Squarefree Sets

Jacob Denson

February 14, 2018

Contents

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 a large sequence of integers

$$1, 2, 3, 4, \dots, N$$

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 $x_*, x_* + 1, x_* + 4, \dots, x_* + m^2$, 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 $n[\sqrt{N-1} + 1]$ integers, hence $n[\sqrt{N-1}+1] \geq N$. It follows that we construct a squarefree subset of the integers with at least

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

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

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

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 provides some refreshing optimisim, showing we can take advantages of the structure of perfect suqares 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.

Remark. *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 how we like to obtain a better polynomial lower bound.*

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

$$A = \left\{ \sum_{k=0}^{n} r_k m^k : 0 \le n < \infty, r_k \in \begin{Bmatrix} R & d \text{ divides } N \\ [1,m] & \text{otherwise} \end{Bmatrix} \right\}$$

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

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

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

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

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

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

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*

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

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. Looking at some sophisticated papers in number theory (I forgot to write down these papers while I was looking them up), it seems this is nowadays quite easy to prove. Thus expanding on Rusza's result in the discrete case requires a new strategy, or perhaps Rusza's result is the best possible...

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 polynomial growth of $D(N,d)$ over asymptotically with respect to N. 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. $\square$

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 $1^k, 2^k, \dots, p^k$ 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. *In the special case of avoiding squarefree numbers, we find*

$$\limsup \frac{\log D(N)}{\log N} \geq \frac{1}{2} + \frac{\log_5 2}{2} = 0.71533\ldots$$

Rusza's leaves the ultimate question of whether one can calculate

$$\alpha = \lim_{N \to \infty} \log D(N) / \log N$$

or even whether it exists at all. The consequence of this would essentially solve the squarefree integers problem, since it would give the exact growth of $D(N) \sim N^\alpha$ in terms of a monomial. Of course, because of how conclusive this problem is, it is much more nontrivial than the work of Rusza's paper.

2 Ideas For New Work

A continuous formulation of the squarefree difference problem is not so clear to formulate, because every positive real number has a square root. Instead, we consider a problem which introduces a similar structure to avoid in the continuous domain rather than the discrete. Unfortunately, there is no direct continuous anology to the squarefree subset problem on the interval $[0, 1]$, because there is no canonical subset of $[0, 1]$ which can be identified as 'perfect squares', unlike in $\mathbf{Z}$. If we only restrict ourselves to perfect squares of a countable set, like perfect squares of rational numbers, a result of Kaletti gives us a set of full

Hausdorff dimension avoiding this set. Thus, instead, we say a set $X \subset [0,1]$ is (continuously) **squarefree** if there are no nontrivial solutions to the equation $x - y = (u - v)^2$, in the sense that there are no $x, y, u, v \in X$ satisfying the equation for $x \neq y, u \neq v$. In this section we consider some blue sky ideas that might give us what we need.

How do we adopt Rusza's power series method to this continuous formulation of the problem? We want to scale up the problem exponentially in a way we can vary to give a better control of the exponentials. Note that for a fixed m, every elements $x \in [0,1]$ has an essentially unique m-ary expansion

$$x = \sum_{n=1}^{\infty} \frac{x_n}{m^n}$$

and the pullback to the Haar measure on $\mathbf{F}_m^\infty$ is measure preserving (with respect to the natural Haar measure on $\mathbf{F}_m^\infty$), so perhaps there is a way to reformulate the problem natural as finding nice subsets of $\mathbf{F}_m^\infty$ avoiding squares. In terms of this expansion, the equation $x - y = (u - v)^2$ can be rewritten as

$$\sum_{n=1}^{\infty} \frac{x_n - y_n}{m^n} = \left(\sum_{k=1}^{\infty} \frac{u_n - v_n}{m^n} \right)^2 = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n-1} (u_k - v_k)(u_{n-k} - v_{n-k}) \right) \frac{1}{m^n}$$

