A NON-LINEAR ROTH THEOREM FOR SETS OF POSITIVE DENSITY

BEN KRAUSE

Abstract. Suppose that $A \subset \mathbb{R}$ has positive upper density,

$$\limsup_{|I| \to \infty} \frac{|A \cap I|}{|I|} = \delta > 0,$$

and $P(t) \in \mathbb{R}[t]$ is a polynomial with no constant or linear term, or more generally a non-flat curve. Then for any $R_0 \leq R$ sufficiently large, there exists some $x_R \in A$ so that

$$\inf_{R_0 \leq T \leq R} \frac{|\{0 \leq t < T : x_R - t \in A, x_R - P(t) \in A\}|}{T} \geq c_P \cdot \delta^2$$

for some absolute constant $c_P > 0$, that depends only on $P$.

1. INTRODUCTION

A beautiful result in Euclidean Ramsey theory, due to Furstenberg, Katznelson, and Weiss, [7], concerns the presence of additive structure in sets of positive density inside of the plane, i.e. those (measurable) sets for which

$$d^*(A) := \limsup_{|Q| \to \infty} \frac{|A \cap Q|}{|Q|} > 0;$$

here $\{Q\}$ are axis-parallel cubes.

**Theorem 1.1.** Suppose that $A \subset \mathbb{R}^2$ has $d^*(A) > 0$. Then there exists a threshold $r_0(A)$ so that for every $r \geq r_0$ there exists $x_r, y_r \in A$ so that

$$|x_r - y_r| = r.$$

This result was first proven by ergodic theoretic techniques, but it has since been recovered by different methods: geometric [6] and probabilistic [19], and – most significant for this paper – Fourier analytic [1].

Indeed, in his paper, *A Szemeredi type theorem for sets of positive density in $\mathbb{R}^k*** [1], Bourgain developed a powerful but elementary Fourier analytic approach through which he not only recovered Theorem 1.1, but also was able to reveal the presence of additional additive structure inside of dense subsets of the plane, via the following pinned variant of Theorem 1.1.

---

*Date: January 8, 2019.*
Theorem 1.2. Suppose that $A \subset \mathbb{R}^2$ has $d^*(A) > 0$. Then there exists a threshold $r_0$ so that for every $r \geq r_0$ there exists $x_r \in A$ so that for all $r_0 \leq s \leq r$, there exists some $y_s \in A$ so that
\[ |x_r - y_s| = s. \]

To the extent that the core of Bourgain’s argument amounted – essentially – to appropriate applications of the uncertainty principle, his methods have proven quite robust in addressing other problems in Euclidean Ramsey theory, see for instance [3], [5], [9], [17], as well as in the discrete context, [14], [15], [16], [18].

In this note, we apply his method in the bi-linear setting, as we consider the issue of Non-Linear Roth’s Theorem (in the Euclidean context).

The first progress towards understanding non-linear Roth-type patterns
\[ \{x, x + t, x + P(t)\}, \ P \in \mathbb{R}[-] \]
inside non-trivial subsets of the real line was made by Bourgain, [2]. In particular, he proved the following theorem.

Theorem 1.3. Let $N \geq 1$, $\delta > 0$ be small, and let $A \subset [0, N]$ have $|A| \geq \delta N$, and $d \geq 2$ be arbitrary. Then there exists $x, x - t, x - t^d \in A$ for some $t \geq c_\delta \cdot N^{1/d}$, for some absolute constant $c_\delta$.

In [4], this result was extended to handle polynomial curves $P(t) \in \mathbb{R}[t]$ without constant or linear terms.

Both results are a consequence of the following proposition. With
\[ B_r(f, g)(x) := B^P_r(f, g)(x) := \frac{1}{r} \int_0^r f(x - t)g(x - P(t)) \ dt, \]

Proposition 1.4. Let $\delta > 0$ be small, and let $A \subset [0, 1]$ have $|A| \geq \delta$. If $P(t)$ is a polynomial as above, then there exists a lower bound
\[ \langle 1_A, B_1(1_A, 1_A) \rangle \geq c_\delta \|P\|, \]
for some $c_\delta, \|P\| > 0$ which depends on $\delta$ and $\|P\|$, the $\ell^1$ sum of the coefficients of $P$.

