On uniformly distributed dilates of finite integer sequences

S. V. Konyagin, I. Z. Ruzsa, W. Schlag

April 20, 2011

Abstract

Given N nonzero real numbers $a_1 < \ldots < a_N$, we consider the problem of finding a real number α so that $\alpha a_1, \ldots, \alpha a_N$ are close to be uniformly distributed modulo one (this question is attributed to Komlos in [13]). Firstly, it turns out that it suffices to consider integers $a_1, \ldots, a_N$. Given various quantities that measure how close a sequence is to being uniformly distributed, e.g., the size of the largest gap between consecutive points on the circle, discrepancy, or the number of points falling into any interval of size $1/N$ ("concentration"), we provide upper bounds for the optimal dilate. These bounds depend only on N and they are attained by typical α, i.e., up to α belonging to some set of small measure. We also provide lower bounds for these quantities. Some of our examples are constructed for this purpose by means of probabilistic methods. In case of the discrepancy, the lower and upper bounds match up to logarithms ($\sqrt{N/\log N}$ vs. $\sqrt{N}\log N$). However, in case of the largest gap ($\log N/N$ vs. $N^{-1/2}$) and the concentration ($\exp(c\log N/\log\log^2 N)$ vs. $N^{1/3+\epsilon}$) the lower and upper bounds do not match and the question about the correct asymptotic behavior in terms of N remains open. Finally, we improve on a recent result of Noga Alon and the second author by showing that every set of N integers contains a non-averaging subset of size at least $N^{1/5}$.

1 Introduction

In this note we consider the following question, stated first informally:

Question 1.1 *Given a set of N distinct integers $\mathcal{A} = \{a_1, \ldots, a_N\}$, does there exist an $\alpha \in \mathbb{R}$ so that $\alpha a_1, \ldots, \alpha a_N$ is "well–distributed" on the circle modulo 1 ? In fact, can this be achieved for all α up to some set of small measure ?*

There are various ways of measuring whether a given sequence $x_1, \dots, x_N \in [0,1]$ is "well–distributed", such as: discrepancy, the size of the largest gap between any two adjacent points, and the largest number of points falling into any interval of size $\frac{1}{N}$ (we will refer to this latter quantity as the "concentration"). Recall that the discrepancy is defined to be

$$D_N(\{x_j\}_1^N) = \sup_{I \subset \mathbb{T}} \Big| \mathrm{card}\{j : x_j \in I\} - N|I| \Big|,$$

where the supremum is taken over all intervals in the torus. For each of these we seek a real number α (depending on $\mathcal{A}$) which minimizes the respective quantity with $x_j = \{\alpha a_j\}$ (here $\{\cdot\}$ denotes the fractional part). We are interested in bounds for these minima that depend *only on* N, and not on the particular choice of $\mathcal{A}$. As far as the second part of Question 1.1 is concerned, the methods discussed below show that the bounds we obtain are achieved by all α up to some set of small measure. However, it is possible that our bounds are not optimal.
Question 1.1 appears in Montgomery's book [13], see the problem section on uniform distribution. Some partial results are stated there, which are also discussed below.

It is natural to ask why we restrict ourselves to *integers* $a_1, \dots, a_N$. Indeed, (1.1) makes perfect sense for real a_j. It turns out that the integer case is the hardest. More precisely, any bound *depending only on* N that holds for integer sequences also holds for general real sequences. For example,

$$\sup_{\lambda_1 < \dots < \lambda_N \in \mathbb{R}} \inf_{\alpha \in \mathbb{R}} D_N(\{\alpha \lambda_j (\mathrm{mod}\ 1)\}_1^N) = \sup_{a_1 < \dots < a_N \in \mathbb{Z}} \min_{\alpha \in [0,1]} D_N(\{\alpha a_j (\mathrm{mod}\ 1)\}_1^N). \quad (1)$$

The underlying principle is as follows: Clearly, question (1.1) concerns some property of the orbit $\{(\alpha a_1, \dots, \alpha a_N) \in \mathbb{T}^N : \alpha \in \mathbb{R}\}$ on the N–dimensional torus. If this orbit was dense, then in particular it would come arbitrarily close to the point $(0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}) \in \mathbb{T}^N$. This would imply that

$$\inf_{\alpha \in \mathbb{R}} D_N(\{\alpha a_j (\mathrm{mod}\ 1)\}_1^N) = 1,$$

which is optimal. However, it is clear that the orbit will be periodic for integer a_j, and thus not dense. On the other hand, Kronecker's theorem implies that a generic choice (in measure) of real numbers $a_1, \dots, a_N$ will have a dense orbit. What is required for (1) and comparable statements for the largest gap and the concentration is the following lemma due to D. Campbell, H. Ferguson, and R. Forcade, see Lemma 2 in [7].

Lemma 1.2 *Let $\lambda_1, \dots, \lambda_N$ be distinct (positive) real numbers. Let $\Lambda\mathbb{R}$ denote the line in $\mathbb{R}^N$ given by $\{(\lambda_1 t, \dots, \lambda_N t) : t \in \mathbb{R}\}$. Then $\overline{\Lambda\mathbb{R} + \mathbb{Z}^N}$ contains a line generated by $(a_1, \dots, a_N)$ where all the entries a_j are distinct (positive) integers.*

Denote the right–hand side of (1) by Δ_N. Given any sequence $\lambda_1, \ldots, \lambda_N$ of distinct reals, let $a_1, \ldots, a_N$ be as in the lemma. Any point on the orbit of $(a_1, \ldots, a_N)$ that minimizes the discrepancy can be approximated arbitrarily closely by points of the orbit of $(\lambda_1, \ldots, \lambda_N)$ by Lemma 1.2. Thus the left–hand side of (1) will be no larger than Δ_N, as claimed.

This paper is organized as follows. First we consider the size of the smallest gap, then the discrepancy, the largest gap, and finally the concentration. In the last section we take the opportunity to point out an improvement of the lower bound on the size of non-averaging sets obtained recently by Alon and the second author [5]. The methods used in [5] are very closely related to some of the arguments in this paper.

2 The smallest gap

In this section we discuss the smallest gap between any two points αa_i and αa_j modulo 1. More precisely, we ask how large

$$\mu(\mathcal{A}) = \max_{\alpha} \min_{i \neq j} \|\alpha(a_i - a_j)\|$$

is in terms of N. Here $\|\cdot\|$ denotes the distance to the closest integer. It is shown below that $\mu(\mathcal{A})$ cannot be made bigger than $\asymp N^{-2}$ (the notation $A \asymp B$ will mean throughout that $c_1 A \leq B \leq c_2 A$ for suitable constants c_1 and c_2). This is already stated in [13], however without proof and without specific constants.

