FINITE FIELD ANALOGUES OF THE ERDOS, FALCONER, AND FURSTENBURG PROBLEMS

TERENCE TAO

ABSTRACT. In a recent paper with Nets Katz [7] we showed some connections between the Erdos, Falconer, and Furstenburg problems in Euclidean space. Here we show the same connections in the finite field setting, which is technically much simpler.

1. INTRODUCTION

Let q be a prime power, and let $F = F_q$ be the finite field of order q. We consider the following three problems:

Problem 1.1 (Finite field Erdos ring problem)**.** *Does there exist a subset $A \subset F$ of cardinality $|A| \sim \sqrt{q}$ such that $|A+A| \sim |A \cdot A| \sim \sqrt{q}$? In other words, is there an "approximate half-dimensional sub-ring" of F_q?*

Problem 1.2 (Finite field Falconer distance problem)**.** *For any set $E \in F \times F$, define the distance set[1] $\Delta(E) \subset F$ to be the set*

$$\Delta(E) = \{(x-x')^2 + (y-y')^2 : (x,y),(x',y') \in E\}.$$

Does there exist a set $E \subset F \times F$ with cardinality $|E| \sim q$ with $|\Delta(E)| \sim \sqrt{q}$?

Problem 1.3 (Finite field Furstenburg problem)**.** *Do there exist a collection P of q points in $F \times F$, and a collection L of q lines in $F \times F$, such that we have the incidence bound*

$$|\{(p,l) \in P \times L : p \in L\}| \sim q^{3/2}?$$

The first problem is an analogue of the Erdös ring problem [4], which asks whether $\mathbf{R}$ contains any sub-rings of Hausdorff dimension $1/2$. The corresponding discrete problem in $\mathbf{R}$ (in which one works with finite sets) is false; this result is initially due to Erdös, but there is a recent simple proof of Elekes [3] (which shows the inequality $\max(|A+A|, |A \cdot A|) \gtrsim |A|^{5/4}$) which uses the Szemerédi-Trotter theorem [9].

The second problem is an analogue of both the Erdös distance problem and the Falconer distance problem (if E has Hausdorff dimension 1, can we improve the trivial dimension bound $\dim(\Delta(E)) \geq 1/2$, perhaps all the way to $\dim(\Delta(E)) = 1$)? In the finite field case the bound $|\Delta(E)| \gtrsim \sqrt{q}$ is trivial (consider two points x_1, x_2 in E. For any other point $x \in E$, we have $|x-x_1|^2, |x-x_2|^2 \in \Delta(E)$, which implies that E has cardinality at most $2|\Delta(E)|^2$). In the discrete Euclidean version of this problem (change from $F \times F$ to $\mathbf{R}^2$, but keep the sets finite) much better is known;

1991 *Mathematics Subject Classification.* 42B15, 35L05.

[1]Strictly speaking, to be consistent with the Euclidean definition we should stick a square root in the definition of $\Delta(E)$, but this causes some unnecessary complications for finite fields.

for instance Chung-Szemeredi-Trotter have shown $|\Delta(E)| \gtrsim |E|^{4/5}$ (again using the Szemeredi-Trotter theorem), and this bound has recently been improved further.

The third problem is an analogue of the Furstenburg set problem. A Furstenburg set in $\mathbf{R}^2$ is like a Besicovitch set but contains only a 1/2-dimensional subset of a line in each direction (as opposed to Besicovitch sets containing a unit line segment in every dimension). The trivial lower bound on the dimension of Furstenburg sets is 1, the best known upper bound is 5/4. The discrete analogue of a Furstenburg set is a collection of N lines and N^α points such that each line contains $N^{1/2}$ points; here α is a proxy for the dimension of a Furstenburg set. In the plane the Szemerédi-Trotter theorem (again!) shows that we indeed have $\alpha \geq 5/4$, but in the finite field case we cannot improve $\alpha \geq 1$, unless we give a negative answer to Problem 1.3.

We remark that it is easy to obtain the upper bound

$$|\{(p,l) \in P \times L : p \in L\}| \lesssim q^{3/2}$$

just from using the fact that any two distinct lines intersect in at most one point (or dually, that every two distinct points determine at most one line) and using Cauchy-Schwarz. In $\mathbf{R}^2$ this bound can be improved to $q^{2/3}$ which is sharp; this is the Szemerédi-Trotter theorem.

