Math Club Notes: 2005 May 22

1 Web sites

Mehran asked about good math web sites. Here's a few:

http://mathworld.wolfram.com/

> MathWorld. A reference site. Few proofs and little discussion, but good if you basically know what you're looking up and just need to be reminded.

http://www.mathpages.com/home/

> MathPages. A collection of Kevin Brown's posts to sci.math. Lots of interesting stuff.

http://www.dpmms.cam.ac.uk/~wtg10/

> A small collection of short articles by Timothy Gowers. The author tries to show how certain results and ideas could be discovered without having a flash of genius. For a sample, see section 2 below.

http://aleph0.clarku.edu/~djoyce/java/elements/toc.html

> Euclid's *Elements*, in Heath's translation (but, alas, without his notes). I think that if your browser does Java you get diagrams you can fiddle with.

2 Cauchy-Schwarz

We looked again at the Cauchy-Schwarz inequality, this time not at applications but at proofs. We got a nice one in Math 225, using the algebraic properties of the dot product.

Timothy Gowers, whose web site I mentioned above, gives an alternative, very natural derivation. Start with two vectors

$$\vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \qquad \text{and} \qquad \vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} .$$

How, Gowers asks, could we express the fact that these vectors are proportional? A natural first attempt:

$$\exists c\colon \forall i\colon u_i = cv_i$$

This is a little awkward because it introduces the constant c, which we don't really care about. Besides, we know what it is: it's u_i/v_i. Rather than say that all the values u_i/v_i are equal to c, we might as well just say that they're equal to each other:

$$\forall i\colon \forall j\colon \frac{u_i}{v_i} = \frac{u_j}{v_j}$$

There's still a problem: some of the v_i might be zero. So let's cross-multiply and say this instead:

$$\forall i\colon \forall j\colon u_i v_j = u_j v_i$$

This version works, that is, it is indeed equivalent to the statement that $\vec{u}$ and $\vec{v}$ are proportional.

Now, rather than saying that some two expressions are equal, it's often more convenient to say that one expression is equal to zero. (This is a high school maneuver when solving polynomial equations.) So:

$$\forall i\colon \forall j\colon u_i v_j - u_j v_i = 0$$

Now comes the important maneuver, which, as Gowers says, is very common and should be part of your standard inventory of tricks: to say that a bunch of real numbers are zero, say that the sum of their squares is zero.

$$\sum_i \sum_j (u_i v_j - u_j v_i)^2 = 0$$

Now we have an equation which expresses the fact that $\vec{u}$ and $\vec{v}$ are proportional.

It doesn't take much to notice that, whether they're proportional or not, this sum of squares must be at least zero. Thus

$$\sum_i \sum_j (u_i v_j - u_j v_i)^2 \geq 0\,,$$

and we have equality exactly when the vectors are proportional. Now for a little routine sum manipulation:

$$\begin{aligned}
0 &\leq \sum_i \sum_j (u_i v_j - u_j v_i)^2 \\
&= \sum_i \sum_j (u_i^2 v_j^2 + u_j^2 v_i^2 - 2u_i u_j v_i v_j) \\
&= \sum_i \sum_j u_i^2 v_j^2 + \sum_i \sum_j u_j^2 v_i^2 - \sum_i \sum_j 2u_i u_j v_i v_j \\
&= \sum_i \Big(u_i^2 \sum_j v_j^2\Big) + \sum_i \Big(v_i^2 \sum_j u_j^2\Big) - 2\sum_i \Big(u_i v_i \sum_j u_j v_j\Big) \\
&= \Big(\sum_i u_i^2\Big)\Big(\sum_j v_j^2\Big) + \Big(\sum_i v_i^2\Big)\Big(\sum_j u_j^2\Big) - 2\Big(\sum_i u_i v_i\Big)\Big(\sum_j u_j v_j\Big) \\
&= 2\|\vec{u}\|^2\|\vec{v}\|^2 - 2(\vec{u}\cdot\vec{v})^2
\end{aligned}$$

Behold! Cauchy-Schwarz.

