ACTA ARITHMETICA
131.1 (2008)

Non-trivial solutions to a linear equation in integers

by

Boris Bukh (Princeton, NJ)

Introduction. The problems of estimating the size of a largest subset of $[1, N]$ not containing a solution to a given linear equation arise very frequently in combinatorial number theory. For example, the sets with no non-trivial solutions to $x_1 + x_2 - 2x_3 = 0$ and $x_1 + x_2 = x_3 + x_4$ are sets with no arithmetic progressions of length three, and Sidon sets respectively. For several of the more prominent equations, like the two equations above, there are large bodies of results that deal with the structure and size of solution-free sets. The first systematic study of general linear equations was undertaken by Ruzsa [Ruz93, Ruz95]. To ascribe a precise meaning to the concept of a "set with no non-trivial solution" he introduced two definitions of a trivial solution. One of them is that a solution is non-trivial if all the variables are assigned different values.

For a fixed linear equation denote by $R(N)$ the size of a largest set of integers in $[1, N]$ with no solution to the equation in distinct integers. Ruzsa [Ruz93] showed that if $k \geq 2$, then for the symmetric equation

$$a_1x_1 + \cdots + a_kx_k = a_1x_{k+1} + \cdots + a_kx_{2k} \tag{1}$$

one has $R(N) = O(N^{1/2})$. For $k = 2$, the estimate is tight for the Sidon equation. Ruzsa gave the example of the equation

$$x_1 + d(x_2 + \cdots + x_k) = x_{k+1} + d(x_{k+2} + \cdots + x_{2k}) \tag{2}$$

and the set A of integers in $[1, N]$ whose development in base d^2k consists only of digits $0, \ldots, d-1$. Since addition of elements of A in (2) involves no carries in base d^2k, for every solution to (2) with elements of A one necessarily has x_1 and x_{k+1} equal digit by digit. Thus $x_1 = x_{k+1}$, implying that A contains no solution (2) in distinct integers. Since $|A| = \Omega(N^{1/2-c/\log d})$, this example shows that there is no $\varepsilon > 0$ such that the estimate $R(N) = O(N^{1/2-\varepsilon})$ holds for all equations of the form (1). However, Ruzsa asked

2000 *Mathematics Subject Classification*: 11D04, 11D45, 11P99.
Key words and phrases: linear equation, symmetric equation.

whether for $k \geq 3$ there is an $\varepsilon > 0$ depending on the coefficients of (1) such that $R(N) = O(N^{1/2-\varepsilon})$. We answer this in the affirmative.

THEOREM 1. *Let* $\|a\|_1 = \sum_{i=1}^{k} |a_i|$. *If* $k \geq 3$ *and* $a_i \neq 0$ *for* $1 \leq i \leq k$, *then*

$$R(N) = O(N^{1/2-1/c(k)\|a\|_1}).$$

To explain the idea behind the proof of Theorem 1 we examine the proof that $R(N) = O(N^{1/2})$. Let $a_{k+i} = -a_i$. Then equation (1) can be written as $\sum a_i x_i = 0$. Assume that $A \subset [1, N]$ has M elements and contains only trivial solutions to (1). Let $r(m)$ denote the number of solutions to

$$a_1 x_1 + \cdots + a_k x_k = m, \qquad x_i \in A.$$

Then $E = \sum r(m)^2$ is the total number of solutions to (1) in A. Since there are at most $\|a\|_1 N$ values of m for which $r(m) > 0$, by Cauchy–Schwarz it follows that

$$E = \sum r(m)^2 \geq \frac{M^{2k}}{\|a\|_1 N}. \tag{3}$$

The next step is to bound the number of solutions with $x_i = x_j$ from above. Let $1 \leq r \leq 2k$ be any index other than i or j. Since

$$x_r = -\frac{1}{a_r}\Big(\sum_{l \neq i,j,r} a_l x_l + (a_i + a_j) x_i\Big), \tag{4}$$

it follows that equation (1) uniquely determines x_r in terms of the other x's. Thus at most one in every M assignments of variables results in a solution to (1). Since there are $\binom{2k}{2}$ choices of i and j,

$$E \leq \binom{2k}{2} \frac{1}{M} M^{2k-1}. \tag{5}$$

The bound $|A| = O(N^{1/2})$ follows by comparison with (3).

In order for this argument to be tight the linear form on the right side of (4) should often take values in A. Heuristically it corresponds to the sumset $\sum_{l \neq i,j,r} x_l \cdot A + (x_i + x_j) \cdot A$ being not much larger than A itself, where $t \cdot A = \{ta : a \in A\}$ is the t-dilate of A. The Plünnecke–Ruzsa inequalities tell us that if some sumset involving A is not much larger than A, then every sumset involving only A is not much larger than A. In particular we expect $\sum_{l \leq k} x_l \cdot A$ to be of size only M, as opposed to $\sim N$ which was assumed in the derivation of the lower bound (3). Therefore, we expect that if (5) is close to being tight, then (3) is not, and vice versa. The remainder of the paper is a rigorous justification of the heuristic argument above.

