Combinatorica (2024) 44:691–698
https://doi.org/10.1007/s00493-024-00092-5

ORIGINAL PAPER

Note on the Theorem of Balog, Szemerédi, and Gowers

Christian Reiher[1] **· Tomasz Schoen**[2]

Received: 20 August 2023 / Revised: 17 February 2024 / Accepted: 18 February 2024 /
Published online: 23 April 2024

Abstract
We prove that every additive set A with energy $E(A) \geq |A|^3/K$ has a subset $A' \subseteq A$ of size $|A'| \geq (1-\varepsilon)K^{-1/2}|A|$ such that $|A' - A'| \leq O_\varepsilon(K^4|A'|)$. This is, essentially, the largest structured set one can get in the Balog–Szemerédi–Gowers theorem.

Keywords Additive combinatorics · Additive sets of large energy · Balog-Szemerédi-Gowers theorem

Mathematics Subject Classification 11B13 · 11P70

1 Introduction

An *additive set* is a nonempty finite subset of an abelian group. The *energy* of an additive set A is defined to be the number $E(A)$ of quadruples $(a_1, a_2, a_3, a_4) \in A^4$ solving the equation $a_1 + a_2 = a_3 + a_4$. An easy counting argument shows

$$E(A) = \sum_{d \in A-A} r_{A-A}(d)^2\,, \tag{1.1}$$

where $r_{A-A}(d)$ indicates the number of representations of d as a difference of two members of A. So the Cauchy–Schwarz inequality yields $E(A) \geq |A|^4/|A-A|$ and, in particular, every additive set A with small difference set $A - A$ contains a lot of energy. In the converse direction Balog and Szemerédi [2] proved that large energy implies the existence of a substantial subset whose difference set is small. After several

✉ Christian Reiher
Christian.Reiher@uni-hamburg.de

Tomasz Schoen
schoen@amu.edu.pl

1 Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany

2 Department of Discrete Mathematics, A. Mickiewicz University, Poznań, Poland

quantitative improvements (see e.g., Gowers [3] and Balog [1]) the hitherto best version of this result was obtained by the second author [4].

Theorem 1.1 *Given a real $K \geq 1$ every additive set A with energy $E(A) \geq |A|^3/K$ has a subset $A' \subseteq A$ of size $|A'| \geq \Omega(|A|/K)$ such that $|A' - A'| \leq O(K^4|A'|)$.* □

When investigating the question how a quantitatively optimal version of this result might read, there are two different directions one may wish to pursue. First, there is the obvious problem whether the exponent 4 can be replaced by some smaller number. Second, one may try to find "the largest" set $A' \subseteq A$ such that $|A' - A'| \leq O_K(|A'|)$ holds. As the following example demonstrates, there is no absolute constant $\varepsilon_\star > 0$ such that $|A'| \geq (1 + \varepsilon_\star)K^{-1/2}|A|$ can be achieved in general.

Fix an arbitrary natural number n. For a very large finite abelian group G we consider the additive set

$$A = \big\{(g_1, \ldots, g_n) \in G^n \colon \text{ there is at most one index } i \text{ such that } g_i \neq 0\big\}$$

whose ambient group is G^n. Obviously we have

$$|A| = |G|n + O_n(1) \quad \text{and} \quad E(A) = |A|^3/n^2 + O_n(|A|^2),$$

so the real number K satisfying $E(A) = |A|^3/K$ is roughly n^2. However, every $A' \subseteq A$ of size $|A'| \geq (1+\varepsilon)|G|$ satisfies $|A' - A'| \geq \varepsilon^2|G|^2$. Our main result implies that this is, in some sense, already the worst example. More precisely, for every fixed $\varepsilon > 0$ the Balog-Szemerédi-Gowers theorem holds with $|A'| \geq (1 - \varepsilon)K^{-1/2}|A|$. Perhaps surprisingly, we can also reproduce the best known factor K^4.

