DISCREPANCY IN GENERALIZED ARITHMETIC PROGRESSIONS

JAVIER CILLERUELO AND NILS HEBBINGHAUS

Abstract. Estimating the discrepancy of the set of all arithmetic progressions in the first N natural numbers was one of the famous open problem in combinatorial discrepancy theory for a long time, successfully solved by K. Roth (lower bound) and Beck (upper bound). They proved that $D(N) = \min_\chi \max_A |\sum_{x\in A}\chi(x)| = \Theta(N^{1/4})$, where the minimum is taken over all colorings $\chi : [N] \to \{-1, 1\}$ and the maximum over all arithmetic progressions in $[N] = \{0, \ldots, N-1\}$.

Sumsets of k arithmetic progressions, $A_1 + \cdots + A_k$, are called k-arithmetic progressions and they are important objects in additive combinatorics. We define $D_k(N)$ the discrepancy of the set $\{P \cap [N] : P$ is a k-arithmetic progression$\}$. The second author proved that $D_k(N) = \Omega(N^{k/(2k+2)})$ and Přívětivý improved it to $\Omega(N^{1/2})$ for all $k \geq 3$. Since the probabilistic argument gives $D_k(N) = O((N \log N)^{1/2})$ for all fixed k, the case $k = 2$ remained the only case with a large gap between the known upper and lower bound. We bridge this gap (up to a logarithmic factor) by proving that $D_k(N) = \Omega(N^{1/2})$ for all $k \geq 2$.

Indeed we prove the multicolor version of this result.

1. Introduction

Sumsets of k arithmetic progressions, $A_1+\cdots+A_k$, are called k-arithmetic progressions and they are important objects in additive combinatorics.

Let P a k-arithmetic progression and $[N] = \{0, \ldots, N-1\}$. The imbalance of P due to the coloring $\chi : [N] \to \{-1, 1\}$ is defined by $\chi(P) = \sum_{x\in P}\chi(x)$ where $\chi(x) = 0$ if $x \notin [N]$. The discrepancy of the set of k-arithmetic in $[N]$ is defined by

$$D_k(N) = \min_\chi \max_P |\sum_{x\in P}\chi(x)| \tag{1}$$

where the minimum is taken over all possible colorings $\chi : [N] \to \{-1, 1\}$ and the maximum over all k-arithmetic progressions.

Thus, $D_k(N)$ is the least possible imbalance of any k-arithmetic progression that can not be avoided under any coloring $\chi : [N] \to \{-1, 1\}$. For short we write $D(N)$ when $k = 1$.

One of the most famous open problem in (combinatorial) discrepancy theory was to determine the right order for the discrepancy of the set of arithmetic progressions in the first N natural numbers. That is, the order for $D(N)$.

In 1964, Roth [8] proved $D(N) = \Omega(N^{1/4})$. Using a random coloring of $[N]$, one can easily show that $D(N) = O((N \log N)^{1/2})$. The first non-trivial upper bound is due to Sárközy [9]. In 1973 he proved that $D(N) = O((N \log N)^{1/3})$. A sketch of his beautiful proof can be found in [3]. Inventing the famous partial coloring method,

First author is supported by Grant CCG07-UAM/ESP-1814 and DGICYT MTM 2005-04730 (Spain)
.

Beck [1] showed in 1981 that Roth's lower bound is nearly sharp. His upper bound of order $O(N^{1/4}\log^{5/4} N)$ was finally improved by Matoušek and Spencer [6] in 1996. They showed by a refinement of the partial coloring method - the entropy method - that $D(N) = O(N^{1/4})$.

After 32 years, this open problem was solved. In the next years several extensions of this discrepancy problem were studied. For example, Doerr, Srivastav and Wehr [2] determined the discrepancy of cartesian product of arithmetic progressions, those of the form $(A_1, \ldots, A_d) \subset [N]^d$ where all A_i are arithmetic progressions. They proved that, in this case, the discrepancy is $\Theta(N^{d/4})$. Another related discrepancy concerning to 1-dimensional arithmetic progressions in the grid $[N]^d$ was studied by Valkó [10]. He proved for the discrepancy in these sets a lower bound of order $\Omega(N^{d/(2d+2)})$ and an upper bound of order $O(N^{d/(2d+2)}\log^{5/2} N)$.

