Quasi-Orthogonality via Finite-Differencing: An Elementary Approach to Geometric Discrepancy

Bernard Chazelle*
Department of Computer Science
Princeton University
Princeton, NJ 08544, USA
chazelle@princeton.edu

Abstract

It is possible to place n points in d-space so that, given any two-coloring of the points, there exists a halfspace within which one color dominates the other by as much as $cn^{1/2-1/2d}$, for some constant $c > 0$. This result was proven in a slightly weaker form by Beck and the bound was later tightened by Alexander. It was shown to be quasi-optimal by Matoušek, Welzl, and Wernisch. The lower bound proofs are highly technical and do not provide much intuitive insight into the "large-discrepancy" phenomenon. We develop a proof technique which allows us to rederive the same lower bound in a much simpler fashion. We give a probabilistic interpretation of the result and we discuss the connection of our method to Beck's Fourier transform approach. We also provide a quasi-optimal lower bound on the discrepancy of fixed-size rotated boxes, which significantly improves the previous bound.

1 Introduction

The Fourier transform, and more specifically, the classical (e.g., Fejér, Gauss) summability kernels of harmonic analysis have become the standard tools for estimating the discrepancy of geometric objects. This approach, pioneered in the number-theoretical work of Roth and greatly developed and expanded by Beck, has been hugely successful [6]. Sometimes, however, overwhelming technical difficulties arise because of vanishing harmonics, and the method fails. For example, it only yields a vastly suboptimal bound on the discrepancy of fixed-size rotated boxes.

In another major development, Alexander [1] followed a completely different approach to derive quasi-optimal lower bounds on the discrepancy of halfspaces and disk-segments. His proof technique involves a clever –but fairly intricate– mix of algebraic and integral-geometric arguments. See also Stolarsky [20] for work in the same spirit.

For all their mathematical ingenuity, however, none of the techniques above provide much intuitive insight into the "large-discrepancy" phenomenon. This paper is an attempt to remedy this by unifying and simplifying those techniques. In the process, we develop an altogether different proof technique which, if perhaps not nearly as versatile as its predecessors, provides a vastly simpler, more intuitive, approach to discrepancy theory. For example, it is a classical

*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.

result that, given any two-coloring of n points in d-space, there exists a halfspace within which one color outnumbers the other by as much as $\Omega(n^{1/2-1/2d})$. A slightly weaker bound was obtained by Beck [4] in 1983, and it was recently tightened by Alexander [1]. Both proofs rely on complicated analytical and algebraic arguments. By contrast, we give a simple probabilistic proof based on the celebrated Buffon needle problem [7] and a standard finite-differencing technique. We also fill a longstanding gap [6] by providing a quasi-optimal lower bound on the discrepancy of fixed-size rotated boxes.

Although the proof technique is self-contained, it is instructive to interpret it in light of the classical harmonic analysis approach. Orthogonal functions have often been used in discrepancy theory to "simulate" the effect of probabilistic independence [6, 15]. (See, for example, Roth's ingenious use of Rademacher functions for the discrepancy of aligned boxes [17].) Our proof technique is based on the simple observation that often quasi-orthogonality suffices. Thus, we can replace the Fourier transform by a simple diagonally-dominant linear transformation, which we build by forward-differencing the discrepancy function (the discrete analog of taking its higher-order partial derivatives).

The intuitive interpretation of our method (see Section 2) is so simple that at first it seems rather implausible. In Section 3, we place the proof technique in a broader context, by drawing connections to Beck's Fourier transform method. Section 4 gives the complete proof of the two-dimensional case. The differencing technique is described in Section 5 and the higher-dimensional case is treated in Section 6. Finally, we discuss the case of fixed-size rotated boxes in Section 7. To summarize, this paper is concerned with proving the following two results:

Theorem 1.1 *There is a set P of n points in E^d such that, given any two-coloring of P, there exists a halfspace within which one color outnumbers the other by at least $cn^{1/2-1/2d}$, for some constant $c > 0$.*

Theorem 1.2 *There is a set P of n points in $[0,1]^d$ such that, given any two-coloring of P, there exists a rotated cube of any desired volume (larger than an arbitrarily small constant) within which one color outnumbers the other by at least $cn^{1/2-1/2d}$, for some constant $c > 0$.*

