Math Club Notes: 2005 June 6

1 Points on a sphere

Another Putnam problem, this one from 1968:

> Show that for any set of n points on a unit sphere, the sum of the squares of the $n(n-1)/2$ distances between them does not exceed n^2.

(This is a twice-removed paraphrase of the original problem. I found it in John Scholes's Putnam archive, at http://www.kalva.demon.co.uk/putnam/putn68.html. Scholes paraphrases the problems to avoid potential copyright liability; I've paraphrased his paraphrase because I like mine better.)

One thing to understand before beginning work on the problem: why are there $n(n-1)/2$ distances among these n points? Eileen and Ray both recognized this expression from Gauss's famous sum trick:

$$1+2+3+\cdots+(n-2)+(n-1)=\frac{n(n-1)}{2}\,.$$

The LHS counts the number of distances: for if we name the points $P_1,\ldots,P_n$, then from P_1 to all the other points there are $n-1$ distances; from P_2 to all the other points (except P_1, since we just counted that distance) is $n-2$ distances; and so forth. (Another way to look at it: $n(n-1)/2=\binom{n}{2}$, the number of pairs of points.)

That is, the authors of the problem want us to (a) count each segment only once (not once in each direction), and (b) not count the distance from a point to itself. The usual way to do this is to write the sum as

$$S=\sum_{\substack{i,j\\1\le i<j\le n}}\|P_iP_j\|^2\qquad\text{or}\qquad S=\sum_{i=1}^{n}\sum_{j=i+1}^{n}\|P_iP_j\|^2\,.$$

In the first version, the requirement that $i<j$ simultaneously ensures that we don't get $i=j$ and that we only count each distance once (e.g., we count the distance from P_1 to P_2 only as $\|P_1P_2\|$, not also as $\|P_2P_1\|$). The second version is the same thing, written a bit more traditionally.

The first thing I did when solving this problem is fix the annoying asymmetry between i and j in this sum. I'd much rather sum unrestrictedly over all pairs of i and j; so first relate that unrestricted sum to the sum in question:

$$\sum_{i,j}=\sum_{i<j}+\sum_{i=j}+\sum_{i>j}=S+0+S\,,\text{ whence }S=\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|P_iP_j\|^2\,.$$

(The problem can be solved without this maneuver, but symmetry usually makes things easier, so it's natural to do things like this before getting up to your elbows in computations.)

So, we wish to show that

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|P_iP_j\|^2 \le n^2 \,,$$

given that all the points P_i lie on a unit sphere. It's natural (again, on grounds of symmetry) to centre that sphere at the origin; then, letting each point P_i have coordinates (x_i, y_i, z_i), the fact that they're all on that sphere is expressed by

$$\forall i\colon\, x_i^2 + y_i^2 + z_i^2 = 1 \,.$$

In order to make use of this fact, we write out the distance formula in full:

$$\begin{aligned}\|P_iP_j\|^2 &= (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 \\ &= x_i^2 + y_i^2 + z_i^2 + x_j^2 + y_j^2 + z_j^2 - 2x_ix_j - 2y_iy_j - 2z_iz_j \\ &= 2 - 2x_ix_j - 2y_iy_j - 2z_iz_j \,.\end{aligned}$$

Thus

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|P_iP_j\|^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}(1 - x_ix_j - y_iy_j - z_iz_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}1 - \sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j - \sum_{i=1}^{n}\sum_{j=1}^{n}y_iy_j - \sum_{i=1}^{n}\sum_{j=1}^{n}z_iz_j \\ &= n^2 - \sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j - \sum_{i=1}^{n}\sum_{j=1}^{n}y_iy_j - \sum_{i=1}^{n}\sum_{j=1}^{n}z_iz_j \\ &= n^2 - \Big(\sum_{i=1}^{n}x_i\Big)^2 - \Big(\sum_{i=1}^{n}y_i\Big)^2 - \Big(\sum_{i=1}^{n}z_i\Big)^2 \,,\end{aligned}$$

which is, of course, at most n^2. (In the last step we've used the trick mentioned in the notes for May 30, factoring the sum of all possible combinations of elements from two sets.)

