Arithmetic Progressions

Jacob Denson

October 13, 2017

1 Difference Sets Without Squares

Recall that if X and Y are subsets of integers, we let

$$X \pm Y = \{x \pm y : x \in X, y \in Y, x \pm y > 0\}$$

The differences of a set X are elements of $X - X$. The *difference set problem* asks to consider how large a subset of the integers whose differences do not contain the square of any positive integer can be. We let D_n denote the maximum number of integers which can be selected from $[1, n]$ whose differences do not contain a square.

Example. *The set $X = \{1, 3, 6, 8\}$ is squarefree, because $X - X = \{2, 3, 5, 7\}$, and none of these elements are perfect squares. On the other hand, $\{1, 3, 5\}$ is not a squarefree subset, because $5 - 1 = 4$ is a perfect square.*

It is easily to greedily construct fairly large subsets of the integers by a 'sieve'-type algorithm. We start by writing out the sequence of integers

$$1, 2, \ldots, n$$

between 1 and n. Then, while we still have numbers to pick, we greedily select the smallest number x_* we haven't crossed out of the list, cross it out, and then cross out all integers y such that $y - x_*$ is a positive square. Since we cross out $x_*, x_*+1, x_*+4, \ldots, x_*+m^2$, where m is the largest integer with $x_*+m^2 \leq n$, we find $m \leq \sqrt{n - x_*} \leq \sqrt{n-1}$, hence we cross out at most $\sqrt{n-1}+1$ integers. When the algorithm terminates, all integers must be crossed out, and if the algorithm runs N iterations, a union bound gives that we cross out at most $N[\sqrt{n-1}+1]$ iterations, hence $N[\sqrt{n-1}+1] \geq n$. It follows that we construct a squarefree subset of the integers with at least

$$\frac{\sqrt{n-1}+1}{n} = \Omega(\sqrt{n})$$

elements. What's more, this algorithm generates an increasing family of squarefree subsets of the integers as n increases, so we may take the union of these subsets over all n to find an infinite squarefree subset X with $|X \cap [1, n]| = \Omega(\sqrt{n})$.

In 1978, Sárközy proved an upper bound on the size of squarefree subsets of the integers, showing $D_n = O(n(\log n)^{-1/3+\varepsilon})$ for every $\varepsilon > 0$. This shows that D_n grows at most slightly slower than linearly. In particular, this proves a conjecture of Lovász that every infinite squarefree subset has density zero, because if X is any infinite squarefree subset, then

$$\frac{|X \cap [1,n]|}{n} \leq \frac{D(n)}{n} = O(\log(n)^{-1/3+\varepsilon})$$

This tends to zero. Sárközy even conjectured that $D_n = O(n^{1/2+\varepsilon})$ for all $\varepsilon > 0$, which essentially says that the greedy technique of selecting squarefree subsets of the integers is asymptotically optimal. This is incredibly pessimistic, because the sieve selection method doesn't depend on any properties of the set of square integers. In general, if $X = \{x_1 < x_2 < \dots\}$ is any sequence of positive integers, the sieve strategy on $[1,n]$ produces a set containing no 'X differences' with at least $n(1 + K_n)^{-1}$ elements, where K_n is the greatest integer with $x_{K_n} \leq n-1$. Ruzsa's paper shows that we can take advantage of the properties of the perfect squares to obtain quadratically better squarefree subsets of the integers, constructing an infinite squarefree subset X with $|X \cap [1,n]| = \Omega(n^{0.73})$. The method reduces the problem to maximizing squarefree subsets modulo a squarefree integer m.

Theorem 1. *If m is a squarefree integer, then*

$$D_n \geq m^{-1} n^{\gamma_m}$$

where

$$\gamma_m = \frac{1}{2} + \frac{\log_m |R^*|}{m}$$

and R^* denotes the maximal subset of $[1,m]$ whose differences contain no squares modulo m. Setting $m = 65$ gives

$$\gamma_m = \frac{1}{2}\left(1 + \frac{\log 7}{\log 65}\right) = 0.733077\dots$$

which is the required result. For $m = 2$, we find $D_n \geq \sqrt{n}/2$, which is only slightly worse than the sieve result.

Let us look at the analysis of the sieve method backwards. Rather than fixing n and trying to find optimal solutions of $[1,n]$, let's fix a particular strategy (to start with, the sieve strategy), and think of varying n and seeing how the size of the solution given by the strategy on $[1,n]$ increases over time. In our analysis, the size of a solution is directly related to the number of iterations the stategy can produce before it runs out of integers to add to a solution set. Because we apply a union bound in our analysis, the cost of each particular new iteration is the same as the cost of the other iterations. If the cost of each iteration was independant of n, we could increase the solution size by increasing n by a fixed constant, leading to family of solutions which increases on the order of n.

