Discrete Comput Geom 13:363–381 (1995)

Discrete & Computational
Geometry

An Elementary Approach to Lower Bounds in Geometric Discrepancy*

B. Chazelle,[1] J. Matoušek,[2] and M. Sharir[3]

[1]Department of Computer Science, Princeton University, Princeton, NJ 08544, USA chazelle@cs.princeton.edu

[2]Department of Applied Mathematics, Charles University, 11800 Praha 1, Czech Republic matousek@kam.mff.cuni.cz

[3]School of Mathematical Sciences, Tel Aviv University, Ramat Aviv 69 978, Israel sharir@math.tau.ac.il
and
Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

Abstract. For each $d \geq 2$, it is possible to place n points in d-space so that, given any two-coloring of the points, a half-space exists within which one color outnumbers the other by as much as $cn^{1/2-1/2d}$, for some constant $c > 0$ depending on d. This result was proven in a slightly weaker form by Beck and the bound was later tightened by Alexander. It was recently shown to be asymptotically optimal by Matoušek. We present a proof of the lower bound, which is based on Alexander's technique but is technically simpler and more accessible. We present three variants of the proof, for three different cases, to provide more intuitive insight into the "large-discrepancy" phenomenon. We also give geometric and probabilistic interpretations of the technique.

* Work by Bernard Chazelle has been supported in part by NSF Grant CCR-90-02352 and The Geometry Center, University of Minnesota, an STC funded by NSF, DOE, and Minnesota Technology, Inc. Work by Jiří Matoušek has been supported by Charles University Grant No. 351, by Czech Republic Grant GAČR 201/93/2167 and in part by DIMACS. Work by Micha Sharir has been supported by NSF Grant CCR-91-22103, by a Max-Planck Research Award, and by grants from the U.S.–Israeli Binational Science Foundation, the Fund for Basic Research administered by the Israeli Academy of Sciences, and the G.I.F., the German–Israeli Foundation for Scientific Research and Development.

1. Introduction

The following result was established by Alexander [1]. A slightly weaker bound was derived earlier by Beck [4].

Theorem 1.1 [1]. *For each $d \geq 2$ and for arbitrarily large values of n, there are sets P of n points in E^d such that, given any two-coloring of P, a half-space exists within which one color outnumbers the other by at least $cn^{1/2-1/2d}$, where $c = c(d)$ is a positive constant.*

A recent result of Matoušek [8] asserts that this lower bound is optimal. That is, for any set P of n points in E^d a two-coloring of P exists such that, within any half-space, no color outnumbers the other by more than $c'n^{1/2-1/2d}$, for some other constant $c' = c'(d) > 0$. Slightly higher upper bounds were established earlier by Matoušek *et al.* [9], and, for the case $d = 2$, by Beck [3]. Alexander also established a "continuous" version of the theorem. Given a half-space τ, let $\mathrm{Vol}(\tau)$ denote the Lebesgue measure of $\tau \cap [0,1]^d$.

Theorem 1.2 [1], [2]. *For each $d \geq 2$, given any set P of n points in $[0,1]^d$, a half-space τ exists such that*

$$||P \cap \tau| - n \cdot \mathrm{Vol}(\tau)| > cn^{1/2-1/2d},$$

for some constant $c = c(d) > 0$.

We present alternative proofs of these theorems. We borrow Alexander's basic approach and add a few ideas to produce proofs which are technically simpler than the original ones and completely elementary. A crucial new ingredient of the proof is a finite-differencing technique that appears to be quite general. We break down our exposition into three sections. First, we treat the two-dimensional case of Theorem 1.1. We give a short, self-contained proof, and we provide an intuitive probabilistic interpretation of the technique.

Although the proof can be generalized directly to higher dimensions, we simplify our basic approach still further, and obtain, in Section 3, a slightly more straightforward proof for the general case. Strictly speaking, treating the two-dimensional case separately is unnecessary (so the reader may skip Section 2 if he or she so desires), but we do so because it has an illuminating geometric interpretation. In Section 4 we treat the continuous case (Theorem 1.2). Here yet another variation of the technique is used, which is maybe somewhat less intuitive compared with the one from Section 3, but gives considerably simpler calculations. We end the paper, in Section 5, with a discussion comparing our proof technique with the classical Fourier transform approach (see, e.g., [4]).

2. Red–Blue Discrepancy in the Plane

Given n points in a two-dimensional $\sqrt{n} \times \sqrt{n}$ square grid, we show that, no matter how we color them red and blue, a half-plane within which one color outnumbers

the other by at least $cn^{1/4}$, for certain appropriate constant c, always exists. Actually, we prove the stronger result that $\Omega(n^{1/2})$ is a lower bound on the mean square discrepancy. To explain what this means, we must first define the notion of a random half-plane or, equivalently, of a random line.

As is well known (see [11]), a measure ω exists on the set of all lines in the plane, which is invariant under the group of motions in E^2. A line h (assumed from now on not to pass through the origin) is uniquely specified by its closest point to the origin. If ρ, θ are the polar coordinates of that point, the desired invariant measure at h is $d\omega(h) = d\rho\, d\theta$ [11]. Cauchy's theorem says that the measure of the set of lines intersecting a convex set A is equal to the perimeter of A.

Let $\mathscr{U} = [0, \frac{1}{4}]^2$. By Cauchy's theorem, the measure of the set $\mathscr{U}^*$ of lines crossing $\mathscr{U}$ is exactly 1, so, from now on, a *random* line h is to be understood as a random element of $\mathscr{U}^*$, with respect to the probability distribution ω.