Here we deal with the discrepancy of k-arithmetic progressions in $[N]$. We observe that, since any k-arithmetic progression is a $(k+1)$-arithmetic progression, we have

$$D(N) = D_1(N) \leq D_2(N) \leq D_3(N) \leq \cdots \leq D_k(N) \leq D_{k+1}(N) \leq \cdots \tag{2}$$

The second author [4] proved that $D_k(N) = \Omega(N^{k/(2k+2)})$. But there remained a large gap between this bound and the upper bound $D_k(N) = O((N\log N)^{1/2})$ obtained from the random coloring. In 2006 Přívětivý [7] almost closed this gap for $k \geq 3$ by proving $D_3(N) = \Omega(N^{1/2})$. This lower bound clearly implies $D_k(N) = \Omega(N^{1/2})$ for all $k \geq 3$. Thus the case $k = 2$ was the last case with a large gap between the lower and the upper bound for $D_k(N)$.

In this paper we improve the lower bound for $D_2(N)$ from $\Omega(N^{1/3})$ to $\Omega(N^{1/2})$.

The multicolor version of discrepancies has only been recently investigated. We state our main result in its general multicolor version.

Theorem 1. *For all $c \geq 2$ and all $k \geq 2$ we obtain the bound*

$$D_k(N, c) = \Omega(N^{1/2})$$

for

$$D_k(N, c) = \min_{\chi} \max_{i=1,\ldots,c} \max_{A} \left| |\chi^{-1}(i) \cap A| - \frac{|A \cap [N]|}{c} \right|,$$

where the minimum is taken over all colorings $\chi : [N] \to \{1, \ldots, c\}$ and the maximum is taken over all colors and k-arithmetic progressions.

It should be noted that $D_k(N) = 2D_k(N, 2)$. Theorem 1 above shows that the upper bound $D_k(N, c) = O((N\log N)^{1/2})$, coming from probabilistic arguments, is nearly sharp for all fixed $k \geq 2$. Theorem 1 above follows immediately from (2) and Theorem 2 below.

Theorem 2. *For any coloring $\chi : [N] \to \{1, \ldots, c\}$ there exists a 2-arithmetic progression P and some $i \in \{1, \ldots, c\}$ such that*

$$\left| |\chi^{-1}(i) \cap P| - \frac{|P \cap [N]|}{c} \right| \geq \frac{N^{1/2}}{800c^{1/2}}.$$

Acknowledgements: This paper is a follow up of Hebbinghaus [5] where the main result was already stated and proved for $c = 2$. The present version contains a simplified version of the original proof and the extension for all $c \geq 2$.

2. Proof of theorem 2

2.1. **Discrete Fourier Analysis in $\mathbb{Z}_p$.** : Let p be a prime. For any function $f : \mathbb{Z} \to \mathbb{C}$ we define $\hat{f} : \mathbb{Z}_p \to \mathbb{C}$ by

$$\hat{f}(a) = \sum_{x\in\mathbb{Z}} f(x)\omega^{ax}$$

where $\omega = e^{\frac{2\pi i}{p}}$. The convolution of two functions $f * g$ is defined by

$$(f * g)(x) = \sum_{y\in\mathbb{Z}} f(y)g(x-y)$$

and it satisfies $\widehat{f * g} = \hat{f}\hat{g}$.

Lemma 1 (Folklore). *If supp$(f) \subset \{-\frac{p-1}{2}, \dots, \frac{p-1}{2}\}$ then*

$$\sum_{x\in\mathbb{Z}} |f(x)|^2 = \frac{1}{p}\sum_{x\in\mathbb{Z}_p} |\hat{f}(x)|^2.$$

2.2. **Proper 2-arithmetic progressions.** A 2-arithmetic progression is a set of the form

$$P = \{a + \delta_1 j_1 + \delta_2 j_2 :\ j_1 \in [L_1],\ j_2 \in [L_2]\}$$

for some $a \in \mathbb{Z}$ and some $\delta_1, \delta_2, L_1, L_2 \in \mathbb{N}$. We say that P is proper if all elements $a + \delta_1 j_1 + \delta_2 j_2$ are distinct.

