Large sets in finite fields are sumsets

Noga Alon *

Abstract

For a prime p, a subset S of Z_p is a *sumset* if $S = A + A$ for some $A \subset Z_p$. Let $f(p)$ denote the maximum integer so that every subset $S \subset Z_p$ of size at least $p - f(p)$ is a sumset. The question of determining or estimating $f(p)$ was raised by Green. He showed that for all sufficiently large p, $f(p) \geq \frac{1}{9} \log_2 p$ and proved, with Gowers, that $f(p) < cp^{2/3} \log^{1/3} p$ for some absolute constant c. Here we improve these estimates, showing that there are two absolute positive constants c_1, c_2 so that for all sufficiently large p,

$$c_1 \frac{\sqrt{p}}{\sqrt{\log p}} \leq f(p) < c_2 \frac{p^{2/3}}{\log^{1/3} p}.$$

The proofs combine probabilistic arguments with spectral techniques.

1 The main result

For a prime number $p > 2$, let Z_p denote the abelian group with p elements. A subset S of Z_p is called a *sumset* if there is a set $A \subset Z_p$ so that $A + A = \{a_1 + a_2 : a_1, a_2 \in A\} = S$. Trivially, Z_p itself is a sumset, as $Z_p = Z_p + Z_p$ (and in fact $Z_p = A + A$ for many other choices of $A \subset Z_p$.) It is also not difficult to check that if $p \geq 7$, then every subset $S \subset Z_p$ of cardinality $p - 1$ is a sumset. Indeed, if $S = Z_p - \{x\}$ then $S = A + A$ for

$$A = \{ \frac{x+2}{2}, \frac{x+4}{2}, \ldots, \frac{x+p-5}{2}, \frac{x+p-3}{2}, \frac{x+p+1}{2} \},$$

(where the operations are modulo p). Similarly, for $p \geq 3$ every $S \subset Z_p$ of cardinality $p - 2$ is also a sumset. To see this, note that $S = Z_p - \{0, 1\}$ is $A + A$ for $A = \{1, 2, \ldots, \frac{p-1}{2}\}$, and every set of size $p - 2$ is an affine image of this S. Ben Green (see [5], [4]) showed that if p is large then every subset S of Z_p that consists of nearly all elements is a sumset. Let $f(p)$ denote the maximum integer f so

*Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by a USA-Israel BSF grant, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

that every $S \subset Z_p$ of size at least $p - f$ is a sumset. In this notation, Green proved that there is a p_0 so that for all $p > p_0$

$$f(p) \geq \frac{1}{9} \log_2 p,$$

and further proved, with Gowers, (see [5]) that there is an absolute constant c so that for all p

$$f(p) \leq c\, p^{2/3} \log^{1/3} p.$$

Green raised the problem of determining or estimating this function more accurately. In this note we improve the above estimates, as follows.

Theorem 1.1 *There exist two positive constants c_1, c_2 and an integer p_0 so that for all $p > p_0$*

$$c_1 \frac{\sqrt{p}}{\sqrt{\log p}} \leq f(p) < c_2 \frac{p^{2/3}}{\log^{1/3} p}.$$

Therefore, for large p, every subset of cardinality at least $p - c_1 \frac{\sqrt{p}}{\sqrt{\log p}}$ of Z_p is a sumset, whereas there exists a subset of cardinality at least $p - c_2 \frac{p^{2/3}}{\log^{1/3} p}$ which is not a sumset. We suspect that $f(p) = p^{1/2+o(1)}$, where the $o(1)$-term tends to zero as p tends to infinity.

The rest of this note is organized as follows. In Section 2 we prove the lower bound in Theorem 1.1 using probabilistic arguments. Section 3 contains the proof of the upper bound, which is based on some properties of random Cayley sum graphs, derived from their spectral properties. Section 4 contains several comments on variants and extensions of the main result, that support the belief that the correct asymptotic behaviour of $f(p)$ is $p^{1/2+o(1)}$. Throughout the paper, we make no attempt to optimize the absolute constants and omit all floor and ceiling signs whenever these are not crucial. All logarithms are in the natural base e, unless otherwise specified.

2 Large sets are sumsets

In this section we prove the following result, which provides the lower bound in Theorem 1.1.