Without loss of generality, assume that $n = m^2$ for some integer $m > 0$. We define the point set P to be the set of interior vertices of an $(m + 1) \times (m + 1)$ grid placed inside $\mathscr{U}$, so that its four corners coincide with those of $\mathscr{U}$. Fix a two-coloring, $P = R \sqcup B$, once and for all. Given a nonvertical line h, we define

$$D(h) = |R \cap h^+| - |B \cap h^+|$$

to be the *discrepancy* of the (closed) half-plane h^+ lying above h. We prove the following lower bound on the expected value of $D^2(h)$, from which Theorem 1.1 follows immediately:

$$\mathbf{E}_h D^2 = \Omega(\sqrt{n}), \tag{1}$$

where $\mathbf{E}_h$ denotes expectation over h. (Actually, (1) holds for more general "well-spread" sets of points; see a remark at the end of this section for details.)

2.1. *The Intuition*

Before we give a proof of (1), let us informally address the simple question: why can't the discrepancy be zero everywhere? Suppose, for the sake of contradiction, that $D(h^+)$ is zero for all $h \in \mathscr{U}^*$. (Of course, we know that it must be at least ± 1 somewhere, so the reader might want to think of zero as just meaning "very small.")

Let slab(h) denote a slab of fixed width w (where w is about $1/m$) centered along the line h. Zero discrepancy implies that, within each of the two half-slabs, $\text{slab}^-(h) = h^- \cap \text{slab}(h)$, $\text{slab}^+(h) = h^+ \cap \text{slab}(h)$, red and blue points match perfectly, which we write as

$$R_h^+ = B_h^+ \quad \text{and} \quad R_h^- = B_h^-, \tag{2}$$

where $R_h^+ = |R \cap \text{slab}^+(h)|$, and the other three quantities are defined analogously.

Pick two random distinct points $p, q \in P \cap \text{slab}(h)$ (see Fig. 1). If pq crosses h, then, by (2) p and q are equally likely to be bichromatic or monochromatic; we write this as $BC = MC$, which stands for: "probability of bichromatic crossing =

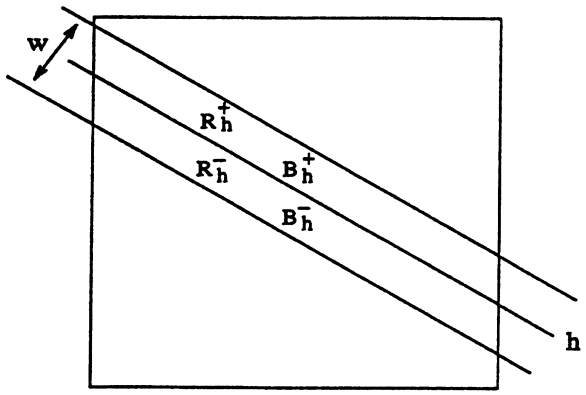

Fig. 1. Assuming no discrepancy: $R_h^+ = B_h^+$ and $R_h^- = B_h^-$.

probability of monochromatic crossing." On the other hand, if pq does not cross h, then $BN > MN$, meaning that the probability that pq is bichromatic noncrossing exceeds that of being monochromatic noncrossing. This is the key inequality driving the whole argument, so it is crucial to understand it. Without loss of generality, suppose that p is blue and lies in h^+. Then q is picked randomly among $B_h^+ - 1$ blue points (-1 because $p \neq q$) and R_h^+ red points. Since $B_h^+ = R_h^+$, the point q is thus a tiny bit more likely to be red than blue.

Adding the two relations $BC = MC$ and $BN > MN$ yields $BN + MC > BC + MN$, and, hence, $BN > BC$ or $MC > MN$. Suppose that $BN > BC$ (the same argument applies to the other case as well). The inequality says that a random bichromatic segment pq ($p, q \in P$) that falls inside a random slab(h) is more likely to avoid h than to cross it. The difference in probabilities is a positive number δ that is relatively large. (At least let us think of it that way.) However, throwing a random slab on top of pq is similar to throwing a random needle of length $|p - q|$ inside the slab. This is the Buffon needle experiment [6], which tells us that long (random) segments in slab(h) are almost as likely to cross the line h as to avoid it. This immediately leads to a contradiction. Indeed, since pq are grid points, a random pq is, on the average, very long. So, by Buffon's result, its probability of crossing h cannot exceed that of avoiding h by as much as δ, and we have a contradiction.

It might seem truly amazing that such a weak argument can produce any kind of nontrivial lower bound, let alone an optimal one. The argument rests totally on the fact that the trivial segment pp is forbidden from the probabilistic experiment (and on the fact that the other segments pq are "sufficiently long"). We show in the next section that splitting the slab in half is a special case of a general differencing scheme.

Why the bound $\Omega(n^{1/4})$? Intuitively, the discrepancy of a half-plane is "concentrated" along its bounding line within a thin strip of width roughly $1/\sqrt{n}$. Indeed, it is possible to subdivide the grid into $O(n)$ squares, each containing a large but constant number of points, and require that the discrepancy within each such square be zero: this restriction does not limit much the freedom the adversary has in

coloring the points. On the other hand, it creates a jagged strip along the bounding line that contains roughly $\sqrt{n}$ points and has exactly the same discrepancy as the half-plane. Now, if the coloring were random, then with high probability the discrepancy within the strip in question would be roughly $n^{1/4}$ (the standard deviation for $\sqrt{n}$ independent coin tosses). Of course, the coloring is not random, the half-plane is. The crux of the argument is to show that a random half-plane provides the same effect as a random coloring (at least as far as the standard deviation is concerned). This is quite remarkable because there is much "less" randomness in choosing the half-plane than in choosing a random coloring. Indeed, there are only $\binom{n}{2}$ combinatorially distinct configurations obtained by intersecting the set with a half-plane, whereas the number of colorings is 2^n.