Proof. This section is organized as follows. First we state the tools from additive combinatorics that we need. Then we introduce the notation that

is going to be used in the proof of Theorem 1. Finally, after a couple of preliminary lemmas the Theorem 1 is proved.

LEMMA 2 (Ruzsa's triangle inequality, [TV06, Lemma 2.6]). *For any finite $A, B, C \subset \mathbb{Z}$ we have*

$$|A-C| \le \frac{|A-B|\,|B-C|}{|B|}.$$

LEMMA 3 (Plünnecke's inequality, [TV06, Corollary 6.26]). *For any finite sets $A, B \subset \mathbb{Z}$, if $|A+B| \le K|A|$, then $|kB| \le K^k|A|$.*

We will also use the hypergraph version of the Balog–Szemerédi–Gowers theorem due to Sudakov, Szemerédi and Vu [SSV05]. Let $A_1, \dots, A_k$ be sets of integers. If H is a subset of $A_1 \times \cdots \times A_k$, then $\sum_H A_i$ is the collection of all sums $a_1 + \cdots + a_k$ where $(a_1, \dots, a_k) \in H$.

LEMMA 4 ([SSV05, Theorem 4.3]). *For any integer $k \ge 1$ there are positive-valued polynomials $f_k(x,y)$ and $g_k(x,y)$ such that the following holds. Let n, C, K be positive numbers. If $A_1, \dots, A_k$ are sets of n positive integers, $H \subset A_1 \times \cdots \times A_k$ with $|H| \ge n^k/K$ and $|\sum_H A_i| \le Cn$, then one can find subsets $A_i' \subset A_i$ such that*

$$|A_i'| \ge n/f_k(C,K) \quad \text{for all } 1 \le i \le k, \qquad |A_1' + \cdots + A_k'| \le g_k(C,K)n.$$

From the heuristic argument above it is clear that we will need to count the number of solutions of various equations. It is therefore advantageous to introduce appropriate notation. Let $r(A_1, \dots, A_k; m)$ denote the number of solutions to

$$a_1 + \cdots + a_k = m, \qquad a_i \in A_i.$$

Then $E(A_1, \dots, A_k; B_1, \dots, B_l) = \sum_m r(A_1, \dots, A_k; m)\, r(B_1, \dots, B_l; m)$ counts the number of solutions to

$$a_1 + \cdots + a_k = b_1 + \cdots + b_l, \qquad a_i \in A_i,\ b_j \in B_j.$$

We write $E(A_1, \dots, A_k)$ to denote $E(A_1, \dots, A_k; A_1, \dots, A_k)$.

LEMMA 5. *There is a positive-valued polynomial $h_k(x)$ such that if $A_1, \dots, A_k$ are sets of integers with n elements each, and $E(A_1, \dots, A_k) \ge cn^{2k-1}$, then there are subsets $A_i' \subset A_i$ such that*

$$|A_i'| \ge h_k(c)n \quad \text{for all } 1 \le i \le k, \qquad |A_1' + \cdots + A_k'| \le n/h_k(c).$$

Proof. For brevity write $E = E(A_1, \dots, A_k)$ and $r(m) = r(A_1, \dots, A_k; m)$. Let $S = \{m : r(m) > cn^{k-1}/2\}$ and $H = \{(a_1, \dots, a_k) \in A_1 \times \cdots \times A_k : a_1 + \cdots + a_k \in S\}$. Note that $\sum_H A_i = S$ and

$$|S| \le \frac{n^k}{cn^{k-1}/2} = \frac{2}{c}\, n.$$

Since $r(m) \le n^{k-1}$ for every m, and $\sum_{m\notin S} r(m) = n^k - |H|$, it follows that

$$E = \sum_{m\in S} r(m)^2 + \sum_{m\notin S} r(m)^2 \le |H|n^{k-1} + (n^k - |H|)cn^{k-1}/2.$$

Therefore $|H| \ge cn^k/2$, and we deduce from Lemma 4 the existence of subsets A_i' with the desired properties. ∎

LEMMA 6. *For every $k \ge 2$ there is a positive-valued polynomial $p_k(x)$ such that if $t_1, \dots, t_k$ are any k positive integers, and $A \subset \mathbb{Z}$ is an n-element set satisfying $E(t_1 \cdot A, \dots, t_k \cdot A) \ge cn^{2k-1}$, then for every l-tuple of positive integers $s_1, \dots, s_l$ we have $E(s_1 \cdot A, \dots, s_l \cdot A) \ge p_k(c)^{\|s\|_1} n^{2l-1}$.*

Proof. Apply the lemma above with $A_i = t_i \cdot A$ to obtain A_i'. Ruzsa's triangle inequality applied to $-A_1'$ and $A_2' + \dots + A_k'$ yields

$$|A_1' - A_1'| \le \frac{|A_1' + A_2' + \dots + A_k'|^2}{|A_2' + \dots + A_k'|} \le \frac{n}{h_k(c)^3}.$$

