Non-averaging subsets and non-vanishing transversals

Noga Alon * Imre Z. Ruzsa †

Abstract

It is shown that every set of n integers contains a subset of size $\Omega(n^{1/6})$ in which no element is the average of two or more others. This improves a result of Abbott. It is also proved that for every $\epsilon > 0$ and every $m > m(\epsilon)$ the following holds. If $A_1, \ldots, A_m$ are m subsets of cardinality at least $m^{1+\epsilon}$ each, then there are $a_1 \in A_1, \ldots, a_m \in A_m$ so that the sum of every nonempty subset of the set $\{a_1, \ldots, a_m\}$ is nonzero. This is nearly tight. The proofs of both theorems are similar and combine simple probabilistic methods with combinatorial and number theoretic tools.

1 Introduction

In this paper we consider two problems in additive number theory. The problems are not directly related, but the methods we use in tackling them are similar. The first problem deals with the existence of large non-averaging subsets in sets of integers. A set of integers is called *non-averaging* if no member of the set is the average of two or more others. Answering a problem of Erdős, Abbott proved in [4] that every set of n integers contains a non-averaging subset of cardinality $\Omega(n^{1/13}/(\log n)^{1/13})$. His method, together with the result of Bosznay [8] mentioned in the next section, can be used to get an $\Omega(n^{1/7-\epsilon})$ bound, for any $\epsilon > 0$. Here we improve this estimate and prove the following.

Theorem 1.1 *Every set of n integers contains a non-averaging subset of cardinality $\Omega(n^{1/6})$.*

*School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540 and Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Email: noga@math.tau.ac.il. Research supported in part by a State of New Jersey grant, by a United States Israeli BSF grant and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†Mathematical Institute of the Hungarian Academy of Sciences, Budapest, P.O.B. 127, H-1364, Hungary. Email: ruzsa@math-inst.hu. Research supported in part by the Hungarian National Foundation for Scientific Research, Grants No. 17433 and 25617.

The second problem we consider deals with non-vanishing transversals. Let $A_1, \ldots, A_m$ be m sets of integers. A *transversal* (for the sets A_i) is a sequence $(a_1, \ldots, a_m)$, where $a_i \in A_i$. It is a *non-vanishing transversal* if for every $\emptyset \neq I \subset \{1, \ldots, m\}$, $\sum_{i \in I} a_i \neq 0$. Let $g(m)$ denote the minimum number g so that for every m sets of cardinality g each there is a non-vanishing transversal. It is not difficult to see that $g(m) > m$. Indeed, if $b_1 < \ldots < b_m$ is an arbitrary set of integers, and $A_i = \{b_i - b_1, b_i - b_2, \ldots, b_i - b_m\}$ then the sets A_i are of size m each, and contain no non-vanishing transversal. To see this, note that for any choice of elements $a_i = b_i - b_{j_i} \in A_i$, the directed graph on the vertices $1, \ldots, m$ whose directed edges are all edges (i, j_i) for $1 \leq i \leq m$ has all outdegrees 1 and hence contains a directed cycle, giving a nontrivial subset of the numbers a_i whose sum is 0.

L. Goddyn and M. Tarsi (private communication) conjectured that this is best possible.

Conjecture 1.2 (Goddyn and Tarsi) *For every $m \geq 1$, $g(m) = m + 1$. That is: for every family of m sets of $m + 1$ integers each, there is a non-vanishing transversal.*

Here we prove the following weaker statement.

Theorem 1.3 *For every $\epsilon > 0$ there exists an $m_0 = m_0(\epsilon)$ such that for every $m > m_0$, $g(m) \leq m^{1+\epsilon}$.*

The proofs of Theorems 1.1 and 1.3 are similar, and apply the second moment method. The proof of Theorem 1.1 is simpler, and is presented in Section 2. The basic idea in it is a simplified version of the method of Komlós, Sulyok and Szemerédi in [13] (see also [14]). The proof of Theorem 1.3 is more complicated, and is presented (in a somewhat stronger form) in Section 3. The proof combines the second moment method with some number theoretic tools and graph theoretic arguments. The basic approach resembles the one in [5], but several new ingredients are incorporated.