2.2. *The Proof*

Given a (nonvertical) line h crossing $\mathscr{Z}$, let h_1 and h_2 be the two lines parallel to h at a distance $w/2 = c_0/m$ from h, for a small constant $c_0 > 0$, with h_1 lying above h_2; these lines bound slab(h), as defined above. Define

$$D_w(h) \stackrel{\text{def}}{=} D(h_2) - 2D(h) + D(h_1). \tag{3}$$

This measures the difference in discrepancy between the half-slabs slab$^-(h)$ and slab$^+(h)$. We consider the average of $D_w^2(h)$ over all positions of the slab obtained by varying h in $\mathscr{Z}^*$. By the Cauchy–Schwarz inequality,

$$D_w^2(h) \le 6(D^2(h_2) + D^2(h) + D^2(h_1)),$$

and therefore

$$\mathbf{E}_h D_w^2 \le 18\mathbf{E}_h D^2. \tag{4}$$

It thus suffices to establish a lower bound on the L^2-norm of D_w, which, as it turns out, is much easier. By expressing $D_w(h)$ in unary (giving red points weight $+1$ and blue points weight -1), we can pair and classify the individual components of $D_w^2(h)$. That is, we have

$$D_w^2(h) = \left(\sum_{R_1} 1 - \sum_{B_1} 1 - \sum_{R_2} 1 + \sum_{B_2} 1\right)^2,$$

where R_1 is the set of red points in the upper half-slab, B_2 is the set of blue points in the lower half-slab, etc. By expanding the right-hand side, we see that companion pairs sum up to

$$|R_1| + |B_1| + |R_2| + |B_2| = |P \cap \text{slab}(h)|,$$

while mixed pairs contribute twice the number of segments $pq \subset \text{slab}(h)$ $(p \neq q)$ that are either monochromatic noncrossing or bichromatic crossing minus twice the

number of segments that are bichromatic noncrossing or monochromatic crossing. In other words,

$$D_w^2(h) = |P \cap \text{slab}(h)| + 2(|MN| + |BC| - |BN| - |MC|),$$

where $|MN|$ denotes the number of nontrivial segments $pq \subset \text{slab}(h)$ that avoid h and have monochromatic endpoints, etc. We can rewrite this as

$$D_w^2(h) = |P \cap \text{slab}(h)| + \sum_{p \neq q} I_{pq}^{MN}(h) + \sum_{p \neq q} I_{pq}^{BC}(h) - \sum_{p \neq q} I_{pq}^{BN}(h) - \sum_{p \neq q} I_{pq}^{MC}(h),$$

where $I_{pq}^{MN}(h)$ is the indicator function that is equal to 1 if pq is contained in $\text{slab}(h)$ and avoids h and has monochromatic endpoints, and 0 otherwise, and similarly for all the other functions. By using linearity of expectation, we find that

$$\mathbf{E}_h D_w^2 \geq \sum_{p \in P} \text{Prob}_h[p \in \text{slab}(h)] - \sum_{\substack{p,q \in P \\ p \neq q}} g(p,q),$$

where

$$\begin{aligned} g(p,q) &= |\text{Prob}_h[pq \text{ lies in slab}(h) \text{ and crosses } h] \\ &\quad - \text{Prob}_h[pq \text{ lies in slab}(h) \text{ and avoids } h]| \\ &= |\text{Prob}_h[pq \text{ lies in slab}(h)] \\ &\quad - 2\,\text{Prob}_h[pq \text{ lies in slab}(h) \text{ and avoids } h]|. \end{aligned}$$

By Cauchy's theorem, a random line h is at a distance $w/2$ or less from any given point p with probability πw. (For this, we need to assume that all points of P lie at distance $\geq w$ from $\partial\mathscr{Q}$; our choice of w below does indeed have this property.) Moreover, pq lies in $\text{slab}(h)$ and avoids h if and only if pq lies in either $\text{slab}^+(h)$ or in $\text{slab}^-(h)$. Hence

$$\mathbf{E}_h D_w^2 \geq \pi w n - \sum_{\substack{p,q \in P \\ p \neq q}} \left| f(w, |p-q|) - 4f\left(\frac{w}{2}, |p-q|\right)\right|, \tag{5}$$

where $f(x,y)$ is the probability that a random slab of width $x \in \{w/2, w\}$ contains a fixed segment of length y.

Lemma 2.1. *For any $y, w > 0$ with $w/y \leq c_0$, where $c_0 > 0$ is a sufficiently small constant, we have*

$$\left| f(w,y) - 4f\left(\frac{w}{2}, y\right)\right| \leq \frac{w^4}{y^3}.$$

Proof. In view of the fact that $x \le w \le y$, the maximum angle α between a segment of length y and the axis of an enclosing slab of width x is equal to $\arcsin(x/y)$. Hence, if we first integrate over the displacement of h and then over its orientation, the outer integral is easily seen to be

$$f(x, y) = \int_{-\alpha}^{\alpha} (x - y \sin|\theta|)\, d\theta = 2\alpha x - 2(1 - \cos\alpha)y,$$

therefore

$$f(w, y) - 4f\left(\frac{w}{2}, y\right) = 2w\left(\arcsin\frac{w}{y} - 2\arcsin\frac{w}{2y}\right) + 2y\left[3 + \sqrt{1 - \left(\frac{w}{y}\right)^2} - 4\sqrt{1 - \left(\frac{w}{2y}\right)^2}\right].$$

