec{OB}$, and $\vec{OC}$.

To find the Markoff number associated with a primitive vector $\vec{OX}$, take the union R of all the triangles that segment OX passes through. The underlying lattice provides a triangulation of R. E.g., for the vector $\mathbf{u} = \vec{OC}$ from the previous page, the triangulation is



Turn this into a planar bipartite graph as in Part I, let $G(\mathbf{u})$ be the graph that results from deleting vertices O and C, and let $M(\mathbf{u})$ be the number of perfect matchings of $G(\mathbf{u})$. (If $\mathbf{u}$ is a shortest vector in the lattice, put $M(\mathbf{u}) = 1$.)

Theorem (Gabriel Carroll, Andy Itsara, Ian Le, Gregg Musiker, Gregory Price, and Rui Viana): If $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ is a lax superbase of the triangular lattice, then $(M(\mathbf{u}),M(\mathbf{v}),M(\mathbf{w}))$ is a Markoff triple. Every Markoff triple arises in this fashion.

In particular, if $\mathbf{u}$ is a primitive vector, then $M(\mathbf{u})$ is a Markoff number, and every Markoff number arises in this fashion.

(The association of Markoff numbers with the topograph is not new; what's new is the combinatorial interpretation of the association, by way of perfect matchings.)

Proof: The base case, with

$$(M(\mathbf{e}_1), M(\mathbf{e}_2), M(\mathbf{e}_3)) = (1, 1, 1),$$

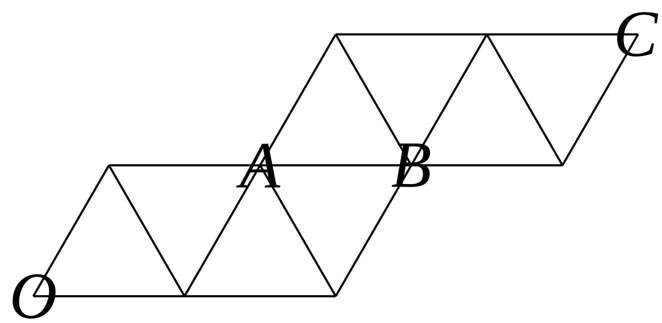
is clear.
The only non-trivial part of the proof is the verification that

$$M(\mathbf{u}+\mathbf{v}) = (M(\mathbf{u})^2 + M(\mathbf{v})^2)/M(\mathbf{u}-\mathbf{v}).$$

(proof of theorem, concluded)

E.g., in the picture below, we need to verify that

$$M(\vec{OC})M(\vec{AB}) = M(\vec{OA})^2 + M(\vec{OB})^2.$$



But if we rewrite the desired equation as

$$M(\vec{OC})M(\vec{AB}) =$$
$$M(\vec{OA})M(\vec{BC}) + M(\vec{OB})M(\vec{AC})$$

we see that this is just Kuo's lemma!

Remarks: Some of the work done by the REACH students used a square lattice picture; this way of interpreting the Markoff numbers combinatorially was actually discovered first, in 2001–2002 (Itsara, Le, Musiker, and Viana).

Also, the original combinatorial model for the Conway-Coxeter numbers (found by Price) involved paths, not perfect matchings. Carroll turned this into a perfect matchings model, which made it possible to arrive at the matchings model of Itsara, Le, Musiker, and Viana via a different route.

See `www.math.wisc.edu/~propp/reach/newback.jpg`.

IV. Geometric implications

Let S be the one-holed torus with a point removed (a 2-manifold with 1-point boundary).

Just as a natural covering space of the one-holed torus is the plane $\mathbf{R}^2$, a natural covering space of S is $\mathbf{R}^2 \setminus L$, the plane minus a lattice.

S can be given a hyperbolic structure that gives it constant curvature -1. There are many ways to do this (a two-parameter family's worth, in fact). The deleted point becomes a "cusp at infinity".

Fix such a hyperbolic structure h (and for technical reasons, a horocycle passing through the cusp). For each primitive vector $\mathbf{u}$ in the lattice, there is a unique simple closed geodesic on S whose lift up to the-plane-minus-a-lattice runs parallel to $\mathbf{u}$. Let $L_h(\mathbf{u})$ be the length of this geodesic. If we define $M(\mathbf{u}) = \alpha \cosh^{-1}(\beta L_h(\mathbf{u}))$ (for suitable α, β that don't depend on h), we get positive real numbers satisfying the relation

$$M(\mathbf{u}+\mathbf{v})M(\mathbf{u}-\mathbf{v}) = M(\mathbf{u})^2 + M(\mathbf{v})^2.$$

E.g., if h is the unique hyperbolic structure on S that gives it three-fold rotational symmetry about the cusp (call it h_0), then $M(\mathbf{u})$ is exactly the Markoff number associated with $\mathbf{u}$.