However, as we increase n, the cost of each iteration increases on the order of $\sqrt{n}$, leading to us only being able to perform $n/\sqrt{n} = \sqrt{n}$ iterations for a fixed n. Rusza's method applies the properties of the perfect squares to perform a similar method of expansion. At an exponential cost, Rusza's method increases the solution size exponentially. The advantage of exponentials is that, since Rusza's is based on a particular parameter, a squarefree integer m, we can vary m to make the exponentials match up to give the best possible bound.

1.1 Rusza's Method

The idea of Rusza's construction is to break the problem into exponentially large intervals, upon which we can solve the problem modulo an integer. More generally, Rusza's method works on the problem of constructing subsets of the integers whose differences are d'th powers-free. Let $R \subset [1, m]$ be a subset of integers such that no difference is a power of d modulo m, where m is a *squarefree integer*. Construct the set

$$A = \left\{ \sum_{k=0}^{n} r_k m^k : 0 \leq n < \infty, r_k \in [1, m], r_k \in R \text{ when } d \text{ divides } k \right\}$$

we claim that A is squarefree. Suppose that we can write

$$\sum (r_k - r'_k) m^k = N^d$$

Set s to be the smallest index with $r_s \neq r'_s$. Then

$$(r_s - r'_s) m^s + M m^{s+1} = N^d$$

where M is some positive integer. If $s = ds_0$, then

$$(N/m^{s_0})^d = (r_s - r'_s) + Mm$$

and this contradicts the fact that $r_s - r'_s$ cannot be a d'th power modulo m. On the other hand, we know m^s divides N^d, but m^{s+1} does not. This is impossible if s is not divisible by d, because primes in N^d occur in multiples of d, and m is squarefree. For any n, we find

$$A \cap [1, m^n] = \left\{ \sum_{k=0}^{n-1} r_k m^k : r_k \in [1, m], r_k \in R \text{ when } d \text{ divides } k \right\}$$

which therefore has cardinality

$$\begin{aligned} |R|^{1+\lfloor n-1/d \rfloor} m^{n-1-\lfloor n-1/d \rfloor} &= m^n \left(\frac{|R|}{m} \right)^{1+\lfloor n-1/d \rfloor} \\ &\geq m^n \left(\frac{|R|}{m} \right)^{n+1/d} = \frac{(m^{n+1})^{1-1/d+\log_m |R|/d}}{m} \\ &= \frac{(m^{n+1})^{\gamma(m,d)}}{m} \end{aligned}$$

where $\gamma(m,d) = 1 - 1/d + \log_m |R|/d$. Therefore, for $m^{n+1} \geq k \geq m^n$

$$A \cap [1,k] \geq A \cap [1,m^n] \geq \frac{(m^{n+1})^{\gamma(m,d)}}{m} \geq \frac{k^{\gamma(m,d)}}{m}$$

This completes Rusza's construction. Thus we have proved a more general result than was required.

Theorem 2. *For every d and squarefree integer m, we can construct a set X whose differences contain no dth powers and*

$$|X \cap [1,n]| \geq \frac{n^{\gamma(d,m)}}{m} = \Omega(n^{\gamma(d,m)})$$

where $\gamma(d,m) = 1 - 1/d + \log_m |R^*|/d$, and R^* is the largest subset of $[1,m]$ containing no d'th powers modulo m.

For $m = 65$, the group $\mathbf{Z}_{65} \cong \mathbf{Z}_5 \times \mathbf{Z}_{13}$ has a set of squarefree residues of the form $\{(0,0),(0,2),(1,8),(2,1),(2,3),(3,9),(4,7)\}$, which gives the required result. Rusza believes that we cannot choose m to construct squarefree subsets of the integers growing better than $\Omega(n^{3/4})$, and he claims to have proved this assuming m is squarefree and consists only of primes congruent to 1 modulo 4.

1.2 Logarithmic comparison of D_n's growth

If we let $D_n(d)$ denote the largest subset of $[1,n]$ containing no dth powers of positive integer, The last part of Rusza's paper is devoted to lower bounding the growth of D_n over time relative to the logarithm. In general, let $D_n(d)$ denote the largest subset of $[1,n]$ containing no dth powers. Rusza proves

Theorem 3. *If p is the least prime congruent to one modulo $2d$, then*

$$\limsup_{n\to\infty} \frac{\log D_n(d)}{\log n} \geq 1 - \frac{1}{d} + \frac{\log_p d}{d}$$

Proof. The set X we constructed in the last theorem shows that for any m,

$$\frac{\log D_n(d)}{\log n} \geq \gamma(d,m) - \frac{\log m}{\log n} = 1 - \frac{1}{d} + \frac{\log_m |R^*|}{d} - \frac{\log m}{\log n}$$