In this paper, we amplify this proposition as follows. First, we allow for the presence of non-flat $P$, functions introduced and studied by Lie in his treatment of bi-linear Hilbert transforms with curvature, [12] and [13], in the definition of $B_r = B^P_r$. Roughly speaking, non-flat curves are locally differentiable curves which do not “resemble a line” near the origin or $\infty$. For instance, in addition to the types of polynomials discussed above, real analytic functions which vanish to degree two at the origin are non-flat, as are real laurent polynomials of the form
\[ \sum_{-n}^m a_j t^j, \ a_{-m}, a_n \neq 0, \ n, m \geq 2, \]
or even linear combinations of functions of the form

\[ |t|^\alpha \log |t|^\beta, \quad \alpha, \beta \in \mathbb{R}, \quad \alpha \neq 0, 1. \]

For a more precise definition, see \cite[§2]{12}.

Our result is then the following.

**Proposition 1.5.** Suppose that \((0, 1] = \bigcup_{J \in \mathcal{P}} J\) is an admissible partition of intervals, in that each \(J\) contains at least one dyadic rational of the form \(2^{-k}\). If \(A \subset [0, 1]\) has \(|A| \geq \delta\), then there exists a subset \(Q \subset \mathcal{P}\) so that

\[
\inf_{J \in Q} \inf_{r \in J} |\langle 1_A, B_r(1_A, 1_A) \rangle| \geq c_P \cdot \delta^3
\]

for some absolute \(c_P > 0\). Moreover, there exists some absolute constant \(C_P\) so that

\[
|\{J : J \in \mathcal{P} \setminus Q\}| \leq C_P \cdot \delta^{-5} \log(\delta^{-1}).
\]

**Remark 1.7.** For polynomial \(P\), both \(c_P, C_P\) depend only on \(\|P\|\).

The following corollary, a quantitative improvement over \cite{4}, immediately presents.

**Corollary 1.8.** In the setting of Proposition 1.5, for any \(\epsilon > 0\), there exists an absolute constant \(c_{\epsilon, P}\) so that

\[
\inf_{1 \geq r > 0} |\langle 1_A, B_r(1_A, 1_A) \rangle| \geq \frac{c_{\epsilon, P}}{2^{\delta^{-5-\epsilon}}} \cdot c_P \cdot \delta^3,
\]

where \(c_P\) is as in (1.6), and similarly \(c_{\epsilon, P}\) is determined by \(\|P\|\) for polynomial \(P\).

Our main result follows from Proposition 1.5 by arguing by contradiction and rescaling as in \cite{1}; this argument is by now standard, and has appeared in the above-discussed works, and so we omit it.

**Theorem 1.9.** Suppose \(A \subset \mathbb{R}\) has positive upper density \(d^*(A) = \delta\). Then for every \(R \geq R_0\) sufficiently large there exists an \(x_R \in A\) so that

\[
\inf_{R_0 \leq t \leq R} \frac{|\{0 \leq t \leq T : x_R - t \in A, \ x_R - P(t) \in A\}|}{T} \geq c_P \cdot \delta^2.
\]

For ease of presentation, we will establish our main results only in the case of \(P(t) = t^2\), as the complications that arise in increasing the generality are essentially notational.

1.1. **Acknowledgement.** This paper, like many of my papers, was inspired by the work of Jean Bourgain; his impact on my mathematics has been profound.
1.2. **Notation.** Here and throughout, \( e(t) := e^{2\pi i t} \). Throughout, \( C \) will be a large number which may change from line to line.

We will henceforth re-define

\[
B_k(f, g)(x) := \int f(x - t)g(x - t^2)\rho_k(t) \, dt
\]

where \( \rho_k(t) := 2^k \rho(2^k t) \), and \( \rho \) is an appropriate bump function.