One problem with this expansion is that the sums of the differences of each element do not remain in $\{0, \dots, m-1\}$, so the sum on the right cannot be considered an equivalent formal expansion to the expansion on the left. Perhaps $\mathbf{F}_m^\infty$ might be a simpler domain to explore the properties of squarefree subsets, in relation to Ruzsa's discrete strategy. What if we now consider the problem of finding the largest subset X of $\mathbf{F}_m^\infty$ such that there do not exist $x, y, u, v \in \mathbf{F}_m^\infty$ such that if $x, y, u, v \in X$, $x \neq y$, $u \neq v$, then for any n

$$x_n - y_n \neq \sum_{k=1}^{n-1} (u_k - v_k)(u_{n-k} - v_{n-k})$$

What if we consider the problem modulo m, so that the convolution is considered modulo m, and we want to avoid such differences modulo m. So in particular, we do not find any solutions to the equation

$$\begin{aligned} x_2 - y_2 &= (u_1 - v_1)^2 \\ x_3 - y_3 &= 2(u_1 - v_1)(u_2 - v_2) \\ x_4 - y_4 &= (u_1 - v_1)(u_3 - v_3) + (u_2 - v_2)^2 \\ &\vdots \end{aligned}$$

which are considered modulo m. The topology of the p-adic numbers induces a power series relationship which 'goes up' and might be useful to our analysis, if the measure theory of the p-adic numbers agrees with the measure theory of normal numbers in some way, or as an alternate domain to analyze the squarefree problem as with $\mathbf{F}_m^\infty$.

The problem with the squarefree subset problem is that we are trying to optimize over two quantities. We want to choose a set X such that the number of distinct differences $x - y$ as small as possible, while keeping the set as large as possible. This double optimization is distinctly different from the problem of finding squarefree difference subsets of the integers. Perhaps a more natural analogy is to fix a set V, and to find the largest subset X of $[0,1]$ such that $x-y = (u-v)^2$, where $x \neq y \in X$, and $u \neq v \in V$. Then we are just avoiding subsets of $[0,1]$ which avoid a particular set of differences, and I imagine this subset has a large theory. But now we can solve the general subset problem by finding large subsets X such that $(X - X)^2 \subset V$ and X containing no differences in V. Does Rusza's method utilize the fact that the problem is a single optimization? Can we adapt Rusza's method work to give better results about finding subsets X of the integers such that $X - X$ is disjoint from $(X - X)^2$?

3 Modulus Techniques For Finding High Dimensional Continuous Squarefree Sets

We now try to adapt Ruzsa's idea of applying congruences modulo m to avoid squarefree differences on the integers to finding high dimensional subsets of $[0,1]$ which satisfy a continuous analogy of the integer constraint. One problem with the squarefree problem is that solutions are non-scalable, in the sense that if $X \subset [N]$ is squarefree, αX may not be squarefree. This makes sense, since avoiding solutions to $\alpha(x - y) = \alpha^2(u - v)^2$ is clearly not equivalent to the equation $x - y = (u - v)^2$. As an example, $X = \{0, 1/2\}$ is squarefree, but $2X = \{0, 1\}$ isn't. On the other hand, if X avoids squarefree differences modulo N, it *is* scalable by a number congruent to 1 modulo N. More generally, if α is a rational number of the form p/q, then αX will avoid nontrivial solutions to $q(x-y) = p(u-v)^2$, and if p and q are both congruent to 1 modulo N, then X is squarefree, so modulo arithmetic enables us to scale down. Since the set of rational numbers with numerator and denominator congruent to 1 is dense in $\mathbf{R}$, *essentially* all scales of X are continuously squarefree. Since X is discrete, it has Hausdorff dimension zero, but we can 'fatten' the scales of X to obtain a high dimension continuously squarefree set. To initially simplify the situation, we now choose to avoid nontrivial solutions to $y - x = (z - x)^2$, removing a single degree of freedom from the domain of the equation.