2 Non-averaging subsets

Let $f(n)$ denote the maximum possible size of a non-averaging subset of $N = \{1, \ldots, n\}$. Several papers ([16], [11], [1], [2], [3], [8]) deal with the problem of determining or estimating $f(n)$. The best known lower estimate is due to Bosznay [8], who modified the constructions of Abbott in [1], [2], [3] and constructed a non-averaging subset of cardinality $\Omega(n^{1/4})$ of N. His construction, as well as those in [1], [2], [3], is based on the clever (and simple) method of Behrend [6], in his construction of a dense subset of N that contains no three-term arithmetic progression. The best known upper bound for $f(n)$ follows from the results in [9], which supply an upper bound of $O(n \log n)^{1/2}$. Therefore

$$\Omega(n^{1/4}) \leq f(n) \leq O((n \log n)^{1/2}). \tag{1}$$

Let $h(n)$ denote the maximum h so that every set of n integers contains a non-averaging subset of cardinality h. The following proposition, together with the lower bound in (1), implies the assertion of Theorem 1.1.

Proposition 2.1 *For every $n \geq k$,*

$$h(n) \geq min\{\ k,\ \frac{1}{2}f(\lfloor n/k^2 \rfloor)\ \}.$$

Proof. Let $A = \{a_1, \ldots, a_n\}$ be an arbitrary set of n integers. Our objective is to show it contains a large non-averaging subset. Put $r = \lfloor n/k^2 \rfloor$. By the definition of the function f there is a non-averaging subset S of cardinality $f(r)$ of $\{1, \ldots, r\}$. For each $s \in S$, let I_s denote the interval $[\frac{s-1}{rk}, \frac{s-1}{rk} + \frac{1}{n})$. We claim that if $b_0, b_1, \ldots, b_p$ is any set of $p+1 \leq k+1$ reals, where the points b_i are in some $p+1$ *distinct* intervals I_s from the intervals above, then the equation $pb_0 \equiv \sum_{i=1}^{p} b_i\ (mod\ 1)$ is *not* satisfied. This is because if the last equation holds, then, in fact $pb_0 = \sum_{i=1}^{p} b_i$ since both sides of the last equation are smaller than 1. Moreover, by the definition of the intervals and as $p/n \leq k/n \leq 1/kr$, the last equation contradicts the fact that S is non-averaging, proving the claim.

It follows that if there are two reals α, β so that the set $\alpha A + \beta\ (mod\ 1)$ intersects at least q of the intervals I_s, $s \in S$, then A contains a non-averaging subset of size $min\{k, q\}$. Indeed, choose $min\{k, q\}$ of the intervals that intersect $\alpha A + \beta\ (\ mod\ 1)$, and for each of them choose some $a \in A$ for which $\alpha a + \beta\ (mod\ 1)$ is in the interval. The set of all the chosen elements is clearly non-averaging. This is because otherwise $pa_0 = a_1 + \ldots + a_p$ for some chosen elements a_i, implying that $p(\alpha a_0 + \beta) \equiv \sum_{i=1}^{p}(\alpha a_i + \beta)\ (mod\ p)$, which is impossible, by the discussion above.

To complete the proof it remains to show that there are α, β for which $\alpha A + \beta\ (mod\ 1)$ intersects sufficiently many intervals I_s. To do so we choose, randomly and independently, α and β in $[0, 1)$, according to a uniform distribution. Fix an interval $I = I_s$ for some $s \in S$, and let X denote the random variable counting the number of elements a of A for which $z_a = \alpha a + \beta\ (mod\ 1) \in I$. X is the sum of the n indicator random variables X_a, $a \in A$, where $X_a = 1$ iff $z_a \in I$. The random variables X_a are pairwise independent and $Prob(X_a = 1) = 1/n$ for all $a \in A$. This is because for every two distinct members a, a' of A, the ordered pair $(z_a, z_{a'})$ attains all values in $[0, 1)^2$ according to a uniform distribution, as α and β range over $[0, 1)$. Therefore, the expectation and variance of X satisfy $E(X) = n \cdot 1/n = 1$ and VAR(X) $= n(1/n)(1 - 1/n) \leq 1$. Let p_i denote the probability that $X = i$. By the Cauchy Schwartz inequality

$$(E(X))^2 = (\sum_{i>0} ip_i)^2 = (\sum_{i>0} i\sqrt{p_i}\sqrt{p}_i)^2$$

$$\leq (\sum_{i>0} i^2 p_i)(\sum_{i>0} p_i) = E(X^2)Prob(X > 0) = (\text{VAR(X)} + (E(X))^2)Prob(X > 0).$$

Therefore, $Prob(X > 0) \geq 1/2$, that is; the probability that $\alpha A + \beta$ $(mod\ 1)$ intersects I is at least a half.