A Taylor expansion around $w = 0$ gives

$$f(w, y) - 4f\left(\frac{w}{2}, y\right) = \frac{w^4}{16y^3} + O\left(\frac{w^6}{y^5}\right),$$

from which the claim follows. □

In our setting, we have $|p - q| \ge 1/4(m + 1)$ for any distinct points $p, q \in P$, while we choose $w = c_0/m$, for a constant $c_0 > 0$ which we can make as small as we wish. Therefore, we may bound each term in the sum of the right-hand side of (5) using Lemma 2.1, and we obtain

$$\mathbf{E}_h D_w^2 \ge \pi w n - \sum_{p \ne q} \frac{w^4}{|p - q|^3}. \tag{6}$$

Since the edge length λ_0 of the grid is $1/4(m + 1)$ and there are at most $8k$ grid points at L^∞-distance $\lambda_0 k$ from a given grid point, we obtain

$$\sum_{p \ne q} \frac{1}{|p - q|^3} \le \sum_{1 \le k \le m-1} \frac{4n}{\lambda_0^3 k^2} \le Cn^{5/2} \tag{7}$$

for an absolute constant C, and, hence, setting $w = c_0/m$, we get

$$\mathbf{E}_h D_w^2 \ge \pi w n - Cw^4 n^{5/2} = c_0 \pi \sqrt{n} - Cc_0^4 \sqrt{n} = \Omega(\sqrt{n})$$

for a small enough c_0. This combined with (4) establishes (1) and thus completes the proof of Theorem 1.1. □

Remark. The proof does not require the points of P to lie on the grid. As is easily checked, it applies equally well to sets $P \subseteq \mathscr{U}$ so that $|p - q| \geq c/m$, for any pair $p \neq q$ of points in P, where c is any fixed number.

3. Red–Blue Discrepancy in Dimension d

Given n points on a d-dimensional cube grid, we now show that, no matter how we color them red and blue, a half-space within which one color outnumbers the other by at least $\Omega(n^{1/2-1/2d})$ always exists. As in the two-dimensional case, we can actually prove this result for more general sets. Let P be a set of n points in E^d, and let δ and δ' be, respectively, the largest and smallest distance between any pair of distinct points in P. We say that P is *well-spread* if the ratio δ/δ' is less than $bn^{1/d}$, for some absolute constant $b > 0$ (clearly, this is asymptotically the smallest ratio one can have). We show that the above lower bound holds for any well-spread set in E^d.

We fix a red–blue coloring of P, which we can regard as a functon $\chi\colon P \mapsto \{-1, 1\}$. The discrepancy of the half-space h^+ bounded below by the (nonvertical) hyperplane h is

$$D(h) \stackrel{\text{def}}{=} \sum_{p \in P \cap h^+} \chi(p). \tag{8}$$

(In what follows, we also use this definition of discrepancy for arbitrary real-valued functions χ.) We want to show that

$$\mathbf{E}_h D^2 = \Omega(n^{1-1/d}).$$

The underlying distribution is given by the motion-invariant measure ω for hyperplanes, normalized so that the measure of hyperplanes crossing any segment pq is equal to its length $|p - q|$. (By Cauchy's theorem, we know that such a normalization is possible.) By appropriate scaling we can always assume that the point set P lies comfortably within a cube whose set of intersecting hyperplanes has measure 1. We assume that the distance of each point of P from the cube boundary is at least some absolute constant, because we shift slightly the points of P in the following analysis.

First we recall a formula (due to Stolarsky) expressing the above quadratic average in terms of sums of distances. We define the indicator function

$$I_{pq}(h) = \begin{cases} 1 & \text{if } h \text{ separates } p \text{ and } q, \\ 0 & \text{otherwise.} \end{cases}$$

For convenience, we assume that P contains as many red points as blue ones. If this is not the case, we can always add new points to P and color them by the deficient color. Note that if we must add too many points (more than $b_0 n^{1/2-1/2d}$, for a suitable small constant $b_0 > 0$), then any half-space that contains the cube has high

discrepancy and we are done. Otherwise, if we prove Theorem 1.1 for the augmented set, it will also hold for the original P, provided b_0 is sufficiently small. Thus, our assumption is not restrictive. We now have $D^2(h) = -D(h)D^*(h)$, where $D^*(h)$ is the discrepancy of the half-space below h, so it follows by elementary calculations that

$$D^2(h) = -\sum_{p,q\in P} I_{pq}(h)\chi(p)\chi(q).$$

By Cauchy's theorem,

$$\int I_{pq}(h)\, d\omega(h) = |p - q|,$$

and therefore

$$\mathbf{E}_h D^2 = -\sum_{p,q\in P} \chi(p)\chi(q)|p - q| \tag{9}$$

(see also [1]). Note that (9) holds for any real-valued function χ on P with $\sum_{p\in P} \chi(p) = 0$, when D is defined as in (8).

It appears difficult to bound the right-hand side directly. Instead, we use a finite-differencing technique, similar to the trick of using half-slabs in the previous section.