So we now fix a subset X of $\{0, \dots, m-1\}$ avoiding squares modulo m. We now ask how large can we make ε such that nontrivial solutions to $x-y = (x-z)^2$ in the set

$$E = \bigcup_{x \in X} [\alpha x, \alpha x + \varepsilon)$$

occur in a common interval, if α is just short of $1/m^n$. This will allow us to recursively place a scaled, 'fattened' version of X in every interval, and then consider a limiting process to obtain a high dimensional continuously squarefree set. If we have a nontrivial solution triple, we can write it as $\alpha x + \delta_1, \alpha y + \delta_2$, and $\alpha z + \delta_3$, with $\delta_1, \delta_2, \delta_3 < \varepsilon$. Expanding the solution leads to

$$\alpha(x-y) + (\delta_1 - \delta_2) = \alpha^2(x-z)^2 + 2\alpha(x-z)(\delta_1 - \delta_3) + (\delta_1 - \delta_3)^2$$

If x, y, and z are all distinct, then, as we have discussed, we cannot have $\alpha(x-y) = \alpha^2(x-z)^2$. if α is chosen close enough to $1/m^n$, then we obtain an approximate inequality

$$|\alpha(x-y) - \alpha^2(x-z)^2| \geq \alpha^2$$

(we require α to be close enough to $1/n$ for some n to guarantee this). Thus we can guarantee at least two of x, y, and z are equal to one another if

$$|2\alpha(x-z)(\delta_1 - \delta_3) + (\delta_1 - \delta_3)^2 - (\delta_1 - \delta_2)| < \frac{1}{m^{2n}}$$

We calculate that

$$2\alpha(x-z)(\delta_1 - \delta_3) + (\delta_1 - \delta_3)^2 - (\delta_1 - \delta_2) < 2\alpha(m-1)\varepsilon + \varepsilon^2 + \varepsilon$$
$$(\delta_1 - \delta_2) - 2\alpha(x-z)(\delta_1 - \delta_3) - (\delta_1 - \delta_3)^2 \leq \varepsilon + 2\alpha(m-1)\varepsilon$$

So it suffices to choose ε such that

$$\varepsilon^2 + [2\alpha(m-1) + 1]\varepsilon \leq \alpha^2$$

This is equivalent to picking

$$\varepsilon \leq \sqrt{\left(\frac{2\alpha(m-1)+1}{2}\right)^2 + \alpha^2} - \frac{2\alpha(m-1)+1}{2} \approx \frac{\alpha^2}{2\alpha(m-1)+1}$$

We split the remaining discussion of the bound we must place on ε into the three cases where two of x, y, and z are equal, but one is distinct, to determine how small ε must be to prevent this from happening. Now

- If $y = z$, but x is distinct, then because we know $\alpha(x-y) = \alpha^2(x-y)^2$ has no solution in X, we obtain that (provided α is close enough to $1/m^n$),

  $$|\alpha(x-y) - \alpha^2(x-y)^2| \geq \alpha^2$$

  and the same inequality that worked for the case where the three equations are distinct now applies for this case.

- If $x = y$, but z is distinct, we are left with the equation

  $$\delta_1 - \delta_2 = \alpha^2(x-z)^2 + 2\alpha(x-z)(\delta_1 - \delta_3) + (\delta_1 - \delta_3)^2$$

  Now $\alpha^2(x-z)^2 \geq \alpha^2$, and

  $$\delta_1 - \delta_2 - 2\alpha(x-z)(\delta_1 - \delta_3) - (\delta_1 - \delta_3)^2 < \varepsilon + 2\alpha(m-1)\varepsilon$$

  so we need the additional constraint $\varepsilon + 2\alpha(m-1)\varepsilon \leq \alpha^2$, which is equivalent to saying

  $$\varepsilon \leq \frac{\alpha^2}{1 + 2\alpha(m-1)}$$

- If $x = z$, but y is distinct, we are left with the equation

$$\alpha(x-y) + (\delta_1 - \delta_2) = (\delta_1 - \delta_3)^2$$

Now $|\alpha(x-y)| \geq \alpha$, and

$$(\delta_1 - \delta_3)^2 - (\delta_1 - \delta_2) < \varepsilon^2 + \varepsilon$$

$$(\delta_1 - \delta_2) - (\delta_1 - \delta_3)^2 < \varepsilon$$

so to avoid this case, we need $\varepsilon^2 + \varepsilon \leq \alpha$, or

$$\varepsilon \leq \frac{\sqrt{1+4\alpha}-1}{2} \approx \alpha$$

Provided ε is chosen as above, all solutions in E must occur in a common interval. Thus, if we now replace the intervals with a recursive fattened scaling of X, all solutions must occur in smaller and smaller intervals. If we choose the size of these scalings to go to zero, these solutions are required to lie in a common interval of length zero, and thus the three values must be equal to one another. Rigorously, we set $\varepsilon \approx 1/m^2$, and $\alpha \approx 1/m$, we can define a recursive construction by setting

