An Extension of Behrend's Theorem

Paul H. Koester[1]
Department of Mathematics
Indiana University
Bloomington, IN 47405
U. S. A.
phkoeste@indiana.edu

Submitted: March 28, 2007; Revised: November 18, 2007; Accepted: December 5, 2007;
Published: Janaury 29, 2008

Abstract

We extend an argument of Felix Behrend to show that fairly dense subsets of the integers exist which contain no solution to certain systems of linear equations.

1 Introduction

A classical argument of Behrend [Beh46] establishes

$$r_3(N) \gtrsim Ne^{-C\sqrt{\log N}} \tag{1}$$

for some $C > 0$ where $r_3(N)$ is the maximum cardinality of any subset $A \subset \{0, 1, \ldots, N-1\}$ which contains no proper 3-term arithmetic progression. We study the problem of finding lower bounds on $r_P(N)$, a generalization of $r_3(N)$.

If $P \subset \mathbb{Z}^r$ is a finite set and $\psi : \mathbb{Q}^r \to \mathbb{Q}$ is an affine map so that $\psi|_P$ is injective, we say $\psi(P)$ is an injective affine image of P, and we let $r_P(N)$ denote the maximum cardinality of any subset $A \subset \{0, 1, \ldots, N-1\}$ which contains no injective affine image of P. In this new notation, $r_{\{0,1,2\}}(N) = r_3(N)$

Ruzsa initiated a systematic study of $r_P(N)$ in [Ruz93]. In particular, he observed that Behrend's argument easily extends to show

$$r_P(N) \gtrsim Ne^{-C\sqrt{\log N}} \tag{2}$$

for some $C = C(P) > 0$ provided P contains a nontrivial convex combination. More recently, Shapira [Sha06] showed (2) holds, unless P which is contained in the zero set of a nonzero quadratic polynomial.

[1]The author was supported in part by the National Science Foundation, Grant No. 0514370, Richard Rochberg PI

We provide a further extension along these lines. In particular, we show (2) holds for the set

$$P_p = \{(0,0),(1,0),(0,1),(3,1),(1,3)\}, \tag{3}$$

among others. This appears to be new, as P_p lies on the parabola defined by $(x-y)^2-(x+y) = 0$, and hence Shapira's Theorem does not apply. Our main theorem shows that (2) holds unless P lies on the zero set of a special type of quadratic polynomial.

The problem of finding affine images of P can be rephrased in terms of solving systems of linear equations. For instance, affine images of $\{0,1,2\}$ correspond to solutions of $x+z = 2y$. Affine images of P_p correspond to solutions of

$$\begin{cases} 3x + w = 3y + z \\ 3x + u = 3z + y \end{cases} \tag{4}$$

Furthermore, an injective affine image of P_p corresponds to a solution of (4) with x, y, z, w, u all distinct. It is perhaps more natural to use the linear equation point of view to define $r_P(N)$, but our argument, and the conditions we find on P, are more natural from the affine image point of view.

In section 3 we provide an exposition of Behrend's argument, with a slightly different interpretation than usual. This new interpretation helps make our later argument more transparent. We then prove a special case of our theorem in section 4, and the full theorem in section 5.

The author thanks the anonymous referee for some helpful points on improving this article.

2 Definitions and Notation

We write $X \lesssim Y$ and $Y \gtrsim X$ if there exists a constant $C > 0$ so that $X \leq CY$, and we write $X \sim Y$ if $X \lesssim Y$ and $Y \lesssim X$.

Let $A \subset \mathcal{Z}$ and $B \subset \mathcal{Z}'$ be finite subsets of abelian groups $\mathcal{Z}$ and $\mathcal{Z}'$ and let $k \in \mathbb{N}$. A *Freiman homomorphism of order k from A to B* is a map $\varphi : A \to B$ so that

$$\sum_{j=1}^{k} \varphi(x_j) = \sum_{j=1}^{k} \varphi(y_j)$$

whenever

$$\sum_{j=1}^{k} x_j = \sum_{j=1}^{k} y_j,$$

provided $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k \in A$. It follows that if $a_1, a_2, \ldots, a_s, b_1, b_2, \ldots, b_t$ are positive integers with $\sum_{j=1}^{s} a_j = \sum_{j=1}^{t} b_t = k$, then

$$\sum_{j=1}^{s} a_j\varphi(x_j) = \sum_{j=1}^{t} b_j\varphi(y_j)$$

whenever

$$\sum_{j=1}^{s} a_j x_j = \sum_{j=1}^{t} b_j y_j,$$

provided $x_1, x_2, \ldots, x_s, y_1, y_2, \ldots, y_t \in A$.