There is a good reason why we cannot disprove any of the above three problems: they are true when q is a square! For instance, if $q = p^2$ for some prime p, then one can set $A = F_p \subset F_{p^2}$ for Problem 1.1, $E = F_p \times F_p$ for Problem 1.2, and $P = F_p \times F_p$ for Problem 1.3, with L equal to the lines with slope and intercept in F_p. (The reason the Szemerédi-Trotter theorem breaks down here is that it crucially relies on the fact that $\mathbf{R}$ is ordered, something which never happens in the finite setting).

On the other hand, we do not know whether any of the problems are true for q prime. Somehow one would have to give a robust reason as to why the field F_p is different from F_{p^2}.

All the counterexamples are sort of similar, all using the same object F_p. In fact the connection is quite strong:

Proposition 1.4. *For any fixed q, the above three problems are equivalent.*

This is the analogue of the main result of [7]. There is already some hints of connections between the three problems above (besides the fact that they all seem to have Erdös's name attached to them somehow :-), especially in that the Szemerédi-Trotter theorem makes an appearance over and over again.

Very roughly, the connections are as follows:

- (Problem 1.1 implies Problem 1.2) If A is a solution to Problem 1.1, then (after some initial refinement) $E = A \times A$ is a solution to Problem 1.2.
- (Problem 1.2 implies Problem 1.3) If E is a solution to Problem 1.2, then $P := E$ and $L := \{$angle bisectors of points in $E\}$ is (after some refinement) a solution to Problem 1.3.
- (Problem 1.3 implies Problem 1.2) If P, L is a solution to Problem 1.3, then $A := \{$slopes of L through a fixed point in $P\}$ is a solution to Problem 1.1.

We now expand upon these three connections. Our arguments shall be mostly informal; see [7] and the other references for the rigorous arguments.

2. From Erdos to Falconer

Let $A \subset F$ be a solution to Problem 1.1. Morally speaking, this means that A is approximately closed under addition and multiplication (perhaps after some affine renormalization). Thus A is something like a sub-ring of F. If we set $E := A \times A$, then E has cardinality q, and the distance set of E is basically $(A-A)^2 + (A-A)^2$, which heuristically has size about $\sqrt{q}$ if we assume A is like a subring of cardinality $\sqrt{q}$.

The question remains, how one can pass from control of $A+A$ and $A \cdot A$ to the more complex object $(A-A)^2 + (A-A)^2$. By expanding out the square, this set is contained inside

$$A \cdot A + A \cdot A + A \cdot A + A \cdot A - A \cdot A - A \cdot A - A \cdot A - A \cdot A,$$

so somehow the idea is to show that $A \cdot A$ is somewhat closed under addition or subtraction.

From additive combinatorics we have the following result: if $A-A$ has about the same cardinality as A, then so does $A+A$, or $A+A+A$, $A+A+A+A-A-A-A-A$, etc. This comes from the theory of commutative graphs; see [**?**]. Thus it would suffice just to make $A \cdot A - A \cdot A$ small.

There is a multiplicative version of the above claim: if $A \cdot A$ (for instance) is roughly the same cardinality as A, then so is A/A, or $A \cdot A/A$, etc. (Let's ignore the problem of division by zero for the sake of discussion).

So the situation is that we can control the cardinality of expressions which are purely additive (e.g. $A + A - A$) or purely multiplicative (e.g. $A \cdot A/A$), but we need to control a hybrid expression $A \cdot A - A \cdot A$.

The idea is to pass from A to a refinement A' (i.e. A' is a subset of A with comparable cardinality), for which $A' \cdot A' - A'cdotA'$ is small. This will ultimately give a solution to Problem 1.2 if we set $E := A' \times A'$.

Actually, it will suffice to find two refinements C, D of A for which $CD - CD$ is small; one can then shift D by some factor to have large intersection with C (this is coming from the fact that $A \cdot A$ is small) to obtain the above result.

The idea of passing to refinement is inspired by the Balog-Szemerédi theorem [1], which asserts that if $A + A$ is "mostly small" (i.e. many pairs in $A + A$ sum to live inside a small set), then by passing to a refinement A' of A one can make $A' + A'$ (or equivalently, $A' - A'$) genuinely small.

