University of Cambridge
Faculty of Mathematics
Summer Research Projects

Glimpses of structure

Jack Bartley
supervised by Sean Eberhard

Introduction

What do we mean by structure? It is common to wonder what local structure, or lack thereof, tells us about an object. One type of local structure we could look for in this setting is strings of numbers separated by equal distance, known as arithmetic sequences. Motivated by this, we may also wish to look for geometric sequences.

By arithmetic and geometric sequences we mean sequences of the form $a, a+r, a+2r, \ldots, a+kr$ and $a, ar, ar^2, \ldots, ar^k$ respectively.

Arithmetic sequences

A lot is known about this case. For example van der Waerden's Theorem states that if we partition $\mathbb{N} = \{1, 2, \ldots\}$ into two sets then one of the sets contains arbitrarily long arithmetic sequences. Further, Szemerédi's Theorem states that given a set of positive upper density this set must contain arbitrarily long arithmetic sequences.

This is, roughly speaking, a measure of the 'size' of a set, defined by bootstrapping from the proportion of the numbers $1, 2, \ldots, N$ in our set.

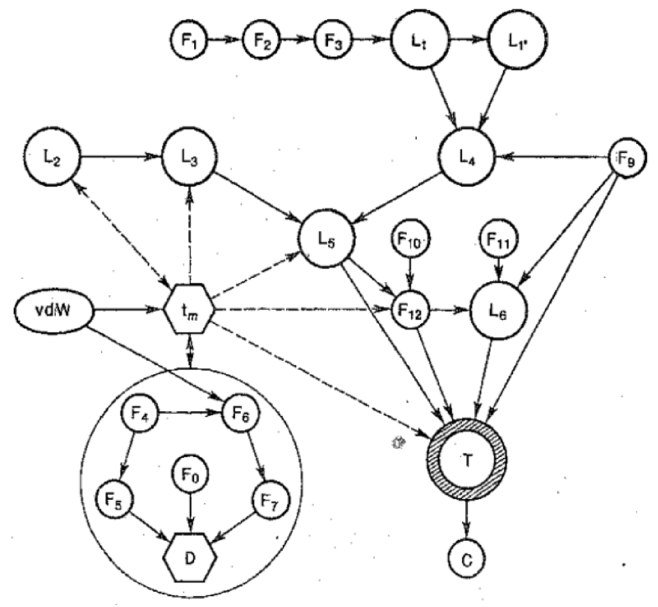

Figure : Diagram outlining various implications between parts of Szemerédi's proof of his celebrated theorem

Arithmetic-free sets An extreme case of note are those sets without any arithmetic sequences. For example, it seems intuitively as though the more we add to a set, the harder it is to avoid arithmetic sequences, but is our intuition correct? The most obvious way to attempt to create a set free of arithmetic sequences is commonly known as the greedy algorithm. Suppose that after k steps of the algorithm we have a set, say $A = \{x_0, x_1, \ldots, x_k\}$. This set will be free of arithmetic sequences by construction, and the greedy algorithm simply tells us to add the least element that doesn't create an arithmetic sequence.

Pesky primes If we're not so bothered about having absolutely no arithmetic sequences, we might instead like to just forbid arithmetic sequences of length k. It's a curious fact that if we run the greedy algorithm for this condition when k is prime, the set that the greedy algorithm spits out is precisely the set of those numbers that don't contain a digit $(k-1)$ in their base k expansion! Arguably even more curious is the fact that little is known about the structure of the set given by the greedy algorithm when k is not prime.

Behrend's construction The upper density of these greedy sets is 0, whereas pushing further in this direction, Behrend gave a construction, which was for sixty years unbeaten. Roughly speaking it uses a cunning embedding into $\mathbb{N}$ and the fact that a sphere in $\mathbb{Z}^d$ doesn't contain any arithmetic sequences (strictly speaking, something analogous to an arithmetic sequence, replacing a and r from our definition with elements of $\mathbb{Z}^d$).

Geometric sequences

A lot less is known about the analogous problems for geometric sequences. To get a feel for the problem it's instructive to see if the results mentioned for arithmetic sequences hold when we replace arithmetic by geometric.

van der Waerden for geometric sequences? This is a consequence of van der Waerden's Theorem for arithmetic sequences. Although not immediately obvious, given a partition, if we look at the powers of 2, then take logs to the base 2, one of the resulting sets under the induced partition must contain an arithmetic sequence, and this corresponds to a geometric sequence in the original setting. Since $a, a+r, a+2r$ corresponds to $2^a, 2^{a+r}, 2^{a+2r}$ which we can see is a geometric sequence by writing as $2^a, 2^a(2^r), 2^a(2^r)^2$.

. . . Yes

Szemerédi for geometric sequences? The set of square-free numbers has density $6/\pi^2 \approx 0.6079$ so Szemerédi's Theorem doesn't carry over.

. . . No

We say a number is square-free if it has no square divisors

Geometric-free sets: Are the square-frees best?

The Fundamental Theorem of Arithmetic tells us that for each number $n \in \mathbb{N}$ there is a unique expression of n as a product of primes. Inspired by this correspondence and the correspondence between an arithmetic sequence in indices and geometric sequences of powers of 2 (from *van der Waerden for geometric sequences*) we may start to think instead about which indices we allow. Notice that the square-frees are exactly those numbers such that if we express $n = 2^{i_1}3^{i_2}\ldots$ as a product of primes, then we allow the is to be 0 or 1. We might then ask, can we do slightly better? If we were to take the set of numbers where the is were restricted to being in some set A, then which set would be best? It's fairly clear that we could allow the is to be 0 or 1 or 3 without any problems, and the set of such ns would contain the square frees.

. . . Not by a long shot

If not 0 and 1, which indices are best? In fact, using this construction, the best choice of A is the so called Cantor set from earlier, those n without a 2 in their base 3 expansion. Showing this requires we look in more detail at how to calculate the densities of such sets, but intuitively, since density is defined in terms of the proportion of those $n \leq N$ which we include, it makes sense that we care a lot more about allowing small numbers in A than we do about adding large numbers, since those n divisible by p^i for some large i come along relatively infrequently. Then it's plausible that greedy does best since greedy is keen to get its hands on the smallest number it can at any stage.

. . . the greedy set from earlier!

This gives the density $\zeta(2)\prod_{t=1}^{\infty}\zeta(3^t)/\zeta(2\cdot 3^t) \approx 0.7197$

What about a greedy algorithm? Surprisingly, the set we just described *is* the set you get if you use a greedy algorithm to construct a set without geometric sequences.

. . . we might as well have used the greedy algorithm!

But surely we can do better? As I understand it, nobody has beaten this example, though this one is well known.

. . . there are no known improvements

Strictly speaking we have implicitly not restricted r to being an integer and have allowed $9, 15, 25$ as a valid geometric sequence (with ratio $5/3$). With this restriction Beiglböcka, Bergelson, Hindman and Strauss have beaten the aforementioned.

Upper bounds

In terms of density, we cannot do arbitrarily well with sets free of geometric sequences. All bounds I've seen proceed by taking a large list of disjoint geometric sequences and observing that any set free of geometric sequences must be missing at least one element of each sequence from the decomposition, where the best of which gives the upper bound ≈ 0.8494.

Bounded gaps and open problems

Another open problem is whether a set with bounded gaps must contain a geometric sequence. In fact, it is not even known whether a set with gaps bounded by 1 must contain a geometric sequence or not. Further the upper bounds for the arithmetic sequences problem are a long way off the best lower bounds, which have been improved upon since Behrend's construction, and it is suspected that Behrend's bound may be the optimal density.