A *Freiman isomorphism of order k between A and B* is a bijective Freiman homomorphism whose inverse is also a Freiman homomorphism[1].

If $P \subset \mathbb{Q}^r$, $\operatorname{AffRank} P$ denotes the dimension of the affine span of P. We usually assume $\operatorname{AffRank} P = r$ since P is affinely isomorphic to some $Q \subset \mathbb{Q}^{\operatorname{AffRank} Q}$.

3 Review of Earlier Arguments

We review the arguments from [SS42] and [Beh46].

We begin by extending the domain of the function $r_3(N)$. Suppose $S \subset \mathcal{Z}$ where $\mathcal{Z}$ is an abelian group and S is a finite subset thereof. A nontrivial three term arithmetic progression in $\mathcal{Z}$ is a triple of distinct elements $(a, b, c) \in \mathcal{Z}^3$ so that $a + c = 2b$. $r_3(S)$ denotes the maximum cardinality of any subset of S which contains no nontrivial arithmetic progressions. When $\mathcal{Z} = \mathbb{Z}$ and $S = \{0, 1, 2, \ldots, N-1\}$ we write $r_3(N)$ rather than $r_3(\{0, 1, 2, \ldots, N-1\})$.

It is useful to study this generalization even if one is only interested in $r_3(N)$. For instance, lower bounds for $r_3(N)$ can be established using lower bounds on $r_3(\{0, 1, \ldots, M-1\}^d)$ for appropriate M and d. The following lemma specifies the required relationship between N, M, and d.

Lemma. *[SS42] Suppose $N \geq \frac{1}{2}[(2M-1)^d - 1]$. Then*

$$r_3(\{0, 1, \ldots, M-1\}^d) \leq r_3(N).$$

Proof. Let $\Phi : \mathbb{Z}^d \to \mathbb{Z}$,

$$\Phi(x_1, x_2, \ldots, x_d) = x_1 + (2M-1)x_2 + \cdots + (2M-1)^{d-1} x_d.$$

Φ is a Freiman isomorphism of order 2 when restricted to the domain $\{0, 1, \ldots, M-1\}^d$.

If $B \subset \{0, 1, \ldots, M-1\}^d$ has no nontrivial arithmetic progressions of length three then $A = \Phi(B)$ has no nontrivial arithmetic progressions of length three and $A \subset \{0, 1, \cdots, \frac{1}{2}[(2M-1)^d - 1]\}$. The estimate

$$r_3(\{0, 1, \ldots, M-1\}^d) \leq r_3(\frac{1}{2}[(2M-1)^d - 1])$$

follows since $|B| = |A|$. The lemma then follows from monotonicity of r_3. □

Let $c_d(M)$ denote the maximum size of any subset $B \subset \{0, 1, \ldots, M-1\}^d$ which contains no nontrivial convex combination. That is, if $p, p_1, p_2, \ldots, p_n$ are in B and

$$p = \sum_{j=1}^{n} \lambda_j p_j$$

[1] We give the standard warning that a bijective Freiman homomorphism need not be an isomorphism.

for some choice of $0 < \lambda_1, \lambda_2, \ldots, \lambda_n \leq 1$ with $\sum_{j=1}^n \lambda_j = 1$ then $p = p_1 = \cdots = p_n$.

Since a nontrivial arithmetic progression of length 3 is a nontrivial convex combination, we obtain

$$c_d(M) \leq r_3(\{0, 1, \ldots, M-1\}^d).$$

Salem and Spencer obtained the lower bound

$$c_d(M) \geq \frac{d!}{\left(\frac{d}{M}\right)!^M}$$

for any pair d, M with $M|d$, using the sets

$$\mathcal{S}_d(M) \subset \{0, 1, \ldots, M-1\}^d$$

consisting of points $(x_1, x_2, \ldots, x_d) \in \mathbb{N}^d$ so that $x_j = a$ for exactly $\frac{d}{M}$ choices of j for each $a \in \{0, 1, \ldots, M-1\}$. Given $N \in \mathbb{N}$ and choosing d and M appropriately (see [SS42] for what appropriate means), Salem and Spencer obtain

$$r_3(N) \gtrsim N e^{-C\frac{\log N}{\log\log N}}$$

for some positive constant C.