Proposition 2.1 *One has*

$$1 \leq \liminf_{N \to \infty} N^2 \min_{\mathrm{card}(\mathcal{A})=N} \mu(\mathcal{A}) \leq \frac{25}{9}.$$

Proof : It is easy to see that $\mu(\mathcal{A}) \geq N^{-2}$ for any set $\mathcal{A}$. Indeed, notice that

$$\int_0^1 \sum_{1 \leq i < j \leq N} \chi_{\{\|\alpha(a_i - a_j)\| \leq N^{-2}\}} \, d\alpha = \frac{1}{2} N(N-1) \, \frac{2}{N^2} < 1.$$

Hence there exists an $\alpha \in (0,1)$ such that $\|\alpha(a_i - a_j)\| > N^{-2}$ for any distinct i, j. To see that this bound can be attained, we use the well–known Singer sequences. In fact, see [10], if m is a prime power there exist integers $1 \leq b_1 < b_2 < \ldots < b_{m+1} \leq m^2 + m + 1$ so that the differences $\{b_i - b_j\}_{i \neq j}$ are all nonzero congruence classes modulo $m^2 + m + 1$. Let $N = 5(m+2)$, $D = \{0, 1, 2, 6, 9\}$, and $\mathcal{A} = \{d(m^2 + m + 1), b_1 + d(m^2 + m +$

$1), \ldots, b_{m+1} + d(m^2+m+1) : d \in D\}$. Notice that any integer between 0 and 9 can be written as difference of two elements from D. By Dirichlet's approximation theorem, for any $\alpha \in (0,1)$ we can choose $q \in [1, 9(m^2+m+1))$ so that $\|\alpha q\| \leq (9(m^2+m+1))^{-1}$. We claim that any such q can be written as the difference of two elements from $\mathcal{A}$. Clearly, $q = q' + \ell(m^2+m+1)$ where $0 \leq q' \leq m^2+m$, and $0 \leq \ell \leq 8$. If $q' = 0$, then the claim is correct in view of the aforementioned property of D. If $1 \leq q' \leq m^2+m$, then there are $i \neq j$ satisfying $q' \equiv b_i - b_j \pmod{m^2+m+1}$. Clearly, either $q' = b_i - b_j$ or $q' = m^2+m+1+b_i-b_j$. This shows that any $q \in [1, 9(m^2+m+1))$ is the difference of two elements from $\mathcal{A}$. ■

Remark : Will's conjecture (see [13], problem 44) asserts that for any integers $1 \leq n_1 < n_2 < \ldots < n_K$ one has

$$\max_{\alpha} \min_{i} \|\alpha n_i\| \geq \frac{1}{K+1}.$$

If this is true, then one can replace 1 with 2 in the lower bound above.

3 Discrepancy

Let $x_1, \ldots, x_N \in [0,1)$. The simplest fact about discrepancy is the standard inequality

$$\Big|\sum_{j=1}^{N} e^{2\pi i k x_j}\Big| \leq 2\pi k D_N(\{x_j\}_{j=1}^N) \quad \text{for} \quad k = 1, 2, \ldots. \tag{2}$$

In fact, let $H(t) = \text{card}\{j : x_j < t\} - Nt$ for $0 \leq t \leq 1$. Notice that $H(0) = H(1) = 0$. Integrating by parts one obtains

$$\begin{aligned} \Big|\sum_{j=1}^{N} e^{2\pi i k x_j}\Big| &= \Big|\int_0^1 e^{2\pi i k t} dH(t)\Big| = 2\pi \Big|\int_0^1 k e^{2\pi i k t} H(t)\, dt\Big| \\ &\leq 2\pi k \sup_{t\in[0,1]} |H(t)| \leq 2\pi k D_N(\{x_j\}_1^N). \end{aligned}$$

A much deeper converse due to Erdös and Turan [13] states that

$$D_N(\{x_j\}_i^N) \leq C\Big(\frac{N}{K} + \sum_{k=1}^{K} \frac{1}{k} |\sum_{j=1}^{N} e^{2\pi i k x_j}|\Big). \tag{3}$$

Since by Cauchy–Schwarz and Plancherel

$$\int_0^1 \Big|\sum_{j=1}^N e^{2\pi i k\alpha x_j}\Big|\, d\alpha \le \sqrt{N},$$

one immediately concludes from (3) with $K = [\sqrt{N}]$ and for $N \ge 2$ that

$$\min_\alpha D_N(\{\alpha a_j (\mathrm{mod}\ 1)\}_1^N) \le \int_0^1 D_N(\{\alpha a_j (\mathrm{mod}\ 1)\}_1^N)\, d\alpha \le C\sqrt{N}\,\log N.$$

Currently we do not know whether this can improved. However, one has the following lower bound.

Theorem 3.1 *There exists an absolute constant $c_0 > 0$ such that for large N*

$$\sup_{0<a_1<\ldots<a_N\in\mathbb{Z}} \min_\alpha D_N(\{\alpha a_j (\mathrm{mod}\ 1)\}_1^N) \ge c_0 (N/\log N)^{\frac{1}{2}}. \tag{4}$$

In fact, a random subset of $[1, 2N]$ of cardinality N will have this property with large probability.

Proof : The idea of the proof is to take a random subset of $\{1, 2, \ldots, 2N\}$ and to show that with positive probability it contains exactly N elements and its discrepancy is $> c_0(N/\log N)^{1/2}$ for some positive c_0.

Let ξ_j $(j = 1, \ldots, 2N)$ be independent random variables with $\mathbb{P}(\xi_j = 1) = \mathbb{P}(\xi_j = 0) = 1/2$. Define the random set $S = \{j : \xi_j = 1\}$. Fix $\alpha \in [0, 1]$. We will prove that for some absolute constant $c_0 > 0$

$$\mathbb{P}\Big(D_N(\{\alpha s (\mathrm{mod}\ 1)\}_{s\in S}) < c_0 (N/\log N)^{1/2}\Big) < N^{-3} \tag{5}$$

for sufficiently large N. Here we use the notation D_N also for sets of points not necessarily containing N elements. To prove (5) we proceed as follows. Let η_j have the same distribution as ξ_j, be independent of ξ_j, and define $S' = \{j : \eta_j = 1\}$. Then (5) is equivalent to

$$\begin{aligned}\mathbb{P}\Big(&D_N(\{\alpha s (\mathrm{mod}\ 1)\}_{s\in S}) < c_0 (N/\log N)^{1/2},\\ &D_N(\{\alpha s' (\mathrm{mod}\ 1)\}_{s'\in S'}) < c_0 (N/\log N)^{1/2}\Big) < N^{-6}.\end{aligned}$$

Suppose that the event

$$\begin{aligned}&D_N(\{\alpha s (\mathrm{mod}\ 1)\}_{s\in S}) < c_0 (N/\log N)^{1/2},\\ &D_N(\{\alpha s' (\mathrm{mod}\ 1)\}_{s'\in S'}) < c_0 (N/\log N)^{1/2}\end{aligned}$$

occurs. It implies that for any $u \in [0,1)$

$$\big|^{\#}\{s \in S : \alpha s (\mathrm{mod}\ 1) \le u\} - {}^{\#}\{s' \in S' : \alpha s' (\mathrm{mod}\ 1) \le u\}\big| < 2c_0(N/\log N)^{1/2}. \qquad (6)$$

Take a permutation $\{n_1, \dots, n_{2N}\}$ of the set $\{1, \dots, 2N\}$ such that the points $\{\alpha n_j\}$ are nondecreasing. Denote $\zeta_j = \xi_{n_j} - \eta_{n_j}$. Clearly, ζ_j are independent random variables with distribution $\mathbb{P}(\zeta_j = 1) = \mathbb{P}(\zeta_j = -1) = 1/4$, $\mathbb{P}(\zeta_j = 0) = 1/2$. Now (6) implies that for any $M = 1, \dots, 2N$

$$\Big|\sum_{j=1}^{M} \zeta_j\Big| < 2c_0(N/\log N)^{1/2}.$$

It is a standard fact of one–dimensional random walk that the probability of the last event is less than N^{-6} if c_0 is sufficiently small. It is easy to prove this with a somewhat wasteful constant c_0. Indeed,

$$\begin{aligned}
& \mathbb{P}\Big(\max_{1\le M\le 2N}\Big|\sum_{j=1}^{M}\zeta_j\Big| < 2c_0(N/\log N)^{\frac12}\Big) \\
\le\ & \mathbb{P}\Big(\max_{1\le M\le [\frac{2N}{\log N}]}\Big|\sum_{j=1}^{M}\zeta_j\Big| < 4c_0(N/\log N)^{\frac12}\Big)^{[\log N]} \\
\le\ & \mathbb{P}\Big(\Big|\sum_{j=1}^{[2N/\log N]}\zeta_j\Big| < 4c_0(N/\log N)^{\frac12}\Big)^{[\log N]} \le \exp(-6\log N) = N^{-6},
\end{aligned}$$

where the last line follows from the central limit theorem provided c_0 is small ($\asymp \exp(-6)$). As argued above, this implies (5). Let $\alpha_j = j/N^2$ for $j = 0, 1, \dots, N^2 - 1$. Summing (5) over all those α_j shows that with probability $> 1 - 1/N$

$$D_N(\{\alpha_j s (\mathrm{mod}\ 1)\}_{s\in S}) \ge c_0(N/\log N)^{\frac12}$$

for all $j = 0, 1, \dots, N^2 - 1$. Therefore, for any $\alpha \in [0,1]$

$$D_N(\{\alpha s (\mathrm{mod}\ 1)\}_{s\in S}) \ge c_0(N/\log N)^{1/2}) - 2. \qquad (7)$$