$$E_1 = \bigcup_{x \in X} [\alpha x, \alpha x + \varepsilon_1)$$

and if we then set X_n to be the set of startpoints of the intervals in E_n, then

$$E_{n+1} = \bigcup_{x \in X_n} (x + \alpha^2 E_n)$$

Then $\bigcap E_n$ is a continuously squarefree subset. But what is it's dimension?

4 Kaletti's Translate Avoiding Sets

Kaletti's two page paper constructs a full dimensional subset X of $[0,1]$ such that X intersects $t + X$ in at most one place for each $t \in \mathbf{R}$. Malabika has adapted this technique to construct high dimensional subsets avoiding nontrivial solutions to differentiable functions, and she thinks we can further exploit these ideas to obtain dimension one squarefree sets. The basic, but fundamental idea of Kaletti is to introduce memory into the Cantor set constructions so the sets avoid progressions, and have dimension one. He starts by creating an initial set of disjoint intervals $X_0 = \{[0,1]\}$, and we let Q be a queue with a history of all intervals created, from left to right, so Q is initially $\{[0,1]\}$. We then perform the following procedure repeatedly, forming a decreasing family of subsets $X_0 \supset X_1 \supset \dots$, each X_k a union of intervals of length l_k:

- Take an interval I off the front of Q.

- For each interval $J \in X_n$, partition J into N_n intervals of length l_n, with the startpoints of each interval separated by $\varepsilon_n \geq l_n$. The intervals start at the beginning of J, except in the case where $J \subset I$, in which case we shift the intervals, placing the initial interval Δ_n from the startpoint. In order to avoid overlap, we must have $N_n \varepsilon_n + \Delta_n \leq |J| = l_{n-1}$. If $J \subset I$, then we shift the initial startpoint by Δ_n.

We claim that $X = \lim X_n$ is a set avoiding translates, which we reduce to showing that the sets avoid a particular equation.

Lemma 2. *A set X avoids translates if and only if there do not exists $x_1 < x_2 \leq x_3 < x_4$ in X with $x_2 - x_1 = x_4 - x_3$.*

Proof. □

First, note that if X intersected $a+X$ in two places $x = a+x'$ and $y = a+y'$, then we would find $x' - x = y' - y$. We may assume $a \leq 0$ by symmetry, so that $x' \leq x$ and $y' \leq y$. We may

TODO: ELABORATE ON KELETI'S METHOD.

5 Pramanik and Fraser's Extension of Kaletti Translation to C^1 functions

Lemma 3. *For any f, M, and $T_1, \dots, T_d$, there exists a constant C depending on these quantities, such that*

Proof. Let J be a cube with sidelengths $1/M$ intersecting the zero set of f. The implicit function theorem can then be applied to conclude that, if we write $J = J' \times I$, where J_1 is a cube in $\mathbf{R}^{d-1}$ and I is an interval, that there is a function $g : J' \to \mathbf{R}$ such that $f(x,y) = 0$ for $(x,y) \in J$ if and only if $y = g(x)$. □

Theorem 4. *Suppose that $f : \mathbf{R}^d \to \mathbf{R}$ is a C^1 function, and there are sets $T_1, \dots, T_d \subset [0,1]$, which each T_k a union of almost disjoint closed intervals of length $1/M$ such that $A \leq |\partial_d f(x)|$ and $|\nabla f(x)| \leq B$ for $x \in T_1 \times \cdots \times T_d$. There there exists a rational constant C and arbitrarily large integers $N \in M\mathbf{Z}$ for which there exist subsets $S_k \subset T_k$ such that*

(i) $f(x) \neq 0$ for $x \in S_1 \times \cdots \times S_d$.

(ii) For each $k \neq d$, If we split each interval T_k into $1/N$ intervals, then S_k contains an interval of length CN^{1-d} of each of these intervals.