Behrend's argument is similar, but uses a better bound on $c_d(M)$. Behrend starts with d considerably smaller than M and uses the pigeonhole principle to find an $R \in (0, (M-1)^2] \cap \mathbb{Z}$ so that

$$|\Sigma(M, d, R)| \geq \frac{M^{d-2}}{d},$$

where

$$\Sigma(M, d, R) = \{(x_1, x_2, \ldots, x_d) \in \{0, 1, \ldots, M-1\}^d : x_1^2 + x_2^2 + \cdots + x_d^2 = R\}.$$

Choosing d and M integers satisfying $d \sim \sqrt{\log N}$ and $M \sim e^{\sqrt{\log N}}$, Behrend concludes

$$r_3(N) \gtrsim N e^{-C\sqrt{\log N}}.$$

Trying to improve Behrend's bound by replacing the spheres with some other collection of strictly convex hypersurfaces is a natural temptation. However, the elementary bound

$$c_d(M) \leq 2M^{d-1}$$

shows that such argument would, at best, improve the value of C in (1).

It is interesting to point out that the very first lower bound, $r_3(N) \gtrsim N^{\log 2/\log 3}$, also fits into the above paradigm, following from $c_d(2) = 2^d$.

The above method easily extends to establish the bound (2) for any $P \subset \mathbb{Z}^r$ which contains a nontrivial convex combination. Let P be any subset of $\mathbb{Z}^r$ with $\operatorname{AffRank} P = r$. Any affine map $\mathbb{Q}^r \to \mathbb{Q}$ defines a Freiman homomorphism of any order of P into $\mathbb{Q}$. Conversely, there exists a positive integer o_P so that any Freiman homomorphism of P into $\mathbb{Q}$ of order at least o_P is the restriction of an affine map. This follows from Konyagin and Lev's linear algebraic method, [KL00]. We call the smallest such o_P the *Freiman order of* P.

Lemma. *Suppose $P \subset \mathbb{Z}^r$ has Freiman order o_P and $N \geq \frac{1}{o_P}[(o_P(M-1)+1)^d - 1]$. Then*

$$r_P(\{0,1,\ldots,M-1\}^d) \leq r_P(N).$$

Proof. This follows as above, except we use

$$\Phi(x_1, x_2, \ldots, x_d) = x_1 + (o_P(M-1)+1)x_2 + \cdots + (o_P(M-1)+1)^{d-1}x_d,$$

which is a Freiman isomorphism of order o_P between $\{0,1,\ldots,M-1\}^d$ and its image. Suppose $\varphi : \mathbb{Q}^r \to \mathbb{Q}$ is an affine map giving an injective affine image of P into $\Phi(\{0,1,\ldots,M-1\}^d)$. Then $\psi = \Phi^{-1} \circ \varphi$ is a Freiman homomorphism of order o_P from P into $\mathbb{Z}^d$. By the choice of o_P, ψ therefore has an extension to an affine map $\mathbb{Q}^r \to \mathbb{Q}^d$. In the contrapositive, if $B \subset \{0,1,\ldots,M-1\}^d$ has no injective affine image of P then $\Phi(B)$ has no injective affine image of P. □

As with r_3, to establish lower bounds on $r_P(N)$ it suffices to do so for $r_P(\{0,1,\ldots,M-1\}^d)$. If P contains a nontrivial convex combination then

$$r_P(\{0,1,\ldots,M-1\}^d) \geq c_d(M) \tag{5}$$

since convex combinations are preserved by affine maps. Taking $d \sim \sqrt{\log N}$ and $M \sim e^{\sqrt{\log N}}$ gives (2) for some $C > 0$. Furthermore, $C = 2 + \log o_P + \epsilon$ is admissible for any positive ϵ, provided N is sufficiently large.

The following proposition is the culmination of the above discussion. The novelty of this proposition is its emphasis on affine structures of P, letting convexity play a subsidiary role. Our later argument exploits this extra flexibility.

Proposition. *Let $P \subset \mathbb{Z}^r$ be a finite set with* $\operatorname{AffRank} P = r$ *and Freiman order o_P.*

- *If P has no injective affine image on any sphere $\Sigma(M,d,R)$ with $0 < R \leq (M-1)^2$, and $d \sim \log M$ then $r_P(N) \gtrsim Ne^{-C\sqrt{\log N}}$.*
- *If P has no injective affine image into any Salem-Spencer set $S(M,d)$, with $M|d$, then $r_P(N) \gtrsim Ne^{-C\frac{\log N}{\log\log N}}$.*

In the next two sections we find obstructions for P to have an injective affine image on a sphere. Although the Salem-Spencer sets provide a worse bound, they may be able to provide additional obstructions. We have not been able to find any additional obstructions yet.

4 Planar Sets

Our method is best illustrated for $P \subset \mathbb{Z}^2$ since this case avoids some annoying degenerate cases that appear in higher dimensions.

