Additive correlation and the inverse problem for the large sieve

BY BRANDON HANSON

Dept. of Mathematics, Boyd Graduate Studies Research Center, University of Georgia, Athens, GA 30602, USA.
e-mail: brandon.w.hanson@gmail.com

(Received 4 June 2018; revised 11 June 2018)

Abstract

Let $A \subseteq [1, N]$ be a set of integers with $|A| \gg \sqrt{N}$. We show that if $A$ avoids about $p/2$ residue classes modulo $p$ for each prime $p$, then $A$ must correlate additively with the squares $S = \{n^2 : 1 \leq n \leq \sqrt{N}\}$, in the sense that we have the additive energy estimate

$$E(A, S) \gg N \log N.$$ 

This is, in a sense, optimal.

1. Introduction

Let $A \subseteq [1, N]$ be a set of integers. For a positive integer $v > 1$, let

$$A_v = \{a \pmod{v} : a \in A\} \subseteq \mathbb{Z}/v\mathbb{Z}$$

denote the set of residue classes covered by $A$ modulo $v$, and let

$$A(v; h) = \{a \in A : a \equiv h \pmod{v}\} \subseteq A$$

denote the set of elements of $A$ congruent to $h$ modulo $v$. The large sieve of Linnik (developed further by others, notably Montgomery [M]) and the larger sieve of Gallagher are useful tools when trying to estimate the cardinality of sets $A$ under the assumption that $|A_p|$ is small when $p$ is a prime. See, for instance, [FI, chapter 9] for more on this. In fact, under the hypothesis that for each prime $p$ we have $|A_p| \leq \alpha p$, the larger sieve tells us the remarkably strong result $|A| \ll N^\alpha$, and in a fairly elementary way. When $\alpha$ is about $1/2$, the estimate $|A| \ll N^{1/2}$ is sharp. This can be seen by taking $A$ to consist of the perfect squares
up to $N$, or any other dense quadratic sequence. The quadratic case of the “inverse conjecture for the large sieve” states that such quadratic sequences are the unique extremisers. Here is a fairly crude form of such a conjecture.

**Conjecture 1.1 (Quadratic inverse large sieve).** 
Let $A \subseteq [1, N]$ be a set of integers such that $|A_p| \leq (p + 1)/2$ for each $p$ and $|A| \geq N^{1/2}$. Then there is a subset $A' \subseteq A$ of size $|A'| \geq 9|A|/10$ and such that $A' \subseteq q(\mathbb{Z})$ for some quadratic polynomial $q(x) \in \mathbb{Q}[x]$.

Further reading about this conjecture can be found in [CL, GH, HV, W1, W2].

To proceed, we need some notation. Let $X$ and $Y$ be two finite sets of integers. The additive energy between $X$ and $Y$ is the quantity

$$E(X, Y) = \left| \{ (x_1, x_2, y_1, y_2) \in X \times X \times Y \times Y : x_1 + y_1 = x_2 + y_2 \} \right|.$$  

This is a quantity intimately related with the sumset $X + Y = \{ x + y : x \in X, y \in Y \}$.

For context, the trivial estimates for additive energy are

$$|X||Y| \leq E(X, Y) \leq |X||Y| \min\{|X|, |Y|\}.$$

We will write $r_{X+Y}(n)$ and $r_{X-Y}(n)$ for the number of solutions to $x + y = n$ and $x - y = n$, respectively, with $x \in X$ and $y \in Y$. A few moments thought reveals the following formulas:

$$E(X, Y) = \sum_n r_{X+Y}(n)^2$$  

and

$$E(X, Y) = \sum_n r_{X-Y}(n)r_{Y-Y}(n).$$  

The main theorem of this paper is the following.

**Theorem 1.2.** Suppose $A$ is a set of positive integers in the interval $[1, N]$ satisfying the condition

$$|A_p| \leq \frac{p}{2} + \varepsilon(p)$$

for all primes $p$, and some uniformly bounded sequence $\varepsilon(p) \geq 0$. Let $S$ denote the set of perfect squares up to $N$. We have

$$E(A, S) \gg |A||S| + |A|^2 \log N.$$  

In particular, if $|A| \gg \sqrt{N}$, then

$$E(A, S) \gg |A||S| \log N.$$

At first glance, a factor of $\log N$ above the trivial bound appears quite weak, so before proceeding, we make the following remarks about this theorem.