By linearity of expectation we conclude that the expected number of intervals I_s containing a member of $\alpha A + \beta$ $(mod\ 1)$ is at least $|S|/2$ and hence there is a choice for α and β for which at least $|S|/2 = f(r)/2 = \frac{1}{2} f(\lfloor n/k^2 \rfloor)$ intervals I_s contain members of $\alpha A + \beta$ $(mod\ 1)$. By the above discussion, this implies the assertion of the proposition, and completes the proof of Theorem 1.1. □

3 Non-vanishing transversals

In this section we prove Theorem 1.3 in the following sharper form.

Proposition 3.1 *There exists a positive constant c so that for every m*

$$g(m) \leq m e^{c\sqrt{\log m \log\log m}}.$$

The basic probabilistic approach in the proof is similar to the one in the previous section, but there are various additional ideas. For any real x, let $\{x\} = x(\ mod\ 1)$ denote the fractional part of x. Given m sets $A_1, \ldots, A_m$ of size $n \geq m e^{c\sqrt{\log m \log\log m}}$ each, our objective is to show that there exists a non-vanishing transversal. To do so, we prove the existence of a real γ and $a_i \in A_i$ so that $\{\gamma a_i\} > 0$ for all i and $\sum_{i=1}^{m} \{\gamma a_i\} < 1$. This clearly implies that $\{a_1, \ldots, a_m\}$ is a non-vanishing transversal. For each set A_i, define $f_i(\gamma) = min_{a \in A_i}\{\gamma a\}$. Let γ be chosen randomly in $[0, 1)$, according to a uniform distribution. Since the probability that $f_i(\gamma) = 0$ is zero for every i, it suffices to show that with positive probability $\sum_{i=1}^{m} f_i(\gamma) < 1$. To do so, it is enough to show that the expected value of the last sum is less than 1. By linearity of expectation it is sufficient to show that the expected value of each f_i is less than $1/m$. It thus suffices to prove the following.

Lemma 3.2 *There exists a positive constant c so that for every m, if A is set of* $n \geq m e^{c\sqrt{\log m \log\log m}}$ *nonzero integers, then the expected value of* $f(\gamma) = min_{a \in A}\{\gamma a\} < 1/m$.

We derive the last lemma from the following result.

Lemma 3.3 *There exists a positive constant* c' *so that for every* n *and* $t > 0$, *if* A *is set of* n *nonzero integers, and* γ *is randomly chosen in* $[0, 1)$, *then the probability that* $\{\gamma a\} > t$ *for all* $a \in A$ *is at most*

$$\frac{e^{c'\sqrt{\log n \log\log n}}}{tn}.$$

Although this is not really essential, it is convenient to assume in the proof of the last two lemmas, that all members of A have the same sign (since at least half of them have the same sign, and we may replace n by $n/2$ without any change in the estimates). Note that if k is a positive integer and t is a real between 0 and 1, then the set $\{\gamma : \{\gamma k\} \le t\}$ is precisely the set

$$X = \bigcup_{j=0}^{k-1} [j/k, j/k + \alpha],$$

where $\alpha = t/k$. The probability that a random γ in $[0, 1)$ lies in this set is precisely the measure $\mu(X)$ $(= t)$ of the set X. In order to apply the second moment method we need to compute the measure of the intersection of two such sets. This is done in the following two lemmas.

Lemma 3.4 *Let k, l be coprime positive integers, $t, z \in [0, 1]$, $\alpha = t/k$, $\beta = z/l$. Consider the sets*

$$X = \bigcup_{j=0}^{k-1} [j/k, j/k + \alpha], \qquad Y = \bigcup_{j=0}^{l-1} [j/l, j/l + \beta].$$

Then

$$\mu(X \cap Y) = tz + \frac{1}{kl}\Delta(\{tl\}, \{zk\}),$$

where $\Delta(x, y) = \min(x, y) - xy$.