Theorem 1.2 *Given real numbers $K \geq 1$, $\varepsilon \in (0, 1/2)$, and an additive set A with energy $E(A) \geq |A|^3/K$ there is a subset $A' \subseteq A$ such that*

$$|A'| \geq (1 - \varepsilon)K^{-1/2}|A| \quad \textit{and} \quad |A' - A'| \leq 2^{33}\varepsilon^{-9}K^4|A'| = O_\varepsilon(K^4|A'|).$$

Our proof has two main cases and in one of them (see Lemma 3.1 below) we even get the stronger bound $|A' - A'| \leq O_\varepsilon(K^3|A'|)$. It would be interesting to prove this in the second case as well. Using examples of the form $A = \{x \in \mathbb{Z}^d \colon \|x\| \leq R\}$ one can show that the exponent 4 cannot be replaced by any number smaller than $\log(4)/\log(27/16) \approx 2.649$ (see [5]).

2 Preliminaries

This section discusses two auxiliary results we shall require for the proof of Theorem 1.2. The first of them is similar to [6, Lemma 6.19].

Lemma 2.1 *If $\delta, \xi \in (0, 1]$ and $R \subseteq A^2$ denotes a binary relation on a set A such that $|R| \geq \delta|A|^2$, then there is a set $A' \subseteq A$ of size $|A'| \geq \delta(1 - \xi)|A|$ which possesses the following property: For every pair $(a_1, a_2) \in A'^2$ there are at least $2^{-7}\delta^4\xi^4|A|^2|A'|$ triples $(x, b, y) \in A^3$ such that $(a_1, x), (b, x), (b, y), (a_2, y) \in R$.*

Proof Set $N(x) = \{a \in A\colon (a, x) \in R\}$ for every $x \in A$. Since $\sum_{x\in A} |N(x)| = |R| \geq \delta|A|^2$, the Cauchy–Schwarz inequality yields

$$\sum_{x\in A} |N(x)|^2 \geq \delta^2|A|^3 \,. \tag{2.1}$$

Setting $K(a, a') = \{x \in A\colon a, a' \in N(x)\}$ for every pair $(a, a') \in A^2$ and

$$\Omega = \big\{(a, a') \in A^2\colon |K(a, a')| \leq \delta^2\xi^2|A|/8\big\}$$

a double counting argument yields

$$\sum_{x\in A} |N(x)^2 \cap \Omega| = \sum_{(a,a')\in\Omega} |K(a, a')| \leq \delta^2\xi^2|A||\Omega|/8 \leq \delta^2\xi^2|A|^3/8.$$

Together with (2.1) we obtain

$$\sum_{x\in A}\big(|N(x)|^2 - 8\xi^{-1}|N(x)^2 \cap \Omega|\big) \geq \delta^2(1-\xi)|A|^3$$

and, hence, there exists some $x_\star \in A$ such that the set $A_\star = N(x_\star)$ satisfies

$$|A_\star|^2 - 8\xi^{-1}|A_\star^2 \cap \Omega| \geq \delta^2(1-\xi)|A|^2 \,. \tag{2.2}$$

We shall prove that the set

$$A' = \{a \in A_\star\colon \text{ the number of all } a' \in A_\star \text{ with } (a, a') \in \Omega \text{ is at most } |A_\star|/4\}$$

has all required properties. By (2.2) we have

$$|A_\star \smallsetminus A'||A_\star|/4 \leq |A_\star^2 \cap \Omega| \leq \xi|A_\star|^2/8,$$

for which reason

$$|A'| \geq (1-\xi/2)|A_\star| \geq (1-\xi)^{1/2}|A_\star| \overset{(2.2)}{\geq} \delta(1-\xi)|A|,$$

meaning that A' is indeed sufficiently large. To conclude the proof we need to show

$$\sum_{b\in A} |K(a_1, b) \times K(b, a_2)| \geq 2^{-7}\delta^4\xi^4|A|^2|A'|$$

for every pair $(a_1, a_2) \in A'^2$. This follows from the fact that due to the definition of A' there are at least $|A_\star|/2$ elements $b \in A_\star$ such that the sets $K(a_1, b)$ and $K(b, a_2)$ both have at least the size $\delta^2\xi^2|A|/8$. □