Let $P \subset \mathbb{Z}^2$ be a finite set with $|P| \geq 3$ and suppose P is not collinear. Suppose there is an affine map $\psi : \mathbb{Q}^2 \to \mathbb{Q}^d$ for some $d \in \mathbb{N}$ so that $\psi|_P$ is injective and $\psi(P) \subset S \cap \mathbb{Z}^d$ for some $d-1$-dimensional sphere $S \subset \mathbb{R}^d$. By continuity there is a unique continuous affine extension $\psi : \mathbb{R}^2 \to \mathbb{R}^d$. $H = \psi(\mathbb{R}^2)$ is a translate of a subspace of $\mathbb{R}^d$ and $\dim H \in \{0,1,2\}$, so $H \cap S$ is either empty, one point, two points, or a circle. However $\psi(P) \subset H \cap S$ and $|\psi(P)| \geq 3$; therefore $H \cap S$ is a circle C. In particular, $\dim H = 2$ so ψ is an affine isomorphism onto H. It follows that $P \subset \psi^{-1}(C) = \mathcal{E}$, an ellipse, establishing the planar case of our theorem.

Theorem. *Let $P \subset \mathbb{Z}^2$ be a finite set. If P is not contained on any ellipse in $\mathbb{R}^2$ then there exists a constant $C > 0$ so that* (2) *holds.*

It is instructive to investigate $r_P(N)$ for $P \subset \mathbb{Z}^2$ as $|P|$ increases. Throughout this discussion we assume P is not collinear. First, $r_P(N) = 3$ for any 3-element set, since any such set is affinely independent. Next, there is a one parameter family of quadratic curves passing through any four element set. A tedious, but elementary, calculation with the quadratic formula shows this family is devoid of ellipses if and only if P contains a nontrivial convex combination. Thus, for $|P| = 4$ the only case where we know (2) is if P has a nontrivial convex combination.

Now suppose $|P| = 5$. First, if P has 4 points on a single line ℓ then there is a one parameter family of quadratic curves passing though P, and each such curve is the union of ℓ with one other line. Such a set obviously does not lie on an ellipse. If P has 3, but not more, points on a single line ℓ then there is a unique quadratic curve passing through P, and it is the union of ℓ and the unique line through the other 2 points of P. Again, such a set does not lie on an ellipse.

If P contains no collinear triple then there is a unique irreducible quadratic curve passing through P. In particular, P cannot lie on an ellipse if it lies on either a parabola or a hyperbola.

Corollary. *Let $P \subset \mathbb{Z}^2$ be a 5 element set. There exists a constant $C > 0$ so that* (2) *holds if any of the following is satisfied:*

1. *P contains at least 3 points on line;*
2. *P lies on a hyperbola, and intersects both components;*
3. *P lies on a single component of a hyperbola;*
4. *P lies on a parabola.*

Conditions 1 and 2 are well known, as they follow from the existence of nontrivial convex combinations in P. Conditions 3 and 4 appear to be new. The result in the introduction concerning P_p follows since this set lies on a parabola.

If $|P| \geq 6$ then (2) holds for "almost all" P, in the sense that the typical 6-element set does not lie on any ellipse. However, there exist arbitrarily large R so that $|\Sigma(M, 2, R)| \gtrsim \log R$, provided $M \gtrsim \sqrt{R}$, (See Proposition 17.6.1 of [IR90].). Taking $P = \Sigma(M, 2, R)$ for such M and R provide arbitrarily large P to which our theorem does not apply. On the other hand, the only planar P which are known to satisfy a power-type upper bound, $r_P(N) \lesssim N^{1-\alpha}$ for some positive α, have $|P| = 4$ and are related to symmetric equations. This is discussed in more detail in the final section of this work.

5 General Case

Our main theorem from the last section generalizes to $P \subset \mathbb{Z}^r$ but there are extra degeneracies. We write $\mathcal{E}^{s-1}$ for an ellipsoid in $\mathbb{R}^s$. An ellipsoidal cylinder is the product of an ellipsoid with a Euclidean space.