First we review the definition and a basic property of finite differencing; see, e.g., [7] for proofs and for more details. The *forward differencing operator* Δ acts on a real function f as follows:

$$\Delta f(x) \stackrel{\text{def}}{=} f(x + 1) - f(x).$$

Its t-fold iteration can be expressed as

$$\Delta^t f(x) = \sum_{i=0}^{t} (-1)^{t-i}\binom{t}{i} f(x + i). \tag{10}$$

For a function $f \in C^t(0, t)$ (that is, f has a continuous tth derivative in $(0, t)$), there is a number $\xi \in (0, t)$ with

$$\Delta^t f(0) = f^{(t)}(\xi). \tag{11}$$

Let w be a small positive real, and let v be the vector $(w, 0, \ldots, 0)$. We form the union of $t = \lceil d/2 \rceil + 1$ copies of P, each translated by a multiple of v:

$$P_w = \bigcup_{j=0}^{t} (P + jv).$$

To avoid notational confusion, it is convenient to assume that no line connecting a pair of distinct points in P is parallel to the x_1-axis. This way, no point appears multiply in P_w, and we can extend the coloring of P to P_w by writing

$$\chi(p + jv) = (-1)^j \binom{t}{j} \chi(p) \qquad \text{for} \quad p \in P, \quad j = 0, \ldots, t.$$

Note that this is not a coloring in the standard sense, but, rather, an assignment of real weights to the points of P_w. Note also that the sum of these weights is 0. We easily see that the discrepancy $D_w(h)$ for the new set P_w (defined by (8) for the new "coloring") is given by

$$D_w(h) = \sum_{j=0}^{t} (-1)^j \binom{t}{j} D(h - jv),$$

where $h - jv$ is the hyperplane h shifted by the vector jv.

We should point out the similarity with the analysis of the previous section. Instead of making three copies of h and assigning them the weights 1, -2, 1, which is what we effectively did in (3), we generalize this idea and make t copies of P (not of h), to which we assign weights derived from the rows of Pascal's triangle. This has the effect that the new discrepancy D_w becomes a th-order finite differencing of the original discrepancy (see (10)).

It is interesting to note that the technique of the preceding section (specifically, the need to use slabs of fixed width w) cannot be interpreted in terms of shifting the points of P by a few fixed vectors. (Alexander [1] uses a translation vector in one dimension higher and weights defined algebraically, with no reference to finite differencing. Explicit use of the finite-differencing calculus is the main simplifying factor in our approach.) In terms of slabs, the difference between the approaches of this section an the previous one is that, instead of having slabs of fixed width, we are now dealing with slabs of fixed x_1-width. However, no explicit reference to these slabs is needed in the present analysis.

By the Cauchy–Schwarz inequality,

$$D_w^2(h) \le \sum_{j=0}^{t} \binom{t}{j}^2 \cdot \sum_{j=0}^{t} D^2(h - jv),$$

and therefore

$$\mathbf{E}_h D_w^2 \le a \mathbf{E}_h D^2, \tag{12}$$

where

$$a = (t + 1) \sum_{j=0}^{t} \binom{t}{j}^2$$

is a constant depending on d. We have a relation similar to (9):

$$\begin{aligned}\mathbf{E}_h D_w^2 &= -\sum_{p',q' \in P_w} \chi(p')\chi(q')|p'-q'| \\ &= -\sum_{p,q \in P} \chi(p)\chi(q) G_w(p,q),\end{aligned}$$

where

$$G_w(p,q) = \begin{cases} \sum_{j=-t}^{t} (-1)^j \binom{2t}{t+j} |p-q+jv| & \text{if } p \neq q, \\ -\binom{2t-2}{t-1} w & \text{if } p = q. \end{cases} \tag{13}$$

To verify these last two identities is straightforward. For example, in the case where $p \neq q$, the length $|p-q+jv|$ is weighted by the factor

$$(-1)^{2i+j} \sum_{i=0}^{t-j} \binom{t}{i}\binom{t}{i+j} = (-1)^j \sum_{i=0}^{t-j} \binom{t}{t-i}\binom{t}{i+j} = (-1)^j \binom{2t}{t+j}.$$

We now simplify $G_w(p,q)$ for $p \neq q$. For fixed p, q, the length function

$$f(x) \stackrel{\text{def}}{=} |p-q+xv| = \sqrt{(p_1-q_1+xw)^2 + (p_2-q_2)^2 + \cdots + (p_d-q_d)^2}$$

is such that all its derivatives exist at a sufficiently small neighborhood of $x = 0$. Comparing (13) with (10), we see that

$$G_w(p,q) = (\Delta^{2t} f)(-t),$$

so $\xi \in (-t, t)$ exists such that

$$G_w(p,q) = f^{(2t)}(\xi),$$

from which it is easily deduced that

$$|G_w(p,q)| \leq \frac{C_1 w^{2t}}{|p-q+\xi v|^{2t-1}} \leq \frac{C_1 w^{2t}}{|p-q| - tw},$$

for some constant $C_1 > 0$ (depending on d). We can now repeat the previous argument. At most $O(k^{d-1})$ points of P lie at distance roughly $k/n^{1/d}$ from any point, so, as long as w is much smaller than $n^{-1/d}$, we have, for a fixed p,

$$\sum_q \frac{1}{(|p-q|-tw)^{2t-1}} = O(1) \cdot \sum_{k>0} \frac{k^{d-1}}{(k/n^{1/d})^{2t-1}} = O(n^{(2t-1)/d}),$$

and, hence,

$$\sum_{\substack{p,q\in P\\ p\neq q}} |G_w(p,q)| = O(1)\cdot w^{2t}n^{1+(2t-1)/d}.$$