Lemma 2.2 *Suppose that the real numbers* $x_1, \dots, x_n \in [0, 1]$ *do not vanish simultaneously. Denote their sum by* S *and the sum of their squares by* T. *For every* $\alpha \in (0, 1)$ *there exists a set* $I \subseteq [n]$ *such that*

$$\sum_{i \in I} x_i \geq \max\left\{\alpha T, \left(\frac{(1-\alpha)^5 |I|^4 T^4}{2^{10} S^2}\right)^{1/6}\right\}.$$

Proof For reasons of symmetry we may assume $x_1 \geq \cdots \geq x_n$. Set $S_i = \sum_{j=1}^{i} x_j$ for every nonnegative $i \leq n$. Due to $T \leq x_1 S$ and $x_1 \leq 1$ we have $T \leq S = S_n$ and thus there exists a smallest index $k \in [n]$ satisfying $S_k \geq \alpha T$. Notice that

$$\sum_{i=1}^{k-1} x_i^2 \leq \sum_{i=1}^{k-1} x_i = S_{k-1} \leq \alpha T.$$

Moreover $x_1 \geq T/S$ implies the existence of a largest index ℓ such that $x_\ell \geq (1-\alpha)T/(2\,S)$. Due to

$$\sum_{i=\ell+1}^{n} x_i^2 \leq \frac{(1-\alpha)T}{2S} \sum_{i=\ell+1}^{n} x_i \leq \frac{(1-\alpha)T}{2},$$

we have

$$\sum_{i=k}^{\ell} x_i^2 \geq \frac{(1-\alpha)T}{2}, \tag{2.3}$$

whence, in particular, $\ell \geq k$. Next,

$$\ell\left(\frac{(1-\alpha)T}{2S}\right)^2 \leq \sum_{i=1}^{\ell} x_i^2 \leq T$$

entails

$$(1-\alpha)^2 \ell T \leq 4S^2. \tag{2.4}$$

Now assume for the sake of contradiction that our claim fails. Every $i \in [k, \ell]$ satisfies $S_i \geq S_k \geq \alpha T$ and thus the failure of $I = [i]$ discloses

$$S_i < \left(\frac{(1-\alpha)^5 i^4 T^4}{2^{10} S^2}\right)^{1/6}.$$

Combined with $i x_i \leq S_i$ this entails

$$\sum_{i=k}^{\ell} x_i^2 \leq \left(\frac{(1-\alpha)^5 T^4}{2^{10} S^2}\right)^{1/3} \sum_{i=k}^{\ell} i^{-2/3}.$$

In view of (2.3) we are thus led to

$$\left(\frac{2^7 S^2}{(1-\alpha)^2 T}\right)^{1/3} \le \sum_{i=k}^{\ell} i^{-2/3} \le \int_0^{\ell} x^{-2/3}\mathrm{d}x = 3\ell^{1/3},$$

i.e., $2^7 S^2 \le 27(1-\alpha)^2 \ell T$, which contradicts (2.4). □

3 The proof of Theorem 1.2

Let us fix two real numbers $K \ge 1$ and $\varepsilon \in (0, 1/2)$ as well as an additive set A satisfying $E(A) \ge |A|^3/K$. We consider the partition

$$A - A = P \mathbin{\dot\cup} Q$$

defined by

$$P = \left\{d \in A - A\colon r_{A-A}(d) \ge K^{-1/2}|A|\right\}$$
$$\text{and} \quad Q = \left\{d \in A - A\colon r_{A-A}(d) < K^{-1/2}|A|\right\}.$$

According to (1.1) at least one of the cases

$$\sum_{d\in P} r_{A-A}(d)^2 \ge \frac{\varepsilon|A|^3}{4K} \quad \text{or} \quad \sum_{d\in Q} r_{A-A}(d)^2 \ge \frac{(4-\varepsilon)|A|^3}{4K} \tag{3.1}$$

needs to occur and we begin by analysing the left alternative.

