PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 143, Number 7, July 2015, Pages 2857–2863
S 0002-9939(2015)12545-2
Article electronically published on February 27, 2015

ON THE HEREDITARY DISCREPANCY OF HOMOGENEOUS ARITHMETIC PROGRESSIONS

ALEKSANDAR NIKOLOV AND KUNAL TALWAR

(Communicated by Harm Derksen)

Abstract. We show that the hereditary discrepancy of homogeneous arithmetic progressions is bounded from below by $n^{1/O(\log\log n)}$. This bound is tight up to a constant in the exponent. Our lower bound goes via an exponential lower bound on the discrepancy of set systems of subcubes of the boolean cube $\{0,1\}^d$.

1. Introduction

Circa 1932 Paul Erdős made the following conjecture:

Conjecture 1.1 ([4]). *For any function $f : \mathbb{N} \to \{-1,+1\}$ and for any constant C, there exist positive integers n and a such that*

$$|\sum_{i=1}^{\lfloor n/a \rfloor} f(ia)| > C.$$

This question can be phrased in the language of discrepancy theory as follows. For a positive integer parameter n, we consider the set system of subsets of $\{1,\dots,n\}$ given by arithmetic progressions of the form $(ia)_{i=1}^k$ for all positive integers $a \le n$ and $k \le \lfloor n/a \rfloor$. As is customary, we shall call such arithmetic progressions *homogeneous*. The *discrepancy* of a function $f : \{1,\dots,n\} \to \{-1,1\}$ for this set system is the maximum value of $|\sum_{i=1}^k f(ia)|$ over all a and k as above. The discrepancy of the set system of homogeneous arithmetic progressions over $\{1,\dots,n\}$ is the minimum achievable discrepancy over all functions $f : \{1,\dots,n\} \to \{-1,1\}$. In this language, Conjecture 1.1 states that the discrepancy of homogeneous arithmetic progressions is unbounded as n goes to infinity.

This problem is now known as the Erdős Discrepancy Problem, and stands as a major open problem in discrepancy theory and combinatorial number theory. Relying on a computer-aided proof, Konev and Lisitsa recently reported [8] that the discrepancy of homogeneous arithmetic progressions over $\{1,\dots,n\}$ is at least 3 for large enough n, and this remains the best known lower bound (a lower bound of 2 for $n \ge 12$ was well known). On the other hand, the function f which takes value $f(i) = -1$ if and only if the last nonzero digit of i in ternary representation is 2 has discrepancy $O(\log n)$. For references and other partial results related to the Erdős Discrepancy Problem, see [1, 5].

The Erdős Discrepancy Problem recently also received attention as the subject of the fifth polymath project [1]. Our note is motivated by results of Alon and

Received by the editors September 23, 2013 and, in revised form, March 27, 2014.
2010 *Mathematics Subject Classification.* Primary 11K38; Secondary 11B25.

Kalai, announced and sketched in the weblog post [6]. Using the Beck-Fiala theorem, they showed that even for homogeneous arithmetic progressions restricted to an *arbitrary subset* of the integers up to n, the discrepancy is no more than $n^{1/\Omega(\log\log n)}$. Also, for infinitely many n, they constructed a set of integers W_n all bounded by n, so that there is a set of homogeneous arithmetic progressions which, when restricted to W_n, form a known high discrepancy set system (the Hadamard set system). This construction showed that the minimum discrepancy for homogeneous arithmetic progressions restricted to W_n is at least $\Omega(\sqrt{\log n}/\sqrt{\log\log n})$. Since their discrepancy upper bound only uses a bound on the number of distinct homogeneous arithmetic progressions any integer less than n belongs to, it was reasonable to guess that the lower bound was closer to the truth.

In this note we show that in fact it is the upper bound of Alon and Kalai which is tight up to the constant in the exponent. Our main result is given by the following theorem.

