

Squarefree Sets

Jacob Denson

January 9, 2018

Contents

I	Discrete Background	1
1	Difference Sets Without Squares	1
1.1	Rusza's Method	3
1.2	Logarithmic comparison of D_n 's growth	5
II	Continuous Background	6
2	Kaletti's Translate Avoiding Sets	6
3	Fraser's Extension of Kaletti Translation	8
III	New Work	9
4	A Solution To A Continuous Squarefree	11
5	Dimensions of Sets Uniformly Avoiding Arithmetic Progressions	14
5.1	Constructing Sets Avoiding Arithmetic Progressions	16

Part I

Discrete Background

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 integers 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 each time. 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

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

which 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 strategy 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 independent 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 optimize the polynomial power of the quotient.

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

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

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

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

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 d th 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 d th 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 d th powers. Rusza proves

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

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

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

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

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 don't contain d th powers 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 k th power modulo p free subset of size greater than or equal to

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

as $p \rightarrow \infty$, 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 $b_1, \dots, b_k \in \mathbf{Z}_p$ such that $b_i - b_j \notin Q$ for any i, j and

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

Given b_1, \dots, 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 $p \equiv 1 \pmod{2k}$), 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 $j \geq k$. \square

Corollary.

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

It appears that using more sophisticated number theoretic techniques, one can show that Rusza's method is essentially bounded in growth by $n^{0.75}$, so it will require another technique to improve this bound. The problem of calculating

$$\lim_{n \rightarrow \infty} \frac{\log D_n}{\log n}$$

if it exists, remains open, which would give a polynomial answer to the sizes of sets constructed.

Part II

Continuous Background

2 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 $t \neq 0$. 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 lets Q be a queue containing all intervals created, from left to right, which we initially set to just contain the interval $[0, 1]$. 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 exist $x_1 < x_2 \leq x_3 < x_4$ in X with $x_2 - x_1 = x_4 - x_3$.*

Proof. If the equation $x_2 - x_1 = x_4 - x_3$ was satisfied on X , and we shifted X by the common value $a = x_2 - x_1 = x_4 - x_3$, then the equation says that x_1 moves to x_2 under shifting, and x_3 moves to x_4 . Conversely, if X and $a + X$ intersected in more than one location then there are $x_1, x_2, x_3, x_4 \in X$ with $a + x_1 = x_2$ and $a + x_3 = x_4$, so $x_4 - x_3 = x_2 - x_1$. We may assume $a > 0$ by symmetry, so $x_1 < x_2$ and $x_3 \leq x_4$, and we may also assume that $x_1 \leq x_3$. If $x_2 \geq x_3$, then moving the equation around gives $x_3 - x_1 = x_4 - x_2$, and so we may also assume $x_2 \leq x_3$. \square

Assuming that $l_n \rightarrow 0$, for any particular solution x_1, x_2, x_3, x_4 to the equation above, we may assume that in the construction, x_1 is eventually contained in an interval that x_2, x_3 , and x_4 are not. If this occurs on the N 'th iteration, then there are startpoints x_k° to intervals in X_N and values $0 \leq dx_k \leq l_N$ such that $x_k = x_k^\circ + dx_k$. The equation we solved then reads

$$(x_4^\circ - x_3^\circ) + (dx_4 - dx_3) = (x_2^\circ - x_1^\circ) + (dx_2 - dx_1)$$

If we assume that each x_k° is an integer multiple of ε_N , then provided $L_N < \varepsilon_N$, we conclude that $dx_4 - dx_3 = dx_2 - dx_1$, working modulo ε_N . But if x_1° is shifted in a certain way (which for a large enough N it will be, because every interval is eventually considered in the queue constructed above), which shows the equation is impossible.

In order to get the x_k° to lie at integer multiples of some $M_k > l_k$, we may assume by induction that previous start points lie on points $M_{n-1} > l_{n-1}$. Then the start points of the newly divided intervals are of the form

$$AM_{n-1} + B\varepsilon_n + \delta\Delta_n$$

so provided M_n divides M_{n-1} , ε_n , and Δ_n , we conclude these points lie at the l_n . To ensure that non-shifted values lie at even points, we also assume $2M_n$ divides M_{n-1} and ε_n , but does not divide Δ_n . To summarize, our construction requires the following parameters:

- ε_n : The gap between endpoints of an interval, and startpoints of the next.
- Δ_n : The size of the shift of intervals $J \subset I$.

- l_n : The length of an interval.
- N_n : The number of partitioned intervals at depth n .
- M_n : The integer multiples of startpoints at depth n .

The construction gives a translation free set provided that the following constraints hold.