Theorem 2.1 *Let p be a large prime, and let $F \subset Z_p$ satisfy $|F| \leq \frac{1}{4000}\sqrt{\frac{p}{\log p}}$. Then there exists an $A \subset Z_p$ so that $Z_p - F = A + A$, that is, $S = Z_p - F$ is a sumset.*

Proof: Throughout the proof we assume, whenever this is needed, that p is sufficiently large. Given F as above, consider the following procedure for generating a random set $A \subset Z_p$. Set $P = \frac{10\sqrt{\log p}}{\sqrt{p}}$. Let $C \subset Z_p$ be a random set obtained by letting each $y \in Z_p$ for which $2y \notin F$ be a member of C with probability P, where all choices are independent. Now define

$$A = C - \{y \in C \ : \ \exists z \in C \text{ so that } y + z \in F \ \}.$$

By definition, $A+A$ contains no element of F. (Note that to ensure this property it suffices to omit, while defining A using C, only one element from each pair of members of C whose sum lies in F, but since we are not optimizing the constants here, we actually omit both.) In order to complete the proof it suffices to show that with positive probability $A+A$ does contain all elements of S. To do so, it suffices to show that for each fixed $g \notin F$, the probability that g does not lie in $A+A$ is less than $1/p$. We thus fix $g \notin F$ and prove three simple lemmas leading to the required conclusion.

Lemma 2.2 *The probability that there are less than* $17 \log p$ *disjoint pairs* $y_i, z_i \in C$ *so that* $y_i + z_i = g$ *is* $o(\frac{1}{p})$.

Proof: Fix a set of $p/4$ pairs y, z in Z_p satisfying $y + z = g$ and $2y, 2z \notin F$. (Since p is large and $|F| < \sqrt{p}$ there are certainly that many pairs, recalling our convention of omitting floor and ceiling signs). The number of these pairs that lie in C is a binomial random variable with parameters $p/4$ and $P^2 = \frac{100 \log p}{p}$, and hence its expectation is $25 \log p$. By the known standard estimates for binomial distributions (c.f., e.g., [3], Theorem A.1.13, page 268), the probability that there are less than $17 \log p$ pairs in C is at most $e^{-8^2 \log^2 p/(50 \log p)} = o(\frac{1}{p})$, as needed. □

Lemma 2.3 *The probability that* C *contains 9 distinct elements* $x, y_1, z_1, y_2, z_2, \ldots, y_4, z_4$ *so that for all* $1 \leq i \leq 4$, $x + y_i \in F$ *and* $y_i + z_i = g$ *is* $o(\frac{1}{p})$.

Proof: There are p choices for x, then less than $|F|^4$ choices for $x + y_1, x + y_2, x + y_3, x + y_4 \in F$ and this determines the elements y_i as well as z_i since $y_i + z_i = g$ for all i. Thus, the probability that there is such a configuration of 9 elements in C is less than

$$p|F|^4 P^9 \leq p\Big(\frac{1}{4000}\sqrt{\frac{p}{\log p}}\Big)^4 \Big(\frac{10\sqrt{\log p}}{\sqrt{p}}\Big)^9 = o(\frac{1}{p}),$$

as required. □

Lemma 2.4 *The probability that* C *contains at least* $s = 1.7 \log p$ *pairwise disjoint triples* x_i, y_i, z_i *so that for all* i, $x_i + y_i \in F$ *and* $y_i + z_i = g$ *is* $o(\frac{1}{p})$.

Proof: The number of triples x, y, z in Z_p so that $x + y \in F$ and $y + z = g$ is at most $p|F|$, hence the probability that C contains s pairwise disjoint triples of this type is at most

$$\binom{p|F|}{s} P^{3s} \leq (\frac{ep|F|P^3}{s})^s \leq (\frac{ep\sqrt{p}}{4000\sqrt{\log p}} \frac{1000(\log p)^{3/2}}{p^{3/2}s})^s = (\frac{e \cdot 1000}{1.7 \cdot 4000})^{1.7 \log p} = o(\frac{1}{p}),$$

as claimed. □

Returning to the proof of Theorem 2.1, we next observe that the three lemmas together imply that the probability that g does not lie in $A + A$ is $o(\frac{1}{p})$. Indeed, with probability at least $1 - o(\frac{1}{p})$ none of the events described in the three lemmas happens. We claim that in this case g does lie in $A + A$. To prove this claim note that as the event in Lemma 2.2 does not happen, there are at