We considered generalizations of this problem. Eileen noted that we can instantly generalize to other numbers of dimensions; e.g., in two dimensions, we have n points on a circle, and everything is the same except that we don't have z. It turns out that we can generalize even further, to any inner product space (where we understand "distance" — both between pairs of points and between each point and the centre of the "sphere" — in terms of the norm induced by the inner product). Try it.

We briefly considered generalizing in a different direction: rather than taking 2 points at a time and considering the distance between them, what if we take 3 points at a time and consider the area of the triangle they form? It looks like the algebra for this would be pretty hairy.

2 Triangle decomposition

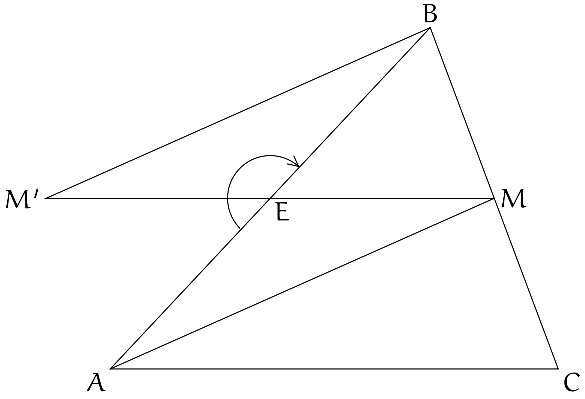
Last session, I mentioned this problem from the 1982 Putnam:

> Let M be the midpoint of side BC of a general $\triangle ABC$. Using the *smallest possible* n, describe a method for cutting $\triangle AMB$ into n triangles which can be reassembled to form a triangle congruent to $\triangle AMC$.

Ray gave us his solution today. It goes like this:

Obviously $n = 1$ is too small; that only works in the special case that $\triangle ABC$ is isosceles, with $AB = AC$, and we're supposed to give a method that works for any triangle.

But $n = 2$ suffices: cut $\triangle AMB$ into two triangles by bisecting AB at E and joining EM; then rotate $\triangle AEM$ about E so that A comes to coincide with B. (This is possible because $AE = BE$ by construction.) This rotation takes A to B, leaves E fixed, and takes M somewhere, say, M'. Then the union of $\triangle BEM'$ and $\triangle BEM$ is a triangle congruent to $\triangle AMC$.



First, we show that M, E, and M' are collinear. This is simple: $\angle BEM'$ is the rotated image of $\angle AEM$, so $\angle BEM' + \angle BEM = \angle AEM + \angle BEM = 180°$.

Therefore $\triangle BEM'$ and $\triangle BEM$ together form $\triangle M'BM$. Now, since

$$\begin{aligned}\angle M'BM &= \angle M'BE + \angle EBM \\ &= \angle MAE + \angle EBM && (\angle M'BE \text{ is rotated image of } \angle MAE) \\ &= \angle AMC && (\angle AMC \text{ is exterior angle of } \triangle ABM)\end{aligned}$$

we have

$$\begin{aligned} M'B &= AM && \text{($M'B$ is rotated image of AM)} \\ \angle M'BM &= \angle AMC && \\ BM &= MC && \text{(M is midpoint of BC)} \end{aligned}$$

so by SAS, $\triangle M'BM$ is congruent to $\triangle AMC$, as claimed.

3 Upper bounds on the integers

Consider these three small theorems:

Theorem 1 $\mathbb{Z}$ has no upper bound in $\mathbb{Z}$.

Proof For any $n \in \mathbb{Z}$, we have $n+1 \in \mathbb{Z}$ and $n+1 > n$. Therefore no $n \in \mathbb{Z}$ is an upper bound for $\mathbb{Z}$. □

Theorem 2 $\mathbb{Z}$ has no upper bound in $\mathbb{Q}$.