Theorem 1.2. *For infinitely many positive integers n, there exists a set $W_n \subseteq \{1,\dots,n\}$ of square-free integers such that the following holds. For any $f : W_n \to \{-1,+1\}$ there exists a positive integer a so that*

$$|\sum_{\substack{b\in W_n\\ a|b}} f(b)| = n^{1/O(\log\log n)}.$$

Our construction of the sets W_n is inspired by the construction of Alon and Kalai. Instead of the Hadamard set system, we embed a set system of subcubes of the boolean cube inside the set of homogeneous arithmetic progressions. Such systems of boolean subcubes were previously considered in computer science in the context of private data analysis [7, 11], and by Chazelle and Lvov [2] as a tool to prove a polynomial lower bound on the discrepancy of axis-aligned boxes in high dimension. We give a new simpler proof of an improved lower bound on the discrepancy of boolean subcubes, using elementary Fourier analysis and the determinant lower bound on hereditary discrepancy due to Lovász, Spencer, and Vesztergombi [9].

Our construction produces sets W_n of square-free integers with a large number of prime divisors, suggesting that such integers are a chief obstacle in achieving bounded discrepancy for homogeneous arithmetic progressions.

2. Preliminaries

For a positive integer n, let $[n]$ be the set $\{1,\dots,n\}$. Given a set S, let $\binom{S}{k}$ be the set of cardinality k subsets of S. The expression $\langle\cdot,\cdot\rangle_2$ denotes the standard inner product over the vector space $\mathbb{F}_2^d$. We identify elements of $\mathbb{F}_2^d$ with the boolean cube $\{0,1\}^d$ in the natural way. We use $|v|$ for the Hamming weight of a vector v, i.e., $|v| = |\{i : v_i = 1\}|$.

A *set system* is defined as a pair $(\mathcal{S}, U)$, where $\mathcal{S} = \{S_1,\dots,S_m\}$ and $\forall j \in [m] : S_j \subseteq U$. The *restriction* $(\mathcal{S}|_W, W)$ of a set system $(\mathcal{S}, U)$ to some $W \subseteq U$ is defined by $\mathcal{S}|_W = \{S_1 \cap W,\dots,S_m \cap W\}$.

The discrepancy and hereditary discrepancy of a set system $\mathcal{S}$ are defined as

$$\operatorname{disc}(\mathcal{S}) = \min_{f:U\to\{-1,+1\}} \max_{j\in[m]} |\sum_{i\in S_j} f(i)|,$$
$$\operatorname{herdisc}(\mathcal{S}) = \max_{W\subseteq U} \operatorname{disc}(\mathcal{S}|_W).$$

The definitions of discrepancy and hereditary discrepancy can be extended to matrices $A \in \mathbb{R}^{m\times n}$ in a natural way. Analogously to the definition of a restriction of a set system, we define a restriction $A|_W$ of $A \in \mathbb{R}^{m\times n}$ for $W \subseteq [n]$ as the submatrix of columns of A indexed by elements of W. Then discrepancy and hereditary discrepancy for matrices are defined as

$$\operatorname{disc}(A) = \min_{x\in\{-1,+1\}^n} \|Ax\|_\infty,$$
$$\operatorname{herdisc}(A) = \max_{W\subseteq[n]} \operatorname{disc}(A|_W).$$

We will need the determinant lower bound on hereditary discrepancy, due to Lovász, Spencer, and Vesztergombi.

Theorem 2.1 ([9]). *For any real $m \times n$ matrix A,*

$$\operatorname{herdisc}(A) \geq \frac{1}{2} \max_k \max_B |\det(B)|^{1/k}$$

where, for any k, B ranges over all $k \times k$ submatrices of A.

3. Proof of the main theorem

Theorem 1.2 is a consequence of a lower bound on the hereditary discrepancy of the set system of subcubes of the boolean cube. Next we define this set system formally. For a positive integer d, we define the set system $(\mathcal{S}^d, \{0,1\}^d)$, where $\mathcal{S}^d = \{S_v\}_{v\in\{0,1,*\}^d}$ is defined by

$$S_v = \{u \in \{0,1\}^d : v_i \neq * \Rightarrow u_i = v_i\}.$$

Similar set systems were studied in computer science in relation to computing conjunction queries on a binary database under the constraint of differential privacy [7,11]. The system $\mathcal{S}^d$ was also considered by Chazelle and Lvov in their study of the discrepancy of high-dimensional axis-aligned boxes [2]. They used the trace bound [3] to prove that $\operatorname{herdisc}(\mathcal{S}^d) = \Omega(2^{cd})$ where c is a constant approximately equal to $c \approx 0.0477$. Here we slightly improve the constant in the exponent, and give a simpler proof using elementary Fourier analysis and the determinant lower bound.