Lemma 3.1 *If $\sum_{d\in P} r_{A-A}(d)^2 \ge \varepsilon|A|^3/(4K)$, then there exists a set $A' \subseteq A$ of size $|A'| \ge (1-\varepsilon)K^{-1/2}|A|$ such that $|A' - A'| \le 2^{10}\varepsilon^{-4}K^3|A'|$.*

Proof For every difference $d \in P$ we set $A_d = A \cap (A + d)$. Due to $|A_d| = r_{A-A}(d)$ the hypothesis implies

$$\sum_{d\in P} |A_d|^2 \ge \varepsilon|A|^3/(4K)\,. \tag{3.2}$$

For every pair $(x, y) \in A^2$ the set $L(x, y) = \{d \in P\colon x, y \in A_d\}$ has at most the cardinality $|L(x, y)| \le r_{A-A}(x - y)$, because every difference $d \in L(x, y)$ corresponds to its own representation $x - y = (x - d) - (y - d)$ of $x - y$ as a difference of two members of A. Applying this observation to all pairs in

$$\Xi = \left\{(x, y) \in A^2\colon r_{A-A}(x - y) \le \varepsilon^2|A|/(16K)\right\}$$

we obtain

$$\sum_{d\in P} |A_d^2 \cap \Xi| = \sum_{(x,y)\in\Xi} |L(x, y)| \le \sum_{(x,y)\in\Xi} r_{A-A}(x - y) \le \frac{\varepsilon^2|A||\Xi|}{16K} \le \frac{\varepsilon^2|A|^3}{16K}.$$

Together with (3.2) this yields

$$\sum_{d\in P}\big(\varepsilon|A_d^2| - 4|A_d^2\cap\Xi|\big)\ge 0$$

and, consequently, for some element $d(\star)\in P$ the set $A_\star = A_{d(\star)}$ satisfies $|A_\star^2\cap\Xi|\le \varepsilon|A_\star|^2/4$. We contend that the set

$$A' = \big\{a\in A_\star\colon \text{ There are at most } |A_\star|/4 \text{ pairs of the form } (a,x) \text{ in } \Xi\big\}$$

has the required properties. As in the proof of Lemma 2.1 we obtain

$$|A'|\ge(1-\varepsilon)|A_\star| = (1-\varepsilon)r_{A-A}(d(\star))\ge(1-\varepsilon)K^{-1/2}|A|;$$

so it remains to derive the required upper bound on $|A'-A'|$.

To this end we consider an arbitrary pair (a,a') of elements of A'. Owing to the definition of A' there are at least $|A_\star|/2$ elements $x\in A_\star$ such that $(a,x)\notin\Xi$ and $(a',x)\notin\Xi$. For each of them we have $a-a' = (a-x)-(a'-x)$, there are at least $\varepsilon^2|A|/(16K)$ pairs $(a_1,a_2)\in A^2$ solving the equation $a-x = a_1-a_2$ and at least the same number of pairs $(a_3,a_4)\in A^2$ such that $a'-x = a_3-a_4$. Altogether there are at least

$$\varepsilon^4|A|^2|A_\star|/(2^9K^2)\ge 2^{-9}\varepsilon^4K^{-5/2}|A|^3$$

possibilities of writing $a-a' = (a_1-a_2)-(a_3-a_4)$ and for this reason we have

$$|A'-A'|\le\frac{|A|^4}{2^{-9}\varepsilon^4K^{-5/2}|A|^3} = 2^9\varepsilon^{-4}K^{5/2}|A|\le 2^{10}\varepsilon^{-4}K^3|A'|.$$

□

We conclude the proof of Theorem 1.2 by taking care of the right case in (3.1).

Lemma 3.2 *If $\sum_{d\in Q} r_{A-A}(d)^2\ge(1-\varepsilon/4)|A|^3/K$, then there is a set $A'\subseteq A$ of size $|A'|\ge(1-\varepsilon)K^{-1/2}|A|$ such that $|A'-A'|\le 2^{33}\varepsilon^{-9}K^4|A'|$.*