Proof. The differences of the form $u/k - v/l$ are identical modulo one with the numbers $j/(kl)$. Consider the pair with difference $j/(kl)$, where

$$-\frac{1}{2} < \frac{j}{kl} \le \frac{1}{2}.$$

Assume $\alpha \le \beta$. The length of the intersection of the corresponding intervals is

$$\min(\alpha, \beta - \frac{j}{kl})$$

for $0 \le j \le \beta kl$,

$$\alpha - \frac{|j|}{kl}$$

for $j < 0$, $|j| \le \alpha kl$, and 0 otherwise. So

$$s = \mu(X \cap Y) = \sum_{0 \le j \le \beta kl} \min(\alpha, \beta - \frac{j}{kl}) + \sum_{0 < j < \alpha kl} (\alpha - \frac{j}{kl}).$$

Put $\alpha kl = tl = p + \varepsilon$, $\beta kl = zk = q + \delta$ with $0 \le \varepsilon, \delta < 1$. Then

$$skl = \sum_{j=0}^{q} \min(p + \varepsilon, q + \delta - j) + \sum_{j=1}^{p} (p + \varepsilon - j)$$

$$= \sum_{j=0}^{q-p+1} (p+\varepsilon) + p + \min(\varepsilon,\delta) + \sum_{j=q-p+1}^{q} (q+\delta-j) + \sum_{j=1}^{p} (p+\varepsilon-j)$$
$$= (p+\varepsilon)(q+\delta) + \min(\varepsilon,\delta) - \varepsilon\delta$$

as claimed. □

Lemma 3.5 *Let k and l be two not necessarily coprime positive integers, and define $k' = k/(k,l)$, $l' = l/(k,l)$, where (k,l) is the greatest common divisor of k and l. Then for the sets X and Y defined in Lemma 3.4 we have*

$$\mu(X \cap Y) = tz + \frac{(k,l)^2}{kl}\Delta(\{tl'\},\{zk'\}).$$

Proof. The systems of intervals in the definition of X and Y can be obtained from (k,l) copies of the corresponding system for k', l' after shrinking it by a factor of (k,l), hence the result is the same as that of Lemma 3.4 for k', l'. □

Lemma 3.6 *Let A be a set of n nonzero integers of the same sign, and write*

$$f(\gamma) = \min_{a\in A}\{\gamma a\}.$$

If γ is chosen randomly and uniformly in $[0,1)$ then

$$Prob(f(\gamma) > t) \le \frac{S}{n^2 t},$$

where

$$S = \sum_{a\in A}\sum_{b\in A} \frac{(|a|,|b|)}{\max(|a|,|b|)}.$$

Proof. We may assume that all members of A are positive, since otherwise we can replace each $a \in A$ by $-a$ and replace γ by $1-\eta$ to deduce the result from the positive case. For each $a \in A$ let Z_a denote the indicator random variable whose value of 1 iff $\{\gamma a\} \le t$, and define $Z = \sum_{a\in A} Z_a$. Clearly, the expectation of each Z_a is t and hence the expected value of Z is $E(Z) = nt$.

By Lemma 3.5, for each $a, b \in A$ the expectation of the product $Z_a Z_b$ is

$$t^2 + \frac{(a,b)^2}{ab}\Delta(\{a't\},\{b't\}).$$

We estimate Δ by $t\min(a',b')$ and conclude that

$$E(Z_a Z_b) \le t^2 + t\frac{(a,b)}{\max(a,b)}.$$

Hence, the variance of Z satisfies

$$\text{VAR(Z)} = E(Z^2) - (E(Z))^2 = \sum_{a\in A}\sum_{b\in A} E(Z_a Z_b) - n^2t^2 \leq tS.$$

By Chebyshev's inequality it thus follows that

$$Prob(f(\gamma) > t) = Prob(Z = 0) \leq \frac{\text{VAR(Z)}}{(E(Z))^2} \leq \frac{S}{n^2 t},$$

as needed. □

The next task is to bound S. We first need the following simple lemma.

Lemma 3.7 *Let a/b be a reduced fraction, and let r and s be positive integers. Then the number of solutions of the equation*

$$\frac{x_1 x_2 \ldots x_s}{y_1 y_2 \ldots y_s} = \frac{a}{b}, \tag{2}$$

with x_i, y_j integers $|x_i| \leq r$ and $0 < y_j \leq r$ for all i, j is at most $2^{s-1}(r(1+\log r)^{s-1})^s$.