The Balog-Szemerédi theorem is not directly useful here, however there is a wonderful proof of this theorem by Gowers [6] (later utilized by Bourgain [2]) which is adaptable to our setting. The main tool in Gowers' argument is the following lemma:

Lemma 2.1. [6] *Let A, B have size $\sim \sqrt{q}$. If $A + B$ mostly lives inside a set of size $\sim \sqrt{q}$ (in that a generic pair has probability ~ 1 of summing into this set), then there exist refinements A', B' of A and B respectively such that, for every $a \in A'$, $b \in B'$, the number of solutions to the problem*

$$a - b = (a_1 - b_1) - (a_2 - b_2) + (a_3 - b_3); \quad a_1, a_2, a_3 \in A; b_1, b_2, b_3 \in B \qquad (1)$$

is near-maximal (i.e. the number of solutions is $\sim \sqrt{q}^5$).

Proof [2] If $A + B$ is mostly small, then there are a near-maximal number of solutions to the equation

$$a + b = a' + b'; \quad a, a' \in A; b, b' \in B$$

and hence a near-maximal number of solutions to

$$a - b' = a' - b; \quad a, a' \in A; b, b' \in B.$$

Thus many differences $a - b$ are "popular" in the sense that $a - b = a' - b'$ has a near-maximal number of solutions.

We say that two numbers $a, a' \in A$ "communicate" if there are a near-maximal number of elements $b \in B$ such that $a - b$ and $a' - b$ are both popular. It turns out that we can find a refinement A' of A such that almost all pairs in A' communicate. The idea is to pick a random element b and define A' to be the set of all elements $a \in A$ for which $a - b$ is popular. (Actually one cannot pick a completely random element; one should condition to ensure that A' is large). If there were a large number of pairs in A' which did not communicate, then b would have been very unlikely to select these pairs in the first place.

Having picked A', we refine B to B' so that for every $b \in B'$, there are a near maximal number of elements $a \in A'$ for which $a - b$ is popular.

Now consider a generic pair $a \in A'$, $b \in B'$. We do not know whether $a - b$ is itself popular, but by the previous paragraph we can find a near-maximal number of $a' \in A'$ such that $a' - b$ is popular. Of these, almost all of them will communicate with a, so that we can find a near-maximal number of b' such that $a - b'$ and $a' - b'$ are both popular. Now consider the identity

$$(a - b) = (a - b') - (a' - b') + (a' - b).$$

Every parenthetical expression on the right is popular, so we can generate a near maximal number of solutions to (1) as desired. ∎

From this Lemma the Balog-Szemerédi theorem is quite easy to deduce: the point is that there are only about $\sqrt{q}^6$ sextuples $(a_1, a_2, a_3, b_1, b_2, b_3)$ overall, so the number of possible values $a' - b'$ can take is at most $\sim \sqrt{q}$. Thus $A' - B'$ has cardinality $\sim \sqrt{q}$, and by fooling around with sums and differences one can turn this into the Balog-Szemerédi theorem.

How does this Lemma help us for our specific problem? Well, it gives us refinements C, D of A such that there are a near-maximal number of solutions to

$$c - d = a_1 + a_2 + a_3 - a_4 - a_5 - a_6; \quad a_1, \ldots, a_6 \in A$$

for each $c \in C$, $d \in D$.

Thus $C - D$ is small. However, we can say a lot more, for instance $A \cdot (C - D)$ is small. This is because for any $a \in A$, $c \in C$, $d \in D$ there are a near-maximal number of solutions to

$$a(c - d) = e_1 + e_2 + e_3 - e_4 - e_5 - e_6; \quad e_1, \ldots, e_6 \in A \cdot A$$

(to see this, multiply the previous equation by a and use the fact that $a \cdot A$ is a large subset of $A \cdot A$). Indeed, we can keep going and say for instance that the set

$$X := A \cdot A \cdot A \cdot A \cdot (C - D)/(A \cdot A)$$

is small (comparable in size to A). (Now do you see how powerful the Lemma is?)