least $17 \log p$ disjoint pairs y_i, z_i of elements in C so that $y_i + z_i = g$. If this is the case and yet g is not in $A + A$, it means that at least one element in each of these pairs has been omitted in the process of generating A from C. Therefore, for each of the pairs y_i, z_i there is some element $x_i \in C$ so that (possibly after swapping y_i and z_i), $x_i + y_i \in F$, where, by the construction of C, $x_i \neq y_i, z_i$. However, the same element of Z_p cannot serve as x_i for more than 3 such pairs, as the event in Lemma 2.3 fails to occur. It follows that every triple $\{x_i, y_i, z_i\}$ can intersect at most 9 other triples of this kind (as each of its elements can serve as x_j at most 3 times and as y_j or z_j at most once). Hence there is a set of at least $\frac{17 \log p}{10} > 1.7 \log p$ pairwise disjoint triples $\{x_i, y_i, z_i\}$ of elements of C, so that for all i, $x_i + y_i \in F$ and $y_i + z_i = g$, contradicting the assumption that the event in Lemma 2.4 does not occur. This proves the claim, and completes the proof of the theorem. □

3 Sets which are not sumsets

This section contains the proof of the upper bound in Theorem 1.1. There is a somewhat simpler variant of this proof, but as it provides a weaker bound by a $\log p$ factor, we prefer the proof below. The basic idea is simple. Our objective is to prove the existence of a set $F \subset Z_p$ of size $|F| \leq O(\frac{p^{2/3}}{\log^{1/3} p})$ so that $S = Z_p - F$ is not a sumset. It is convenient to choose F as the disjoint union of two sets, $F = T \cup T'$, where each of the sets T, T' is of size $O(\frac{p^{2/3}}{\log^{1/3} p})$. The set T is chosen first. Once it is chosen, it already follows that, for any T', if $S = Z_p - (T \cup T')$ is a sumset of the form $A + A$, then $A + A$ cannot intersect T. This gives an upper bound for the number of sets A which are possible candidates. It is convenient to establish this bound using an appropriate Cayley sum graph. Let $G = G(Z_p, T)$ be the graph whose vertices are all elements of Z_p, where two are adjacent iff their sum lies in T. Note that A satisfies $(A + A) \cap T = \emptyset$ iff A is an independent set in G. Therefore, we need to bound the number of independent sets in G. This can be done using the eigenvalues of G. These eigenvalues can be expressed in terms of the characters of Z_p, and the resulting character sums for a random choice of T can be estimated using standard large deviation inequalities. This enables us to prove the existence of a set T of the required size, so that all nontrivial eigenvalues of $G = G(Z_p, T)$ are, in absolute value, at most $O(\sqrt{|T| \log p})$. In [2] the authors obtain an upper estimate for the number of independent sets in a regular graph provided all its nontrivial eigenvalues are, in absolute value, considerably smaller than its degree. Using this bound we get that the number of possible sets A is at most

$$I = e^{O(p^{2/3} \log^{2/3} p)}.$$

We can now choose an additional set T' of size $t' = c\frac{p^{2/3}}{\log^{1/3} p}$ to define the final set $S = Z_p - (T \cup T')$. There are $\binom{p-|T|}{t'}$ possible choices for T', and for an appropriately large constant c, this number exceeds I. Hence there are more sets S than potential sets A, and thus at least one of these sets S is not a sumset, implying the desired result. The detailed proof is described in the rest of this section.

3.1 Cayley sum graphs and their spectra

For an abelian group B and a subset $T \subset B$, the *Cayley sum graph* $G = G(B,T)$ of B with respect to T is the graph whose set of vertices is B, in which yz is an edge for each $y, z \in B$ satisfying $y + z \in T$. Clearly, this is a $|T|$-regular graph (which may contain up to one loop at each vertex). Let D be the adjacency matrix of G. It is well known that the eigenvalues of D (which are also called the eigenvalues of G) can be expressed in terms of T and the characters of B. Indeed, for every character χ of B and every $y \in B$, the entry indexed by y of the product of D with the vector $\chi = \chi(z)_{z \in B}$ is $\sum_{s \in T} \chi(s-y) = (\sum_{s \in T} \chi(s))\overline{\chi(y)}$. Applying D again, it follows that the entry indexed by y of $D^2\chi$ is $|\sum_{s \in T} \chi(s)|^2 \chi(y)$. Therefore, the eigenvalues of the symmetric matrix D^2 are precisely the expressions $|\sum_{s \in T} \chi(s)|^2$, where the characters are the corresponding eigenvectors, and as the characters are orthogonal, these are all eigenvalues. It follows that each nontrivial eigenvalue of the graph $G = G(B,T)$ is, in absolute value, $|\sum_{s \in T} \chi(s)|$ for some nontrivial character χ of B (it is not difficult to determine the signs as well, but these are not needed here). In particular, if p is a prime and $T \subset Z_p$, then every nontrivial eigenvalue of the Cayley graph of Z_p with respect to T is, in absolute value, $|\sum_{s \in T} \omega^s|$, where ω is a nontrivial p-th root of unity.