Proof Any rational number has a representation a/b, where $a, b \in \mathbb{Z}$ and $b > 0$. By the division algorithm, there exist integers q, r such that $a = qb + r$ and $0 \le r < b$. Thus

$$\frac{a}{b} = \frac{qb+r}{b} < \frac{qb+b}{b} = q+1 \in \mathbb{Z}\,.$$

So no rational is an upper bound for $\mathbb{Z}$. □

Theorem 3 $\mathbb{Z}$ has no upper bound in $\mathbb{R}$.

Proof Suppose it did; then it would have a least upper bound, say, $M = \sup \mathbb{Z}$. Then, since $M - 1 < M$, we know that $M - 1$ is not an upper bound for $\mathbb{Z}$, so there exists an integer n such that $n > M - 1$. But then $n + 1 \in \mathbb{Z}$ and $n + 1 > M$, so M is not an upper bound for $\mathbb{Z}$ after all. □

The theorems are not surprising. What is perhaps a bit surprising is that each proof proceeds on different principles, specially suited to the algebraic structure under question. For $\mathbb{R}$, we use the completeness axiom; for $\mathbb{Q}$, the fact that it is the field of quotients of $\mathbb{Z}$; for $\mathbb{Z}$, the inductive fact that $n + 1$ is always an integer.

What we conspicuously don't have here is a proof that $\mathbb{Z}$ has no upper bound, based purely on the properties of $\mathbb{Z}$ itself. So, one wonders: is it possible to embed $\mathbb{Z}$ in a larger structure in which it does have an upper bound?

In one sense, this problem is trivial: just take the set $\mathbb{Z} \cup \{\infty\}$, where ∞ is a formal symbol with the property that $n < \infty$ for any $n \in \mathbb{Z}$. Then $\mathbb{Z}$ is bounded above by ∞. But this answer is unsatisfying; introducing this object ∞ breaks the nice algebraic properties of $\mathbb{Z}$.

So, to ask the question more precisely: does there exist an ordered, commutative ring with unity which contains $\mathbb{Z}$ as a subring, and in which $\mathbb{Z}$ is bounded above? (More precisely still, we really just want a subring which is isomorphic to $\mathbb{Z}$. It doesn't have to "really" be $\mathbb{Z}$.) We might drop some of these requirements if necessary, e.g., it would be interesting enough if we found a noncommutative ring with the other properties.

4 Converse of IVT

Recall the Intermediate Value Theorem:

> If f is continuous on $[a, b]$, and $f(a) < 0 < f(b)$, then there exists a $c \in (a, b)$ such that $f(c) = 0$.

This is what is usually proved first; then one immediately generalizes from "crossing 0" to "crossing y", obtaining:

> If f is continuous on $[a, b]$, and for some y we have $f(a) < y < f(b)$, then there exists a $c \in (a, b)$ such that $f(c) = y$.

The generalization step is easy: let $g(x) = f(x) - y$. Then g satisfies the conditions of the first version of the theorem. Similarly, by considering $g(x) = -f(x)$, we can generalize to

> If f is continuous on $[a, b]$, and for some y we have either $f(a) < y < f(b)$ or $f(b) < y < f(a)$, then there exists a $c \in (a, b)$ such that $f(c) = y$.

Let's abuse notation a bit. Normally, when one writes $[s, t]$, one means the set of values $\{x \colon s \leq x \leq t\}$, and it is implicitly assumed that $s < t$. Instead, let's define $[s, t]$ to mean the set of values between s and t (inclusive), whether $s < t$ or $s > t$ (or, indeed, $s = t$). That is:

$$[s, t] = \begin{cases} \{x \colon s \leq x \leq t\} & \text{if } s \leq t \\ \{x \colon t \leq x \leq s\} & \text{if } s > t \end{cases}$$

Further, we will write

$$f([a, b]) = \{y \colon y = f(x) \text{ for some } x \in [a, b]\} ,$$

that is, $f([a, b])$ the image of $[a, b]$ under f.

