Constructing small sets that are uniform in arithmetic progressions

A. Razborov*
Steklov Mathematical Institute
Moscow, Russia

E. Szemerédi
Rutgers University
New Brunswick, N.J., USA

A. Wigderson
Hebrew University
Jerusalem, Israel

August 15, 1993

1 Introduction

Let Z_N be the set of residue classes mod N, identified with the set $[N] = \{0, 1, \ldots, N-1\}$. For sets $A, S \subseteq [N]$ denote the *discrepancy* of A in S by

$$D_A(S) = \left| \frac{|A \cap S|}{|A|} - \frac{|S|}{N} \right|.$$

Similarly we define $D_A(S)$ for the case when A is a multiset. Note that this definition is not symmetric in A and S.

For a family $\mathcal{F}$ of subsets of Z_N we denote

$$D_A(\mathcal{F}) = max_{S \in \mathcal{F}} D_A(S).$$

*The research was done while this author was visiting Department of Mathematics at MIT partially supported by the Sloan foundation.

We shall be interested in the family $\mathcal{I}_N$ of all intervals of Z_N. For $\epsilon > 0$ we define $A \subseteq Z_N$ to be ϵ *-uniform* (mod N) if for all $x \in Z_N^*$, $D_{xA}(\mathcal{I}_N) \leq \epsilon$, where $xA = \{xa \,|\, a \in A\}$ (with arithmetic mod N).[1]

A simple counting argument shows that for every N and $\epsilon = \epsilon(N)$ there exist sets $A_{\epsilon,N}$ that are ϵ-uniform mod N and satisfy

$$|A_{\epsilon,N}| \leq (\epsilon^{-1}\log N)^{O(1)}.$$

The main result of this note is an explicit construction of such small sets. Moreover, the sets $A_{\epsilon,N}$ we construct satisfy the following stronger property: for every divisor M of N, the reduced multiset $A_{\epsilon,N}$ (mod M) is ϵ-uniform mod M. Our construction is a variant of a construction in [AIKPS]. The main result is stated and proved in Section 3. In Section 2 we refer, somewhat informally, to related results of similar flavour that partly motivated this work.

2 Related Work

2.1 Arithmetic Progressions

Let $\mathcal{AP}_N = \{yI \,|\, y \in Z_N^*,\, I \in \mathcal{I}_N\}$ be the set of arithmetic progressions mod N, with invertible difference. Since for every $x \in Z_N^*$ and $A, S \subseteq Z_N$ we have $D_{xA}(S) = D_A(x^{-1}S)$, an equivalent definition of the ϵ-uniformity of A is $D_A(\mathcal{AP}_N) \leq \epsilon$. Thus the small sets $A_{\epsilon,N}$ we construct are uniformly distributed in all arithmetic progressions in $\mathcal{AP}_{\mathcal{N}}$ (which explains the title).

The classical interest within Number Theory and Discrepancy Theory regarding arithmetic progressions is in the "dual" problem of how uniformly can arithmetic progressions $A \in \mathcal{AP}_{\mathcal{N}}$ be distributed in sets $S \subseteq [N]$. Tight bounds on $min_{|S|=N/2} max_{A\in\mathcal{AP}_{\mathcal{N}}} D_A(S)$ were given by [Roth, Beck].

2.2 Fourier Transforms

The discrete Fourier transform of (the characteristic function of) a set $A \subseteq Z_N$ is the function

$$f_A(t) = \sum_{a\in A} e^{2\pi i a t/N}$$

[1] The restriction to invertible $x \in Z_N^*$ is natural since if e.g. $x|N$, $x > \epsilon N$, the elements of xA take on less than $1/\epsilon$ values, and thus miss intervals of measure ϵ for any set A.

defined for every $t \in Z_N$. Clearly, $f_A(0) = |A|$. The parameter $\lambda(A) = max_{t \neq 0}|f_A(t)|/|A|$ gives some measure of the randomness of the set A; the smaller it is, the more "random" A is. This parameter has a variety of applications in Additive Number Theory (e.g. [Ruzsa, AIKPS]), Graph Theory (e.g. [Chung]) and Complexity Theory (e.g. [ABK*, GKS]).

The connection of this parameter with our problem stems from the fact that there are many cancelations in the sum of unit vectors that are almost uniformly distributed on the unit cycle. An easy calculation (which for completeness is given in the appendix) yields

$$\left| \sum_{a \in A} e^{2\pi i a/N} \right| \leq O(|A| D_A(\mathcal{I}_N)).$$

From it we conclude that if $A \pmod M$ is ϵ-uniform mod M for every divisor M of N, then $\lambda(A) \leq \epsilon$.