Hence

$$\limsup_{n\to\infty} \frac{\log D_n(d)}{\log n} \geq 1 - \frac{1}{d} + \frac{\log_m |R^*|}{d}$$

The claim is then proven by the following lemma. □

Lemma 1. *If p is a prime congruent to 1 modulo $2d$, then we can construct a set $R \subset [1,p]$ whose differences do not contain a dth power modulo p with $|R| \geq d$.*

Proof. Let $Q \subset [1,p]$ be the set of powers $1^k, 2^k, \ldots, p^k$ modulo p. We have

$$|Q| = \frac{p-1}{k} + 1$$

This follows because the nonzero elements of Q are the images of the group homomorphism $x \mapsto x^k$ from $\mathbf{Z}_p^*$ to itself. Since $\mathbf{Z}_p^*$ is cyclic, the equation $x^k = 1$ has the same number of solutions as the equation $kx = 0$ modulo $p-1$, and since $p \equiv 1$ modulo $2k$, there are exactly k solutions to this equation. The sieve method yields a kth power modulo p free subset of size greater than or equal to

$$p/q = \frac{p}{1 + \frac{p-1}{k}} = \frac{pk}{p+k-1} \to k$$

as $p \to \infty$, which is greater than $k-1$ for large enough p (this shows the theorem is essentially trivial for large enough primes, because we don't need to use any particularly interesting properties of the squares to prove the theorem). However, for smaller primes a more robust analysis is required. We shall construct a sequence $b_1, \ldots, b_k \in \mathbf{Z}_p$ such that $b_i - b_j \notin Q$ for any i, j and

$$|B_j + Q| \leq 1 + j(q-1)$$

Given $b_1, \ldots, b_j$, let b_{j+1} be any element of

$$(B_j + Q + Q) - (B_j + Q)$$

Since $b_{j+1} \notin B_j + Q$, $b_{j+1} - b_i \notin Q$ for any i. Since $b_{j+1} \in B_j + Q + Q$, the sets $B_j + Q$ and $b_{j+1} + Q$ are not disjoint (note $Q = -Q$ because $p \equiv 1 \mod 2k$), and so

$$\begin{aligned} |B_{j+1} + Q| &= |(B_j + Q) \cup (b_{j+1} + Q)| \\ &\leq |B_j + Q| + |b_{j+1} + Q| - 1 \\ &\leq 1 + j(q-1) + q - 1 \\ &= 1 + (j+1)(q-1) \end{aligned}$$

This procedure ends when $B_j + Q + Q = B_j + Q$, and this can only happen if $B_j + Q = \mathbf{Z}_p$, because we can obtain all integers by adding elements of Q recursively, so $1 + j(q-1) \geq p$, and thus $j \geq k$. □

Corollary.

$$\limsup \frac{\log D_n}{\log n} \geq \frac{1}{2} + \frac{\log_5 2}{2} = 0.71533\ldots$$

Rusza's leaves the open question of whether $\lim \log D_n / \log n$ is a number that exists. I don't entirely understand why calculating this value is important, other than that it is an interesting challenge.

2 Dimensions of Sets Uniformly Avoiding Arithmetic Progressions

A k length arithmetic progression with gap length Δ is a sequence of the form

$$\{a, a+\Delta, a+2\Delta, \ldots, a+(k-1)\Delta\}$$

It is of interest to analyze the dimension of subsets of $\mathbf{R}$ avoiding k length arithmetic progressions, for some $k \geq 3$ (the case $k = 2$ doesn't really make sense). However, such discrete subsets are very difficult to analyze the dimension of. More importantly, Keleti has constructed subsets of the real line with full Hausdorff dimension not containing any k length arithmetic progressions. In this paper, we upper bound the dimension of subsets containing a 'wider' family of sequences than arithmetic progressions.

If we consider the arithmetic progression above, then an *almost arithmetic progression* with error $\varepsilon > 0$, length k and gap length Δ is a sequence of the form $b_0, b_1, \ldots, b_{k-1}$, which uniformly approximates a k length arithmetic progression, so that $|a+k\Delta-b_k| < \varepsilon\Delta$. We will often abbreviate this discussion by saying a sequence is a (k, ε) progression The advantage of an almost arithmetic progression is that it restricts our sets from occuring in intervals, rather than points, which leads to a decrease in the Hausdorff dimension of sets.

Theorem 4. *If $X \subset \mathbf{R}$ contains no (k, ε) almost arithmetic progressions, for $\varepsilon \in (0, 1)$, then*

$$\dim_A(X) \leq 1 - \frac{\log(\frac{k}{k-1})}{\log k\lceil 1/2\varepsilon \rceil}$$

where $\dim_A(X)$ is the Assoud dimension of X, the smallest number such that if $s > \dim_A(X)$, then for radius R and $x \in \mathbf{R}$, the number of balls of radius $r \leq R$ required to cover $B_x(R) \cap X$ is bounded up to a constant by $(R/r)^s$.