It is believed, but unproved, that for h_0, no two simple closed geodesics have the same length unless they are related by an automorphism of S that preserves h_0.

This is the **unicity conjecture** for Markoff numbers: No positive integer "is a Markoff number for two distinct reasons."

Equivalently, $M_{h_0}(\mathbf{u}) = M_{h_0}(\mathbf{v})$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are in the same orbit of L under the action of the 6-element dihedral group generated by permutations of $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$.

Our group used algebraic and combinatorial methods to prove a weak version of this conjecture.

"Generic unicity": For a dense G_δ set of hyperbolic structures on S, no two simple closed geodesics have the same length.

Sketch of proof: Put $M(\mathbf{e}_1) = x$, $M(\mathbf{e}_2) = y$, and $M(\mathbf{e}_3) = z$ (with $x, y, z > 0$) and recursively define

$$M(\mathbf{u}+\mathbf{v}) = (M(\mathbf{u})^2 + M(\mathbf{v})^2)/M(\mathbf{u}-\mathbf{v}).$$

Then for all primitive vectors $\mathbf{u}$, $M(\mathbf{u})$ is a **Laurent polynomial** in x, y, z; that is, it can be written in the form $P(x,y,z)/x^a y^b z^c$, where $P(x,y,z)$ is an ordinary polynomial in x, y, z (with non-zero constant term).

(sketch of proof of theorem, concluded)

If $\mathbf{u}$ inside the cone generated by $+\mathbf{e}_1$ and $-\mathbf{e}_3$, then $a < b > c$ and $(c+1)\mathbf{e}_1 - (a+1)\mathbf{e}_3 = \mathbf{u}$. (Likewise for the other sectors of L.)

Hence all the "Markoff polynomials" $M(\mathbf{u})$ are distinct (aside from the fact that $M(\mathbf{u}) = M(-\mathbf{u})$), and thus $M(\mathbf{u})(x,y,z) \neq M(\mathbf{v})(x,y,z)$ for all primitive vectors $\mathbf{u} \neq \pm\mathbf{v}$ as long as (x,y,z) lies in a dense G_δ set of real triples.

(The numerator of each Markoff polynomial is the sum of the weights of all the perfect matchings of the graph $G(\mathbf{u})$, where edges have weight x, y, or z according to orientation.)

V. Other directions for exploration

Neil Herriot (another member of REACH) showed that if we replace the triangular lattice used above by the tiling of the plane by isosceles right triangles (generated from one such triangle by repeated reflection in the sides), superbases of the square lattice correspond to triples (x, y, z) of positive integers satisfying either

$$x^2 + y^2 + 2z^2 = 4xyz$$

or

$$x^2 + 2y^2 + 2z^2 = 4xyz.$$

So, is there some more general combinatorial approach to ternary cubic equations of similar shape?

Gerhard Rosenberger ("Uber die diophantische Gleichung $ax^2+by^2+cz^2=dxyz$," *J. Reine Angew. Math.* **305** (1979), 122–125) showed that there are exactly three ternary cubic equations of the shape $ax^2+by^2+cz^2=(a+b+c)xyz$ for which all the positive integer solutions can be derived from the solution $(x,y,z)=(1,1,1)$ by means of the obvious exchange operations $(x,y,z)\to(x',y,z)$, $(x,y,z)\to(x,y',z)$, and $(x,y,z)\to(x,y,z')$, namely:

$$x^2+y^2+z^2=3xyz,$$

$$x^2+y^2+2z^2=4xyz,$$

and

$$x^2+2y^2+3z^2=6xyz.$$

The third Diophantine equation “ought” to be associated with some combinatorial model involving the reflection-tiling of the plane by 30-60-90 triangles, but the most obvious approach (based on analogy with the 60-60-60 and 45-45-90 cases) does not work.

What about the equation $w^2 + x^2 + y^2 + z^2 = 4wxyz$? (Such equations are called Markoff-Hurwitz equations.)

The Laurent phenomenon applies here too: The four exchange operations convert an initial formal solution (w, x, y, z) into a quadruple of Laurent polynomials. (This is a special case of Theorem 1.10 in Fomin and Zelevinsky's paper "The Laurent phenomenon," `math.CO/0104241`.)

The numerators of these Laurent polynomials ought to be weight-enumerators for some combinatorial model, but we have no idea what this model looks like. We can't even prove that the coefficients are positive, although they appear to be.