Proof Let $Q=\{d_1,\dots,d_{|Q|}\}$ enumerate Q. By the definition of Q there are real numbers $x_1,\dots,x_{|Q|}\in[0,1]$ such that

$$r_{A-A}(d_i) = x_iK^{-1/2}|A| \quad \text{holds for every } i\in[|Q|].$$

Owing to $\sum_{d\in A-A} r_{A-A}(d) = |A|^2$ and the hypothesis we have

$$\sum_{i=1}^{|Q|} x_i\le K^{1/2}|A| \quad\text{as well as}\quad \sum_{i=1}^{|Q|} x_i^2\ge(1-\varepsilon/4)|A|.$$

By Lemma 2.2 applied with $\alpha = 1 - \varepsilon/4$ there exist an index set $I \subseteq [|Q|]$ such that

$$\sum_{i \in I} x_i \geq \max\left\{(1-\varepsilon/2)|A|, \left(2^{-21}\varepsilon^5 K^{-1}|I|^4|A|^2\right)^{1/6}\right\}. \tag{3.3}$$

Now we set $Q' = \{d_i : i \in I\}$, consider the relation

$$R = \{(a_1, a_2) \in A^2 : a_1 - a_2 \in Q'\}$$

and define $\delta \in (0, 1]$ by $|R| = \delta|A|^2$. Due to

$$\delta = |A|^{-2}\sum_{i \in I} r_{A-A}(d_i) = \frac{1}{K^{1/2}|A|}\sum_{i \in I} x_i$$

the bounds in (3.3) imply both

$$\delta \geq (1-\varepsilon/2)K^{-1/2} \quad \text{and} \quad \frac{|I|^4}{\delta^6|A|^4} \leq 2^{21}\varepsilon^{-5}K^4. \tag{3.4}$$

By Lemma 2.1 applied to $\xi = \varepsilon/2$ and R there exists a set $A' \subseteq A$ of size

$$|A'| \geq (1-\varepsilon/2)\delta|A| \geq (1-\varepsilon)K^{-1/2}|A|$$

such that for every pair $(a_1, a_2) \in A'^2$ there are at least $2^{-11}\varepsilon^4\delta^4|A|^2|A'|$ triples $(x, b, y) \in A^3$ with $(a_1, x), (b, x), (b, y), (a_2, y) \in R$. Due to the equation

$$(a_1 - a_2) = (a_1 - x) - (b - x) + (b - y) - (a_2 - y)$$

this means that every difference $a_1 - a_2 \in A' - A'$ has at least $2^{-11}\varepsilon^4\delta^4|A|^2|A'|$ representations of the form $q_1 - q_2 + q_3 - q_4$ with $q_1, q_2, q_3, q_4 \in Q'$, whence

$$|A' - A'| \leq \frac{|Q'|^4}{2^{-11}\varepsilon^4\delta^4|A|^2|A'|} \overset{(3.4)}{\leq} 2^{32}\varepsilon^{-9}K^4(\delta|A|/|A'|)^2|A'|.$$

Due to $|A'| \geq (1-\varepsilon/2)\delta|A| \geq \delta|A|/\sqrt{2}$ the result follows. □

Acknowledgements We would like to thank the referees for reading our article very carefully.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Balog, A.: Many additive quadruples, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, pp. 39–49 (2007). https://doi.org/10.1090/crmp/043/03. MR2359466
2. Balog, A., Szemerédi, E.: A statistical theorem of set addition. Combinatorica **14**(3), 263–268 (1994). https://doi.org/10.1007/BF01212974
3. Gowers, W.T.: A new proof of Szemerédi's theorem for arithmetic progressions of length four. Geom. Funct. Anal. **8**(3), 529–551 (1998). https://doi.org/10.1007/s000390050065
4. Schoen, T.: New bounds in Balog-Szemerédi-Gowers theorem. Combinatorica **35**(6), 695–701 (2015). https://doi.org/10.1007/s00493-014-3077-4
5. Shao, X.: Large values of the additive energy in $\mathbb{R}^d$ and $\mathbb{Z}^d$. Math. Proc. Cambridge Philos. Soc. **156**(2), 327–341 (2014). https://doi.org/10.1017/S0305004113000741
6. Tao, T., Vu, V.: Additive combinatorics, Cambridge studies in advanced mathematics, vol. 105. Cambridge University Press, Cambridge (2006)