Firstly, a logarithmic factor is non-trivial. This can be observed by considering dense Sidon subsets of $[1, N]$. Recall that a set $X$ is called Sidon if all of its pairwise sums are distinct. It is well known that there are Sidon subsets $X$ of $[1, N]$ of cardinality about $\sqrt{N}$,

---

1 In this paper we make frequent use of the asymptotic notation $X \ll Y$ which means that $|X| \ll c|Y|$ for some constant $c$, or equivalently $X = O(Y)$. 

Additive correlation and the inverse problem for the large sieve

the existence of which was proved in [BC]. For such sets \(X, r_{X-X}(n) \leq 1 \) for \(n \neq 0\), so that by (1.2)

\[
E(X, S) = |X||S| + \sum_{n \neq 0} r_{X-X}(n)r_{S-S}(n) \leq |X||Y| + \sum_{n} r_{S-S}(n) = |S|(|X| + |S|) \leq N,
\]

which shows that \(X\) has next to no additive energy with the perfect squares. Now, if we consider arbitrary subsets \(X, Y \subseteq [1, N]\) of integers then \(X + Y \subseteq [2, 2N]\), and we have the obvious estimate \(|X + Y| \leq 2N - 1\). The Cauchy–Schwarz inequality gives

\[
|X||Y| \leq (2N - 1) \sum_{2 \leq n \leq 2N} r_{X+Y}(n)^2 = (2N - 1)E(X, Y),
\]

showing that sets \(X\) and \(Y\) which are a little larger than \(\sqrt{N}\) in size necessarily have some additive energy between them. So, while sets at the \(\sqrt{N}\)-threshold need not have any substantial additive correlation, they only just fail to do so. However, the squares and other sets which avoid residue classes are biased in arithmetic progressions. Since arithmetic progressions are, after all, cosets of subgroups of \(\mathbb{Z}\), this bias hints at a bit of underlying additive structure.

Secondly, this logarithmic factor is interesting in that it is what should be expected if Conjecture 1.1 were to hold, and so is in that sense best possible. Indeed, a well-known theorem of Ramanujan\(^2\) states

\[
E(S, S) = \sum_{1 \leq n \leq N} r_{S+S}(n)^2 \sim \frac{1}{4}N \log N. \tag{1.3}
\]

By (1.2), Cauchy–Schwarz and (1.1), we have for any sets of integers \(X\) and \(Y\) the estimate

\[
E(X, Y)^2 \leq E(X, X)E(Y, Y). \tag{1.4}
\]

If we believe Conjecture 1.1, then \(A\) should look like a quadratic sequence \(a = sx_a^2 + t\) for rational \(s\) and \(t\) (upon completing the square), and so \(A\) is an affine transform of the squares. The inequality (1.4) shows that for \(u \neq 0\) and \(v\) arbitrary

\[
E(X, uX + v)^2 \leq E(X, X)E(uX + v, uX + v) = E(X, X)^2,
\]

so that the logarithmic factor of Theorem 1.2 is the most one could hope prove in such generality.

Finally, we mention that the inequality (1.4) also shows that among all sets \(Y\) with comparable of additive energy to \(X\), \(X\) is essentially the best “additive partner” for itself in that \(E(X, Y)\) is maximised when \(Y = X\), provided \(Y\) has additive structure comparable to \(X\)’s. From this perspective, we see Theorem 1.2 as evidence in support of Conjecture 1.1.

One corollary from Theorem 1.2 is that, at the very least, there are at least \(\log N\) elements of \(A\) in the image of a single quadratic. Thus we have proved an \(\omega\)-result. This is far weaker than Conjecture 1.1 would give, but it is a positive result in the right direction.