We will also make use of certain Fourier projection operators: we let \( \phi \) denote a Schwartz function which satisfies

\[
1_{|\xi| \leq 1/8} \leq \hat{\phi} \leq 1_{|\xi| \leq 1/2},
\]

and set \( \phi_k(t) := 2^k \phi(2^k t) \).

We will make use of the modified Vinogradov notation. We use \( X \lesssim Y \), or \( Y \gtrsim X \), to denote the estimate \( X \leq CY \) for an absolute constant \( C \). We use \( X \approx Y \) as shorthand for \( Y \lesssim X \lesssim Y \). We also make use of big-O notation: we let \( O(Y) \) denote a quantity that is \( \lesssim Y \). If we need \( C \) to depend on a parameter, we shall indicate this by subscripts, thus for instance \( X \lesssim_{\delta} Y \) denotes the estimate \( X \leq C_\delta Y \) for some \( C_\delta \) depending on \( \delta \). We analogously define \( O_\delta(Y) \).

## 2. Preliminaries

Before turning to the proof, we need to collect various results from Euclidean harmonic analysis. The first is essentially a martingale inequality, and was observed by Bourgain in \([2]\); see \([4]\) for a proof.

**Lemma 2.1.** Suppose that \( 1 \geq f \geq 0 \) is supported in \([0, 1]\). Then for any \( r, s > 0 \),

\[
\int f \cdot \rho_r \ast f \cdot \rho_s \ast f \gtrsim \left( \int f \right)^3.
\]

The second is a special case of a result of Li and Xiao, \([11]\); it’s extension to non-flat curves is announced there, but a full proof can be found in forthcoming work \([8]\).

**Theorem 2.2.** Suppose that \( P \) is as above. Then

\[
\| \sup_k B_k(f, g) \|_1 \lesssim_P \| f \|_2 \| g \|_2.
\]

The key ingredient in proving this theorem is obtaining so-called “scale-type” decay. Some notation: with \( \Psi \) a smooth approximation of the indicator function of an annulus, define via the Fourier transform

\[
\hat{f}_m(\xi) := \hat{f}(\xi) \cdot \Psi(2^{-m} \xi).
\]

The following is the key proposition; it first appeared essentially as in \([12, \text{Theorem 2}]\), see also \([11, \text{Propositions 3,4}]\).
Proposition 2.3. The following estimate holds for any \(|p| \lesssim 1\)
\[
\|B_k(f_{k+m}, g_{2k+m+p})\|_1 \lesssim 2^{-cm}\|f\|_2\|g\|_2.
\]

With these preliminaries in mind, we turn to the proof.

3. Proof of Proposition 1.5

Proposition 1.5 will follow directly from the following proposition.

Proposition 3.1. Suppose \(A \subset [0, 1]\) has \(|A| = \delta, \) and \(k \geq l \gg \log(\delta^{-1})\) is sufficiently large. Then there exists an absolute constant \(1 \gg c_0 > 0\) so that if the following upper bound holds,
\[
\int_{A} \inf_{l \leq r \leq k} B_r(1_A, 1_A)(x) \ll c_0\delta^3,
\]
then
\[
\|\hat{1}_A(\xi) \cdot 1_{\delta C 2^l \leq |\xi| \leq \delta - 2^k}\|_2 + \|\hat{1}_A(\xi) \cdot 1_{\delta C 2^l \leq |\xi| \leq \delta - 2^k}\|_2 \gtrsim \delta^3.
\]

Before turning to the proof, we will make use of the following splitting. Set \(k_\delta := k + C \log \delta^{-1}\) and \(l_\delta := l - C \log \delta^{-1},\) and assume that we are interested in decomposing \(B_r(f, g)\) for some \(l \leq r \leq k.\) We split
\[
f = f_L + f_M + f_H,
\]
where
\[
\hat{f}_L := \hat{f} \cdot \hat{\phi}_{l_\delta} \text{ and } \hat{f}_H := \hat{f} \cdot (1 - \hat{\phi}_{k_\delta}),
\]
and
\[
g = g_L + g_M + g_H,
\]
where
\[
\hat{g}_L := \hat{g} \cdot \hat{\phi}_{2l_\delta} \text{ and } \hat{g}_H := \hat{g} \cdot (1 - \hat{\phi}_{2k_\delta}).
\]

Furthermore, for ease of presentation, we will drop all terms which contribute \(\lesssim \delta^C\) as negligible.