(iii) If we split T_d into $1/N$ intervals, then S_d contains a length C/N portion of at least a fraction $1 - 1/M$ of the total number of these intervals (but this portion need not be a complete interval, like in the last property).

Proof. Choosing the sets S_k for $k \neq d$ is easy. We split the intervals T_k into length $1/M$ segments, and define S_k as the union of the first CN^{1-v} portions of them. This satisfies property (ii) of the theorem automatically. Now if $a \in \mathbf{R}^{d-1}$ is chosen, with each $a_k \in T_k$, then the total number of points $x \in T_d$ for which $f(a,x) = 0$ is M. This follows from the fact that T_d can contain at most M intervals of length $1/M$, and as we vary x in the equation, because the partial derivative $\partial_d f(a,x)$ is non-vanishing on the interval, the function is either increasing or decreasing on this interval. Define

$$\mathbf{A} = \{a : a_k \text{ is a startpoint of a length } 1/N \text{ segment in } T_k\}$$

Then $|\mathbf{A}| \leq N^{d-1}$, since T_k, contained in $[0,1]$, can contain at most N almost disjoint intervals of length $1/N$. This means that if

$$\mathbf{B} = \{x \in T_d : \text{there is } a \in \mathbf{A}, f(a,x) = 0\}$$

is the set of 'bad points' in T_d, then $|\mathbf{B}| \leq MN^{d-1}$. We now pick to include a portion of a length N interval I in T_d in the set S_d if it intersects relatively few bad points, i.e. $\#(\mathbf{B} \cap I) \leq M^3 N^{d-2}$. Each interval we don't pick has more than $M^3 N^{d-2}$ bad points, and since there can only be MN^{d-1} bad points, this means that we do not pick at most $MN^{d-1}/M^3 N^{d-2} = N/M^2$ intervals. Since there are at most N/M intervals in T_d, this implies that we are picking a fraction $1-1/M$ of the total number of intervals in T_d. Now we split each interval I picked into intervals of length D/N^{d-1}, for some integer D to be chosen later. Assume N is large enough that N^{d-2}/D is an integer. All intervals in this division that intersect $\mathbf{B}$ are discarded, and the remaining intervals form the set S_d. The total length of the discarded intervals is $2DM^3N^{d-2}/N^{d-1} = 3M^3D/N$, and so provided

Since there are at most N^{d-1} choices for a, where each a_k is the beginning of an interval of length $1/N$, it follows that there are at most MN^{d-1} choices of x for which there is a □

How do we use this lemma to construct a set avoiding solutions to f? We set

$$I_k[0] = \left[(k-1)\frac{\eta}{\nu}, \frac{k\eta}{\nu}\right]$$

6 Full Dimensional Squarefree Subsets Using Interval Dissection Methods

The main idea of the last section was that, for a function f, given a method that takes a sequence of disjoint unions of sets $J_1, \dots, J_N$, each a union of almost disjoint closed intervals of the same length, and gives large subsets $J'_n \subset J_n$, each a union of almost disjoint intervals of a much smaller length, such that $f(x_1, \dots, x_n) \neq 0$ for $x_n \in J'_n$. Then one can find high dimensional subsets K of the real line such that $f(x_1, \dots, x_n) \neq 0$ for a sequence of distinct $x_1, \dots, x_n \in$

K. The larger the subsets J'_n are compared to J_n, the higher the Hausdorff dimension of K. We now try and apply this method to construct large subsets avoiding solutions to the equation $f(x,y,z) = (x-y) - (x-z)^2$. In this case, since solutions to the equation above satisfy $y = x - (x-z)^2$, given J_1, J_2, J_3, finding J'_1, J'_2, J'_3 as in the method above is the same as choosing J'_1 and J'_3 such that the image of $J'_1 \times J'_3$ under the map $g(x,z) = x - (x-z)^2$ is small in J_2. We begin by discretizing the problem, splitting J_1 and J_3 into unions of smaller intervals, and then choosing large subsets of these intervals, and finding large intervals of J_2 avoiding the images of the startpoints to these intervals.