A (p,t,λ)-graph is a t-regular graph on p vertices, in which the absolute value of each nontrivial eigenvalue is at most λ. This notion was introduced by the author in the 80s, motivated by the observation that such graphs in which λ is much smaller than t exhibit strong pseudo-random properties.

An old result of Hoffman [7] implies that the maximum independent set in any (p,t,λ)-graph is of size at most $\frac{p\lambda}{t+\lambda}$. Note that this already suffices to provide a nontrivial bound for the number of independent sets in such a graph, namely, the number of all subsets of vertices of cardinality at most the above bound. One can give, however, a better upper bound, as proved in [2].

Lemma 3.1 ([2]) *Let G be a (p,t,λ)-graph, and suppose $m \geq \frac{2p \log p}{t}$. Then, the number of independent sets of size m in G is at most*

$$\Big(\frac{emt^2}{4\lambda p \log p}\Big)^{\frac{2p \log p}{t}} \Big(\frac{2e\lambda p}{mt}\Big)^m.$$

3.2 Cayley sum graphs with small nontrivial eigenvalues

In this subsection we prove the existence of a Cayley sum graph $G = G(Z_p, T)$ with a relatively small number of independent sets. We need the following.

Lemma 3.2 *For every integer $t \leq p^{2/3}$ there exists a subset $T \subset Z_p$ of cardinality t so that for every nontrivial $p-th$ root of unity ω*

$$|\sum_{s \in T} \omega^s| \leq 3\sqrt{t}\sqrt{\log(10p)}.$$

Proof: Let $(s_1, s_2, \ldots, s_t)$ be a random sequence of not necessarily distinct elements of Z_p obtained by choosing each s_i randomly, uniformly and independently among all elements of Z_p. For each fixed nontrivial root of unity ω, the real part of the sum $\sum_{i=1}^{t} \omega^{s_i}$ is a sum of t mutually independent random variables, each having expectation 0 and each bounded in absolute value by 1. Therefore, by a standard Chernoff type inequality (c.f., e.g., [3], Theorem A.1.16, page 269), the probability that the absolute value of such a sum exceeds a is bounded by $2e^{-a^2/(2t)}$. Setting $a = \sqrt{2t}\sqrt{\log(10p)}$ we conclude that the probability that the absolute value of the real part of the sum exceeds a is at most $\frac{2}{10p}$. As the same reasoning applies to the imaginary part of the sum, it follows that with probability at least $1 - 2(p-1)\frac{2}{10p} > 0.6$ all these absolute values are at most a. The expected number of pairs of equal elements in the sequence is $\binom{t}{2}/p < \frac{t^2}{2p} \leq \frac{\sqrt{t}}{2}$, where here we used the fact that $t \leq p^{2/3}$. Thus, by Markov's Inequality, with probability at least 0.5 there are at most $\sqrt{t}$ such pairs. Hence, with positive probability, there are at most $\sqrt{t}$ such pairs, and all real and imaginary parts of all the sums above are, in absolute value, at most a, implying that the absolute value of each of these sums is at most $\sqrt{2}a = 2\sqrt{t}\sqrt{\log(10p)}$. Fix a sequence satisfying these properties, omit an element from each pair of equal elements, and replace the omitted elements by arbitrary other elements of Z_p. This yields a set T of t distinct elements of Z_p. By the triangle inequality, for each nontrivial $p-th$ root of unity ω

$$|\sum_{s \in T} \omega^s| \leq 2\sqrt{t}\sqrt{\log(10p)} + 2\sqrt{t} \leq 3\sqrt{t}\sqrt{\log(10p)}.$$

This completes the proof. □

Corollary 3.3 *There exists a Cayley sum graph $G = G(Z_p, T)$ with $|T| = t = 9\frac{p^{2/3}}{\log^{1/3} p}$, that has at most*

$$e^{(2+o(1))p^{2/3}\log^{2/3} p}$$

independent sets.