It follows that

$$\mathbf{E}_h D_w^2 \geq -\sum_p G_w(p,p) - O(1)\cdot w^{2t}n^{1+(2t-1)/d}$$

$$\geq n\binom{2t-2}{t-1}w - O(1)\cdot w^{2t}n^{1+(2t-1)/d}.$$

Choosing $w = c_0 n^{-1/d}$, for a sufficiently small constant $c_0 > 0$, yields $\mathbf{E}_h D_w^2 = \Omega(n^{1-1/d})$. Hence, by (12),

$$\mathbf{E}_h D^2 = \Omega(n^{1-1/d}),$$

which establishes Theorem 1.1. □

4. Discrepancy for the Lebesgue Measure

Let P be any set of n points in $[0,1]^d$, and let $\mathrm{Vol}(\tau)$ denote, as in the Introduction, the Lebesgue measure of $\tau \cap [0,1]^d$. Given a nonvertical hyperplane h, we define its discrepancy by

$$D_0(h) \stackrel{\text{def}}{=} |P \cap h^+| - n\cdot \mathrm{Vol}(h^+).$$

Clearly, if h misses $[0,1]^d$, then $D_0(h) = 0$.

We want to show that

$$\mathbf{E}_h D_0^2 \geq cn^{1-1/d}, \tag{14}$$

for some constant $c = c(d) > 0$, where the expectation is taken with respect to the probability distribution ω, over all hyperplanes intersecting $[0,1]^d$. As above, this will readily imply Theorem 1.2.

To make the proof formally closer to the previous one, we first replace the continuous distribution by its approximation by the points of a sufficiently fine grid Q. Namely, we choose a number m, of a much larger order of magnitude than n, and we let Q be the points of the rectangular grid with step $m^{-1/d}$ inside the cube $[0,1]^d$.

We define $D(h)$ as the discrepancy of "P with respect to Q" in a natural way: We consider the set $S = P \cup Q$ (we assume, for convenience, that no point of P coincides with any point of Q), and define a "coloring" χ on S by

$$\chi(p) = \begin{cases} +1 & \text{if } p \in P, \\ -\dfrac{n}{m} & \text{if } p \in Q. \end{cases}$$

Clearly, we have $\sum_{p \in S} \chi(p) = 0$. Then we define

$$D(h) \stackrel{\text{def}}{=} \sum_{p \in S \cap h^+} \chi(p).$$

It is easy to check that if m was chosen sufficiently large (considerably larger than n^d, say), then we have, for any h,

$$|D(h) - D_0(h)| = o(n^{1/2 - 1/2d}), \tag{15}$$

and thus it suffices to prove that

$$\mathbf{E}_h D^2 = \Omega(n^{1-1/d}).$$

Let us remark that we could also deal with the continuous case directly, writing integrals instead of sums, with no conceptual difference.

We now apply a finite-differencing scheme that is somewhat different from the one used above. Let w be a fixed positive real, and let t be a positive integer constant (depending on d); the value of these parameters are determined later. We embed E^d in E^{d+t}, which we regard as the Cartesian product $E^d \times E^t$, by the map $(u_1, \ldots, u_d) \mapsto (u_1, \ldots, u_d, 0, \ldots, 0)$ (that is, $u \mapsto (u, 0)$), so we regard S as a set of points in E^{d+t}. We now create 2^t copies of S of the form $S + (0, wb)$, one for each vector $b = (b_1, \ldots, b_t) \in \{0,1\}^t$. Let S_w denote the union of all these copies. We extend the coloring χ to S_w by defining

$$\chi(p + (0, wb)) = (-1)^{\sum_j b_j} \chi(u), \tag{16}$$

for $p \in S$, $b \in \{0,1\}^t$.

We remark that this scheme could also be used in the proof from the previous section, and, conversely, the present proof could be performed using the differencing scheme from the previous section. The current scheme yields considerably simpler calculations, especially for the continuous case we treat here, while the scheme of the previous section can be considered more elementary, as it does not add extra dimensions to the ambient space.

The discrepancy $D_w(h)$ for the new set S_w, where now h is a hyperplane in E^{d+t}, is easily seen to be

$$D_w(h) = \sum_{b \in \{0,1\}^t} (-1)^{\sum_j b_j} D(h - (0, wb)),$$

where $h - (0, wb)$ is the displacement of h by the vector $(0, wb)$, and where $D(h)$ is the discrepancy of h with respect to the original embedded set S. By the Cauchy–Schwarz inequality, we have

$$D_w^2(h) \le 2^t \sum_{b \in \{0,1\}^t} D^2(h - (0, wb)), \tag{17}$$

so $\mathbf{E}_h D_w^2 \le 2^{2t} \mathbf{E}_h D^2$, and it thus suffices to show that

$$\mathbf{E}_h D_w^2 = \Omega(n^{1-1/d}).$$

It is easily seen that $\sum_{p \in S_w} \chi(p) = 0$, so equality (9) holds for S_w as well. With the invariant measure ω on the hyperplanes scaled appropriately, we thus have

$$\mathbf{E}_h D_w^2 = -\sum_{p', q' \in S_w} \chi(p')\chi(q')|p' - q'| = -\sum_{p, q \in S} \chi(p)\chi(q) G_w(p, q), \tag{18}$$

where

$$G_w(p, q) = 2^t \sum_{j=0}^{t} (-1)^j \binom{t}{j} \sqrt{|p - q|^2 + jw^2}\,. \tag{19}$$

To derive this formula, we consider all differences of the form $[p + (0, wb)] - [q + (0, wb')]$, for $b, b' \in \{0,1\}^t$. If $b - b'$ has exactly j nonzero components (which must be ± 1), then the length of the corresponding difference is $\sqrt{|p - q|^2 + jw^2}$. The number of such pairs (b, b') is $2^t \binom{t}{j}$, and the product of the "colors" of $p + (0, wb)$ and of $q + (0, wb')$ is $(-1)^j \chi(p)\chi(q)$.