Proof. Consider some interval $I \subset \mathbf{R}$ of length R, and fix $r < R$. If $1/2\varepsilon$ was an integer, we could divide I evenly into $k/2\varepsilon$ intervals of length $(2\varepsilon/k)R$. If we partition these intervals into k classes on which the midpoints of the intervals are separated by a multiple of $1/2\varepsilon$, then a given X cannot intersect all the intervals in a given class, because this would give a (k, ε) progression. We conclude that $X \cap I$ intersects at most

$$\frac{k-1}{2\varepsilon}$$

intervals of length $(2\varepsilon/k)R$. Repeating this argument on each of these intervals recursively, we conclude that $X \cap I$ intersects at most

$$\left(\frac{k-1}{2\varepsilon}\right)^m$$

intervals of length $(2\varepsilon/k)^m R$. If we choose $(2\varepsilon/k)^m R \leq r$, then the number of intervals of radius r to cover $I \cap X$ is bounded by $\left(\frac{k-1}{2\varepsilon}\right)^m$. This occurs, in particular, if

$$m = \left\lceil \log_{2\varepsilon/k} r/R \right\rceil$$

For any $\delta > 0$, we can write

$$m = \left\lceil \frac{\log R/r}{\log k/2\varepsilon} \right\rceil \leq (1+\delta)\frac{\log R/r}{\log k/2\varepsilon}$$

Provided R/r is sufficiently large, which we may assume. Thus $I \cap X$ is covered by at most

$$\left(\frac{k-1}{2\varepsilon}\right)^{(1+\delta)\frac{\log R/r}{\log k/2\varepsilon}} = \left(\frac{R}{r}\right)^{(1+\delta)\frac{\log k-1/2\varepsilon}{\log k/2\varepsilon}}$$

Since I was arbitrary, it follows that the Assoud dimension of X is less than or equal to

$$(1+\delta)\frac{\log \frac{k-1}{2\varepsilon}}{\log \frac{k}{2\varepsilon}} = (1+\delta) - \frac{\log k/k-1}{\log k/2\varepsilon}$$

We then let $\delta \to 0$ to get the required result. If $1/2\varepsilon$ is not an integer, we may simply replace ε by $\varepsilon' = 1/2\lceil 1/2\varepsilon \rceil$ to get the required result, since any set containing no (k,ε) arithmetic progressions also contains no (k,ε') arithmetic progressions since $\varepsilon' \leq \varepsilon$. □

2.1 Constructing Sets Avoiding Arithmetic Progressions

Now we have upper bounded the dimension of sets avoiding uniform almost arithmetic progressions, we construct subsets avoiding uniform almost arithmetic progressions with a high enough dimension. We first provide a technical lemma showing that if we remove a wide enough chunk from the middle of an interval, (k,ε) arithmetic progressions must occur on one side of the chunk or the other.

Lemma 2. *Let I be a closed interval of length $|I|$, and $J \subset I$ an open interval of length $|J|$, which is sufficiently wide, so that*

$$\frac{1+2\varepsilon}{k-1-2\varepsilon}|I| < |J|$$

Then a (k,ε) arithmetic progression in $I - J$ must occur either entirely to the left of J or entirely to the right.

Proof. If I contains a (k,ε) arithmetic progression with some gap length Δ, then

$$|I| \geq (k-1-2\varepsilon)\Delta$$

On the other hand, if the progression is in $I - J$, and occurs on both sides of J, it must 'bridge the gap' between I and J, so

$$\Delta(1+2\varepsilon) \geq |J|$$

But this implies

$$|J| \leq \frac{1+2\varepsilon}{k-1-2\varepsilon}|I|$$

So the hypothesis of the theorem is contradicted. □

We use this lemma to construct a set avoiding (k, ε) arithmetic progressions of dimension

$$\frac{\log 2}{\log \frac{2k-2-4\varepsilon}{k-2-4\varepsilon}}$$

Suppose $\varepsilon < \min(1, k-2/4)$, and take an increasing sequence c_m which converges to $(k-2-4\varepsilon)/(2k-2-4\varepsilon)$. Set $X_0 = [0,1]$, and then

$$X_{m+1} = c_m X_m \cup (c_m X_m + 1 - c_m)$$

Then $\bigcap X_m$ does not contain any (k, ε) arithmetic progressions, because for each interval I in X_{m+1} we add a hole of length

$$|I|(1-2c_m) > |I|\frac{1+2\varepsilon}{k-1-2\varepsilon}$$

References

[1] I. Z. Ruzsa *Difference Sets Without Squares*

[2] Jonathan M. Fraser, Kota Saito, Han Yu *Dimensions of Sets Which Uniformly Avoid Arithmetic Progressions*