Furthermore, the probability that S contains exactly N elements is not too small ($\gg N^{-1/2}$) and that it contains about N elements is large. ■

As far as deterministic examples are concerned, we remark that the increase of D_N at a polynomial rate can be shown by means of a concrete example. In fact, for every N

taken from an appropriate sequence of integers tending to infinity there exist N integers for which the discrepancy of every dilate is at least $N^{0.1399}$. To see this one uses the polynomial

$$P_{12}(z) = 1 + z + z^2 + z^3 + z^4 + z^7 + z^8 + z^{10} + z^{12}.$$

In fact, it was shown in [7] that $\min_{|z|=1} |P_{12}(z)| = c_0 > 1.36$. Following C. Smyth [14] one then defines inductively

$$P_{13^k-1}(z) = P_{13^{k-1}-1}(z^{13})P_{12}(z).$$

These polynomials have $0,1$–coefficients (so called Newman polynomials) and $P_{13^k-1}(z) = \sum_{j=1}^{N_k} z^{a_j^{(k)}}$ where $a_j^{(k)}$ are all numbers that can be written in base 13 with k digits taken from $\{0,1,2,3,4,7,8,10,12\}$, cf. P_{12} above. In particular, $N_k = 9^k$. Hence

$$\min_{|z|=1} |P_{13^k-1}(z)| \geq c_0^k = N_k^{\delta},$$

where $\delta = \frac{\log c_0}{\log 9} > 0.1399$. In view of (2) with $k = 1$ this establishes the claim. We do not know whether one can improve the exponent δ by finding better Newman polynomials of some fixed (small) degree and applying the same construction as above. In view of Littlewood's conjecture on trigonometric polynomials with ± 1 coefficients, see [13], it seems reasonable to ask whether one can construct Newman polynomials $P_N(\alpha) = \sum_{j=1}^{N} e^{2\pi i \alpha a_j^{(N)}}$ so that $\min_\alpha |P_N(\alpha)| \geq C\sqrt{N}$ for some sequence $N \to \infty$. However, it seems rather hard to construct such large Newman polynomials. In fact, the following proposition shows that a generic (in an appropriate sense) Newman polynomial has minima less than one. This answers a question of Montgomery's, see [13], problem 59.

Theorem 3.2 *Let $\{\xi_j\}_{j\in\mathbb{Z}}$ be i.i.d. where $\xi_0 = 0, 1$ with probability $\frac{1}{2}$ each. For any $c > \frac{1}{2}$*

$$\mathbb{P}\Big(\min_{0\leq x\leq 1}\Big|\sum_{j=-N}^{N} \xi_j e^{2\pi i j x}\Big| < c\Big) \to 1$$

as $N \to \infty$.

Proof : This is an immediate consequence of a result of the first author [11]. In fact, it was shown there that for any *fixed* $t \in (0, \frac{1}{2})$ and $\delta > 0$

$$\mathbb{P}\Big(\min_{t\leq x\leq \frac{1}{2}}\Big|\sum_{j=-N}^{N} \pm e^{2\pi i j x}\Big| < N^{-\frac{1}{2}+\delta}\Big) \to 1 \tag{8}$$

as $N \to \infty$. Here the signs are chosen independently with probability $\frac{1}{2}$ each. Strictly speaking, [11] only contains the case $t = 0$, but it is straightforward to check that any $0 < t < \frac{1}{2}$ works (cf. page 947 in [11]). If

$$D_N(x) = \frac{\sin[(2N+1)\pi x]}{\sin(\pi x)}$$

denotes the usual Dirichlet kernel, then clearly

$$\begin{aligned} & \mathbb{P}\Big(\min_{0\le x\le 1}\Big|\sum_{j=-N}^{N}\xi_j e^{2\pi i j x}\Big| < c\Big) \\ \ge\ & \mathbb{P}\Big(\min_{t\le x\le \frac{1}{2}}\Big|\sum_{j=-N}^{N}\pm e^{2\pi i j x}\Big| < N^{-1/4} \text{ and } \max_{t\le x\le\frac{1}{2}}|D_N(x)| < 2c - N^{-1/4}\Big). \end{aligned}$$

The second condition on the right–hand side is trivially satisfied provided t is sufficiently close to $\frac{1}{2}$. Thus the left–hand side will tend to one because of (8). ■

4 The largest gap

Let $\mathcal{A}$ be an arbitrary set of N nonzero integers. The upper bound on the discrepancy from the previous section implies that for an appropriate choice of α every interval of size $\asymp N^{-\frac{1}{2}} \log N$ has to contain at least one element from $\alpha\mathcal{A}$ (mod 1). In other words, the largest gap between any two consecutive points $\alpha\mathcal{A}$ (mod 1) is no larger than $N^{-\frac{1}{2}} \log N$ (in fact, this is true for "most" choices of α). It turns out that one can remove the $\log N$ factor from this bound by very elementary means.

Proposition 4.1 *Given an arbitrary set $\mathcal{A}$ of N nonzero integers there is an $\alpha \in (0,1)$ such that the largest gap between any two consecutive elements of $\alpha\mathcal{A}$ (mod 1) is less than $2N^{-\frac{1}{2}}$.*

Proof : Let $h = N^{-1/2}$. Define the 1-periodic function

$$\psi(x) = \max(1 - |x|/h, 0) \quad \text{for} \quad |x| \le 1/2.$$

Let

$$\psi(x) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i k x}$$

be the Fourier expansion of ψ. It is easy to calculate c_k. We have $c_0 = h$, $c_k \geq 0$, $\sum_k c_k = 1$. Our aim is to prove that for some α

$$\sum_{a\in\mathcal{A}} \psi(x-\alpha a) > 0$$

for all x. Denote $T(x) = \sum_{a\in A} e^{2\pi i a x}$. Then

$$\sum_{a\in\mathcal{A}} \psi(x-\alpha a) = \sum_k c_k e^{2\pi i k x} T(-k\alpha).$$

As

$$\int_0^1 \sum_{k\neq 0} c_k |T(-k\alpha)| d\alpha \leq N^{1/2} \sum_{k\neq 0} c_k < N^{1/2},$$

there is α so that

$$\sum_{k\neq 0} c_k |T(-k\alpha)| < N^{1/2}.$$

Therefore,

$$\sum_{a\in\mathcal{A}} \psi(x-\alpha a) = hN + \sum_{k\neq 0} c_k e^{2\pi i k x} T(-k\alpha) > hN - N^{1/2},$$

as required. ■

It is possible to replace $2N^{-\frac{1}{2}}$ with $cN^{-\frac{1}{2}}$ for some $c < 2$. This can be achieved by using an appropriate substitute for the function ψ above, see [2]. It is natural to apply the second moment method to bound the largest gap. I.e, divide the circle into m intervals of length $1/m$ each and bound the probability that one of them does not contain a point in $\alpha\mathcal{A}$ via the second moment method. We leave it to the reader to check that this gives nothing better than $N^{-1/2}$ (this requires the methods from the following section). Next we show that for some (generic) sets the largest gap cannot be made any smaller than $c \log N/N$.