Proof. The number of possible sign-patterns of the numbers x_i is clearly 2^{s-1} and hence we restrict our attention to bounding the number $M_{a,b}$ defined as the number of solutions of (2) in which x_i, y_j are positive integers which do not exceed r. For any integer m, let $\tau_{r,s}(m)$ be the number of solutions of

$$m = x_1 x_2 \ldots x_s, \qquad 1 \leq x_i \leq r.$$

If (2) holds then $x_1 \ldots x_s = aX$ and $y_1 \ldots y_s = bX$ for some integer X and hence, by Cauchy Schwartz

$$M_{a,b} = \sum_{X\geq 1} \tau_{r,s}(aX)\tau_{r,s}(bX) \leq (\sum_{X\geq 1} \tau^2_{r,s}(aX))^{1/2}(\sum_{X\geq 1} \tau^2_{r,s}(bX))^{1/2} \leq \sum_{X\geq 1} \tau^2_{r,s}(X) = M_{1,1}.$$

However,

$$M_{1,1} = \sum_{x_1,\ldots,x_s=1}^{r} \tau_{r,s}(x_1\ldots x_s) \leq \sum_{x_1,\ldots,x_s=1}^{r} \tau_{r,s}(x_1)\ldots\tau_{r,s}(x_s) = (\sum_{x=1}^{r} \tau_{r,s}(x))^s.$$

Clearly

$$\sum_{x=1}^{r} \tau_{r,s}(x) = \sum_{x_1,\ldots,x_{s-1}=1}^{r} \lfloor \frac{r}{\prod_{i=1}^{s-1} x_i} \rfloor \leq r(\sum_{i=1}^{r} 1/i)^{s-1} \leq r(1+\log r)^{s-1}.$$

Therefore, $M_{1,1} \leq (r(1+\log r)^{s-1})^s$ and the desired result follows. □

In the proof of the next lemma we need some graph theoretic arguments. A *walk* of length s in an undirected, simple graph $G = (V, E)$ is a sequence $v_0, v_1, \ldots, v_s$ of (not necessarily distinct) vertices of G such that $v_{i-1}v_i \in E$ for all $1 \leq i \leq s$. The following lemma is proved in [10] using some linear algebra tools. (The proof for even values of s is attributed in [10] to Godsil; the general case follows from an earlier result in linear algebra, first proved in [7].) Here we present a more elementary proof of the same result.

Lemma 3.8 ([10]) *The number of walks of length s in any graph $G = (V, E)$ on n vertices and e edges with average degree $d = 2e/n$ is at least nd^s.*

Proof. Let $F(n, e)$ denote the minimum possible number of walks of length s in a graph of n vertices and e edges. We first prove the weaker estimate

$$F(n,e) \geq \left(\frac{e}{2n}\right)^s n,$$

4^{-s} times the claimed bound. We do this by induction on n; for $n \leq 2$ this is trivially true.

Let δ be the minimum degree. If $\delta \geq d/4$, the claim is obvious.

Assume $\delta < d/4 = e/(2n)$. Omit a vertex of minimum degree from the graph. By the induction hypothesis we have

$$F(n,e) \geq F(n-1, e-\delta) \geq (n-1)\left(\frac{e-\delta}{2(n-1)}\right)^s \geq n\left(\frac{e}{2n}\right)^s.$$

To show the last inequality, rearrange it as

$$\left(1 - \frac{\delta}{e}\right)^{s/(s-1)} \geq 1 - \frac{1}{n}.$$

Since $s \geq 2$, the left side is at least

$$\left(1 - \frac{\delta}{e}\right)^2 \geq \left(1 - \frac{1}{2n}\right)^2 > 1 - 1/n.$$

Now we get rid of the 4^s term. Consider the m'th direct power of our graph, that is, the graph whose vertex set is the Cartesian product of m copies of V in which the vertices $(v_1, v_2, \ldots, v_m)$ and $(u_1, u_2, \ldots, u_m)$ are connected iff $v_i u_i \in E$ for all $1 \leq i \leq m$. This graph has n^m vertices, average degree d^m ($2^{m-1}e^m$ edges), and p^m walks of length s if the original graph had p. Thus

$$p^m \geq n^m (d^m/4)^s.$$

Taking the m'th root and making $m \to \infty$ we get the desired result. $\square$

Remark: An *embedding* of a graph $H = (U, F)$ in a graph $G = (V, E)$ is a (not necessarily injective) mapping $f : U \mapsto V$ such that for every edge uv of H, $f(u)f(v)$ is an edge of G. The above proof easily extends and implies that for every tree H with $s + 1$ vertices, and for every graph G with n vertices and average degree d, the number of embeddings of H in G is at least $d^s n$. This was first proved by Sidorenko [15] (see also [12]) using a different method.