Proof: Put $t = 9\frac{p^{2/3}}{\log^{1/3} p}$, and let $T \subset Z_p$ be a set of cardinality t satisfying the assertion of Lemma 3.2. Let $G = G(Z_p, T)$ be the corresponding Cayley graph. By the discussion in subsection 3.1 and by Lemma 3.2, G is a (p, t, λ)-graph with $\lambda = 3\sqrt{t}\sqrt{\log(10p)}$. By Lemma 3.1, the number of independent sets of cardinality m in G, for each $m \geq \frac{2p\log p}{t}$, is at most

$$\Big(\frac{emt^2}{4\lambda p \log p}\Big)^{\frac{2p\log p}{t}} \Big(\frac{2e\lambda p}{mt}\Big)^m \leq p^{O(p^{1/3}\log^{4/3} p)} \Big(\frac{2e3 \cdot 3p^{1/3}\sqrt{\log(10p)}p\log^{1/3} p}{m9(\log^{1/6} p)p^{2/3}}\Big)^m = e^{O(p^{1/3}\log^{7/3} p)} F_p(m),$$

where

$$F_p(m) = \Big(\frac{(2e + o(1))p^{2/3}\log^{2/3} p}{m}\Big)^m.$$

Differentiating the logarithm of the function $f(x) = (r/x)^x$, where r is a positive real, it is easy to check that its maximum for $x > 0$ is $e^{r/e}$, attained at $x = r/e$. It follows that for every $m \geq \frac{2p \log p}{t}$,

$$F_p(m) \leq e^{(2+o(1))p^{2/3} \log^{2/3} p}.$$

Summing over all values of m (and observing that the number of independent sets of size $m \leq \frac{2p \log p}{t} = O(p^{1/3} \log^{4/3} p)$ is trivially at most $p^{O(p^{1/3} \log^{4/3} p)}$), the desired result follows. □

3.3 The proof of the upper bound

The following result implies the upper bound in Theorem 1.1.

Theorem 3.4 *For all sufficiently large p there exists an $F \subset Z_p$ of cardinality $16\frac{p^{2/3}}{\log^{1/3} p}$ so that $S = Z_p - F$ is not a sumset.*

Proof: Let $T \subset Z_p$, $G = G(Z_p, T)$ and t be as in Corollary 3.3. Put $t' = 7\frac{p^{2/3}}{\log^{1/3} p}$. There are

$$\binom{p-t}{t'} = e^{(7/3-o(1))p^{2/3} \log^{2/3} p}$$

subsets T' of cardinality t' in $Z_p - T$. As this number exceeds the number of independent sets A in G, it follows that there exists such a set T' so that there is no independent set A in G for which $A+A = Z_p-(T \cup T')$. As explained in the beginning of the section this implies that $S = Z_p-(T \cup T')$ is not a sumset, as needed. □

4 Concluding remarks

Theorem 1.1 can be extended to other abelian groups. A set S in an abelian group B is called a *sumset* if there is an $A \subset B$ so that $S = A+A$. The proof of Theorem 1.1, with essentially no changes, implies that for every finite abelian group B of **odd** order n, the maximum possible cardinality of a subset of B which is not a sumset is at least $n - O(\frac{n^{2/3}}{\log^{1/3} n})$ and at most $n - \Omega(\frac{\sqrt{n}}{\sqrt{\log n}})$. For (some) abelian groups of even order the situation is a bit different, because of the existence of elements of order 2. Consider, for example, $B = Z_2^k$, where $n = 2^k$. Here every element is of order 2, and hence every nonempty sumset must contain 0. The proof of Theorem 2.1, with essentially no change, implies that for every set F of at most $\frac{1}{4000}\frac{\sqrt{n}}{\sqrt{\log n}}$ **nonzero** elements of Z_2^k, $Z_2^k - F$ is a sumset. Moreover, here it is easy to see that the exponent $1/2$ cannot be improved. Indeed, let F consist of all nonzero vectors of $B = Z_2^k$ in which the first $\lfloor k/2 \rfloor$ coordinates are 0. For every set A of more than $2^{\lfloor k/2 \rfloor}$ members of B, $A+A$ must contain a member of F, by the pigeonhole principle. However, if $B - F$ is a sumset of the form $A + A$, then $|A|(|A| - 1)/2 + 1 \geq |A + A| \geq 2^k - |F|$ implying that A must be of size exceeding $2^{\lfloor k/2 \rfloor}$. Thus $B - F$ is not a sumset.

Call a subset S of an abelian group B a *diffset* if $S = A - A$ for some $A \subset B$. (We do not use the term *difference set* as this has another meaning). Clearly every nonempty diffset must satisfy $0 \in S$ and $S = -S$. The proof of Theorem 2.1 can be easily modified to prove the following.