Lemma 2. *If $(\delta_1, \delta_2) = 1$ and $L_1 \le \delta_2$ then P is proper.*

Proof. Otherwise, $\delta_1 j_1 + \delta_2 j_2 = \delta_1 j_1' + \delta_2 j_2' \implies \delta_1(j_1 - j_1') = \delta_2(j_2' - j_2)$ and then (since $(\delta_1, \delta_2) = 1$) $\delta_2 | (j_1 - j_1')$, in particular $\delta_2 < L_1$. □

Lemma 3. *For all $a \in \mathbb{Z}_p$ there exists a proper 2-arithmetic progression*

$$P_a = \{\delta_1 j_1 + \delta_2 j_2 :\ j_i \in [L_i],\ i = 1, 2\} \subset [N]$$

such that $|\hat{1}_{-P_a}(a)| \ge p/400$.

Proof. For $a = 0$ we take $P_0 = [N]$ and it is clear that $|\hat{1}_{-P_0}(0)| = N \ge p/4$. For $a \not\equiv 0 \pmod p$, let δ_1 be the least positive integer such that

$$a\delta_1 = r_1 + a_1 p, \qquad 1 \le r_1 < \sqrt{p} \tag{3}$$

for some integer a_1. Using the pigeonhole principle we can check that $1 \le \delta_1 \le \sqrt{p}$. Then $m = \max\{r_1, \delta_1\} \le \sqrt{p}$. Sometimes we will use that $m \le p/m$.

Let δ_1^* be the solution of the congruence $a_1 x \equiv -1 \pmod{\delta_1}$ in $[\delta_1]$. Then

$$a_1^*\delta_1 - \delta_1^* a_1 = 1, \qquad 0 \le \delta_1^* < \delta_1 \tag{4}$$

for some positive integer a_1^*. We define $L_1 = \lceil \frac{p}{16m} \rceil$, $L_2 = \lceil \frac{m}{16} \rceil$, $k = \left\lceil \frac{p}{\delta_1 m} \right\rceil$ and

$$\delta_2 = \delta_1^* + \delta_1 k. \tag{5}$$

We claim that the 2-progression $P_a = \{\delta_1 j_1 + \delta_2 j_2 :\ j_i \in [L_i],\ i = 1, 2\}$ satisfies the conditions of Lemma 3. To see that $P_a \subset [N]$ we observe that

$$0 \le \delta_2 \le \delta_1^* + \delta_1\left(\frac{p}{\delta_1 m} + 1\right) \le \frac{p}{m} + 2\delta_1 \le \frac{p}{m} + 2m \le \frac{p}{m} + \frac{2p}{m} = \frac{3p}{m},$$

so the largest element in P_a is

$$\delta_1(L_1-1)+\delta_2(L_2-1)\le m\frac{p}{16m}+\frac{3p}{m}\frac{m}{16}\le\frac{p}{4}<N.$$

To see that P_a is proper we observe that relations (4) and (5) imply that $(\delta_1,\delta_2)=1$. On the other hand if $L_1>\delta_2$ then $1+\frac{p}{16m}\ge\delta_1 k\ge\frac{p}{m}\implies 1\ge\frac{15p}{16m}\ge\frac{15\sqrt{p}}{16}\implies p\le 1$. So $L_1\le\delta_2$ and we use Lemma 2 to conclude that P_a is proper.