So suppose that J_1, J_2, and J_3 are unions of intervals of length $1/M$, for which we may find subsets $A, B \subset [M]$ of the integers such that

$$J_1 = \bigcup_{a \in A} \left[\frac{a}{M}, \frac{a+1}{M} \right] \qquad J_3 = \bigcup_{b \in B} \left[\frac{b}{M}, \frac{b+1}{M} \right]$$

If we split J_1 and J_3 into intervals of length $1/NM$, for some $N \gg M$ to be specified later (though we will assume it is a perfect square), then

$$J_1 = \bigcup_{\substack{a \in A \\ 0 \leq k < N}} \left[\frac{Na+k}{NM}, \frac{Na+k}{NM} + \frac{1}{NM} \right] \quad J_3 = \bigcup_{\substack{b \in A \\ 0 \leq l < N}} \left[\frac{Nb+l}{NM}, \frac{Nb+l}{NM} + \frac{1}{NM} \right]$$

We now calculate g over the startpoints of these intervals, writing

$$\begin{aligned} g\left(\frac{Na+k}{NM}, \frac{Nb+l}{NM} \right) &= \frac{Na+k}{NM} - \left(\frac{N(a-b)+(k-l)}{NM} \right)^2 \\ &= \frac{a}{M} - \frac{(a-b)^2}{M^2} + \frac{k}{NM} - \frac{2(a-b)(k-l)}{NM^2} + \frac{(k-l)^2}{(NM)^2} \end{aligned}$$

which splits the terms into their various scales. If we write $m = k - l$, then m can range on the integers in $(-N, N)$, and so, ignoring the first scale of the equation, we are motivated to consider the distribution of the set of points of the form

$$\frac{k}{NM} - \frac{2(a-b)m}{NM^2} + \frac{m^2}{(NM)^2}$$

where k is an integer in $[0, N)$, and m an integer in $(-N, N)$. To do this, fix $\varepsilon > 0$. Suppose that we find some value $\alpha \in [0,1]$ such that S intersects

$$\left[\alpha, \alpha + \frac{1}{N^{1+\varepsilon}} \right]$$

Then there is k and m such that

$$0 \leq \frac{kNM - 2N(a-b)m + m^2}{(NM)^2} - \alpha \leq \frac{1}{N^{1+\varepsilon}}$$

Write $m = q\sqrt{N} + r$ (remember that we chose N so it's square root is an integer), with $0 \leq r < \sqrt{N}$. Then $m^2 = qN + 2qr\sqrt{N} + r^2$, and if $2qr = Q\sqrt{N} + R$, where $0 \leq R < \sqrt{N}$, then we find

$$-\frac{R}{M^2 N^{3/2}} - \frac{r^2}{(NM)^2} \leq \frac{kM - 2(a-b)m + q + Q}{NM^2} - \alpha \leq \frac{1}{N^{1+\varepsilon}} - \frac{R}{M^2 N^{3/2}} - \frac{r^2}{(NM)^2}$$

Thus

$$d(\alpha, \mathbf{Z}/NM^2) \leq \max\left(\frac{1}{N^{1+\varepsilon}} - \frac{R}{\sqrt{N}} - \frac{r^2}{N}, \frac{R}{M^2 N^{3/2}} + \frac{r^2}{(NM)^2}\right)$$

If we now restrict our attention to the set S consisting of the expressions we are studying where $R \leq (\delta_0/2)\sqrt{N}$, $r \leq \sqrt{\delta_0 N/2}$, then if the interval corresponding to α intersects S, then

$$d(\alpha, \mathbf{Z}/NM^2) \leq \max\left(\frac{1}{N^{1+\varepsilon}}, \frac{\delta_0}{NM^2}\right)$$

If $N^\varepsilon \geq M^2/\delta_0$, then we can force $d(\alpha, \mathbf{Z}/NM^2) \leq \delta_0/NM^2$ for all α intersecting S. Thus, if we split J_2 into intervals starting at points of the form

$$\frac{k+1/2}{NM^2}$$

each of length $1/N^{1+\varepsilon}$, then provided $\delta_0 < 1/2$, we conclude that these intervals do not contain any points in S, since

$$d\left(\frac{k+1/2}{NM^2}, \mathbf{Z}/NM^2\right) = \frac{1}{2NM^2} > \frac{\delta_0}{NM^2}$$

So we're well on our way to using Pramanik and Fraser's recursive result, since this argument shows that, provided points in J_1 and J_3 are chosen carefully, we can keep $O_M(1/N^{1+\varepsilon})$ of each interval in J_2, which should lead to a dimension bound arbitrarily close to one.