Comparing the formula for $G_w(p, q)$ with (10), we see that $G_w(p, q) = \Delta^t f(0)$, where

$$f(x) \stackrel{\text{def}}{=} 2^t (-1)^t \sqrt{|p - q|^2 + xw^2}\,.$$

Considering p, q, and w fixed, we get from (11) the existence of a number $\xi \in (0, t)$ with

$$G_w(p, q) = f^{(t)}(\xi)$$

(since f is clearly t times continuously differentiable everywhere except possibly at 0). The tth derivative is easy to calculate, obtaining

$$G_w(p,q) = -C_1 \frac{w^{2t}}{(|p-q|^2 + \xi w^2)^{t-1/2}}, \tag{20}$$

where $O < \xi < t$ and $C_1 = C_1(t)$ is a *positive* constant.

We now split the sum in the right-hand side of (18) into several parts, as follows:

$$\mathbf{E}_h D_w^2 = E_{P2} + E_{PP} + E_{Q2} + E_{QQ} + E_{PQ},$$

where

$$E_{P2} = -\sum_{p \in P} \chi(p)^2 G_w(p,p),$$

$$E_{PP} = -\sum_{\substack{p,q \in P \\ p \neq q}} \chi(p)\chi(q) G_w(p,q),$$

$$E_{PQ} = -\sum_{p \in P, q \in Q} \chi(p)\chi(q) G_w(p,q)$$

and similarly for E_{QQ}, E_{Q2}. We estimate these terms one by one. We use the fact, established above, that $-G_w(p,q) > 0$ for all p, q, w. For E_2, formula (19) for $G_w(p,p)$ simplifies to

$$G_w(p,p) = -c_2 w,$$

where

$$c_2 = -2^t \sum_{j=1}^{t} (-1)^j \binom{t}{j} \sqrt{j}$$

is a constant, which must be positive by (20). Therefore, the sum E_{P2} is $c_2 wn$. For E_{PP}, we observe that $\chi(p)\chi(q) = 1$ for $p, q \in P$, and so $E_{PP} \geq 0$. Similarly, we find that E_{Q2} and E_{QQ} are both positive, which suffices for our purposes.

Finally, the sum E_{PQ} of the mixed terms has negative sign, and here we have to estimate its magnitude. We have

$$-E_{PQ} = \sum_{p \in P} \sum_{q \in Q} \chi(p)\chi(q) G_w(p,q) = -\frac{n}{m} \sum_{p \in P} \sum_{q \in Q} G_w(p,q).$$

We bound the inner sum (over Q), using the fact that Q is a grid with side $m^{-1/d}$, and thus that there are $O(k^{d-1})$ points q with $k/m^{1/d} \leq |p-q| < (k+1)/m^{1/d}$. For points $q \in Q$ at distance smaller than w, we use the estimate $G_w(p,q) \leq 2^{2t}|p-q| \leq 2^{2t} w$ (which is clear from (19)), and only for points further apart we

apply the derivative-based estimate (20). As long as $t \geq d/2 + 1$, we can estimate this sum by

$$\begin{aligned}
-\sum_{q\in Q} G_w(p,q) &= -\sum_{q\in Q,\,|q-p|\leq w} G_w(p,q) - \sum_{q\in Q,\,|q-p|>w} G_w(p,q) \\
&\leq 2^{2t} w \left(\frac{w}{m^{-1/d}}\right)^d + O(1)\cdot \sum_{q\in Q,\,|q-p|>w} \frac{w^{2t}}{|p-q|^{2t-1}} \\
&= O(w^{d+1}m) + O(w^{2t}) \sum_{k\geq wm^{1/d}} \frac{k^{d-1}}{(k/m^{1/d})^{2t-1}} \\
&\leq O(w^{d+1}m) + O(w^{2t}m^{(2t-1)/d}) \sum_{k\geq wm^{1/d}} \frac{1}{k^{2t-d}} \\
&\leq C_3 w^{d-1} m,
\end{aligned}$$

for some constant C_3 (depending on d). Multiplying this by the number of points in P and the weighting factor n/m, we get

$$|E_{PQ}| \leq C_3 n^2 w^{d+1}.$$

We now choose $w = c_0 n^{-1/d}$, as before, with c_0 sufficiently small, thereby obtaining

$$\mathbf{E}_h D_w^2 \geq E_{P2} - E_{PQ} \geq c_2 wn - C_3 w^{d+1} n^2 = n^{1-1/d}(c_2 c_0 - C_3 c_0^{d+1}) = \Omega(n^{1-1/d}).$$

This establishes Theorem 1.2. □

5. Discussion

Methods for proving lower bounds in discrepancy theory are often based on eigenvalues and Fourier analysis; see [5 and 10]. In this section we compare this with our (and Alexander's) approach.