Proposition 4.2 *For any sufficiently large N there exist sets $\mathcal{A}$ of cardinality N so that the largest gap between consecutive elements from $\alpha\mathcal{A}$ (mod 1) is at least $\frac{\log_2 N}{3N}$ for any α. In fact, a random subset of $[1, N]$ of cardinality $\asymp N$ has this property with large probability.*

Proof : Divide the circle into $\nu = \left[\frac{3N}{\log_2 N}\right]$ congruent, disjoint intervals $\{I_j\}_{j=1}^{\nu}$ of length $1/\nu$. Let the random set $\mathcal{A} \subset [1, 2N]$ be defined via $\mathcal{A} = \{j \in [1, 2N] : \xi_j = 1\}$ where ξ_j

are i.i.d. with $\mathbb{P}(\xi_j = 1) = \mathbb{P}(\xi_j = 0) = \frac{1}{2}$. Now fix some $\alpha \in [0,1]$. We say that $j \in [1,\nu]$ is *good* if

$$\operatorname{card}\{\ell : \alpha a_\ell \in I_j \pmod 1\} \le \frac{2}{3}\log_2 N.$$

Clearly, at least half of all intervals are good. Now

$$\begin{aligned}
& \mathbb{P}\Big(\text{for every } j\text{: } I_j \cap (\alpha\mathcal{A} \pmod 1) \neq \emptyset\Big) \\
\le\ & \mathbb{P}\Big(\text{for every good } j\text{: } I_j \cap (\alpha\mathcal{A} \pmod 1) \neq \emptyset\Big) \\
\le\ & \Big(1 - 2^{-\frac{2}{3}\log_2 N}\Big)^{\nu/2} \le \exp\Big(-\frac{N}{\log_2 N}\cdot N^{-2/3}\Big) \\
=\ & \exp(-N^{1/3}/\log_2 N).
\end{aligned}$$

Let $\alpha_j = \frac{j}{N^2}$ for $j = 0, 1, \dots, N^2 - 1$. Summing the previous line over all these j yields for large N

$$\begin{aligned}
& \mathbb{P}\Big(\text{for some } \alpha \text{ the largest gap in } \alpha\mathcal{A} \pmod 1 \text{ is smaller than } \tfrac{\log_2 N}{3N}\Big) \\
\le\ & N^2 \exp(-N^{1/3}/\log_2 N).
\end{aligned}$$

Since $\operatorname{card}(\mathcal{A}) = N$ with probability $\asymp N^{-\frac{1}{2}}$ and $\operatorname{card}(\mathcal{A}) \asymp N$ with large probability, the proposition follows. ■

5 Concentration

Given a set of N distinct integers $\mathcal{A} = \{a_1, \dots, a_N\}$ let

$$\kappa(\mathcal{A}) = \min_{\alpha\in[0,1]} \max_{x\in[0,1]} \operatorname{card}\{j \in [1,\dots,N] : \|\alpha a_j - x\| < N^{-1}\}.$$

In other words, for any fixed α consider the largest number of $\alpha\mathcal{A} \pmod 1$ contained in any interval of size N^{-1}. Then minimize this quantity in α. Before discussing upper bounds on the concentration κ we show that it cannot be made too small. It is fairly simple to construct an example by probabilistic methods. Although this does not give a very good bound, the following result shows that any dilate of a "typical" set of N integers has concentration at least $\log N/\log\log N$. In the next section we present a deterministic example that is more involved, but which gives a much better lower bound.

Proposition 5.1 *For sufficiently large N there exist positive integers $\mathcal{A} = \{a_1 < \ldots < a_N\}$ so that*

$$\kappa(\mathcal{A}) > \frac{\log N}{2 \log\log N}.$$

In fact, with probability $> \frac{1}{2}$ a random subset of size $\asymp N$ of the first $N[\log N]$ positive integers has this property.

Proof : Take $M = N[\log N]$ and let ξ_j be i.i.d. with $\mathbb{P}(\xi_j = 1) = p = \frac{1}{\log N}$ and $\mathbb{P}(\xi_j = 0) = 1 - p$. Define $\mathcal{A} = \{j : \xi_j = 1\}$. Clearly, $\mathrm{card}(\mathcal{A}) \asymp N$ with high probability. We divide the circle into the intervals $I_j = [\frac{j}{N}, \frac{j+1}{N})$ where $j = 0, \ldots, N-1$. Now fix some $\alpha \in [0,1]$ and consider the points

$$\mathcal{P} = \Big\{ \{k_1\alpha\}, \{k_2\alpha\}, \ldots, \{k_M\alpha\} \Big\}$$

where $\{k_1, \ldots, k_M\}$ is a permutation of $\{1, \ldots, M\}$ that arranges the points in nondecreasing order. Call I_j *unpopular* if it contains no more than $\frac{1}{2}\log N$ points from $\mathcal{P}$ and *popular* otherwise. Obviously, at least $M/3$ points belong to popular intervals. The proposition follows from the fact that with high probability the sequence $\xi_{k_1}, \ldots, \xi_{k_M}$ has blocks of ones of length $> [\frac{1}{2}\frac{\log N}{\log\log N}] = L$ that lie entirely inside popular intervals (this is basically the Erdös–Renyi law of long head runs). I.e., with high probability there is some j_0 so that for all $j \in [j_0, j_0 + L]$ one has $k_j \in \mathcal{A}$ and the points $\{k_j\alpha\}$ belong to the same popular interval. More precisely,

$$\begin{aligned} & \mathbb{P}\Big(\mathrm{card}\{I_j \cap (\alpha\mathcal{A} \pmod 1)\} < \frac{\log N}{2\log\log N} \ \text{ for all } j\Big) \\ \leq\ & \mathbb{P}\Big(\text{no popular interval contains a row of } L \text{ ones}\Big) \\ \leq\ & (1-p^L)^{[M/4L]} \leq \exp\Big(-N(\log N)^{-\frac{\log N}{2\log\log N}}\Big) = \exp(-\sqrt{N}). \end{aligned} \tag{9}$$

Now let α run over all fractions with denominator $N^3[\log N]$ and sum the contributions from (9) for each of these α. Thus

$$\begin{aligned} & \mathbb{P}\Big(\text{for some } \alpha:\ \mathrm{card}\{I_j \cap (\alpha\mathcal{A} \pmod 1))\} < \frac{\log N}{2\log\log N} \ \text{ for all } j\Big) \\ \leq\ & N^3 \log N \exp(-\sqrt{N}). \end{aligned} \tag{10}$$

The proposition follows since $\mathrm{card}(\mathcal{A}) = N$ with positive probability and $\mathrm{card}(\mathcal{A}) \asymp N$ with large probability. ■

We now discuss upper bounds. It is easy to see that $\kappa(\mathcal{A}) \le C\sqrt{N}$ for any $\mathcal{A} \subset \mathbb{Z}$ with $|\mathcal{A}| = N$, and we will give an argument that yields $N^{\frac{1}{3}+\epsilon}$. One way of showing $N^{\frac{1}{2}}$ is to let $X_{ij}(\alpha) = 1$ if and only if $\|\alpha(a_i - a_j)\| < 2N^{-1}$ and $X_{ij}(\alpha) = 0$ otherwise. We consider $X_{ij}(\alpha)$ as random variables by taking α to be uniform on $[0,1]$. One checks that $\mathbb{E}(X_{ij}) = 4N^{-1}$ and thus $\mathbb{E}(\sum_{i=1}^N \sum_{j\neq i} X_{ij}) \le CN$. Fix some α with $\sum_{i=1}^N \sum_{j\neq i} X_{ij}(\alpha) \le CN$. With this choice of α there are at most $\asymp N$ pairs of points αa_j in any interval of size N^{-1}.

