Contemporary Mathematics
Volume **223**, 1999

Discrepancy Bounds for Geometric Set Systems with Square Incidence Matrices

Bernard Chazelle

ABSTRACT. Alexander has proven the existence of a set of n points in the plane such that, given any two-coloring of the points, there exists a halfplane within which one color outnumbers the other by $\Omega(n^{1/4})$. We strengthen this result by showing that the halfplane can be chosen among n fixed ones. In other words, we build a point/halfplane set system of discrepancy $\Omega(n^{1/4})$, whose incidence matrix is $n \times n$. By a result of Matoušek, this lower bound is tight.

The second result is an $n \times n$ variant of a classical lower bound of Roth on the discrepancy of arithmetic progressions. Stated in dual form, our result asserts the existence of a set system of discrepancy $\Omega(n^{1/4})$, whose $n \times n$ incidence matrix $(a_{i,j})$ is formed as follows: each row (resp. column) corresponds to a line segment (resp. horizontal line) in the plane, and $a_{i,j} = 1$ if segment i and line j intersect in an integer point. Matoušek and Spencer have shown the lower bound to be tight.

1. Introduction

Schmidt [**S**] has shown the existence of n points in the plane such that, given any two-coloring of the points, there is always an axis-parallel box within which one color outnumbers the other by $\Omega(\log n)$. It is natural to ask the question: can the box be chosen among a small set of prespecified boxes? Given an incidence matrix A of a set system, we define the discrepancy of A,

$$D(A) = \Big\{ \min \|Ax\|_\infty \,:\, x \in \{-1, 1\} \Big\}.$$

Is there an $n \times n$ incidence matrix for boxes, $A = (a_{i,j})$, of discrepancy $\Omega(\log n)$, such that $a_{i,j} = 1$ if the box associated with row i contains the point associated with column j? As it happens, the answer is trivially affirmative. Here is why: the set of candidate boxes for Schmidt's bound can obviously be restricted to $O(n^4)$. So, by applying the bound for $n' = O(n^{1/4})$ points, we can make the number of candidates equal to n. Next, we form an $n \times n$ incidence matrix A by filling in $n - n'$ columns of zeroes derived from the addition of $n - n'$ dummy points. The

1991 *Mathematics Subject Classification.* Primary 68U05, 68Q15.

Key words and phrases. Discrepancy theory, geometric set systems, eigenvalue bounds.

Supported in part by NSF Grant CCR-96-23768 and by the US Army Research Office under Grant DAAH04-96-1-0181.

discrepancy of the resulting set system is $\Omega(\log n') = \Omega(\log n)$. Of course, this trick does not work if the discrepancy is of the form n^α, for some constant α. By using a more subtle argument, we prove the following:

THEOREM 1.1. *There exist n points and n halfplanes in $\mathbf{R}^2$, such that the $n\times n$ incidence matrix $A = (a_{i,j})$ has discrepancy $D(A) = \Omega(n^{1/4})$, where $a_{i,j} = 1$ if and only if the i-th halfplane contains the j-th point.*

THEOREM 1.2. *There exist n horizontal lines and n segments in $\mathbf{R}^2$, such that the $n \times n$ incidence matrix $A = (a_{i,j})$ has discrepancy $D(A) = \Omega(n^{1/4})$, where each row (resp. column) corresponds to a segment (resp. horizontal line), and $a_{i,j} = 1$ if and only if segment i and line j intersect in an integer point.*

Both theorems are optimal. This follows from results by Matoušek [**M**] and Matoušek and Spencer [**MS**], respectively. It remains an open question whether the same bounds hold for $m \times n$ matrices, where m is asymptotically smaller than n.

2. The proof of Theorem 1.1

Let P be the set of n integer points in $[1, \sqrt{n}]^2$ (assume that n is a large square), and let x be a vector in $\mathbf{R}^n$ whose i-th coordinate x_i is associated with $p_i \in P$. Given a closed halfplane h bounded above by a nonvertical line, we define $f(h) = \sum_{p_i\in h} x_i$. Let ω be the motion-invariant measure for lines, normalized so as to provide a probability measure for the lines crossing the square $[1, \sqrt{n}]^2$. Alexander [**A**] has proven that if $x_1 + \cdots + x_n = 0$, then