With this in mind, we turn to the proof.

Proof of Proposition 3.1. Suppose that (3.2) holds. Then, with \(B := [0, 1] \setminus A\) we have the following lower bound, provided that \(l \gg \log(\delta^{-1})\) is sufficiently large:
\[
(3.3) \int_{A} \sup_{l \leq r \leq k} |B_r(1_B, 1_B)| + \sup_{l \leq r \leq k} |B_r(1_A, 1_B)| + \sup_{l \leq r \leq k} |B_r(1_B, 1_A)| \geq (1 - \frac{c_0}{1000} \cdot \delta^2) \cdot \delta.
\]
For each \( f, g \), we decompose
\[
B_r(f, g) = B_r(f_L, g_L) + B_r(f_L, g_M + g_H) + B_r(f_M, g) + B_r(f_H, g_L) + B_r(f_H, g_H).
\]
For \( l \leq r \leq k \), \( B_r(f_L, g_L) = f_L \cdot g_L \) up to pointwise errors of \( \delta^C \); by Lemma 2.1 and trivial geometric considerations, taking into account the large size of \( l \gg \delta \), we deduce that
\[
\sum_{C, C' \in \{A, B\}^2 \setminus \{A, A\}} \int_{A} \sup_{A \leq r \leq k} |B_r((1_C)_L, (1_{C'})_L)| \leq \delta - c\delta^3,
\]
where \( c \) is essentially given by Lemma 2.1. We also observe that
\[
B_r(f_L, g_M + g_H) = f_L \cdot \mathcal{M} g_M,
\]
again up to pointwise errors on the order of \( \delta^C \) on \([0, 1]\); here, \( \mathcal{M} \) denotes an appropriate maximal function, pointwise bounded by the Hardy-Littlewood maximal function. Again using the large size of \( l \), we have that \((1_B)_M = - (1_A)_M\) up to errors which have \( L^1 \) norm \( \lesssim \delta^C \) when restricted to \([0, 1]\), by Cauchy-Schwartz. Next,
\[
B_r(f_H, g_L + g_M) = 2^{-\epsilon(k-r)} \cdot \delta^C \cdot \Pi_{r,k}(f_H, g_L + g_M),
\]
for some paraproduct \( \Pi_{r,k} \), where \( \|\Pi_{r,k}\|_{L^2 \times L^2 \to L^1} \leq C_{k-r} \leq C \) for all \( r, k \) and some absolute \( C \), see [12]. Thus
\[
\int_{A} \sup_{A \leq r \leq k} |B_r(f_H, g_L + g_M)| \ll \delta^C.
\]
It remains to analyze \( B_r(f_H, g_H) \). If we express
\[
B_r(f_H, g_H) = \sum_{m,n \gg \log(\delta^{-1}), |m-n| \leq 1} B_r(f_{k+m}, g_{2k+n}) + \sum_{m,n \gg \log(\delta^{-1}), |m-n| \gg 1} B_r(f_{k+m}, g_{2k+n}),
\]
then the second term is a sum of rapidly decaying paraproducts the sum of whose \( L^1 \) norms are \( O(2^{-\epsilon(k-r)} \cdot \delta^C) \), while the first term has \( L^1 \) norm bounded by \( 2^{-\epsilon(k-r)} \cdot \delta^C \) by Proposition 2.3. In particular, the lower bound (3.3) forces
\[
\sum_{C, C' \in \{A, B\}^2 \setminus \{A, A\}} \int_{A} (1_C)_L \cdot \mathcal{M} ((1_{C'})_M) + \int_{A} \sup_{r} |B_r((1_C)_M, 1_{C'})| \gtrsim \delta^3.
\]
By pointwise considerations, we may refine this to
\[
\int_{A} \mathcal{M} ((1_A)_M) + \int_{A} \sup_{r} |B_r((1_A)_M, 1_A)| + \int_{A} \sup_{r} |B_r((1_A)_M, 1_B)| \gtrsim \delta^3,
\]
which yields the result by Cauchy-Schwartz and Theorem 2.2. \( \square \)
NON-LINEAR ROTH