Alternatively, pick a smooth, positive, one–periodic bump function ϕ with support of size N^{-1} and height 1. We need to estimate the average in α of $\|\sum_{i=1}^N \phi(\alpha a_i - x)\|_{L^\infty(dx)}$. This can be done by bounding the average in α of the ℓ^1 norm of the Fourier coefficients of this sum:

$$\begin{aligned}\int_0^1 \|\sum_{j=1}^N \phi(\alpha a_j - x)\|_{L^\infty(dx)}\, d\alpha &\le \sum_{k\in\mathbb{Z}} |\hat{\phi}(k)| \Big(\int_0^1 |\sum_{j=1}^N e^{2\pi i k a_j}|^2\, d\alpha\Big)^{\frac{1}{2}} \\ &\le \sum_{k\neq 0} |\hat{\phi}(k)|\sqrt{N} + |\hat{\phi}(0)|N \le C\sqrt{N}.\end{aligned}$$

The last inequality follows since $|\hat{\phi}(k)| \le C\min(N^{-1}, N|k|^{-2})$.

It is easy to deal with lacunary sequences $\{a_j\}$, i.e., $a_{j+1} > qa_j$ where $q > 1$ is fixed. Indeed, assume first that $q > N$. Split $[0,1]$ into N disjoint intervals I_j of size N^{-1} centered at x_j and let $Y_j(\alpha) = \sum_{i=1}^N \chi_{I_j}(\alpha a_i)$. Clearly,

$$\begin{aligned}&\mathbb{P}(\alpha : Y_j(\alpha) > N^\epsilon) \le N^{-p\epsilon}\, \mathbb{E}(Y_j^p) \\ = \;& N^{-p\epsilon} \sum_{n_1=1}^N \dots \sum_{n_p=1}^N \mathbb{P}(\alpha : \|\alpha a_{n_1} - x_j\| < N^{-1}, \dots, \|\alpha a_{n_p} - x_j\| < N^{-1}). \qquad (11)\end{aligned}$$

It is easy to see that $q > N$ implies that for j fixed and α uniformly distributed in $[0,1]$ the random variables $\chi_{I_j}(\alpha a_i)$ are basically independent. More precisely, one checks that the probability on the right–hand side is $\le CN^{-k}$ if k is the number of distinct integers among $n_1, \dots, n_p$. Hence $\mathbb{E}(Y_j^p) \le C_p$ independently of N and thus the left–hand side of (11) is bounded by $C_\epsilon\, N^{-2}$ if $p\epsilon > 2$. For general $q > 1$, one applies this estimate to each of the $\asymp \log N$ many subsequences $k + \ell\nu$ where $q^\nu > N$. Summing over j and k yields $\kappa(\mathcal{A}) \le C_{\epsilon,q} N^\epsilon$ for q–lacunary sequences.

Generally speaking, the variables $\chi_{I_j}(\alpha a_i)$ will not be independent and the sum on the right–hand side of (11) will be too large. Nevertheless, we will show below that one can use (11) summed in j and with $p = 3$. The point is that for the *majority* of distinct triples i, j, k the random variables $\chi_{[\|\alpha(a_i - a_j)\| < \delta]}$ and $\chi_{[\|\alpha(a_i - a_k)\| < \delta]}$ where $\delta = N^{-1}$ behave like

independent variables. Unfortunately, this completely fails for fourth or higher moments ($p \geq 4$), see the discussion in Section 7.

Our proof of Theorem 5.2 below uses ideas from the paper by Alon and Peres [4] and the recent preprint of Alon and the second author [5]. For the sake of simplicity we do not state various bounds in an optimal form, but refer the reader to [5] for further details, see Lemmas 3.6 to 3.10. In particular, their work allows one to prove Theorem 5.2 below with $N^{\frac{1}{3}} e^{c\sqrt{\log N \log\log N}}$.

Theorem 5.2 *For any $\epsilon > 0$ there exists a constant C_ϵ so that*

$$\kappa(\mathcal{A}) \leq C_\epsilon \, N^{\frac{1}{3}+\epsilon}$$

for all N and any set $\mathcal{A}$ of N distinct integers.

Proof : To obtain the $N^{\frac{1}{3}+\epsilon}$ bound let $X_{ijk}(\alpha) = 1$ if and only if $\|\alpha(a_i - a_j)\| < N^{-1}$ and $\|\alpha(a_i - a_k)\| < N^{-1}$ and $X_{ijk}(\alpha) = 0$ otherwise. As in the case of second moments considered above, it suffices to show that $\mathbb{E}\sum_{i,j,k} X_{ijk} \leq C_\epsilon \, N^{1+\epsilon}$, where the summation runs over triples of distinct indices. We will show that

$$\begin{aligned} \mathbb{E}(X_{ijk}) &= \mathbb{P}(\alpha : \|\alpha(a_i - a_j)\| < \delta, \|\alpha(a_i - a_k)\| < \delta) \\ &\leq C\Big(N^{-2} + N^{-1}\frac{\gcd(|a_k - a_i|, |a_j - a_i|)}{|a_k - a_i| + |a_j - a_i|}\Big). \end{aligned} \tag{12}$$

First notice that we may assume that $\gcd(|a_k - a_i|, |a_j - a_i|) = 1$. Secondly, given positive, relatively prime integers $n < m$ and any $\delta > 0$,

$$\begin{aligned} \mathbb{P}(\alpha : \|\alpha m\| < \delta, \|\alpha n\| < \delta) &\leq \operatorname{card}\{\ell \in [1, \ldots, m] : \ell n \in [-[2\delta m], \ldots, [2\delta m]] \bmod m\}\frac{2\delta}{m} \\ &\leq C(1 + \delta m)\frac{\delta}{m} = C(\delta^2 + \frac{\delta}{m}). \end{aligned} \tag{13}$$

Indeed, the set on the left–hand side of (13) is the union of all intervals $[\frac{\ell}{m} - \frac{\delta}{m}, \frac{\ell}{m} + \frac{\delta}{m}]$ which come $\frac{\delta}{n}$ close to fractions with denominator n. The number of such intervals is given by the first factor on the right hand–side of (13). Setting $\delta = N^{-1}$ now yields (12). We will show that $\max_i \mathbb{E}(\sum_{j,k} X_{ijk}) \leq C_\epsilon \, N^\epsilon$ where the summation runs over distinct $j, k \in [1, \ldots, N]$. In view of (12) this follows from the inequality

$$\begin{aligned} &\max_{i=1,\ldots,N} \operatorname{card}\{(j,k) : M \gcd(|a_k - a_i|, |a_j - a_i|) \geq |a_k - a_i| + |a_j - a_i|\} \\ \leq\ & C_\epsilon \, \min\{N^2, (MN)^{1+\epsilon}\} \end{aligned} \tag{14}$$

for any positive integer M. In fact,

$$\mathbb{E}\sum_{j,k} X_{ijk} \le C_\epsilon\, N^{-1} \sum_{2^\ell \le N} 2^{-\ell}(2^\ell N)^{1+\epsilon} + \sum_{j,k} CN^{-2} \le C_\epsilon\, N^\epsilon.$$