Set $\widetilde{A} = (1/t_1) \cdot A_1'$. Then $\widetilde{A}$ is a subset of A satisfying $|\widetilde{A}| \ge h_k(c)n$ and $|\widetilde{A} - \widetilde{A}| \le n/h_k(c)^3$. Plünnecke's inequality then implies that $|\,\|s\|_1\widetilde{A}| \le h_k(c)^{-3\|s\|_1} n$. The inclusion $s_1 \cdot \widetilde{A} + \dots + s_l \cdot \widetilde{A} \subset \|s\|_1 \widetilde{A}$ together with the Cauchy–Schwarz inequality yields

$$\begin{aligned} E(s_1 \cdot A, \dots, s_l \cdot A) \ge E(s_1 \cdot \widetilde{A}, \dots, s_l \cdot \widetilde{A}) &\ge \frac{|\widetilde{A}|^{2l}}{|s_1 \cdot \widetilde{A} + \dots + s_l \cdot \widetilde{A}|} \\ &\ge \frac{h_k(c)^{2l} n^{2l}}{|\,\|s\|_1 \widetilde{A}|} \ge h_k(c)^{3\|s\|_1 + 2l} n^{2l-1}. \end{aligned}$$

Since $s_1, \dots, s_l$ are positive integers, $\|s\|_1 \ge l$, and the lemma follows. ∎

Proof of Theorem 1. Without loss of generality we may assume that $a_1, \dots, a_k$ are positive. Rewrite equation (1) as

$$a_1 x_1 + \dots + a_{2k} x_{2k} = 0, \tag{6}$$

where $a_{k+i} = -a_i$. Let $A \subset [1, N]$ with $|A| = M$ contain only trivial solutions to (6). The number of solutions to (6) with variables in A is $E = E(a_1 \cdot A, \dots, a_k \cdot A)$. Let $T_{i,j}$ be the number of solutions with $x_i = x_j$. By the pigeonhole principle for at least one pair $i \ne j$ we have $T_{i,j} \ge E/\binom{2k}{2}$. Fix such a pair.

Next we partition $\{1, \dots, 2k\} \setminus \{i, j\}$ arbitrarily into $I_1 = \{l_{1,1}, \dots, l_{1,k-1}\}$ and $I_2 = \{l_{2,1}, \dots, l_{2,k-1}\}$ with $k-1$ elements each. For convenience write

$$r_1(m) = r(a_{l_{1,1}} \cdot A_{l_{1,1}}, \dots, a_{l_{1,k-1}} \cdot A_{l_{1,k-1}}; m)$$

and $E_1 = \sum r_1(m)^2$. Define r_2 and E_2 analogously with respect to I_2. Then

by Cauchy–Schwarz,

$$T_{i,j} = \sum_{x\in A}\sum_{m} r_1(m)r_2(-(a_i+a_j)x-m) \le \sum_{x\in A} E_1^{1/2}E_2^{1/2} = ME_1^{1/2}E_2^{1/2}.$$

Either E_1 or E_2 is at least $E/\binom{2k}{2}M$. We can assume it is E_1. Then by Lemma 6 above we have

$$E \ge p_{k-1}\left(\frac{E_1}{M^{2k-3}}\right)^{\|a\|_1} M^{2k-1} \ge p_{k-1}\left(\frac{E}{\binom{2k}{2}M\,M^{2k-3}}\right)^{\|a\|_1} M^{2k-1}.$$

If $\deg p_{k-1} = d$, then we obtain

$$E = O(M^{2k-2-1/(d\|a\|_1-1)}).$$

Since $E \ge M^{2k}/\|a\|_1 N$, the theorem follows. ∎

Conclusion. Theorem 1 gives the estimate $R(N) = O(N^{1/2-1/r})$ with $r = c(k)\|a\|_1$. Ruzsa's example shows that no estimate better than $r = c(k)\log\|a\|_1$ can be true.

After this paper was written, a Plünnecke-type estimate on sums of dilates of the form $s_1\cdot A+\cdots+s_l\cdot A$, which appears in the proof of Lemma 6, was established in [Buk07]. Whereas Plünnecke's inequality and the inclusion $s_1\cdot A+\cdots+s_l\cdot A \subset \|s\|_1 A$ yield the exponent of $\|s\|_1$, the new inequality yields the exponent $C\log\|s\|_1$ for an absolute constant C. Plugging that into the proof of Lemma 6, one finds that the estimate in Theorem 1 is valid with $r = c(k)\log\|a\|_1$, which is sharp in view of Ruzsa's example.

Acknowledgement. I thank Benjamin Sudakov for stimulating discussions.

References

[Buk07] B. Bukh, *Sums of dilates*, arXiv:0711.1610, Nov 2007.

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

[Ruz95] —, *Solving a linear equation in a set of integers. II*, ibid. 72 (1995), 385–397.

[SSV05] B. Sudakov, E. Szemerédi, and V. H. Vu, *On a question of Erdős and Moser*, Duke Math. J. 129 (2005), 129–155.

[TV06] T. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge Stud. Adv. Math. 105, Cambridge Univ. Press, 2006.

Department of Mathematics
Princeton University
Fine Hall, Washington Rd.
Princeton, NJ 08544, U.S.A.
E-mail: bbukh@math.princeton.edu

Received on 14.3.2007
and in revised form on 10.11.2007 (5410)