3 Bernoulli

In the notes for May 16, I used the manipulation

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \geq \sum_{k=0}^{1} \binom{n}{k} x^k = 1 + nx \,.$$

Here n is a positive integer and x is a nonnegative real number. The derivation relies on these facts: since $n \geq 1$, we know that changing from $\sum_{k=0}^{n}$ to $\sum_{k=0}^{1}$ either does nothing or throws terms away; since $x \geq 0$, we know that the discarded terms are nonnegative. Thus we get $\geq$.

The result, that $(1+x)^n \geq 1+nx$, is known as Bernoulli's inequality. It can be extended to real $x \geq -1$, though not, as far as I can see, using the above argument. A simple induction proves this more general version.

Once we have calculus, we can give this argument too: Let $f(x) = (1+x)^n$. Then $f''(x) = n(n-1)(1+x)^{n-2} \geq 0$ when $x \geq -1$ (since $n \geq 1$); thus f is concave up on $[-1, \infty)$, that is, its graph lies above its tangent lines in that interval. One of those tangent lines (the one at $x = 0$) is $y = 1 + nx$.

For even n, this concavity argument works for all of $\mathbb{R}$, not just $[-1, \infty)$. When n is odd, it doesn't, and indeed, when n is odd, Bernoulli's inequality fails somewhere to the left of $x = -1$.

(By the way, the bare statement that $f''(x) = n(n-1)(1+x)^{n-2}$ is arguably a little imprecise. Consider the cases $n = 1$ and $n = 2$ and take $x = -1$ to see why. More on this in the future, perhaps.)

I don't know what this inequality is for.

4 Binet

We discussed the Fibonacci numbers a bit.

Because the recurrence expresses each element in terms of the preceding two, it is in some ways more natural to think of the Fibonacci sequence not as a sequence of numbers but as a sequence of pairs of numbers:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \end{bmatrix}, \ldots$$

Each element of *this* sequence has all the information needed to compute the next one. Indeed, we have

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} F_n + F_{n-1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} .$$

Repeating this expansion n times, we obtain

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} .$$

So now we are thinking of a sequence of vectors in $\mathbb{R}^2$, starting with the standard basis vector $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, and iteratively applying a certain linear transformation, the one expressed by this matrix $\left[\begin{smallmatrix}1&1\\1&0\end{smallmatrix}\right]$. The most natural coordinate system for studying this linear transformation is one built out of its eigenvectors. By routine calculation, we find that

$$\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}\phi\\1\end{bmatrix}=\phi\begin{bmatrix}\phi\\1\end{bmatrix}\qquad\text{and}\qquad\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}\hat\phi\\1\end{bmatrix}=\hat\phi\begin{bmatrix}\hat\phi\\1\end{bmatrix},$$

where

$$\phi=\frac{1+\sqrt{5}}{2}\approx 1.61803\qquad\text{and}\qquad\hat\phi=\frac{1-\sqrt{5}}{2}\approx -0.61803$$

are the roots of the characteristic equation $\lambda^2=\lambda+1$. By expressing the first vector in our sequence, $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, in terms of this coordinate system, we obtain

$$\begin{aligned}\begin{bmatrix}F_{n+1}\\F_n\end{bmatrix}&=\begin{bmatrix}1&1\\1&0\end{bmatrix}^n\begin{bmatrix}1\\0\end{bmatrix}\\&=\begin{bmatrix}1&1\\1&0\end{bmatrix}^n\left(\frac{1}{\sqrt5}\begin{bmatrix}\phi\\1\end{bmatrix}-\frac{1}{\sqrt5}\begin{bmatrix}\hat\phi\\1\end{bmatrix}\right)\\&=\frac{1}{\sqrt5}\left(\begin{bmatrix}1&1\\1&0\end{bmatrix}^n\begin{bmatrix}\phi\\1\end{bmatrix}-\begin{bmatrix}1&1\\1&0\end{bmatrix}^n\begin{bmatrix}\hat\phi\\1\end{bmatrix}\right)\\&=\frac{1}{\sqrt5}\left(\phi^n\begin{bmatrix}\phi\\1\end{bmatrix}-\hat\phi^n\begin{bmatrix}\hat\phi\\1\end{bmatrix}\right).\end{aligned}$$

Therefore

$$F_n=\frac{\phi^n-\hat\phi^n}{\sqrt5}.$$

This closed form for the Fibonacci numbers is known as Binet's formula. It can be proved directly by induction, if you're so inclined; it can also be derived (in a manner similar to the above) using generating functions.