$$\int f^2(h)\, d\omega(h) = \Omega(1/\sqrt{n})\, \|x\|_2^2.$$

For completeness, we repeat (with only a few modifications) the argument we used in [**C**] to discretize Alexander's result and derive Lemma 2.1 below. We subdivide the space of lines crossing $[1, \sqrt{n}]^2$ into $N+O(n^2)$ regions within which $f(h)$ remains invariant. By choosing n and N large enough, say, $N = 2^n$, we can easily ensure that the ω-area σ of N of these regions is exactly the same, i.e., about $1/N$, while the other $O(n^2)$ regions have smaller areas. Thus, the error produced in computing $\int f^2(h)\, d\omega(h)$ by integrating f^2 only over the equal-area regions is $O(n^2/N) \sup f^2$. Because $|f|$ cannot exceed

$$|x_1| + \cdots + |x_n| \le \sqrt{n}\, \|x\|_2,$$

this error is bounded by $O(n^3\|x\|_2^2/N)$. Let B be the $N \times n$ matrix whose rows are indexed by the N equal-area regions $\hat\sigma$ and are the characteristic vectors of the set of x_i's appearing in (the unique form) $f(h)$, for $h \in \hat\sigma$. We have

$$\left| \|Bx\|_2^2 - \frac{1}{\sigma} \int f^2(h)\, d\omega(h) \right| = O(n^3) \frac{\|x\|_2^2}{N\sigma},$$

and because $\sigma = 1/N \pm O(n^2/N^2)$,

$$\left| \|Bx\|_2^2 - N \int f^2(h)\, d\omega(h) \right| = O(n^3\|x\|_2^2).$$

LEMMA 2.1. [**C**]

$$\det B^T B = \Omega\Big(N\Big/\sqrt{n}\Big)^{n-1}.$$

PROOF. Let $\mu_1 \geq \cdots \geq \mu_n \geq 0$ be the eigenvalues of $B^T B$ and let $\{v_i\}$ be an orthonormal eigenbasis, with μ_i corresponding to v_i. Let $(\xi_1, \ldots, \xi_n)$ be the coordinates of x in the basis $\{v_i\}$. The solution space of the system of equations, $x_1 + \cdots + x_n = 0$ and $\xi_j = 0$ $(j < n-1)$, is of dimension at least 1. It lies in the (ξ_{n-1}, ξ_n)-plane, so it intersects the cylinder $\xi_{n-1}^2 + \xi_n^2 = 1$. For any point x of the intersection,

$$\|Bx\|_2^2 = \sum_{i=1}^{n} \mu_i \xi_i^2 = \mu_{n-1}\xi_{n-1}^2 + \mu_n \xi_n^2 \leq \mu_{n-1}.$$

This implies that for the unit vector x,

$$\mu_{n-1} \geq N \int f^2(h)\, d\omega(h) - O(n^3 \|x\|_2^2) \geq \Omega(N/\sqrt{n}) - O(n^3),$$

and hence,

$$\mu_{n-1} \geq \Omega(N/\sqrt{n}). \tag{1}$$

We need a lower bound on the smallest eigenvalue. With N being large enough, we can always assume that, for each point p_i, there exist two lines (adding them on, if necessary, and updating N accordingly), each represented by a distinct row of B, that pass right above and below p_i. The contribution of these two rows to $\|Bx\|_2^2$ is of the form $\Phi^2 + (\Phi + x_i)^2$, which is always at least $x_i^2/2$. It follows that $\|Bx\|_2^2 \geq \frac{1}{2}\|x\|_2^2$, and hence, $\mu_n \geq 1/2$. The lemma follows from (1) and the fact that $\det B^T B$ is the product of the eigenvalues.

By the Binet-Cauchy formula,[1]

$$\det B^T B = \sum_{1 \leq j_1 < \cdots < j_n \leq N} \left| \det B \begin{pmatrix} j_1 & j_2 & \ldots & j_n \\ 1 & 2 & \ldots & n \end{pmatrix} \right|^2.$$