A simple use of the pigeonhole principle (which for completeness is also given in the appendix) shows that every family of sets A_N satisfying $\lambda(A_N) < 1/2$ must have size $|A_N| \geq \Omega(\log N)$. On the other hand, simple probabilistic arguments achieve $\lambda(A_{\epsilon,N}) \leq \epsilon(N)$ with $|A_{\epsilon,N}| = (\log N/\epsilon(N))^{O(1)}$, where $\epsilon = \epsilon(N)$ is any function tending to zero as N grows to infinity (the best current bounds on $|A|$ in terms of N, ϵ appear in [AR]).

There are two known explicit constructions which for

$$\epsilon(N) = \frac{1}{(\log N)^{O(1)}}$$

achieve

$$|A_{\epsilon,N}| = (\log N)^{O(1)}.$$

One is in [Katz], and is based on deep results in number theory. The second is in [AIKPS] and is completely elementary. Our construction is a variant of the second. Note that, due to the remark above, the sets $A_{\epsilon,N}$ we construct in this paper also satisfy $\lambda(A_{\epsilon,N}) \leq \epsilon$.

2.3 Higher dimensional discrepancy

A central problem in Discrepancy Theory (see [BC] and the references within) is to determine the smallest size of sets in $[0,1)^d$ that have small discrepancy with the family $\mathcal{B}$ of all aligned boxes (i.e. cartesian product of d intervals in [0,1)). Defining the discrepancy $D_A(\mathcal{B})$ of a set A in the natural way, the question is how many points A should have as a function of ϵ and d so as to achieve $D_A(\mathcal{B}) \leq \epsilon$.

The classical results focus on constant d and obtain bounds which depend well on ϵ^{-1} but are exponential in d. The recent paper [EGLNV] motivates studying the case where ϵ^{-1} and d are polynomially related. It is not hard to prove that almost all sets A of size $(d\epsilon^{-1})^{O(1)}$ would satisfy this discrepancy bound. However, an explicit construction of such sets A will have strong applications to derandomizing certain probabilistic algorithms. [EGLNV] obtain somewhat weaker explicit bounds, and leave open the problem of explicitly constructing a set whose size is polynomial in both parameters.

A possible connection to our construction is the following. As observed in [EGLNV], it will suffice to replace the universe $[0,1)^d$ with $(Z_m)^d$, where $m = 2/\epsilon$. Using the natural mapping from Z_{m^d} onto $(Z_m)^d$ (writing a number in base m) we ask (but are not ready to conjecture) if the image of the sets we construct (taking $N = m^d$) have small discrepancy with all boxes.

3 The Construction

Fix N and $t \leq N$. Set the parameters

$$\begin{aligned} m_1 &= (t^4 (\log t)^5 \log N)^{1/9} \\ m_2 &= (t^5 (\log t)^4 / \log N)^{1/9} \\ \epsilon = m_1/m_2 &= (\log t (\log N)^2 / t)^{1/9} \end{aligned}$$

and the sets $S = [m_1]$, $P = \{p \in [m_2, 2m_2] \,|\, p \text{ prime}, \ (p, N) = 1\}$. Using $(x)_N$ for x mod N we define multisets $A_{t,N}$ by

$$A_{t,N} = \{s(p^{-1})_N \,|\, s \in S, \ p \in P\}.$$

Our main result gives the discrepancy bound

Theorem 3.1 *The multiset $A_{t,N}$ is $O(\epsilon)$-uniform* mod *N. Moreover, for every divisor M of N, $A_{t,N}$ mod M is $O(\epsilon)$-uniform* mod *M.*

Before we start proving this, note that in the "interesting" range

$$3 \log\log N (\log N)^2 < t < \frac{N}{3 \log N}$$

we have $s_1(p_1^{-1})_N \neq s_2(p_2^{-1})_N \mod N$ unless $(s_1, p_1) = (s_2, p_2)$ that is our multisets $A_{t,N}$ are actually sets. Also, in this range they clearly have size $|A_{t,N}| \leq O(t)$ and are easy to compute in time $(t \log N)^{O(1)}$ by any reasonable model of computation. Hence, taking $A_{\epsilon,N} = A_{t(\epsilon,N),N}$ for $\epsilon = \epsilon(N) \geq \omega\left(((\log N)^4/N)^{1/9}\right)$, where

$$t(\epsilon, N) = C\epsilon^{-9}(\log N)^2 \left(\log \frac{1}{\epsilon} + \log\log N\right)$$

($C > 0$ is a sufficiently large constant), we will get the sets $A_{\epsilon,N}$ promised in the introduction.