Since $|\hat\phi|<1$, the term $\hat\phi^n$ tends to zero as n increases, while $|\phi|>1$, so ϕ^n tends to infinity as n increases. That is, the ϕ^n term dominates for large n. For one thing, this means that, for sufficiently large n, we can just disregard the $\hat\phi^n$ and say that F_n is the nearest integer to $\phi^n/\sqrt5$. (It turns out that "sufficiently large" here means "at least zero".) For another, it means that F_{n+1}/F_n tends to $\phi^{n+1}/\phi^n=\phi$ as n increases.

(Note that

$$\phi^n=\left(\frac{1+\sqrt5}{2}\right)^n=\frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}(\sqrt5)^k.$$

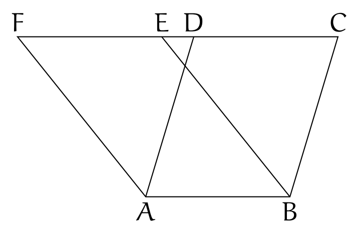
Taking a similar expansion of $\hat\phi^n$, substituting into Binet's formula, and simplifying yields an identity expressing F_n as a sum of binomial coefficients. Try writing out the sums in the case $n=4$ to see how the simplification should work.)

5 Parallelogram area

Euclid I.35 states:

> Parallelograms which are on the same base and in the same parallels are equal to one another.

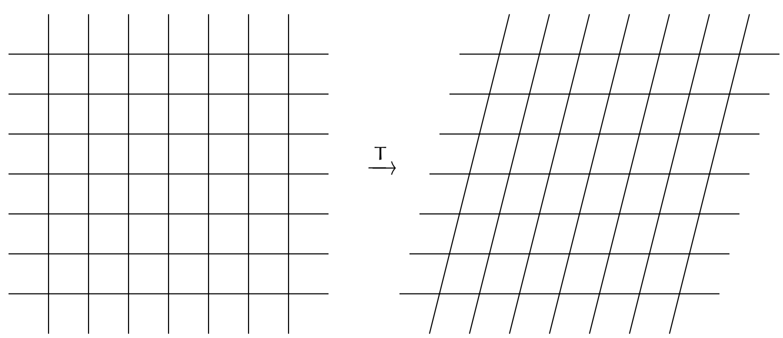
What this means is, if two parallelograms share a side (the "base"), and the sides opposite the base are collinear, then the parallelograms are of equal area.



(Note that this figure shows the case where the upper sides CD and EF don't overlap; they might.)

Start the proof by showing that $\triangle ADF$ and $\triangle BCE$ are congruent; then observe that taking $\triangle ADF$ away from the trapezoid $ABCF$ leaves one of the parallelograms, while taking $\triangle BCE$ away leaves the other. (That's Simson's proof, which is better than Euclid's.)

One way to look at this result: shears preserve area. That is, transformations of the plane such as



preserve the area of all figures in the plane. (The proposition tells us that the squares shown keep their area; other reasonable shapes can be approximated by such squares as in a Riemann sum.)

Shearing in the x direction leaves y-coordinates unchanged. The effect on x-coordinates is an increase in proportion to the y-coordinate: the x-axis is left alone; the line $y = 1$ is shifted, say, c units to the right; the line $y = 2$ is shifted $2c$ units to the right; and so forth. Thus a shear is given by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + cy \\ y \end{bmatrix} = \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} .$$

That is, shearing in the x-direction is associated with a matrix of the form $\left[\begin{smallmatrix} 1 & c \\ 0 & 1 \end{smallmatrix}\right]$; that shearing preserves area appears here in the fact that the determinant of such a matrix is 1, no matter what c is.