Since P_a is proper we can write

$$\hat{1}_{-P_a}(a)=\sum_{x\in P_a}\omega^{-ax}=\left(\sum_{j_1\in[L_1]}\omega^{-a\delta_1 j_1}\right)\left(\sum_{j_2\in[L_2]}\omega^{-a\delta_2 j_2}\right).\tag{6}$$

Since $|r_1(L_1-1)|\le r_1p/(16m)\le p/16$ we have

$$\left|\sum_{j_1\in[L_1]}\omega^{-a\delta_1 j_1}\right|\ge\mathcal{R}\left(\sum_{j_1\in[L_1]}\omega^{-r_1j_1}\right)\ge L_1\min_{j_1\in[L_1]}\cos(2\pi r_1j_1/p)\ge L_1\cos(\pi/8).\tag{7}$$

We observe that

$$\begin{aligned}a\delta_2 &\equiv a(\delta_1^*+\delta_1k)\equiv a\delta_1^*+r_1k\equiv\left(\frac{r_1+a_1p}{\delta_1}\right)\delta_1^*+r_1k\equiv\frac{r_1\delta_1^*+(a_1^*\delta_1-1)p}{\delta_1}+r_1k\\ &\equiv \frac{r_1\delta_1^*-p}{\delta_1}+r_1k\equiv\frac{r_1\delta_1^*-p+r_1p/m}{\delta_1}+r_1\left(1-\left\{\frac{p}{\delta_1m}\right\}\right)\pmod p.\end{aligned}$$

We write r_2 for the last long expression. Since $r_1\le m$ and $\delta_1^*<\delta_1$ we have that $r_2\le 2r_1\le 2m\le 2p/m$. If $m=r_1$ then $0\le r_2$. If $m=\delta_1$ then $r_2\ge -p/\delta_1=-p/m$. In any case we have $|r_2|\le 2p/m$, so $|r_2(L_2-1)|\le(2p/m)(m/16)\le p/8$. Thus,

$$\left|\sum_{j_2\in[L_2]}\omega^{-a\delta_2 j_2}\right|\ge\mathcal{R}\left(\sum_{j_2\in[L_2]}\omega^{-r_2j_2}\right)\ge L_2\min_{j_2\in[L_2]}\cos(2\pi r_2j_2/p)\ge L_2\cos(\pi/4).\tag{8}$$

Finally, (6), (7) and (8) give $\left|\hat{1}_{-P_a}(a)\right|\ge L_1\cos(\pi/8)L_2\cos(\pi/4)\ge p/400$. □

2.3. **End of the proof.** For any coloring $\chi:[N]\to\{1,\dots,c\}$ we consider the functions $f_i:\mathbb{Z}\to\mathbb{C},\ i=1,\dots,c$ defined by

$$f_i(x)=\begin{cases}1-\frac{1}{c} & \text{if } x\in\chi^{-1}(i)\cap[N]\\ -\frac{1}{c} & \text{if } x\in[N]\setminus\chi^{-1}(i)\\ 0 & \text{otherwise.}\end{cases}$$

For any set $P\subset\mathbb{Z}$ we write $f(P)=\sum_{x\in P}f(x)$. We observe that for any set P,

$$f_i(P)=\sum_{x\in P}f_i(x)=|\chi^{-1}(i)\cap P|-\frac{|P\cap[N]|}{c}.$$

If we write 1_P for the characteristic function of the set P, we can see easily that

$$f_i(a+P)=f_i*1_{-P}(a).$$

Now we take a prime p such that $2N<p<4N$. We observe that if $P\subset[N]$ and $a\notin\{-\frac{p-1}{2},\dots,\frac{p-1}{2}\}$ then $f_i(a+P)=0$ and we can apply Lemma 1 to the function

$f_i * 1_{-P}$ to get

$$\sum_{a\in\mathbb{Z}} |f_i * 1_{-P}(a)|^2 = \frac{1}{p}\sum_{a\in\mathbb{Z}_p} |\widehat{f_i * 1_{-P}}(a)|^2 = \frac{1}{p}\sum_{a\in\mathbb{Z}_p} |\hat{f}_i(a)|^2|\hat{1}_{-P}(a)|^2.$$