To prove (14), fix some $i \in [1,\dots,N]$ and a positive integer M. W.l.o.g. $a_i = 0$. Consider the graph $G = (V,E)$ with vertices labeled by a_j, $j \neq i$ and an edge connecting a_j, a_k if and only if $M\gcd(|a_k|,|a_j|) \ge |a_k| + |a_j|$. There exists a vertex, say a_1, so that for any positive integer r the number of closed walks of length $2r$ starting at a_1 is at least $(N-1)^{-1}(2|E|/(N-1))^{2r}$. Indeed, eliminating all vertices of degree less than the average degree $\bar{d} = 2|E|/(N-1)$ together with all their incident edges one obtains a subgraph in which all vertices have degree at least $\bar{d}$. In particular, from any vertex in this subgraph the number of walks of length r is at least $\bar{d}^r$. The claim about the number of closed walks now follows by joining any two of these walks that end at the same point. The stated bound then follows by Cauchy–Schwarz, see [4] for details. On the other hand, suppose that $a_1, a_2, \dots, a_{2r}, a_1$ is a closed walk. By assumption, $\frac{a_{i+1}}{a_i} = \frac{q_i}{p_i}$, where $|q_i|, p_i \in \{1,2,\dots,M\}$ are relatively prime, for $i = 1,\dots,2r$. It is clear that $p_1,\dots,p_{2r},q_1,\dots,q_{2r}$ completely determine the walk. The number of choices of $p_1,\dots,p_{2r}$ is at most M^{2r}. Since $p_1p_2\cdot\ldots\cdot p_{2r} = q_1q_2\cdot\ldots\cdot q_{2r}$, it follows that the number of choices of $q_1,q_2,\dots,q_{2r}$ given $p_1,p_2,\dots,p_{2r}$ is at most $[2d(p_1\cdot\ldots\cdot p_{2r})]^{2r}$, where d is the divisor function. By the elementary estimate $d(n) \le C_\epsilon\, n^\epsilon$, one concludes that the number of closed walks starting at a_1 is at most $M^{2r}(C_\epsilon\, M^{2r})^{2r\epsilon}$. In view of the lower bound, $|E| \le C_\epsilon\, N^{1+1/2r}M^{1+2r\epsilon}$. Choosing r large and then ϵ small yields (14). ■

6 A lower bound for the concentration

Recall that given a set of N distinct integers $\mathcal{A} = \{a_1,\dots,a_N\}$ we let

$$\kappa(\mathcal{A}) = \min_{\alpha\in[0,1]} \max_{x\in[0,1]} \mathrm{card}\{j \in [1,\dots,N] : \|\alpha a_j - x\| < N^{-1}\}.$$

Theorem 6.1 *For sufficiently large N there exist positive integers $\mathcal{A} = \{a_1 < \ldots < a_N\}$ so that*

$$\kappa(\mathcal{A}) > \exp\left(c\frac{\log N}{(\log\log N)^2}\right).$$

Proof : We construct a set $\mathcal{A}$, $|\mathcal{A}| \le N$ and another set V (as large as we can) with the following property. For every $q \le Q = N(\overline{v} - \underline{v})$, where $\overline{v}, \underline{v}$ are the maximal and minimal elements of V, there is a b such that

$$b + qv \in \mathcal{A} \text{ for all } v \in V. \tag{15}$$

Any set of integers $\mathcal{A}'$ containing $\mathcal{A}$ such that $|\mathcal{A}'| = N$ will satisfy $\kappa(\mathcal{A}') \geq |V|$. To show this, take an arbitrary α. We can find $q \leq Q$ such that $\|\alpha q\| \leq 1/Q$. Now take the b satisfying (15). We have

$$\|\alpha q(v - \underline{v})\| \leq (v - \underline{v})\|\alpha q\| \leq (\overline{v} - \underline{v})/Q \leq 1/N,$$

thus all the numbers $\alpha(b+qv)$ are at distance $\leq 1/N \pmod 1$ from $\alpha(b+q\underline{v})$. We construct the set V in the following way. We take an integer k, a subset U of the set of divisors of k such that

$$U \subset [l, l + l/l_1)$$

with certain integers l, l_1 and put

$$V = \{k/u : u \in U\}.$$

In this way we have

$$\overline{v} \leq \frac{k}{l}, \quad \underline{v} \geq \frac{k}{l + l/l_1},$$

so

$$\overline{v} - \underline{v} \leq \frac{k}{ll_1}, \quad Q \leq \frac{kN}{ll_1}. \tag{16}$$

We write the integers $q \leq Q$ in the form

$$q = \lambda_0 + l\lambda_1 + ll_1(\lambda_2 + l\lambda_3) + ... + (ll_1)^{r-1}(\lambda_{2r-2} + l\lambda_{2r-1}),$$

where $\lambda_{2i} \leq l - 1$ and $\lambda_{2i+1} \leq l_1 - 1$. This is possible as long as

$$Q < (ll_1)^r. \tag{17}$$

For $u \in U$ we have

$$\frac{\lambda_0 + l\lambda_1}{u} = \lambda_1 + \frac{\lambda_0 + (l-u)\lambda_1}{u}.$$

Here the first summand is independent of u and, since $-l/l_1 < l - u \leq 0$, the numerator of the second satisfies

$$-l < \lambda_0 + (l-u)\lambda_1 < l.$$

By doing the same for each pair $\lambda_{2i}, \lambda_{2i+1}$ we obtain

$$\frac{q}{u} = (\lambda_1 + ll_1\lambda_3 + ... + (ll_1)^{r-1}\lambda_{2r-1}) + \frac{1}{u}(\mu_0 + ll_1\mu_1 + ... + (ll_1)^{r-1}\mu_{r-1}),$$

with certain integers μ_j satisfying $-l < \mu_j < l$. Thus if $\mathcal{A}$ contains all the integers of the form

$$\frac{k}{u}(\mu_0 + ll_1\mu_1 + ... + (ll_1)^{r-1}\mu_{r-1}),\ u \in U,\ -l < \mu_j < l, \tag{18}$$

then it will have property (15) with b given by

$$b = -k(\lambda_1 + ll_1\lambda_3 + ... + (ll_1)^{r-1}\lambda_{2r-1}).$$

The number of integers of the form (18) is $< |U|(2l)^r$, so this choice of $\mathcal{A}$ is acceptable if

$$N \geq |U|(2l)^r. \tag{19}$$

Now we determine k, l, l_1 and r. Let k be the product of the first s primes, where

$$s = \left[c_1 \frac{\log N}{(\log\log N)^2}\right]$$

with some constant $c_1 > 0$. This k satisfies $\log k \sim s \log s$, so with any $c_2 > c_1$ for large N we have

$$k < N^{c_2 \log N/\log\log N}.$$

This k has altogether 2^s divisors. Write $l_1 = [2^{s/2}]$. Since the interval $[1, k]$ can be covered with $O(l_1 \log k)$ intervals of type $[l, l + l/l_1)$, there will be at least one choice of l for which this contains

$$\gg \frac{2^s}{l_1 \log k}$$

divisors of k. Hence,

$$|V| \gg \frac{2^s}{l_1 \log k} > 2^{(1/2-\varepsilon)s} > \exp\left(c_4 \frac{\log N}{(\log\log N)^2}\right)$$

for any constant $c_4 < (c_1/2)\log 2$. To complete the proof of the theorem, it remains to achieve (17) and (19). We define r by the condition

$$(ll_1)^r < kN \leq (ll_1)^{r+1}.$$

By (16) this choice guarantees that (17) holds. Since $ll_1 < k^2$, this r satisfies