- $\Delta_n + (N_n - 2)\varepsilon_n + l_n \leq l_{n-1}$.
- $l_n \leq \varepsilon_n$.
- $l_n \rightarrow 0$
- $2M_n \mid M_{n-1}, \varepsilon_n$.
- $M_n \mid M_{n-1}, \varepsilon_n, \Delta_n$.
- $l_1 = 1$.
- $2M_n \nmid \Delta_n$.

In particular, Kaletti chooses

- $N_n = n$.
- $l_n = 1/6^{n-1}n!$.
- $M_n = 3l_n$.
- $\Delta_n = 3l_n$.
- $\varepsilon_n = 6l_n$.

and the set he obtains has full Hausdorff dimension.

3 Fraser's Extension of Kaletti Translation

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 $g : J' \rightarrow \mathbf{R}$ such that $f(x, y) = 0$ for $(x, y) \in J$ if and only if $y = g(x)$. \square

Theorem 4. *Suppose that $f : \mathbf{R}^d \rightarrow \mathbf{R}$ 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 $x \in T_1 \times \dots \times T_d$. 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-d} 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 $2DM^3 N^{d-2}/N^{d-1} = 3M^3 D/N$, 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 \square

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

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

Part III

New Work

A continuous formulation of the squarefree difference problem is not so clear to formulate, because every positive integer has a squareroot. Instead, we ask a different problem. We say a set $X \subset [0, 1]$ is **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 $x \neq y, u \neq v$.

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

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

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. Maybe we can consider the problem of sets avoiding particular subsets having nice properties in \mathbf{F}_m^∞ .

What if we now consider the problem of finding the largest subset X of \mathbf{F}_m^∞ 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 map $x \mapsto x^2$ is a ‘squishing’ map on $[0, 1]$, which maybe influence the dimension of the sets which avoid squares?

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.

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 Ruzsa’s method utilize the fact that the problem is a single optimization? Does Ruzsa’s method work to give better results about finding subsets X of the integers such that $X - X$ is disjoint from $(X - X)^2$?

Suppose X is an ‘ α ’ squarefree subset of $[0, 1]$, in the sense that X contains no non-trivial solution to $x - y = \alpha^{-1}(u - v)^2$. Then αX is a squarefree subset of $[0, 1]$. If X and Y are squarefree, then the only possible way for $X \cup Y$ to fail to be squarefree is if t

Is there a cantor’esque technique to ‘pump up’ the dimension of X ? If we scale down X , then the scaled down version

then the scaled down version will be squarefree. Is it possible to then translate X so that the union of X and it’s translation is still squarefree?

4 A Solution To A Continuous Squarefree

We now use 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. The main idea we employ is that, while a squarefree subset of \mathbf{Z} is not scalable, a subset of $\{0, \dots, m - 1\}$ which contains no squarefree differences modulo m *scalable* by a number congruent to 1 modulo m .

Unfortunately, there is no direct continuous analogy 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} (Remark: How well can we avoid squares of rational numbers?). We instead consider a problem which enables us to build the ‘set of perfect squares’ into the choice of a feasible solution. In this paper, we consider the problem of finding subsets $X \subset [0, 1]$ of maximal Hausdorff dimension which avoid *nontrivial* solutions to

$$x - y = (x - z)^2 \tag{1}$$

in the sense that no distinct $x, y, z \in X$ solve this equation. We call a solution

this equation *continuously squarefree*. Our strategy to finding high dimensional subsets of the real numbers avoiding this equation is to employ a Cantor set construction involving large subsets of $\{0, \dots, m-1\}$ whose differences contain no squares modulo m .

One immediate problem with the class of continuously squarefree subsets of the real numbers is that they are not scale invariant. If X is a squarefree subset, then αX is squarefree only when X contains no nontrivial solutions to

$$x - y = \alpha(x - z)^2$$

For instance, $X = \{0, 1/2\}$ is squarefree, but $2X = \{0, 1\}$ is *not* squarefree. An interesting feature of modulo arithmetic to notice is that if $X \subset \{0, \dots, m-1\}$, with $X - X$ avoiding squares modulo m , and α is a rational number of the form p/q , where p and q are both congruent to 1 modulo m , then X does avoid nontrivial solutions to

$$x - y = \alpha(x - z)^2$$

because we can rearrange this equation to read $q(x - y) = p(x - z)^2$, which, when read modulo m , just reads $x - y = (x - z)^2$. 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.

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 (1) 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 its dimension?

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*

5 Dimensions of Sets Uniformly Avoiding Arithmetic Progressions

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

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

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, \dots, 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 occurring in intervals, rather than points, which leads to a decrease in the Hausdorff dimension of sets.

Theorem 5. *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

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

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 $\delta \rightarrow 0$ 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 $\varepsilon' \leq \varepsilon$. \square

5.1 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 4. *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. \square

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

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