Then we can state the IVT in its full generality as follows:

$$f \text{ is continuous on } I \implies \forall a, b \in I \colon [f(a), f(b)] \subseteq f([a, b]) .$$

Now the question: is the converse of this theorem true? That is, are continuous functions the only ones with the property stated in the conclusion here?

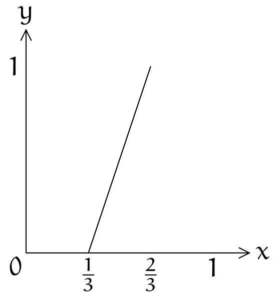
It turns out the answer is no; there are discontinuous functions with this property. We will define such a function on $(0, 1)$.

To compute $f(x)$, first compute the ternary (i.e., base 3) expansion of x. Choose a terminating expansion where possible (that is, when you have the choice between ending with infinitely many 2s or infinitely many 0s, choose the 0s). Now, if there are no 1s in the ternary expansion of x, then set $f(x) = 1$.

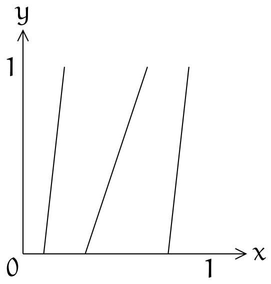
If there is a 1, then remove all the digits up to and including the first 1, and set $f(x)$ to be the number whose ternary expansion is the result.

For example, consider $11/54 = (0.0121111\ldots)_3$. This ternary expansion contains 1s, the first of which occurs in the 3^{-2} place. Removing the digits up to and including that first 1 yields $(0.21111\ldots)_3 = 5/6$. Therefore $f(11/54) = 5/6$.

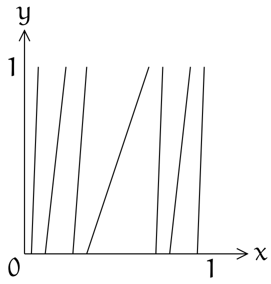
This function is a little weird. To see what it does, consider first those numbers that have a 1 in the 3^{-1} place, that is, those that have ternary expansions starting $(0.1\ldots)_3$. These are the numbers in $[\frac{1}{3}, \frac{2}{3})$. The digits after that initial 1 are then taken to represent a number which lies in $[0, 1)$; it is easy to see that the interval $[\frac{1}{3}, \frac{2}{3})$ is mapped onto $[0, 1)$ by simple scaling and translation. (Indeed, in this interval f is the same as the function $g(x) = 3x - 1$.) Furthermore, at the right endpoint of this interval, we have the number $\frac{2}{3} = (0.20000\ldots)_3$, which has no 1s; thus $[\frac{1}{3}, \frac{2}{3}]$ gets mapped to $[0, 1]$.



Similarly, considering the numbers whose first 1 is in the 3^{-2} place, that is, those in $[\frac{1}{9}, \frac{2}{9}) \cup [\frac{7}{9}, \frac{8}{9})$, we have:



and then, for 3^{-3},



and so forth. (These slashes don't fill up the whole of the domain $(0, 1)$; the missing numbers, those with no 1s in their ternary expansion, were treated specially in the definition of f.)

The idea here is that, if we pick a and b on the same slash, then f is continuous between them and so satisfies the conclusion of IVT. If we pick a and b not on the same slash, then there is a slash between them, on which f attains its full range of values, hence in particular all the values between $f(a)$ and $f(b)$. (Actually proving that takes a bit of work; it's a good exercise.)

Thus we have here a (very) discontinuous function which satisfies the conclusion of IVT, which demonstrates that the converse of IVT is false.

I don't think this is useful for anything. It's just a curiosity.