**Corollary 1.3.** Suppose \(A\) is a set of positive integers in the interval \([1, N]\) satisfying the condition

\[
|A_p| \leq \frac{p}{2} + \epsilon(p)
\]

\(^2\) The earliest appearance of this I could find was in [R], but the result is a bit hard to find in the literature.
for all primes $p$ and some uniformly bounded sequence $\varepsilon(p) \geq 0$. If $|A| \gg \sqrt{N}$, then there is a rational quadratic $q(x) \in \mathbb{Q}[x]$ such that $|q(\mathbb{Z}) \cap A| \gg \log N$.

**Proof.** Since $E(A, S) \gg |A||S|\log N$, we have by (1-1),

$$\log N|A||S| \leq \sum_n r_{A+S}(n)^2 \leq |A||S| \max_n r(n)$$

whence there are at least $c \log N$ solutions $x_a$ to

$$a = n - x_a^2$$

with $a \in A$ for some $c > 0$. The, in fact integral, quadratic $q(x) = n - x^2$ will suffice.

Finally, it is known that dense Sidon subsets of $[1, N]$ are well-distributed modulo primes. This has been asserted in [Li] and [K]. We recover a similar such result here.

**COROLLARY 1.4.** Suppose $A$ is a Sidon set of positive integers in the interval $[1, N]$ satisfying the condition

$$|A_p| \leq \frac{p}{2} + \varepsilon(p)$$

for all primes $p$ and some uniformly bounded sequence $\varepsilon(p) \geq 0$. Then

$$|A| \ll \frac{\sqrt{N}}{\sqrt{\log N}}.$$  

**Proof.** By Theorem 1-2 we have

$$E(A, S) \gg |A|\sqrt{N} + |A|^2 \log N.$$  

On the other hand

$$E(A, S) \leq |A||S| + \sum_{n \neq 0} r_{A-A}(n)r_{S-S}(n) \leq (|A| + \sqrt{N})\sqrt{N}$$

since $A$ is Sidon. Rearranging gives the corollary.

2. **Lemmas and proofs**

One of the main observations in this paper is that one can do a little bit better than the larger sieve by considering composite moduli.

**LEMMA 2.1.** Suppose $A$ is a set of integers such that $|A_p| \leq p/2 + \varepsilon(p)$ for each prime $p$. Let $\Delta : \mathbb{N} \rightarrow \mathbb{R}$ be the multiplicative function such that

$$\Delta(p^k) = p^{k-1} \left(\frac{p}{2} + \varepsilon(p)\right).$$

Then for any $\nu \geq 1$, we have

$$\frac{|A|^2}{\Delta(\nu)} \leq \sum_{h \pmod{\nu}} |A(\nu; h)|^2.$$
Proof. Since $|A_p| \leq p/2 + \epsilon(p)$, then we also have $|A_{p^k}| \leq p^{k-1} (p/2 + \epsilon(p))$. By the Chinese Remainder Theorem, it follows that $|A_v| \leq \Delta(v)$. By Cauchy–Schwarz,

$$|A|^2 = \left( \sum_{h \pmod{v}} |A(v; h)| \right)^2 \leq |A_v| \sum_{h \pmod{v}} |A(v; h)|^2 \leq \Delta(v) \sum_{h \pmod{v}} |A(v; h)|^2.$$ 

In addition, we will use an estimate for averages of multiplicative functions. This particular estimate can be found in [FI, appendix A] as Corollary A·6.

**Lemma 2·2.** Let $g : \mathbb{N} \to \mathbb{R}$ be a multiplicative function supported on square-free integers. Suppose

$$g(p) = \frac{k}{p} + O \left( \frac{1}{p^2} \right)$$

for some integer $k \geq 1$. Then

$$\sum_{n \leq x} g(m) = \frac{\mathcal{S}_g}{k!} (\log x)^k + O \left( (\log x)^{k-1} \right),$$

where

$$\mathcal{S}_g = \prod_p \left( 1 - \frac{1}{p} \right)^k (1 + g(p)) \cdot$$

**Corollary 2·3.** Let $\Delta : \mathbb{N} \to \mathbb{R}$ be the multiplicative function such that

$$\Delta(p) = \frac{p}{2} + \epsilon(p)$$

for all primes $p$ and some uniformly bounded sequence $\epsilon(p) \geq 0$. Then,