Proof of Theorem 3.1: We prove only the first statement of uniformity mod N. The second statement follows exactly the same proof with M replacing N in appropriate places, and is therefore omitted.

Let $\boldsymbol{s}$ be a random member of S and $\boldsymbol{p}$ be a random member of P. Fix an element $x \in Z_N^*$ and interval $I \in \mathcal{I}_N$. In probabilistic terms we need to prove

$$\left|\mathbf{P}\left[x\boldsymbol{s}(\boldsymbol{p}^{-1})_N \in I\right] - |I|/|N|\right| \leq O(\epsilon). \tag{1}$$

We use the Chinese remainder theorem in the following form (assuming p does not divide N)

$$p(p^{-1})_N + N(N^{-1})_p = pN + 1.$$

Dividing by pN, using $(\alpha)_1$ to denote the fractional part (in $[0,1)$) of any real number α, and noting that $x < N$ we obtain

$$\left(\frac{x\boldsymbol{s}(\boldsymbol{p}^{-1})_N}{N} + \frac{x\boldsymbol{s}(N^{-1})_{\boldsymbol{p}}}{\boldsymbol{p}}\right)_1 \leq \frac{m_1}{m_2} = \epsilon.$$

This implies

$$\mathbf{P}\left[\left(\frac{x\boldsymbol{s}(N^{-1})_{\boldsymbol{p}}}{\boldsymbol{p}}\right)_1 \in J_-\right] \leq \mathbf{P}\left[x\boldsymbol{s}(\boldsymbol{p}^{-1})_N \in I\right] \leq \mathbf{P}\left[\left(\frac{x\boldsymbol{s}(N^{-1})_{\boldsymbol{p}}}{\boldsymbol{p}}\right)_1 \in J_+\right]$$

where $J_- \subseteq J_+ \subseteq [0,1)$ and $|J_+| - |J_-| \leq 2\epsilon$.

Hence, in order to prove (1) it suffices to prove

$$\left| \mathbf{P}\left[\left(\frac{x\boldsymbol{s}(N^{-1})_{\boldsymbol{p}}}{\boldsymbol{p}} \right)_1 \in J \right] - |J| \right| \leq O(\epsilon) \qquad (2)$$

for any interval $J \subseteq [0,1)$ (which will apply to J_-, J_+).

Fix J. To prove (2) we may clearly assume without loss of generality (by enlarging J if necessary) that $|J| \geq \epsilon$. Let

$$|\alpha|_1 = \begin{cases} (\alpha)_1 \text{ if } \ 0 \leq (\alpha)_1 < 1/2, \\ 1 - (\alpha)_1 \text{ if } \ 1/2 \leq (\alpha)_1 < 1. \end{cases}$$

Say that $p \in P$ is *bad* if there exists a positive integer $s \leq 1/\epsilon$ such that

$$\left| \frac{xs(N^{-1})_p}{p} \right|_1 \leq \frac{2}{m_1 \epsilon^2}.$$

Then it suffices to show that

$$\mathbf{P}[\boldsymbol{p} \text{ is bad}] \leq O(\epsilon) \qquad (3)$$

and that for each fixed *good* p,

$$\left| \mathbf{P}\left[\left(\frac{x\boldsymbol{s}(N^{-1})_p}{p} \right)_1 \in J \right] - |J| \right| \leq O(\epsilon). \qquad (4)$$

We first prove (3). If p is bad, then for some $s \leq 1/\epsilon$ and some b with $|b| \leq O(\frac{m_2}{m_1 \epsilon^2})$ we have $xs(N^{-1})_p \equiv b \pmod p$, or equivalently $p | (xs - bN)$. But observe that as $(x, N) = 1$ and $s + |b| \leq O(\frac{m_2}{m_1 \epsilon^2}) < N$ we have $xs - bN \neq 0$ and $|xs - bN| \leq N^2$, so for each fixed choices of s and b we have at most $O(\log N)$ bad p's. Using the bounds on s and b we calculate that there are at most $O(\frac{m_2 \log N}{m_1 \epsilon^3})$ bad p's, which implies (3).