Meanwhile, we use the near-multiplicative closure of A (and hence C, D) to refine C, D to C', D' such that for every $c \in C'$, $d \in D'$ there are a near-maximal number of solutions to

$$cd = (c_1 d_1)(c_2 d_2)^{-1}(c_3 d_3); \quad c_1, c_2, c_3 \in C; d_1, d_2, d_3 \in D.$$

Now this implies that

$$C'D' - C'D'$$

is small. This is because for every $c, c' \in C'$ and $d, d' \in D'$, we can write $cd - c'd'$ in a near-maximal number of ways as

$$(c_1 d_1)(c_2 d_2)^{-1}(c_3 d_3) - c'd'$$

which can be re-arranged as

$$x_1 - x_2 + x_3 - x_4$$

where

$$x_1 = \frac{(c_1 - d')d_1 c_3 d_3}{c_2 d_2}$$
$$x_2 = \frac{d'(c' - d_1)c_3 d_3}{c_2 d_2}$$
$$x_3 = \frac{d'c'(c_3 - d_2)d_3}{c_2 d_2}$$
$$x_4 = \frac{d'c'd_2(c_2 - d_3)}{c_2 d_2}.$$

The numbers x_1, x_2, x_3, x_4 lie in X. Thus we have a near-maximal number of solutions to the problem

$$cd - c'd' = x_1 - x_2 + x_3 - x_4; x_1, x_2, x_3, x_4 \in X$$

which, given that X is comparable to A, forces $C'D' - C'D'$ to be comparable to A. And we are done.

One can almost certainly continue these types of arguments and conclude that just about any algebraic expression of A will have size comparable to A after first refining A by a suitable amount[2]. So the statement that A is nearly a ring is not too inaccurate (in fact it is nearly a field). On the other hand, the implicit constants deteriorate rapidly with the complexity of the algebraic expression under consideration (just look at what we had to do just to control $A \cdot A - A \cdot A$!). So in order to disprove Proposition 1.1 for $F = \mathbf{Z}_p$ using these techniques, you need to find an explanation as to why $\mathbf{Z}_p$ does not contain a sub-field of order $\sqrt{p}$, without involving arithmetic expressions of too high a complexity (so e.g. having to iterate some sort of algebraic operation $\sqrt{p}$ times is not going to work. This seems to rule out, e.g. Galois theory, unless one can somehow find a very clever minimalist formulation of this theory that doesn't need to add or multiply too often.).

Another approach (suggested by Gowers) is to develop some analogue of Freiman's theorem for approximate-rings instead of additive approximate-groups. Sounds feasible in principle, but I don't know how to make it rigorous.

3. From Falconer to Furstenburg

Let's move on now, and suppose we have a subset $E \subset F \times F$ of cardinality $\sim q$ but whose distance set is only $\sim \sqrt{q}$.

For two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in $F \times F$, let's use $d(x, y)$ to denote the quantity $(x_1 - y_1)^2 + (x_2 - y_2)^2$. (Yeah, there's a square root missing, but it won't cause any damage).

[2]For instance, one can get the Erdos -¿ Furstenburg implication directly this way; cf. [3]

Given any x in E, the remainder of E is contained in $\sim \sqrt{q}$ circles centered at x. We expect each circle to thus contain about $\sqrt{q}$ elements. Thus the number of pairs y, y' in E which are the same distance from x, i.e. $d(x,y) = d(x,y')$, should be about $\sqrt{q} \times \sqrt{q} \times \sqrt{q} = q^{3/2}$. (To make this rigorous, use Cauchy-Schwarz). Since there are q choices for x, we thus see that the number of *isosceles triangles*

$$(x, y, y') \in E^3 : d(x,y) = d(x,y')$$

is about $q^{5/2}$. We can throw away the degenerate triangles when two or more vertices are equal as these clearly form a minority.

Now consider the set of perpendicular bisectors of edges in E, i.e. lines of the form

$$\{x \in F \times F : d(x,y) = d(x,y')\}$$

for some distinct $y, y' \in E$. At first glance there appear to be $q \times q = q^2$ such bisectors, since this is the number of pairs in E. However, there could be massive repetition of these bisectors. In fact the best lower bound for the number of bisectors one can come up with is q, which comes simply from fixing y and let y' run around E. These bisectors are guaranteed to be distinct (oh, I should have mentioned, let's ignore the case when F has characteristic 2, that case is terminally pathological). In the model case $F = F_{p^2}$, $E = F_p \times F_p$ one can verify that we have captured all the angular bisectors that are out there.

If we fix y then the number of isosceles triangles is down to $q^{3/2}$. Thus the number of incidences between points in E and angular bisectors is at least $q^{3/2}$, which happens to be maximal. Thus we have found our solution to the Furstenburg problem.

We remark that in Euclidean space, this argument, combined with the Szemerédi-Trotter theorem (which is a quantitative way of measuring the failure of Problem 1.3) gives a decent bound ($q^{4/5}$, in our notation) for the distance set problem; see the work of Chung, Szemerédi, and Trotter.