$$\sum_{n \leq x} \frac{n^2}{\Delta(n)} \gg x^2 \log x.$$

**Proof.** Since $\Delta$ is non-negative, we get a lower bound by summing over only those $n$ which are square-free. We have

$$\Delta(p)^{-1} = \frac{2}{p} - \frac{4\epsilon(p)}{p(p + 2\epsilon(p))} = \frac{2}{p} + O \left( \frac{1}{p^2} \right)$$

so that if

$$M(x) = \sum_{n \leq x} \frac{\mu(n)^2}{\Delta(n)}$$

then Lemma 2·2 gives

$$M(x) = \frac{\mathcal{S}_\Delta}{2} (\log x)^2 + O (\log x).$$

Here,

$$\mathcal{S}_\Delta = \prod_p \left( 1 - \frac{1}{p} \right)^2 \left( 1 + \frac{2}{p} + O \left( \frac{1}{p^2} \right) \right) = \prod_p \left( 1 + O \left( \frac{1}{p^2} \right) \right)$$
is a convergent product. Again, since $\Delta(n)$ is non-negative, we are free to include only the terms with $n \geq x/C$ for some $C > 1$. Thus

$$\sum_{n \leq x} \frac{n^2 \mu(n)^2}{\Delta(n)} \geq \sum_{x/C \leq n \leq x} \frac{n^2 \mu(n)^2}{\Delta(n)}$$

$$\geq \left(\frac{x}{C}\right)^2 (M(x) - M(x/C))$$

$$\geq \left(\frac{x}{C}\right)^2 \log x \left(\mathcal{S}_\Delta \log C + O(1)\right).$$

Since $\mathcal{S}_\Delta$ converges, once $C$ is large enough we have $\mathcal{S}_\Delta \log C + O(1) \geq 1$, and the result is proved.

**Lemma 2.4.** Let $A \subseteq [1, N]$ be a set of integers with

$$|A_p| \leq \frac{p}{2} + \varepsilon(p)$$

for all primes $p$ and some uniformly bounded sequence $\varepsilon(p) \geq 0$. Then

$$\sum_{1 \leq u < v \leq \sqrt{N}} r_{A-A}(uv) \gg |A|^2 \log N.$$  

**Proof.** Observe that $a - b = uv$ with $1 \leq u < v$ if and only if $a \equiv b \pmod{v}$ and $b \leq a < b + v^2$, so that

$$\sum_{1 \leq u < v \leq \sqrt{N}} r_{A-A}(uv) = \sum_{1 \leq u < v \leq \sqrt{N}} \sum_{\substack{a, b \in A \atop a \equiv b \pmod{v} \atop b \leq a < b + v^2}} 1. \quad (2.1)$$

For fixed $v$, we divide the interval $[1, N]$ into $J_v = [N/v^2]$ intervals $I_j = [jv^2, (j + 1)v^2)$ of length $v^2$ (the last interval may be shorter). If $a, b$ are elements of $A \cap I_j$ with $b \leq a$ and $a$ congruent $b$ modulo $v$, then they are counted in the inner sum of (2.1). Thus, we produce a partition of $A$ into sets $A_j = A \cap I_j$ and deduce that

$$\sum_{1 \leq u < v \leq \sqrt{N}} r_{A-A}(uv) \geq \sum_{1 \leq v \leq \sqrt{N}} \sum_{1 \leq j \leq J_v} \sum_{h \equiv a \pmod{v}} |A_j(v; h)|^2$$

$$\geq \sum_{1 \leq v \leq \sqrt{N}} \frac{1}{\Delta(v)} \sum_{1 \leq j \leq J_v} |A_j|^2,$$

having applied Lemma 2.1 to each of the sets $A_j$. By Cauchy–Schwarz,

$$\sum_{1 \leq j \leq J_v} |A_j|^2 \geq \frac{|A|^2}{J_v} \geq \frac{v^2 |A|^2}{N}$$

so we have deduced

$$\sum_{1 \leq u < v \leq \sqrt{N}} r_{A-A}(uv) \gg \frac{|A|^2}{N} \sum_{1 \leq v \leq \sqrt{N}} \frac{v^2}{\Delta(v)}.$$

The lemma now follows from Corollary 2.3.