7 Finding Many Startpoints of Small Modulus

To ensure a high dimension corresponding to the recursive construction, it now suffices to show J_1 and J_3 contain many startpoints corresponding to points in S, so that the refinements can be chosen to obtain $O_M(1/N)$ of each of the original intervals. Define T to be the set of all integers $m \in (-N, N)$ with $m = q\sqrt{N} + r$ and $r \leq \sqrt{\delta_0 N/2}$ and $2qr = Q\sqrt{N} + R$ with $R \leq (\delta_0/2)\sqrt{N}$. Because of the uniqueness of the division decomposition, we find T is in one to one correspondence with the set T' of all pairs of integers (q, r), with $q \in (-\sqrt{N}, \sqrt{N})$ and $r \in [0, \sqrt{N})$, with $r \leq \sqrt{\delta_0 N/2}$, $2qr = Q\sqrt{N} + R$, and $R \leq (\delta_0/2)\sqrt{N}$. Thus we require some more refined techniques to better upper bound the size of this set.

Let's simplify notation, generalizing the situation. Given a fixed ε, We want to find a large number of integers $n \in (-N, N)$ with a decomposition $n = qr$, where $r \leq \varepsilon\sqrt{N}$, and $q \leq \sqrt{N}$. The following result reduces our problem to understanding the distribution of the smooth integers.

Lemma 4. *Fix constants A, B, and let $n \leq AN$ be an integer. If all prime factors of n are $\leq BN^{1-\delta}$, then n can be decomposed as qr with $r \leq \varepsilon\sqrt{N}$ and $q \leq \sqrt{N}$.*

Proof. Order the prime factors of n in increasing order as $p_1 \leq p_2 \leq \cdots \leq p_K$. Let $r = p_1 \dots p_m$ denote the largest product of the first prime factors such that $r \leq \varepsilon\sqrt{N}$. If $r = n$, we can set $q = 1$, and we're finished. Otherwise, we know $rp_{m+1} > \varepsilon\sqrt{N}$, hence

$$r > \frac{\varepsilon\sqrt{N}}{p_{m+1}} \geq \frac{\varepsilon\sqrt{N}}{BN^{1-\delta}} = \frac{\varepsilon}{B} N^{\delta-1/2}$$

And if we set $q = n/r$, the inequality above implies

$$q < \frac{nB}{\varepsilon} N^{1/2-\delta} \leq \frac{AB}{\varepsilon} N^{3/2-\delta}$$

But now we run into a problem, because the only way we can set $q < \sqrt{N}$ while keeping A, B, and ε fixed constants is to set $\delta = 1$, and $AB/\varepsilon \leq 1$. $\square$

Remark. *Should we expect this method to work? Unless there's a particular reason why values of (q, r) should accumulate near $Q = 0$, we should expect to lose all but $N^{-1/2}$ of the N values we started with, so how can we expect to get $\Omega(N)$ values in our analysis. On the other hand, if a number n is suitably smooth, in a linear amount of cases we should be able to divide up primes into two numbers q and r such that r is small and q fits into a suitable value of Q, so maybe this method will still work.*

Regardless of whether the lemma above actually holds through, we describe an asymptotic formula for perfect numbers which might come in handy. If $\Psi(N, M)$ denotes the number of integers $n \leq N$ with no prime factor exceeding M, then Karl Dickman showed

$$\Psi(N, N^{1/u}) = N\rho(u) + O\left(\frac{uN}{\log N}\right)$$

This is essentially linear for a fixed u, which could show the set of (q, r) is $\Omega_\varepsilon(N)$, which is what we want.

References

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

[2] Tamás Keleti *A 1-Dimensional Subset of the Reals that Intersects Each of its Translates in at Most a Single Point*

[3] Robert Fraser, Malabika Pramanik *Large Sets Avoiding Patterns*

[4] Karl Dickman *On the Frequency of Numbers Containing Prime Factors of a Certain Relative Magnitude*