REFERENCES

[1] J. Bourgain. A Szemerédi type theorem for sets of positive density in Rk. Israel J. Math. 54 (1986), no. 3, 307-316. pages 1, 3
[2] J. Bourgain. A nonlinear version of Roth’s theorem for sets of positive density in the real line. J. Analyse Math. 50 (1988), 169-181. pages 2, 4
[3] B. Cook, A. Magyar, M. Pramanik. A Roth-type theorem for dense subsets of R^d. Bull. Lond. Math. Soc. 49 (2017), no. 4, 676-689. pages 2
[4] P. Durcik, S. Guo, J. Roos. A polynomial Roth theorem on the real line. To appear in Trans. Amer. Math. Soc., arXiv:1704.01546. pages 2, 3, 4
[5] P. Durcik, V. Kovač, L. Rimanić. On side lengths of corners in positive density subsets of the Euclidean space. To appear in Int. Math. Res. Not. pages 2
[6] K. Falconer, J. Marstrand. Plane sets with positive density at infinity contain all large distances. Bull. Lond. Math. Soc. 18 (1986), 471-474. pages 1
[7] H. Furstenberg, Y. Katznelson, B. Weiss. Ergodic theory and configurations in sets of positive density. Mathematics of Ramsey theory, 184-198, Algorithms Combin., 5, Springer, Berlin, 1990. pages 1
[8] A. Gaitan, B. Krause, V. Lie. Pointwise Convergence of Random Bilinear Maximal Functions. Preprint. pages 4
[9] L. Huckaba, N. Lyall, A. Magyar. Simplices and sets of positive upper density in R^d. Proc. Amer. Math. Soc. 145 (2017), no. 6, 2335-2347. pages 2
[10] X. Li. Bilinear Hilbert transforms along curves, I. Analysis & PDE Vol. 6 (2013), No. 1, 197-220. pages
[11] X. Li, L. Xiao. Uniform estimates for bilinear Hilbert transforms and bilinear maximal functions associated to polynomials. Amer. J. Math. 138 (2016), no. 4, 907-962. pages 4
[12] V. Lie. On the boundedness of the bilinear Hilbert transform along ”non-flat” smooth curves. Amer. J. Math. 137 (2015), no. 2, 313-363. pages 2, 3, 4, 6
[13] V. Lie. On the boundedness of the bilinear Hilbert transform along ”non-flat” smooth curves. The Banach triangle case (L^r, 1 ≤ r < ∞. Rev. Mat. Iberoam. 34 (2018), no. 1, 331-353. pages 2
[14] N. Lyall, A. Magyar. An Optimal Version of Sarközy’s Theorem. Preprint, https://arxiv.org/pdf/1010.3451.pdf pages 2
[15] N. Lyall, A. Magyar. Optimal polynomial recurrence. Canad. J. Math. 65 (2013), no. 1, 171-194. pages 2
[16] N. Lyall, A. Magyar. Distances in Dense Subsets of Z^d. https://arxiv.org/pdf/1509.09298.pdf pages 2
[17] N. Lyall, A. Magyar. Product of simplices and sets of positive upper density in R^d. Math. Proc. Cambridge Philos. Soc. 165 (2018), no. 1, 2551. pages 2
[18] A. Magyar. On distance sets of large sets of integer points, Israel J. Math. 164 (2008), 251-263. pages 2
[19] A. Quas. Distances in Positive Density Sets J. Comb. Th. A 116 (2009) 979-987. pages 1

Department of Mathematics, Caltech, Pasadena, CA 91125
E-mail address: benkrause2323@gmail.com