Lemma 3.1. *For all positive integers d,* $\operatorname{herdisc}(\mathcal{S}^d) = \Omega(2^{d/16})$.

In the remainder of this section we prove that Lemma 3.1 implies Theorem 1.2. We prove Lemma 3.1 in the subsequent section.

Proof of Theorem 1.2. For each positive integer d, we will construct a set of integers B_d such that the hereditary discrepancy of homogeneous arithmetic progressions restricted to B_d is bounded from below by the hereditary discrepancy of $\mathcal{S}^d$. Then Theorem 1.2 will follow from Lemma 3.1.

Let $p_{1,0} < p_{1,1} < \ldots < p_{d,0} < p_{d,1}$ be the first $2d$ primes. We define B_d to be the following set of square free integers:

$$B_d = \{\prod_{i=1}^{d} p_{i,u_i} : u \in \{0,1\}^d\}.$$

In other words, B_d is the set of all integers that are divisible by exactly one prime $p_{i,b}$ from each pair $(p_{i,0}, p_{i,1})$ and no other primes. By the prime number theorem, the largest of these primes satisfies $p_{d,1} = \Theta(d \log d)$. Let $n = n(d)$ be the largest integer in B_d. The crude bound $n(d) = 2^{O(d \log d)}$ will suffice for our purposes. Notice that $d = \Omega(\log n / \log\log n)$.

There is a natural one-to-one correspondence between the set B_d and the set $\{0,1\}^d$: to each $u \in \{0,1\}^d$ we associate the integer $b_u = \prod_{i=1}^{d} p_{i,u_i}$. By this correspondence, we can think of any assignment $f : \{0,1\}^d \to \{-1,+1\}$ as an assignment $f : B_d \to \{-1,+1\}$. We also claim that each set in the set system $\mathcal{S}^d$ corresponds to a homogeneous arithmetic progression restricted to B_d. With any $S_v \in \mathcal{S}^d$ (where $v \in \{0,1,*\}^d$) associate the integer $a_v = \prod_{i:v_i \neq *} p_{i,v_i}$. Observe that for any $b_u \in B_d$, a_v divides b_u if and only if $u \in S_v$. We have the following implication for any assignment f, any $U \subseteq \{0,1\}^d$, and the corresponding $W = \{b_u : u \in U\}$:

$$\exists S_v : |\sum_{u \in S_v \cap U} f(u)| \geq D \quad \Leftrightarrow \quad \exists a \in \mathbb{N} : |\sum_{\substack{b \in W \\ a|b}} f(b)| \geq D. \tag{3.1}$$

Notice again that we treat f as an assignment both to elements of $\{0,1\}^d$ and to integers in B_d by the correspondence $u \leftrightarrow b_u$. Lemma 3.1 guarantees the existence of some U such that the left hand side of (3.1) is satisfied with $D = 2^{\Omega(d)} = n^{1/O(\log\log n)}$ for any f. Theorem 1.2 follows from the right hand side of (3.1). $\square$

4. Lower bounding the discrepancy of $\mathcal{S}^d$

It is convenient to first prove an easier lower bound on the hereditary discrepancy of low-weight characters of $\mathbb{F}_2^d$. Then we show that an exponential (in d) lower bound on the discrepancy of characters of weight $d/8$ implies an exponential lower bound on $\mathcal{S}^d$. This approach is inspired by the noise lower bounds on differential privacy in [7].

As usual, for $v \in \mathbb{F}^d$ we define the character χ_v by

$$\forall u \in \mathbb{F}^d : \chi_v(u) = (-1)^{\langle v,u \rangle_2}.$$

We refer to the Hamming weight $|v|$ of v (taken as a binary vector) as the *weight* of the character χ_v. The matrix of the Walsh-Hadamard transform is defined as $H_d = (\chi_v)_{v \in \mathbb{F}^d}$, where each χ_v is written as a row vector of dimension 2^d. Notice that for any $v \neq w$, $\sum_{u \in \mathbb{F}^d} \chi_v(u)\chi_w(u) = 0$, i.e., H_d is an orthogonal matrix; each row of H_d has squared Euclidean norm $\sum_{u \in \mathbb{F}^d} \chi_v(u)^2 = 2^d$.