(To shear in some other direction, rotate to make that direction into the x-direction, then shear as above, then rotate back. Rotations also preserve area, of course.)

This kind of matrix has a lot to do with Gaussian elimination. We can restate the problem of solving a linear system such as

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

in the following way: find x and y such that

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = x \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + y \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} .$$

In other words, find the coordinates of the point (b_1, b_2) in the coordinate system generated by the columns of the coefficient matrix. Now, left-multiplying both sides by $\left[\begin{smallmatrix} 1 & c \\ 0 & 1 \end{smallmatrix}\right]$ changes the problem into

$$\begin{bmatrix} b_1 + cb_2 \\ b_2 \end{bmatrix} = x \begin{bmatrix} a_{11} + ca_{21} \\ a_{21} \end{bmatrix} + y \begin{bmatrix} a_{12} + ca_{22} \\ a_{22} \end{bmatrix} .$$

Note that this is the same as if we had applied the elementary row operation $R_1 \to R_1 + cR_2$. (For this reason, this kind of matrix is called an elementary matrix, which are often used when developing the basics of linear algebra formally. What matrices correspond to the other elementary row operations?)

Thus, row reduction solves the problem "What are the coordinates of such-and-such a point in such-and-such a coordinate system?" by shearing the space a few times to turn the coordinate system in question into the standard coordinate system. Then the problem is trivial.

(Well, shearing can't always get the standard coordinate system. You might have to scale in some directions too. And, of course, the original set of vectors might not be linearly independent, in which case they generate a somewhat screwed-up coordinate system.)

6 Telescoping

In the notes for May 9, I suggested the following as an exercise: Use the perturbation method to find a closed form for $\sum_{k=0}^{n} F_k$, where F_n is the nth Fibonacci number.

On May 16, I mentioned an alternative method for evaluating this sum, but forgot to put it in the notes. Here it is: from the recurrence we have

$$F_n = F_{n+2} - F_{n+1} \,.$$

Thus

$$\begin{aligned}
\sum_{k=0}^{n} F_k &= \sum_{k=0}^{n} (F_{k+2} - F_{k+1}) \\
&= \sum_{k=0}^{n} F_{k+2} - \sum_{k=0}^{n} F_{k+1} \\
&= \sum_{k=1}^{n+1} F_{k+1} - \sum_{k=0}^{n} F_{k+1} \\
&= F_{n+2} + \sum_{k=1}^{n} F_{k+1} - \sum_{k=1}^{n} F_{k+1} - F_1 \\
&= F_{n+2} - 1 \,.
\end{aligned}$$

The sigma notation can be a little opaque. Here's the case $n = 4$, to illustrate what's going on:

$$\begin{aligned}
0 + 1 + 1 + 2 + 3 &= (-1 + 1) + (-1 + 2) + (-2 + 3) + (-3 + 5) + (-5 + 8) \\
&= -1 + (1 - 1) + (2 - 2) + (3 - 3) + (5 - 5) + 8 \\
&= -1 + 8 \,.
\end{aligned}$$

This kind of sum is called a *telescoping sum*. The second half of each term cancels the first half of the next term; thus every term in the middle is destroyed, and the sum (like a collapsing telescope) is reduced to a couple uncancelled pieces from either end.

Closed forms for the following can be found by the same technique:

$$\sum_{k=1}^{n} \frac{1}{k(k+1)} \qquad \sum_{k=0}^{n} \binom{m+k}{k} \qquad \prod_{k=0}^{n} \cos(2^k\theta)$$

Recall that, in the Fibonacci example, we used the recurrence to massage the expression into a telescoping form. In these three examples, that purpose is served by partial fractions, $\binom{a}{b-1} + \binom{a}{b} = \binom{a+1}{b}$, and $\sin 2a = 2 \sin a \cos a$, respectively.