**Proof of Theorem 1.2.** By passing to a subset of $A$, we may assume that all elements of $A$ lie in a single congruence class modulo 4. The cost of this is that $|A|$ decreases by a factor
of at most 4. The additive energy between $A$ and $S$ is
\[
E(A, S) = \sum_{m^2, n^2 \in S} r_{A-A}(n^2 - m^2) \\
= |A||S| + 2 \sum_{1 \leq m^2 < n^2 \leq N} r_{A-A}((n-m)(n+m)),
\]
where in the second line we have extracted the contribution from $r_{A-A}(0)$ and used that $r_{A-A}(k) = r_{A-A}(-k)$. Now, for $k \not\equiv 0$, $k = n^2 - m^2$ has a solution if and only if $k \equiv 2 \pmod{4}$, and its solutions are indexed by pairs $(u, v)$ where $uv = k$, $u < v$ and $u \equiv v \pmod{2}$. This can be seen by considering the system
\[
n - m = u, \quad n + m = v
\]
which has as its solution
\[
n = \frac{v + u}{2}, \quad m = \frac{v - u}{2}.
\]
Thus
\[
E(A, S) = |A||S| + 2 \sum_{1 \leq u < v \leq \sqrt{N}/2} r_{A-A}(uv).
\]
Because all elements of $A$ are in a single congruence class modulo 4, $r_{A-A}$ is supported only on numbers which are divisible by 4. Thus if $a - b = n = uv$ for some $a, b \in A$, either $u \equiv v \pmod{2}$ or else one of $(u/2, 2v)$ or $(2u, v/2)$ is going to satisfy the necessary congruence condition. Thus, by reducing our count of pairs $(u, v)$ by a factor of at most 4, we can remove the condition $u \equiv v \pmod{2}$, and we obtain a lower bound
\[
E(A, S) \geq \frac{1}{2} \sum_{1 \leq u < v \leq \sqrt{N}/2} r_{A-A}(uv).
\]
The proof is complete after an application of Lemma 2.4.

REFERENCES

[BH] R. C. Bose and S. Chowla. Theorems in the additive theory of numbers. Comment. Math. Helv. 37 (1962–63), 141–147.

[CL] E. S. Croot, III and V. F. Lev. Open problems in additive combinatorics. In Additive combinatorics, volume 43 of CRM Proc. Lecture Notes, pages 207–233 (Amer. Math. Soc., Providence, RI, 2007).

[FI] J. Friedlander and H. Iwaniec. Opera de cribro. American Mathematical Society Colloquium Publications, 57 (American Mathematical Society, Providence, RI, 2010).

[GH] B. Green and A. J Harper. Inverse questions for the large sieve. Geom. Funct. Anal. 24(4) (2014), 1167–1203.

[HV] H. A. Helfgott and A. Venkatesh. How small must ill-distributed sets be? Analytic Number Theory, Essays in honour of Klaus Roth (Cambridge University Press 2009), 224–234.

[K] M. N. Kolountzakis. On the uniform distribution in residue classes of dense sets of integers with distinct sums. J. Number Theory 76 (1999), 147–153.

[La] E. Landau. Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindeszahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate. Arch. Math. Phys. 13 (1908), 305–312.

[Li] B. Lindström. Well distribution of Sidon sets in residue classes. J. Number Theory 69 (1998), 197–200.

[M] H. L. Montgomery. A note on the large sieve. J. London Math. Soc. 43 (1968), 93–98.

[R] S. Ramanujan. Some formulæ in the analytic theory of numbers. Messenger Math. 45 (1916), 81–84.

[W1] M. N. Walsh. The inverse sieve problem in high dimensions. Duke Math. J. 161 (2012), no. 10, 2001–2022.

[W2] M. N. Walsh. The algebraicity of ill-distributed sets. Geom. Funct. Anal. 24 (2014), no. 3, 959–967.