Theorem. *Let $P \subset \mathbb{Z}^r$ be a finite set. If P is not contained on any ellipsoidal cylinder $\mathcal{E}^{s-1} \times \mathbb{R}^{r-s} \subset \mathbb{R}^r$ with $2 \leq s \leq r$ then there exists a constant $C > 0$ so that (2) holds.*

Proof. Assuming there is an affine map $\psi : \mathbb{Q}^r \to \mathbb{Q}^d$ for some $d \in \mathbb{N}$ so that $\psi|_P$ is injective and $\psi(P) \subset S \cap \mathbb{Z}^d$ for some $d-1$-dimensional sphere $S \subset \mathbb{R}^d$, we extend $\psi : \mathbb{R}^r \to \mathbb{R}^d$ as before and $H = \psi(\mathbb{R}^d)$ is a translate of a subspace of $\mathbb{R}^d$ with $\dim H \in \{0, 1, 2, \ldots, r\}$. The cardinality considerations we relied on before can only rule out $\dim H = 0$ and 1. If $s = \dim H$ then $s \in \{2, 3 \ldots, r\}$, and we find $H \cap S$ is an $s-1$-dimensional sphere. Then

$$\psi^{-1}(H \cap S) \cong \mathcal{E}^{s-1} \times \mathbb{R}^{r-s}.$$

□

Corollary. *Let $P \subset \mathbb{Z}^r$ be a finite set and let $P' \subset P$ with $t = \operatorname{AffRank} P' \geq 2$. If P' is not contained on any ellipsoidal cylinder $\mathcal{E}^{s-1} \times \mathbb{R}^{t-s} \subset \mathbb{R}^t$ with $2 \leq s \leq t$ then there exists a constant $C > 0$ so that (2) holds.*

This follows from the monotonicity $r_P(N) \geq r_{P'}(N)$ whenever $P' \subset P$.

6 Conclusion

It may be possible that (2) holds for all $P \subset \mathbb{Z}^2$ with at least 5 elements, but our method is not able to establish this when P lies on an ellipse. Perhaps the simplest elliptic set is

$$P_e = \{(0,0), (1,0), (0,1), (2,1), (1,2)\}, \tag{6}$$

which superficially looks like P_p, but the latter satisfies (2) whereas only trivial lower bounds are currently known for $r_{P_e}(N)$.

A stronger result may in fact hold. Affine images of a four element $P \subset \mathbb{Z}^2$ correspond to solutions to a single linear equation in four variables. It is an open question whether

$$r_P(N) \lesssim N^{1-\alpha}$$

for some $\alpha = \alpha(P)$ if and only if the linear equation corresponding to P is of the form

$$ax + by = az + bw \tag{7}$$

for $a, b \in \mathbb{N}$. We call such equations symmetric equations. In [Ruz93] it is shown that all equations of the form (7) satisfy

$$\sqrt{N} e^{-\beta\sqrt{\log N}} \lesssim r(N) \lesssim \sqrt{N}.$$

The equations

$$3x + y = 2z + 2w \tag{8}$$

$$3x + 6y = z + 8w \tag{9}$$

are equations for which $\alpha = 0$ is unknown. (8) is perhaps the simplest equation in four variables for which $\alpha = 0$ is unknown. The result $\alpha = 0$ for (9) would have ergodic theoretic consequences, [Fra06]. In fact, the weaker $\alpha < \frac{1}{3}$ suffices for ergodic theoretic consequences.

The essential property of spheres we used in our argument is what we call the "unique cross section" property. As an example, let

$$H(d,R) = \{(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d : x_1^2 + x_2^2 + \ldots + x_{d-1}^2 - x_d^2 = R\}$$

and let L be a rank 2 affine subspace. If $L \perp \{x_n = 0\}$ then $H(d,R) \cap L$ is empty, a single point, or a circle, as before. On the other hand, if $\{x_n = 0\} \subset L$ then $H(d,R) \cap H$ is a hyperbola. More generally, if $\mathcal{Q}$ is any quadratic hypersurface in $\mathbb{R}^d$ and r is much smaller than d, then the set of quadratic varieties obtained by slicing $\mathcal{Q}$ with affine subspaces of dimension not exceeding r contain the set of such cross sections for the sphere. Thus, at least among quadratic hypersurfaces, the sphere has the minimal collection of cross sections, and therefore presents the strictest collection of obstructions to injective affine mappings.

References

[Beh46] F. A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. U. S. A.*, 32:331–332, 1946.

[Fra06] Nikos Frantzikinakis. Multiple ergodic averages for three polynomials and applications. preprint, 2006.

[IR90] Kenneth Ireland and Michael Rosen. *A classical introduction to modern number theory*, volume 84 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[KL00] Sergei V. Konyagin and Vsevolod F. Lev. Combinatorics and linear algebra of Freiman's isomorphism. *Mathematika*, 47(1-2):39–51 (2002), 2000.

[Ruz93] Imre Z. Ruzsa. Solving a linear equation in a set of integers. I. *Acta Arith.*, 65(3):259–282, 1993.

[Sha06] Asaf Shapira. Behrend-type constructions for sets of linear equations. *Acta Arith.*, 122(1):17–33, 2006.

[SS42] R. Salem and D. C. Spencer. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. U. S. A.*, 28:561–563, 1942.