In a method using eigenvalues, we represent a set system by an $m \times n$ incidence matrix A, where each of the m rows of A is the characteristic vector of a distinct set in the set system. A coloring is a 0/1 column vector x (the generalized colorings used in the above proofs correspond to arbitrary column vectors x), and the discrepancy is simply the L^∞-norm of the column vector Ax. Just as we did in this paper, the first step of a lower-bound proof is to pass to the L^2-norm, that is, we look at $\|Ax\|_2$ instead and try to derive a lower bound on it. In particular, note that if

$U = \{x \in E^n; \|x\|_2 = \sqrt{n}\}$ denotes the Euclidean sphere of radius $\sqrt{n}$ centered at the origin, the maximum discrepancy $D(A)$ of the set system satisfies

$$|D(A)| \geq \frac{1}{\sqrt{m}} \inf_{x \in U} \|Ax\|_2 .$$

To derive a lower bound on $\|Ax\|_2$, we first observe that the matrix $B = A^T A$ is positive semidefinite, so it can be diagonalized as

$$B = M^T \Lambda M,$$

where Λ is the diagonal matrix of eigenvalues. We have

$$\|Ax\|_2^2 = (Ax)^T Ax = x^T Bx = y^T \Lambda y,$$

where $y = Mx$. The column matrix of eigenvectors M^T is orthogonal, therefore $y \in U$ and

$$D^2(A) \geq \frac{1}{m} \inf_{y \in U} y^T \Lambda y.$$

Finally, we obtain

$$D^2(A) \geq \frac{n}{m} \lambda_1 ,$$

where λ_1 is the smallest eigenvalue of B; note that all the eigenvalues are nonnegative.

Thus, with this approach, estimating the smallest eigenvalue is the key to deriving a lower bound on the discrepancy. Set systems are often defined by means of a convolution. For example, consider the (red–blue) discrepancy $D(q)$ of a fixed-radius disk centered at q. We can write

$$D(q) = \int_{E^2} \tilde{\chi}(x) I_q(x)\, dx,$$

where I_q is the indicator function of the disk centered at q, and

$$\tilde{\chi}(x) = \sum_{p \in P} \chi(p) \delta(x - p),$$

where $\delta(y)$ denotes the two-dimensional Dirac (or impulse) function, that is, $\delta(y)$ corresponds to the density of the mass distribution which is 0 everywhere except for the origin, and in which the origin is a point of unit mass.

We have $I_q(p) = I_0(q - p)$, and therefore

$$D(q) = \int \tilde{\chi}(p) I_0(q - p)\, dp = (\tilde{\chi} * I_0)(q),$$

which is the convolution of $\tilde{\chi}$ and I_0. Because of the convolution theorem this makes Fourier analysis a favorite tool to get a handle on the spectrum of B, and hence, on the discrepancy.

In Alexander's method (and in this paper), a slightly different tack is taken. We give up trying to diagonalize B perfectly. Instead, we settle for a transformation which brings B into diagonally dominant form. Somewhat miraculously, the finite-differencing technique does just that. Going back to our analysis of the two-dimensional case in Section 3, after using the slab-based transformation from D to D_w, we show that the diagonal elements associated with D_w contribute most of the "action." Specifically, let A_w be the matrix associated with the modified discrepancy function D_w. We can reinterpret the various steps of the proof as follows: By the Cauchy–Schwarz inequality, we derive that

$$\|Ax\|_2^2 = \Omega\big(\|A_w x\|_2^2\big).$$

By analogy with our previous discussion, let B_w denote $A_w^T A_w$. Obviously,

$$\|A_w x\|_2^2 = x^T B_w x = \sum_{i,j} \left(\sum_k w_{k,i} w_{k,j} \right) x_i x_j,$$

where x_i is the color of point i and $A_w = (w_{i,j})$. Now, observe that $-\sum_k w_{k,i} w_{k,j}$ is essentially the term $G_w(p, q)$ in the proof, where p is the ith point and q is the jth point. The remainder of the proof shows that, up to a constant factor,

$$x^T B_w x \approx x^T D_w x,$$

where D_w is derived from B_w by zeroing out every nondiagonal element. This means that B_w is diagonally dominant, as desired.

Acknowledgments

We wish to thank József Beck, Joel Friedman, and, especially, Emo Welzl, for helpful discussions.

References

1. Alexander, R. Geometric methods in the study of irregularities of distribution, *Combinatorica* **10** (1990), 115–136.
2. Alexander, R. Principles of a new method in the study of irregularities of distribution, *Invent. Math.* **103** (1991) 279–296.
3. Beck, J. Quasi-random 2-colorings of point sets, *Random Struct. Algorithms* **2** (1991) 289–302.
4. Beck, J., Chen, W. W. L. *Irregularities of Distribution*, Cambridge Tracts in Mathematics, Vol. 89, Cambridge University Press, Cambridge, 1987.

5. Beck, J., Sós, V. T. Discrepancy theory, in *Handbook of Combinatorics*, to appear.
6. Buffon, G. Essai d'arithmétique morale, *Suppl. Hist. Natur.*, **4**, (1777).
7. Burden, R. L., Faires, J. D., Reynolds, A. C. *Numerical Analysis*, PWS-KENT, 1989.
8. Matoušek, J. Tight upper bounds for the discrepancy of half-spaces, KAM Series (Tech. Report), Charles University, Praka, 1994; submitted to *Discrete Comput. Geom.*
9. Matoušek, J., Welzl, E., Wernisch, L. Discrepancy and ε-approximations for bounded *VC*-dimension, *Combinatorica* **13** (1993), 455–466.
10. Roth, K. F. Remark concerning integer sequences, *Acta Arith.* **9** (1964), 257–260.
11. Santaló, L. A. *Integral Geometry and Geometric Probability*, Encyclopedia of Mathematics and Its Applications, Vol. 1, Addison-Wesley, Rading, MA, 1976.

Received April 15, 1994.