Using the last lemma we next prove the following,

Lemma 3.9 *Let $u_1 < u_2 \ldots < u_n$ be an arbitrary set of n nonzero integers and let r be a positive integer. Then the number of pairs $i \leq j$ for which in the expression of u_i/u_j as a reduced fraction a/b with $b > 0$ satisfies $|a|, b \leq r$ is at most*

$$n^{1+1/s} r(1 + \log r)^{s-1},$$

for any positive integer s.

Proof. Define a graph $G = (V, E)$ on the set of vertices $V = \{1, \ldots, n\}$, in which the pair u_i, u_j forms an edge iff in the expression of u_i/u_j as a reduced fraction a/b with $b > 0$, both $|a|$ and b are bounded by r. Let $d = 2|E|/n$ denote the average degree of G. By Lemma 3.8 there are at least nd^s walks of length s in G, and hence at least d^s/n of them starting at the same vertex, say $i = i_0$, and ending at the same vertex, say j. Suppose $u_j/u_i = a/b$, where a/b is reduced. Every walk $i_0 i_1 \ldots i_s$ ending at j defines a solution of the equation (2) by letting x_q/y_q denote the expression of u_q/u_{q-1} as a reduced fraction with positive denominator. By the definition of the graph G, each $|x_q|$ and y_q is bounded by r, and different walks supply different solutions. Hence, by Lemma 3.7, the number of walks is at most $2^{s-1}(r(1 + \log r)^{s-1})^s$. Therefore

$$d^s/n \leq 2^{s-1}(r(1 + \log r)^{s-1})^s,$$

implying the assertion of the lemma. □

Lemma 3.10 *There exists an absolute positive constant c so that the following holds. Let $u_1 < u_2 \ldots < u_n$ be an arbitrary set of n nonzero integers, and define*

$$S = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(|u_i|, |u_j|)}{max\{|u_i|, |u_j|\}}.$$

Then

$$S \leq n e^{c\sqrt{\log n \log \log n}}.$$

Proof. For all r,$1 \leq r \leq n$, let n_r denote the number of ordered pairs (u_i, u_j) for which

$$\frac{(|u_i|, |u_j|)}{max\{|u_i|, |u_j|\}} \geq 1/r.$$

It is easy to see the last fraction is at least $1/r$ iff in the expression of u_i/u_j as a reduced fraction a/b with positive denominator, both $|a|$ and b are at most r. Therefore, by Lemma 3.9,

$$n_r \leq 2n^{1+1/s} r(1 + \log r)^{s-1}$$

for every positive integer s. Put $n_0 = 0$ and observe that the total contribution of the pairs (u_i, u_j) for which $\frac{(|u_i|,|u_j|)}{max\{|u_i|,|u_j|\}} < 1/n$ to the sum S does not exceed n. Therefore, for every $s \geq 1$

$$S \leq n + \sum_{r=1}^{n} \frac{1}{r}(n_r - n_{r-1}) = n + \frac{1}{n+1} n_n + \sum_{r=1}^{n} n_r(\frac{1}{r} - \frac{1}{r+1})$$

$$\leq 2n + \sum_{r=1}^{n} 4n^{1+1/s} r(1+\log r)^{s-1} \frac{1}{r(r+1)} \leq 2n + 4n^{1+1/s}(1+\log n)^s.$$

Taking $s = \lfloor \sqrt{\frac{\log n}{\log\log n}} \rfloor$ we conclude that

$$S \leq n e^{(2+o(1))\sqrt{\log n \log\log n}},$$

implying the desired result. $\square$

Remark: The assertion of the last lemma is nearly tight, as there are n distinct nonzero integers $u_1, \ldots, u_n$ for which the sum S is at least $ne^{c'\sqrt{\log n/\log\log n}}$. To see this let $k = \lfloor \log_2 n \rfloor$, let $2 = p_1 < p_2 \ldots < p_k$ be the k smallest primes and let the numbers u_i contain all numbers of the form $\prod_{i=1}^{k} p_i^{\epsilon_i}$, where each $\epsilon_i \in \{0,1\}$. The number of ordered pairs of products u_i and u_j as above, in which each product contains precisely l primes not in the other one, is

$$\binom{k}{l}\binom{k-l}{l} 2^{k-2l}.$$

For each such product, $(u_i, u_j)/max\{u_i, u_j\}$ is at least the reciprocal of the product of the largest l primes among $p_1, \ldots, p_k$, which is at least, say, $1/(2k\log k)^l$. It follows that for the above set of integers

$$S \geq \binom{k}{l}\binom{k-l}{l} 2^{k-2l}(2k\log k)^{-l} \geq (\frac{k-l}{l})^{2l} 2^{k-2l}(2k\log k)^{-l} = n\frac{(k-l)^{2l}}{l^{2l}2^{3l}(k\log k)^l}.$$

Taking, e.g., $l = \lfloor 0.1 \frac{\sqrt{k}}{\sqrt{\log k}} \rfloor$ the desired estimate follows.