We will be interested in a submatrix of H_d. For the remainder of this note we assume that d is divisible by 8; this is purely for notational convenience: our arguments can easily be adapted to the case when d is not divisible by 8. Let $G_d = (\chi_v)_{v:|v|=d/8}$. Notice that $G_d G_d^T = 2^d I_M$ where $M = \binom{d}{d/8}$ and I_M is the M-dimensional identity matrix. Therefore,

$$\det(G_d G_d^T) = (2^d)^{\binom{d}{d/8}}. \tag{4.1}$$

Given (4.1) and using the determinant lower bound, we can derive a lower bound on the hereditary discrepancy of G_d.

Lemma 4.1. *For positive integers d,*

$$\operatorname{herdisc}(G_d) \geq \frac{2^{3d/16}}{2e}.$$

Proof. Let $N = 2^d$ and let $M = \binom{d}{d/8}$. By (4.1) and the Binet-Cauchy formula for the determinant, we have

$$N^M = \det(G_d G_d^T) = \sum_{W \in \binom{[N]}{M}} \det(G_d|_W)^2.$$

By averaging, there exists a set $W \in \binom{[N]}{M}$ so that

$$|\det(G_d|_W)|^{1/M} \geq \sqrt{N}\binom{N}{M}^{-1/2M} \geq \sqrt{\frac{M}{e}}. \tag{4.2}$$

For the second inequality above we used the bound $\binom{N}{M} \leq (Ne/M)^M$. Plugging in the lower bound $M = \binom{d}{d/8} \geq 2^{3d/8}$ in (4.2), we have $|\det(G_d|_W)|^{1/M} \geq 2^{3d/16}e^{-1}$. The proof is completed by an application of Theorem 2.1. □

We are now ready to prove Lemma 3.1 by exhibiting a connection between the discrepancy of G_d and the discrepancy of $\mathcal{S}^d$.

Proof of Lemma 3.1. By Lemma 4.1, it is enough to prove the following inequality:

$$\operatorname{herdisc}(G_d) \leq 2^{d/8} \operatorname{herdisc}(\mathcal{S}^d). \tag{4.3}$$

The key observation is that when $|v| = d/8$, we can express the character χ_v as a linear combination of the indicator functions of $2^{d/8}$ sets in $\mathcal{S}^d$. Moreover, the coefficients of the linear combination are ± 1. Next we make this observation precise.

Let v be an arbitrary fixed element of $\mathbb{F}^d$ such that $|v| = d/8$, and let $1 \leq i_1 < i_2 < \ldots < i_{d/8} \leq d$ denote the coordinates i such that $v_i = 1$. Given $w \in \{0,1\}^{d/8}$, let its *extension* $v(w) = \{0, 1, *\}^d$ be defined by

$$v(w)_i = \begin{cases} w_\ell & \text{if } i = i_\ell \text{ for some } \ell \in [d/8]; \\ * & \text{otherwise.} \end{cases}$$

We use the notation $\mathbf{1}_{v(w)}$ for the indicator function of the set $S_{v(w)}$. Let $r_{v(w)}$ be a representative from the set $S_{v(w)}$, say one obtained by replacing every $*$ in $v(w)$ with 0. Taking $r_{v(w)}$ and the elements of $S_{v(w)}$ as elements of $\mathbb{F}_2^d$ in the standard way, for each $z \in S_{v(w)}$, $\langle v, z\rangle_2 = \langle v, r_{v(w)}\rangle_2$, since only the coordinates $z_{i_1}, \ldots, z_{i_{d/8}}$ affect the inner product. Thus, we can express $\chi_v(u)$ as the linear combination of these indicator functions:

$$\forall u \in \{0,1\}^d : \chi_v(u) = \sum_{w \in \{0,1\}^{d/8}} (-1)^{\langle v, r_{v(w)}\rangle_2} \mathbf{1}_{v(w)}(u). \tag{4.4}$$

For any set $U \subseteq \{0,1\}^d$ and any $f : U \to \{-1,1\}$, we use (4.4) to write the linear transformation $(G_d|_U)f$ in terms of discrepancy values of sets in $\mathcal{S}^d$ restricted to the set U:

$$\sum_{u\in U} \chi_v(u) f(u) = \sum_{u\in U} \left(\sum_{w\in\{0,1\}^{d/8}} (-1)^{\langle v, r_{v(w)}\rangle_2} \mathbf{1}_{v(w)}(u) \right) f(u)$$

$$= \sum_{w\in\{0,1\}^{d/8}} (-1)^{\langle v, r_{v(w)}\rangle_2} \left(\sum_{u\in S_{v(w)}\cap U} f(u) \right). \tag{4.5}$$