4. From Furstenburg to Erdos

Now suppose that we have about q lines L and about q points P in $F \times F$ in which the number of incidences is the near-maximal $\sim q^{3/2}$. Thus every point expects to be contained in about $q^{1/2}$ lines, and vice versa.

Take a point x in P, and consider the fan of x, by which we mean the union of the lines in L which go through x. This fan consists of about $q^{1/2}$ lines, each of which contain $q^{1/2}$ points. Thus the fan contains a very large fraction of the q points in P. (Of course, the distinct lines of the fan contain distinct points, except at x).

So we can find two distinct points x_0, x_1 in P such that a large portion of P is contained inside the intersection of the two fans from x_0 and x_1.

The intersection of two distinct fans in Euclidean space looks like a distorted Cartesian grid. To remove this distortion one could apply a projective linear transformation to send x_0, x_1 to the horizontal and vertical points at infinity respectively. Note that projective transformations barely affect the incidence problem (OK, so there is some issue with the line at infinity, but this is such a small fraction of the whole (projective) plane that its influence is negligible).

If we apply the same type of reasoning to the finite field case, we can projectively transform our situation so that a large fraction of P is contained inside the

intersection of a "horizontal fan" of $\sim \sqrt{q}$ horizontal lines, and a "vertical fan" of $\sim \sqrt{q}$ vertical lines.

In other words, we can find sets $C, D \subset F$ of cardinality $\sim \sqrt{q}$ such that a large subset of P is contained inside $C \times D$. Does this begin to look familiar?

Thus $C \times D$ is filled quite densely with elements of P. Pick two rows $c \times D$ and $c' \times D$ each of which contain a near maximal (i.e. $\sim \sqrt{q}$) number of elements of P. There are at most $\sim q$ lines connecting these points, but this is the same number as the number of lines in L. Thus we can assume that a near-maximal number of lines in L connect pairs of points in these two rows. (To make this precise one can first ensure that the rows communicate heavily, in the sense of previous sections). We may normalize so that $c = 0$ and $c' = 1$. This implies that for a near-maximal family of points $d, d' \in D$, the line

$$\{(t, (1-t)d + td') : t \in F\}$$

lies in L. Applying this with t a generic element of C, and using the fact that the lines in L must have large intersection with $C \times D$ (as this set contains a large share of P, and every point in P is incident with many lines in L) we thus see that for a near maximal family of triples $(t, d, d') \in C \times D \times D$, the element $(1-t)d + td'$ is also in D.

Applying this for a fixed value of t we see from Balog-Szemeredi that $D + D$ is small (after refining D somewhat). Applying this instead for a fixed value of d we and using multiplicative Balog-Szemeredi we see that $C \times (D - d)$ is small (after refining C and D some more). Putting this together by the same games as previously we can eventually show that $D + D$ and $D \cdot D$ is small (after some more refinement), and we are done.

References

[1] A. Balog, E. Szemerédi, *A statistical theorem of set addition*, Combinatorica, **14** (1994), 263–268.

[2] J. Bourgain, *On the dimension of Kakeya sets and related maximal inequalities*, Geom. Funct. Anal. **9** (1999), no. 2, 256–282.

[3] G. Elekes, *On the number of sums and products*, Acta Arith. **81** (1997), 365–367.

[4] P. Erdös, B. Volkmann, *Additive Gruppen mit vorgegebener Hausdorffscher Dimension*, J. Reine Angew. Math. **221** (1966), 203-208.

[5] K. J. Falconer, *The Hausdorff dimension of distance sets*, Mathematika, **32** (1985), 206–212.

[6] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, to appear in GAFA.

[7] N. Katz, T. Tao, *Some connections between the Falconer and Furstenberg conjectures*, submitted, New York J. Math.

[8] I. Ruzsa, *Sums of finite sets*, Number Theory: New York Seminar; Springer-Verlag (1996), D.V. Chudnovsky, G.V. Chudnovsky and M.B. Nathanson editors.

[9] E. Szemerédi, W. T. Trotter Jr., *Extremal problems in discrete geometry*, Combinatorica **3** (1983), 381–392.

[10] T. Wolff, *Recent work connected with the Kakeya problem*, Prospects in mathematics (Princeton, NJ, 1996), 129–162, Amer. Math. Soc., Providence, RI, 1999.

Department of Mathematics, UCLA, Los Angeles, CA 90024

E-mail address: tao@math.ucla.edu