Returning to the proof of the main result observe, now, that the assertion of Lemma 3.3 follows by applying Lemma 3.6 and Lemma 3.10 (where we consider here either the subset of all positive members or the subset of all negative members of A). The proof of Lemma 3.2 follows easily from Lemma 3.3; indeed

$$E(f(\gamma)) \leq \frac{1}{n} Prob(f(\gamma) \leq 1/n) + \sum_{i=1}^{\lceil \log_2 n \rceil} \frac{2^i}{n} Prob(\frac{2^{i-1}}{n} < f(\gamma) \leq \frac{2^i}{n})$$

$$\leq \frac{1}{n} + \sum_{i=1}^{\lceil \log_2 n \rceil} \frac{2^i}{n} \frac{e^{c'\sqrt{\log n \log\log n}}}{2^{i-1}} \leq \frac{1}{n} + \frac{2\lceil \log_2 n \rceil}{n} e^{c'\sqrt{\log n \log\log n}},$$

implying the desired result. The assertion of Proposition 2.1 and that of Theorem 1.3 follow. □

Acknowledgment. We would like to thanks F. Chung, R. Graham and an anonymous referee for helpful comments.

References

[1] H. L. Abbott, *On a conjecture of Erdős and Straus on non-averaging sets of integers*, Proc. of the Fifth British Combinatorial Conference, Congressus Numerartium XV (1975), 1-4.

[2] H. L. Abbott, *Extremal problems on non-averaging and non-dividing sets*, Pacific J. Math. 91 (1980), 1-12.

[3] H. L. Abbott, *On the Erdős-Straus non-averaging set problem*, Acta Math. Hung. 47 (1986), 117-119.

[4] H. L. Abbott, *On non-averaging sets of integers*, Acta Math. Acad. Sci. Hungar. 40 (1982), 197-200.

[5] N. Alon and Y. Peres, *Uniform dilations*, Geometric and Functional Analysis 2 (1992), 1-28.

[6] F. A. Behrend, *On sets of integers which contain no three terms in arithmetic progression*, Proc. Nat. Acad. Sci. USA 32 (1946), 331-332.

[7] G. R. Blakley and P. Roy, *A Hölder type inequality for symmetric matrices with nonnegative entries*, Proc. AMS 16 (1965), 1244-1245.

[8] A. P. Bosznay, *On the lower estimation of non-averaging sets*, Acta Math. Hungar. 53 (1989), 155-157.

[9] P. Erdős and A. Sárközy, *On a problem of Straus*, in: Disorder in Physical Systems, Oxford Univ. Press, New York (1990), 55-66.

[10] P. Erdős and M. Simonovits, *Compactness results in extremal graph theory*, Combinatorica 2 (1982), 275-288.

[11] P. Erdős and E. G. Straus, *Non-averaging sets II*, Combinatorial Theory and its Applications, Vol. II, Colloquia Mathematica Societatis János Bolyai 4 (1970), 405-411.

[12] C. Jagger, P. Šťovíček and A. Thomason, *Multiplicities of subgraphs*, Combinatorica 16 (1996), 123-141.

[13] J. Komlós, M. Sulyok and E. Szemerédi, *Linear problems in combinatorial number theory*, Acta Math. Acad. Sci. Hungar. 26 (1975), 113-121.

[14] I. Ruzsa, *Solving a linear equation in a set of integers, II*, Acta Arithmetica 72 (1995), 385-397.

[15] A. F. Sidorenko, *Extremal problems in graph theory and inequalities in functional analysis*, Proc. Soviet Seminar on Discrete Math. Appl., Moscow Univ. Press, 1986, 99-105 (in Russian.)

[16] E. G. Straus, *Non-averaging sets*, Proc. of Symposia in Pure Mathematics XIX, AMS (1971), 215-222.