Let f be the function that achieves $\mathrm{disc}(\mathcal{S}^d|_U)$. Each of the $2^{d/8}$ terms on the right hand side of (4.5) is then bounded in absolute value by $\mathrm{disc}(\mathcal{S}^d|_U) \le \mathrm{herdisc}(\mathcal{S}^d)$. Since the choice of U and v was arbitrary, this proves (4.3), and the lemma follows. □

5. Conclusion

We presented a tight (up to the constant in the exponent) lower bound on the hereditary discrepancy of homogeneous arithmetic progressions. Our lower bound instances are given by sets of integers in $\{1,\dots,n\}$ with $\Theta(\log n/\log\log n)$ distinct prime factors. This suggests that integers with many distinct factors are the main obstacle to achieving bounded discrepancy for homogeneous arithmetic progressions.

Our discrepancy lower bound follows from a lower bound on the discrepancy of a set system of subcubes of the boolean cube. Such set systems have applications in the theory of differential privacy. The ideas used in the proof of Lemma 3.1, together with the connection between discrepancy and differential privacy formalized in [10] can be used to give simpler proofs of noise lower bounds of the type considered in [7]. It is an interesting question whether discrepancy bounds on set systems of boolean subcubes can find other applications in combinatorics and computer science. We leave open the question of characterizing the exact discrepancy of such set systems.

References

[1] Multiple Authors, *Erdős discrepancy problem: Polymath wiki*, `http://michaelnielsen.org/polymath1/index.php?title=The_Erd%C5%91s_discrepancy_problem`.

[2] B. Chazelle and A. Lvov, *The discrepancy of boxes in higher dimension*, Discrete Comput. Geom. **25** (2001), no. 4, 519–524, DOI 10.1007/s00454-001-0014-2. The Micha Sharir birthday issue. MR1838418 (2002f:11097)

[3] B. Chazelle and A. Lvov, *A trace bound for the hereditary discrepancy*, Discrete Comput. Geom. **26** (2001), no. 2, 221–231, DOI 10.1145/336154.336179. ACM Symposium on Computational Geometry (Hong Kong, 2000). MR1843438 (2003b:68190)

[4] Paul Erdős, *Some unsolved problems*, Michigan Math. J. **4** (1957), 291–300. MR0098702 (20 #5157)

[5] Steven Finch, *Two-colorings of positive integers*, `http://www.people.fas.harvard.edu/~sfinch/csolve/ec.pdf`.

[6] Gil Kalai, *Erdős discrepancy problem 22*, `http://gowers.wordpress.com/2012/08/22/edp22-first-guest-post-from-gil-kalai/`, 09 2012.

[7] Shiva Prasad Kasiviswanathan, Mark Rudelson, Adam Smith, and Jonathan Ullman, *The price of privately releasing contingency tables and the spectra of random matrices with correlated rows*, STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, ACM, New York, 2010, pp. 775–784. MR2743327 (2011i:94095)

[8] Boris Konev and Alexei Lisitsa, *A sat attack on the erdős discrepancy conjecture*, CoRR **abs/1402.2184** (2014).

[9] L. Lovász, J. Spencer, and K. Vesztergombi, *Discrepancy of set-systems and matrices*, European J. Combin. **7** (1986), no. 2, 151–160, DOI 10.1016/S0195-6698(86)80041-5. MR856328 (88b:05036)

[10] S. Muthukrishnan and Aleksandar Nikolov, *Optimal private halfspace counting via discrepancy*, STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing, ACM, New York, 2012, pp. 1285–1292, DOI 10.1145/2213977.2214090. MR2961660

[11] Mark Rudelson, *Row products of random matrices*, Adv. Math. **231** (2012), no. 6, 3199–3231, DOI 10.1016/j.aim.2012.08.010. MR2980497

Department of Computer Science, Rutgers University, Piscataway, New Jersey 08854
Current address: Microsoft Research, Redmond, Washington 98052
E-mail address: alenik@microsoft.com

Microsoft Research, Mountain View, California 94043
Current address: Google, Mountain View, California 94043
E-mail address: kunal@kunaltalwar.org