$$r \geq \frac{\log N}{2\log k} - 1 > c_3 \log\log N \tag{20}$$

with any $c_3 < 1/(2c_1)$. We have

$$(2l)^r = (ll_1)^r (2/l_1)^r < kN(2/l_1)^r.$$

Using the crude estimate $|U| < k$ we see that to achieve (19) it is sufficient that

$$(l_1/2)^r > k^2. \tag{21}$$

The definition of l_1 and estimate (20) of r show that

$$(l_1/2)^r > \exp\left(\frac{1}{4} - \varepsilon\right) \frac{\log N \log 2}{\log\log N},$$

thus (21) will hold for large N if $c_1 < \log 2/8$, as required. ■

7 Comments and Questions

The arithmetic sequence $a_j = j$ shows that the argument from Section 5 does not carry over to fourth moments. Indeed, let $X_{ijk\ell}(\alpha) = 0, 1$ depending on whether the points αa_i to αa_ℓ lie in an interval of size N^{-1} or not. Then $X_{ijk\ell}(\alpha) = 1$ *for all* $\alpha \in [0, N^{-2}]$ so that $\sum_{i,j,k,\ell} \mathbb{E}(X_{ijk\ell}) \geq N^2$ for that sequence, which gives the trivial $\sqrt{N}$ bound for the concentration. This suggests to seek a different approach for sequences with not too large gaps. One possibility might be to choose α to be a rational number with denominator of size $\asymp N$. This leads to the following problem:

Question 7.1 *Let* $\mathcal{A} \subset \{1, 2, \ldots, N^2\}$ *be a set with* $\mathrm{card}(\mathcal{A}) = N$. *How large is*

$$D(\mathcal{A}) = \min_{N \leq d \leq 2N} \max_{x \in \mathbb{Z}_d} \mathrm{card}\{a \in \mathcal{A} : a \equiv x \,(\mathrm{mod}\ d)\} \ ? \tag{22}$$

It is easy to see that $D(\mathcal{A}) \leq C\sqrt{N}$. Indeed, assume that $D(\mathcal{A}) > D = [\sqrt{2N}]$. Take D distinct primes $p_1, \ldots, p_D$ in $[N, 2N]$ and consider the corresponding residues of $\mathcal{A}$ modulo p_j for $j = 1, \ldots, D$. There are D residues classes, each contains at least D elements. Thus, the total number of the elements is $\geq D^2$. Any two residues classes have at most one element in common, and we have to subtract at most $D(D-1)/2$. The number of distinct elements is at least $D^2 - D(D-1)/2 = D(D+1)/2$. One might conjecture that $D(\mathcal{A}) \leq C \log N$, as in the case of random $\mathcal{A}$. However, we do not know how to prove this. Another — potentially useful — way of viewing this question is by analogy with incidences

between points and lines. To be precise, observe that any two arithmetic progressions whose increments are different prime numbers in $[d, 2d]$ can intersect in at most one point inside the interval $\{1, 2, \ldots, N^2\}$. Suppose now that we have $M \asymp N/\log N$ many primes $p_1, \ldots, p_M \in [N, 2N]$ for which $\max_x \mathrm{card}\{a \in \mathcal{A} \,:\, a \equiv x \mathrm{mod}\, p_j\} \geq C_0\sqrt{N \log N}$, where C_0 is some sufficiently large constant. Let $\gamma_1, \ldots, \gamma_M$ be the associated arithmetic progressions with increments p_j so that $\mathrm{card}(\gamma_j \cap \mathcal{A}) \geq C_0\sqrt{N \log N}$ for each j. Consider now a bipartite graph with vertex sets $\{\gamma_1, \ldots, \gamma_M\}$ and $\mathcal{A}$, respectively and connect a point from $\mathcal{A}$ with an arithmetic progression if and only if the point lies in the progression. By the aforementioned intersection property of arithmetic progressions this graph does not contain a $K_{2,2}$. By some elementary combinatorics, see [3], there are no more than $\asymp N^{3/2}/\sqrt{\log N}$ edges in total, which contradicts our assumption if C_0 is large. The (very simple) combinatorics we used arises frequently. In particular, it leads to the bound $I(n) \leq Cn^{3/2}$ for the number of incidences $I(n)$ between n points and n lines in the plane (two lines intersect in at most one point, which is all we used about the arithmetic progressions above). However, the well–known Szemeredi-Trotter theorem [9] states the optimal bound $I(n) \leq Cn^{4/3}$. This cannot be obtained by graph theory alone, but requires further geometric insight. We refer the reader to [9] and the recent book [3], where the Szemeredi-Trotter theorem is discussed in the context of various probabilistic methods in combinatorics, mainly "ϵ–nets in hypergraphs with finite VC–dimension". It is however not clear whether these methods are relevant for Question 7.1. Finally, we want to point out that the large sieve also yields this bound. In fact, by the large sieve inequality [12],

$$\sum_{p \leq X} p \sum_{h=1}^{p} (\mathrm{card}\{a \in \mathcal{A} \,:\, a \equiv h \,(\mathrm{mod}\, p)\})^2 \leq C(X^2 + N^2) \int_0^1 \Big| \sum_{a \in \mathcal{A}} e^{2\pi i a t} \Big|^2 \, dt.$$

Setting $X = N$ as desired we obtain the same as before. Replacing second with fourth moments in the previous line does not lead to any improvement because of arithmetic progressions.

It is perhaps worth pointing out that the probability on the right–hand side of (11) can be computed in full generality. In fact, given p distinct positive integers $b_1, \ldots, b_p$ and any $\delta \in (0, 1]$ one has

$$\begin{aligned} &|\{\alpha \in [0,1] \,:\, \|\alpha b_1\| < \delta, \ldots, \|\alpha b_p\| < \delta\}| \asymp \\ \asymp \;& \delta^p \,\mathrm{card}\{(x_1, \ldots, x_p) \in [-\delta^{-1}, \delta^{-1}]^p \cap \mathbb{Z}^p \,:\, x_1 b_1 + \ldots + x_p b_p = 0\}. \end{aligned} \tag{23}$$

To prove this, choose a smooth one–periodic bump function ϕ of height one and support $[-\delta, \delta]$. Then the left–hand side of (23) is bounded by

$$\int_0^1 \phi(b_1\alpha)\phi(b_2\alpha) \cdot \ldots \cdot \phi(b_p\alpha) \, d\alpha = \widehat{\phi(b_1 \cdot)} * \ldots * \widehat{\phi(b_p \cdot)}(0).$$

It is an exercise in Fourier series to check that the convolution is (up to some inessential technicalities) equal to the right–hand side of (23). This also provides a rigorous way of checking the independent behavior in the lacunary case.

8 Non–averaging subsets

This section 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. Let $h(n)$ denote the maximum h so that every set of n integers contains a non-averaging subset of cardinality h. Answering a problem of Erdős, Abbott proved in [1] that $h(n) \gg n^{1/13}/(\log n)^{1/13}$ for $n \geq 2$. Alon and the second author [5] improved this estimate to $h(n) \gg n^{1/6}$. Their proof uses the result of Bosznay [6] who established that the set $\{1, \ldots, n\}$ contains a non-averaging subset of cardinality $\gg n^{1/4}$. The method of [5], together with the construction (rather than the result) of [6] gives the following.

Theorem 8.1 *Every set of n integers contains a non-averaging subset of cardinality $\gg n^{1/5}$.*

Let $I_j = [a_j, b_j)$ $(j = 0, \ldots, k-1)$ be k intervals. We say that I_j are non-averaging if for any set $d_0, d_1, \ldots, d_p$ of p reals, where $1 \leq p < k$ and the points d_i are in some $p+1$ *distinct* intervals I_s from the intervals above, the equation $pd_0 = \sum_{i=1}^{p} d_i \pmod 1$ is *not* satisfied. In particular, taking $p = 1$ we see that non-averaging intervals are disjoint.