By Lemma 3 we can select, for any $a \in \mathbb{Z}_p$, a proper 2-arithmetic progression P_a such that $|\hat{1}_{-P_a}(a)| \geq p/400$. Thus,

$$\begin{aligned}
\sum_{x\in\mathbb{Z}_p} \sum_{-\frac{p-1}{2}\leq a\leq\frac{p-1}{2}} |f_i(a+P_x)|^2 &= \sum_{x\in\mathbb{Z}_p}\sum_{a\in\mathbb{Z}} |f_i * 1_{-P_x}(a)|^2 \\
&= \frac{1}{p}\sum_{a\in\mathbb{Z}_p} |\hat{f}_i(a)|^2 \sum_{x\in\mathbb{Z}_p} |\hat{1}_{-P_x}(a)|^2 \geq \frac{1}{p}\sum_{a\in\mathbb{Z}_p} |\hat{f}_i(a)|^2|\hat{1}_{-P_a}(a)|^2 \\
&\geq \left(\frac{p}{400}\right)^2 \frac{1}{p}\sum_{a\in\mathbb{Z}_p} |\hat{f}_i(a)|^2 = \left(\frac{p}{400}\right)^2 \sum_{a\in\mathbb{Z}_p} |f_i(a)|^2 \\
&= \left(\frac{p}{400}\right)^2 \left(\left(1-\frac{1}{c}\right)^2 |\chi^{-1}(i)| + \frac{1}{c^2}(N - |\chi^{-1}(i)|)\right) \\
&= \left(\frac{p}{400}\right)^2 \left(\left(1-\frac{2}{c}\right) |\chi^{-1}(i)| + \frac{1}{c^2}N\right).
\end{aligned}$$

Summing in all colors we obtain

$$\sum_{i=1}^{c}\sum_{x\in\mathbb{Z}_p}\sum_{a\in\{-\frac{p-1}{2},\dots,\frac{p-1}{2}\}} |f_i(a+P_x)|^2 \geq \left(\frac{p}{400}\right)^2\left(1-\frac{1}{c}\right)N \geq \frac{p^3}{8(400)^2}$$

Thus, there exists $a+P_x$ and a color i such that $|f_i(a+P_x)| \geq \frac{\sqrt{p}}{(8c)^{1/2}400} \geq \frac{\sqrt{N}}{c^{1/2}800}$ which completes the proof of Theorem 2.

References

[1] J. Beck, Roth's estimate of the discrepancy of integer sequences is nearly sharp, *Combinatorica*, **1**(4) (1981), 319-325.

[2] B. Doerr, A. Srivastav and P. Wehr, Discrepancies of cartesian products of arithmetic progressions, *Electronic Journal of Combinatorics.* **11** (1) (2004), R5.

[3] P. Erdős and J. Spencer, Probabilistic methods in Combinatorics, *Akademia Kiado, Budapest*, 1974.

[4] N. Hebbinghaus, Discrepancy of sums of arithmetic progressions, *Electronic Notes in Discrete Mathematics*, **17C** (2004), 185-189.

[5] N. Hebbinghaus, Discrepancy of sums of two arithmetic progressions, *Electronic Notes in Discrete Mathematics*, **29** (2007), 547-551.

[6] J. Matoušek, J. Spencer, Discrepancy in arithmetic progressions, *Journal of the American Mathematical Society*, **9**(1) (1996), 195-204.

[7] A. Přívětivý, Discrepancy of sums of three arithmetic progressions, *The Electronic Journal of Combinatorics*, **13**(1) (2006), R5.

[8] K. F. Roth, Remark concerning integer sequences, *Acta Arithmetica*, **9** (1964), 257-260.

[9] A. Sárközy, In P. Erdős and J. Spencer, editors, Probabilistic Methods in Combinatorics. Akadémia Kiadó, Budapest, 1974.

[10] B. Valkó, Discrepancy of arithmetic progressions in higher dimensions, *Journal of Number Theory*, **92** (2002), 117-130.

Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049-Madrid.
E-mail address: franciscojavier.cilleruelo@uam.es

Max-Planck-Institut für Informatik, Saarbrücken, Germany
E-mail address: nhebbing@mpi-inf.mpg.de