Therefore, there exists an $n \times n$ submatrix A of B such that

$$\begin{aligned} (\det A)^2 &= \left| \det B \begin{pmatrix} j_1 & j_2 & \ldots & j_n \\ 1 & 2 & \ldots & n \end{pmatrix} \right|^2 \\ &\geq \binom{N}{n}^{-1} \det B^T B = \Omega(1)^n \left(\frac{n}{eN}\right)^n \left(\frac{N}{\sqrt{n}}\right)^{n-1} \\ &\geq (cn)^{n/2}, \end{aligned}$$

for some fixed $c > 0$. Lovász et al. [**LSV**] define the hereditary discrepancy, $D^H(A)$, of the incidence matrix A to be the maximum value of $D(A')$ over all matrices A' formed by subsets of the columns of A. They prove that

$$D^H(A) = \Omega(|\det A|^{1/n}).$$

In our case, this implies that

$$D^H(A) = \Omega(n^{1/4}).$$

Let A' be the (or any) submatrix of A that achieves the hereditary discrepancy, and let M be the matrix derived from A by zeroing out the columns not in the submatrix A'. By introducing artificial points if necessary, we can make M the incidence matrix of a point/halfplane set system, whose discrepancy is thus $\Omega(n^{1/4})$. This completes the proof of Theorem 1.1. □

[1] The notation following $\det B$ refers to the matrix obtained by picking the rows indexed $j_1, \ldots, j_n$ in B.

For reference it might be useful to make a general lemma out of the technique we just used.

LEMMA 2.2. *If B is an $N \times n$ incidence matrix of a set system, then there exists an $n \times n$ matrix A formed by n rows of B, such that*

$$D^H(A) \geq c\sqrt{\frac{n}{N}}\left(\det B^T B\right)^{1/2n},$$

for some constant $c > 0$.

3. The proof of Theorem 1.2

PROOF. Let B be the $N \times n$ incidence matrix of the following set system: each set is an arithmetic progression modulo m, of length k and difference at most $6k$, where $k = \lfloor\sqrt{n/6}\rfloor$; note that $N = O(n\sqrt{n})$. By adapting an argument of Roth [**R**], Beck and Sós [**BS**] have shown that the matrix $B^T B$ has all its eigenvalues in $\Omega(n)$. This implies that

$$\det B^T B = \Omega(n)^n.$$

By Lemma 2.2, we derive the existence of an $n \times n$ submatrix A, such that

$$D^H(A) = \Omega(n^{1/4}).$$

This result can be interpreted in terms of arithmetic progressions, but it is perhaps better grasped in dual space. Since arithmetic progressions are considered modulo n, a row of A might consist of two distinct progressions. By doubling the number of rows if necessary we can make them into regular arithmetic progressions. Let n denote the new number of rows. If $b, a+b, \ldots, ka+b$ is the progression associated with row i, let us now associate with that row the segment on line $Y = aX + b$ running from $X = 0$ to $X = k$. Column j is associated with line $Y = j$. The lower bound on the hereditary discrepancy implies that the restriction of the set system to a certain subset of lines has discrepancy $\Omega(n^{1/4})$. By zeroing out the leftover columns and adding dummy column lines, we thus create an $n \times n$ incidence matrix of discrepancy $\Omega(n^{1/4})$, where element $a_{i,j}$ is 1 if and only if segment i and line j intersect in an integer point. This proves Theorem 1.2. □

Acknowledgments. I wish to thank the referees for their constructive comments.

References

[A] R. Alexander, *Geometric methods in the study of irregularities of distribution*, Combinatorica, 10 (1990), 115–136.

[BC] J. Beck and W. W. L. Chen, *Irregularities of Distribution*, Cambridge Tracts in Mathematics, 89, Cambridge Univ. Press, Cambridge, 1987.

[BS] J. Beck and V. T. Sós, *Discrepancy theory*, in "Handbook of Combinatorics," eds. R. L. Graham, M. Grötschel and L. Lovász, MIT Press (1995), chap. 26, 1405–1446.

[C] B. Chazelle, *A spectral approach to lower bounds with applications to geometric searching*, SIAM J. Comput., to appear. Preliminary version in Proc. 35th IEEE Symp. Found. Comp. Sci. (1994), 674–682.

[LSV] L. Lovász, J. Spencer and K. Vesztergombi, *Discrepancy of set systems and matrices*, European Journal of Combinatorics, 7 (1986), 151–160.

[M] J. Matoušek, *Tight upper bounds for the discrepancy of halfspaces*, Disc. Comput. Geom., 13 (1995), 593–601.

[MS] J. Matoušek and J. Spencer, *Discrepancy in arithmetic progressions*, J. AMS, 9 (1996), 195–204.

[S] W. M. Schmidt, *Irregularities of distribution, VII*, Acta Arithmetica, 21 (1972), 45–50.
[R] K. F. Roth, *Remark concerning integer sequences*, Acta Arithmetica, 9 (1964), 257–260.

Department of Computer Science, Princeton University Princeton, NJ 08544
E-mail address: chazelle@cs.princeton.edu