Lemma 8.2 *For any $n \in \mathbb{N}$ there exist a set of $\gg n^{1/5}$ non-averaging intervals of length $1/n$.*

Proof : We follow the construction of [6]. Denote $k = [n^{1/5}]$. Let us introduce the intervals

$$[a_j, b_j) = [\frac{j}{k^2} + \frac{j^2}{k^5}, \frac{j}{k^2} + \frac{j^2}{k^5} + \frac{1}{n}) \quad (j = 0, \ldots, k-1).$$

We will show that these intervals are non-averaging. Let $j_0, \ldots, j_p$ be distinct and assume that

$$pd_0 = \sum_{i=1}^{p} d_i \pmod 1. \tag{24}$$

for some $d_i \in I_{j_i}$. As $p < k$ and all intervals I_j are contained in $[0, 1/k)$, we have $0 \leq pd_0 < 1$ and $0 \leq \sum_{i=1}^{p} d_i < 1$. Therefore, (24) implies the equality

$$pd_0 = \sum_{i=1}^{p} d_i. \tag{25}$$

Denote $d_i = \frac{j_i}{k^2} + \frac{r_i}{k^5}$, where $r_i \in [j_i^2, j_i^2 + k^5/n) \subset [j_i^2, j_i^2 + 1)$. The equality (25) can be rewritten as

$$pj_0 - \sum_{i=1}^{p} j_i + \frac{1}{k^3}\left(pr_0 - \sum_{i=1}^{p} r_i\right) = 0. \tag{26}$$

Taking into account that $0 \le r_i < k^2$, we get

$$-k^3 < pr_0 - \sum_{i=1}^{p} r_i < k^3,$$

or, by (26), $-1 < pj_0 - \sum_{i=1}^{p} j_i < 1$. The last inequalities mean that

$$pj_0 - \sum_{i=1}^{p} j_i = 0$$

and, by (26),

$$pr_0 - \sum_{i=1}^{p} r_i = 0.$$

It follows from the last two equalities that

$$p(r_0 - j_0^2) - \sum_{i=1}^{p}(r_i - 2j_i j_0 + j_0^2) = 0.$$

But this is impossible because $r_0 - j_0^2 < 1$ but $r_i - 2j_i j_0 + j_0^2 \ge (j_i - j_o)^2 \ge 1$ for $j = 1, \ldots, p$. ■

Lemma 8.3 *If $n \in \mathbb{N}$ and there exists a set of k non-averaging intervals of length $1/n$, then $h(n) \ge k/2$.*

Actually the lemma was proved in [5]. For completeness, we reproduce the proof here.

Let A be a set of n integers and suppose $I_0, \ldots, I_{k-1}$ are non-averaging intervals of length $1/n$. The crucial idea is the following. If there are two real numbers α, β so that the set $\alpha A + \beta \pmod 1$ intersects at least q of the intervals I_j, then A contains a non-averaging subset of size q. Indeed, choose q of the intervals that intersect $\alpha A + \beta \pmod 1$, and for each of them choose some $a \in A$ for which $\alpha a + \beta \pmod 1$ is in that interval. The set of all the chosen elements is clearly non-averaging. Indeed, otherwise $pa_0 = a_1 + \ldots + a_p$ for some chosen elements a_i, implying that $p(\alpha a_0 + \beta) = \sum_{i=1}^{p}(\alpha a_i + \beta) \pmod p$, which contradicts our assumption.

To complete the proof it remains to show that there are α, β for which $\alpha A+\beta \pmod 1$ intersects sufficiently many intervals I_s. This follows easily from the second moment method. In fact, choose randomly and independently α and β in $[0,1)$, according to a uniform distribution. Fix an interval $I = I_j$ for some $j = 0, \ldots, k-1$, and let X denote the random variable counting the number of elements a of A for which $z_a = \alpha a+\beta \pmod 1 \in I$. 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 $\mathbb{P}(X_a = 1) = 1/n$ 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 $\mathbb{E}(X) = n \cdot 1/n = 1$ and $\mathbb{V}(X) = n(1/n)(1-1/n) \leq 1$. By the Cauchy–Schwartz inequality

$$(\mathbb{E}X)^2 \leq \mathbb{E}(X^2)\mathbb{P}(X > 0) = ((\mathbb{E}X)^2 + \mathbb{V}X)\mathbb{P}(X > 0).$$

Therefore, $\mathbb{P}(X > 0) \geq 1/2$, that is; the probability that $\alpha A + \beta \pmod 1$ intersects I is at least $\frac{1}{2}$.

By linearity of expectation we conclude that the expected number of intervals I_j containing a member of $\alpha A + \beta \pmod 1$ is at least $k/2$ and hence there is a choice for α and β for which at least $k/2$ intervals I_j contain members of $\alpha A + \beta \pmod 1$. By the above discussion, this completes the proof of Lemma 8.3.

Theorem 8.1 immediately follows from Lemmas 8.2 and 8.3.

Acknowledgment : The authors are grateful to Noga Alon and Jean Bourgain for discussions on the subject. The research of Imre Z. Ruzsa was supported by the Hungarian National Foundation of Scientific research, Grant No. 25617. The third author thanks the Erdös center in Budapest for a visit during the summer 98 and Bálint Virág for helpful discussions on this paper.

Addresses : (Konyagin) Dept. of Mechanics and Mathematics, Moscow State University, Moscow 119899, Russia
email: kon@nw.math.msu.su
(Ruzsa) Mathematical Institute of the Hungarian Academy of Sciences, Budapest, P.O.B. 127, H-1364, Hungary, email: ruzsa@math-inst.hu
(Schlag) Princeton University, Mathematics Department, Fine Hall, Princeton, N.J. 08544, U.S.A.
email: schlag@math.ias.edu

References

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

[2] N. N. Andreev, S. V. Konyagin, A. Y. Popov, *Extremal problems for functions with small support.* (Russian) Mat. Zametki 60 (1996), no. 3, 323–332.

[3] P. Agarwal, J. Pach, *Combinatorial Geometry.* John Wiley & Sons Inc., 1995.

[4] N. Alon, Y. Peres, *Uniform dilations.* GAFA 2 (1992), 1–28.

[5] N. Alon, I. Ruzsa, *Non-averaging subsets and non-vanishing transversals.* Preprint, 1998. To appear in J. Combinatorial Theory (A).

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

[7] D. Campbell, H. Ferguson, R. Forcade, *Newman Polynomials on* $|z| = 1$. Ind. Univ. Math. J. 32 (1983), 517–525.

[8] K. Chandrasekharan, *Arithmetical functions.* Springer–Verlag, 1970.

[9] K. L. Clarkson, H. Edelsbrunner, L. J. Guibas, M. Sharir, E. Welzl, *Combinatorial complexity bounds for arrangements of curves and spheres.* Discrete and Comput. Geom. 5 (1990), 99–160.

[10] H. Halberstam, K. Roth, *Sequences.* Oxford, 1966.

[11] S. V. Konyagin, *Minimum of the absolute value of random trigonometric polynomials with coefficients* ± 1. Math. Notes (translation of Matematicheskie Zametki), Vol. 56, No 3–4 (1994), 931–947.

[12] H. Montgomery, *Topics in multiplicative number theory.* LNMS #227, Springer, Berlin–New York, 1971.

[13] H. Montgomery, *Ten lectures on the interface between harmonic analysis and analytic number theory.* CBMS Regional conference series in mathematics #84, AMS 1994.

[14] C. J. Smyth, *Some results on Newman polynomials.* Indiana Univ. Math. J. 34 (1985), 195–200.