Theorem 4.1 *There exists an absolute positive constant c so that for every abelian group B of order n and every subset $S \subset B$ satisfying $|S| \geq n - c\frac{\sqrt{n}}{\sqrt{\log n}}$, $0 \in S$ and $S = -S$, S is a diffset.*

Here, too, the exponent $1/2$ cannot be improved, as stated in the following simple claim.

Proposition 4.2 *Every abelian group B of order n contains a subset S satisfying $0 \in S$, $S = -S$ and $|S| \geq n - 4\sqrt{n}$ which is not a diffset.*

Proof: Suppose $B = Z_{d_1} \oplus Z_{d_2} \oplus \cdots \oplus Z_{d_k}$, with $d_1 | d_2 | d_2 \cdots | d_k$. Let i be the minimum index so that $d_1 \cdot d_2 \cdots d_i \geq \sqrt{n}$. Let F be the set of all nonzero elements of B whose representation in the above direct sum has 0 in the first $i-1$ coordinates, and has absolute value at most $\frac{d_1 d_2 \cdots d_i}{\lfloor \sqrt{n} \rfloor}$ in coordinate number i. By the pigeonhole principle, for every set A of at least $\lfloor \sqrt{n} \rfloor$ elements of B, $A - A$ must intersect F, implying that $B - F$ is not a diffset.

For two functions $f(n)$ and $g(n)$ we write $f(n) = \tilde{\Theta}(g(n))$ if there are two (not necessarily positive) reals c_1, c_2 and an integer n_0 so that $g(n)(\log n)^{c_1} \leq f(n) \leq g(n)(\log n)^{c_2}$ for all $n > n_0$. The following conjecture seems plausible.

Conjecture 4.3 *The function $f(p)$ considered in this paper satisfies $f(p) = \tilde{\Theta}(p^{1/2})$.*

A similar conjecture should hold for all abelian groups of odd order. A stronger conjecture is the following.

Conjecture 4.4 *There are constants c_1, c_2 so that the following holds. Let B be an abelian group of odd order n. Then for every integer t between 1 and n there is a subset $T \subset B$ of size t so that the independence number of the Cayley sum graph $G = G(B, T)$ is at most $c_1 \frac{n}{t} (\log n)^{c_2}$.*

This is trivial for $t \leq (\log n)^{O(1)}$ and is essentially proved in [1] (with $c_2 = 2$) for $t = \Omega(n)$, and in [6] with the optimal $c_2 = 1$ for cyclic groups and $t = \Omega(n)$. The proof in [1] deals, in fact with usual Cayley graphs, and not with sum graphs, but can be easily modified for this case as well. The proof in [6] deals with t around $n/2$ but the same proof works for a wider range of t. The validity of this conjecture for $t = \sqrt{n}(\log n)^{c_3}$ for some $c_3 > c_2$ would imply the assertion of Conjecture 4.3 (and its analog for any abelian group of odd order n) by the reasoning in the previous paragraphs. Note that the assertion of the conjecture may well hold for a random choice of T. This would mean that the independence number of random Cayley sum graphs is similar to that of random regular graphs, discussed in [8].

Acknowledgment: This research was initiated during a visit as an Aisenstadt Chair holder at the CRM, Montreal; the hospitality of my hosts at the CRM is gratefully acknowledged. I also thank Andrew Granville for helpful discussions.

References

[1] N. Alon and A. Orlitsky, Repeated communication and Ramsey graphs, IEEE Transactions on Information Theory 41 (1995), 1276-1289.

[2] N. Alon and V. Rödl, Sharp bounds for some multicolor Ramsey numbers, Combinatorica 25 (2005), 125-141.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method*, Second Edition, Wiley, New York, 2000.

[4] E. Croot and V. Lev, Open problems in additive combinatorics, to appear.

[5] B. J. Green, Essay submitted for the Smith's Prize, Cambridge University, 2001.

[6] B. J. Green, Counting sets with small sumset, and the clique number of random Cayley graphs, Combinatorica 25 (2005), 307-326.

[7] A. J. Hoffman, On eigenvalues and colorings of graphs, B. Harris Ed., Graph Theory and its Applications, Academic, New York and London 1970, 79-91.

[8] M. Krivelevich, B. Sudakov, V. H. Vu and N. Wormald, Random regular graphs of high degree, Random Structures and Algorithms 18 (2001), 346-363.