Now consider a good p and prove (4). By the pigeonhole principle, there is a positive integer $s_0 \leq 1/\epsilon$ for which

$$\left| \frac{xs_0(N^{-1})_p}{p} \right|_1 \leq \epsilon. \qquad (5)$$

For this choice of s_0 denote the left hand side of (5) by γ. Since p is good, we also have a lower bound of $\frac{2}{m_1\epsilon^2}$ on γ. Define $m_3 = m_1/s_0$, and observe that $m_3 \geq m_1\epsilon$. We shall prove (4) separately for each residue class $i < s_0$ of $\boldsymbol{s}$, i.e. taking $\boldsymbol{r}$ at random from $[m_3]$ we prove for each i

$$\left|\mathbf{P}\left[\left(\frac{x(\boldsymbol{r}s_0+i)(N^{-1})_p}{p}\right)_1 \in J\right] - |J|\right| \leq O(\epsilon). \tag{6}$$

But the lower bound on γ guarantees that as we go through all possible values of $\boldsymbol{r}$ we will go around the interval $[0,1)$ at least $1/\epsilon$ times, and the upper bound on γ guarantees that in each such cycle we will visit J the correct number of times (as $|J| \geq \epsilon$), which completes the proof. $\Box$

4 Appendix

Here we give, for completeness, the proofs of two statements made in Section 2.2. The propositions below imply these statements. We remark that both propositions are folklore. We provide simple proofs, rather than try to obtain the best constants.

Proposition 4.1 $\left|\sum_{a\in A} e^{2\pi i a/N}\right| \leq 2\pi|A|D_A(\mathcal{I}_N)$.

Proof: (Suggested by Alon and Sudakov). Set $D = D_A(\mathcal{I}_N)$, and let $A = \{a_1, a_2, \ldots, a_m\}$, with the a_i's sorted in increasing order. Note that for every $j \in [m]$, $|a_j/N - j/m| \leq D$, by considering the discrepancy of A on the interval $[0, a_j]$. Since $|e^{ix} - e^{iy}| \leq |x-y|$, this implies $\left|e^{2\pi i a_j/N} - e^{2\pi i j/m}\right| \leq 2\pi D$ and thus $\left|\sum_{a\in A} e^{2\pi i a/N} - \sum_{j=1}^m e^{2\pi i j/m}\right| \leq 2\pi m D$. We are only left to note that $\sum_{j=1}^m e^{2\pi i j/m} = 0$. $\Box$

Proposition 4.2 *If* $\lambda(A) < 1/2$ *then* $|A| > (\log N)/3$.

Proof: For contradiction assume $|A| = m \leq (\log N)/3$. Define the map $\chi : Z_N \to \{0,1,2,3,4,5\}^A$ by letting, for every $t \in Z_N$ and $a \in A$, $\chi(t)_a = j$ iff $at \bmod N \in [jN/6, (j+1)N/6)$. As $6^m < N$, by the pigeonhole principle there must be $t_1, t_2 \in Z_N$ with $\chi(t_1) = \chi(t_2)$. Set $t = t_1 - t_2$. Then for every $a \in A$, $at \bmod N \in (-N/6, N/6)$. We get

$$|A|\lambda(A) \geq f_A(t) = |\sum_{a \in A} e^{2\pi iat/N}| \geq |A|cos(\pi/3) = |A|/2.$$

$\square$

Acknowledgements

We are greatful to Noga Alon and Benny Sudakov for careful reading, helpful comments, and the proof of Proposition 4.1.

References

[ABK*] M. Ajtai, L. Babai, J. Komlós, P. Pudlák, V. Rödl, E. Szemerédi, G. Turán, Two lower bounds for branching programs, *Proc. 18th STOC*, pp. 30–38, 1986.

[AIKPS] M. Ajtai, H. Iwaniec, J. Komlós, J. Pintz, E. Szemerédi, Construction of a thin set with small Fourier coefficients, *Bull. London Math. Soc.* 22, pp. 583–590, 1990.

[AR] N. Alon, Y. Roichman, Random Cayley graphs and expanders, Manuscript, 1991.

[Beck] J. Beck, Roth's estimate of discrepancy of integer sequences is nearly sharp, *Combinatorica* 1, pp. 319–325, 1981.

[BC] J. Beck, W. Chen, *Irregularities of Distributions*, Cambridge University Press, 1987.

[Chung] F. R. K. Chung, Diameters and eigenvalues, *J. AMS* 2, pp. 187–196, 1989.

[EGLNV] G. Even, O. Goldreich, M. Luby, N. Nisan, B. Veličković, Approximations of general independent distributions, *Proc. 24th STOC*, pp. 10–16, 1992.

[GKS] Z. Galil, R. Kannan, E. Szemerédi, On nontrivial separators for k-page graphs, and simulations by nondeterministic one-tape Turing machines, *Proc. 18th STOC*, pp. 39–49, 1986.

[Katz] N. M. Katz, An estimate for character sums, *J. AMS* 2, pp. 197–200, 1989.

[Roth] K. F. Roth, Remark concerning integer sequences, *Acta Arith.* 9, pp. 257–260, 1964.

[Ruzsa] I. Ruzsa, Essential components, *Proc. London Math. Soc.* 54, pp. 38–56, 1987.