A result of Matoušek, Welzl, and Wernisch [13] shows that both lower bounds are optimal up to within a factor of $O(\sqrt{\log n})$. Moreover, the corresponding upper bounds hold for *any* point set and not simply grid points. (Note, however, that the lower bounds are obviously false if the points are allowed to be very close to each other.) A slightly weaker upper bound was also derived by Beck [5] for the case $d = 2$ of Theorem 1.1. Here are the basic steps in Matoušek et al.'s proof: By the spanning path result of [10], we know how to match the n points into vertex-disjoint edges, so that no halfspace cuts more than $O(\kappa)$ edges, where $\kappa = n^{1-1/d}$. Now, for each edge, color one endpoint at random and assign the other color to the other end. In this way, the discrepancy of a halfspace can be entirely attributed to the $O(\kappa)$ edges cutting the bounding hyperplane. Because of the random selection, with high probability, no color will outnumber the other by more than $O(\sqrt{\kappa \log \kappa})$, which establishes the upper bound.

Discrepancy Theory and Derandomization. The construction of ε-approximations and ε-nets has been a focus of research in computational geometry for several years. The principal motivation has been to derandomize probabilistic algorithms. For example, ε-approximations

can be used to simulate random sampling and generate pseudo-random permutations deterministically, in a manner vastly superior to what other derandomization techniques alone can offer in geometry [2, 3, 12, 14, 16, 19]. In addition, nets and approximations provide the tools for deterministic quasi-Monte Carlo integration (which, for example, played a key role in solving the convex hull problem [8]).

Matoušek et al. [13] recently demonstrated the close connection between ε-approximations and discrepancy theory. They showed that good discrepancy bounds imply good ε-approximations and vice-versa. This was used in [9] to give a simplified construction of ε-approximations. Another link between derandomization and discrepancy was exhibited in Even et al. [11], where approximating the uniform distribution for general random variables was viewed as achieving low rectangle-discrepancy. Theorem 1.2 establishes the optimality (within a sublogarithmic factor) of the ε-approximation construction of [13] for the range space formed by the group of isometries acting on *any* cube of constant volume.

Preliminaries. We begin with the proof of Theorem 1.1 in dimension $d = 2$. Let $\mathcal{U} = [0, 1/4]^2$ be a square of unit perimeter. We use the invariant measure of lines under motions to define a *random* line h. By Crofton's theorem[1] the measure of the set of lines crossing $\mathcal{U}$ is exactly 1, so from now on, a *random* line h is to be understood as a random line among those crossing $\mathcal{U}$.

Without loss of generality, assume that $n = m^2$, for some integer $m > 0$. Make $\mathcal{U}$ into an $(m+1)$-by-$(m+1)$ chessboard and let $P = \{p_1, \ldots, p_n\}$ be the set of interior grid points. Fix a two-coloring of $P = R \sqcup B$ once and for all. The discrepancy of a halfplane s, denoted $D(s)$, is defined as $|s \cap R| - |s \cap B|$. Let h^+ (resp. h^-) denote the open halfplane bounded by the line h that contains (resp. does not contain) the origin. We want to show that $\sup_h |D(h^+)| = \Omega(n^{1/4})$.

Intuitively, it appears that the discrepancy should be "concentrated" around the boundary of h^+. For this reason we will focus on a thin slab centered along h of width w. Analytically, if q_h is the point of h closest to the origin, then

$$\mathrm{slab}(h) = \left\{ p \in E^2 \,:\, |\langle p, q_h \rangle - |q_h|^2| < \frac{w}{2}|q_h| \right\}.$$

For technical convenience we choose $w = c_1/m$, for an appropriately small constant $c_1 > 0$. This choice implies, in particular, that any positioning of $\mathrm{slab}(h)$ that encloses a point of P is such that h belongs to the set of random lines, i.e., crosses $\mathcal{U}$. Let $R_h^+ = |R \cap \mathrm{slab}(h) \cap h^+|$ be the number of red points in the half-slab near the origin, and let $R_h^- = |R \cap \mathrm{slab}(h) \cap h^-|$ be the number of red points on the other side. We define B_h^+ and B_h^- similarly for the blue points, and turn our attention to

$$\Delta(h) = (R_h^+ - B_h^+) - (R_h^- - B_h^-).$$

We call $\Delta(h)$ the *differential discrepancy* of h because it can be expressed in the form

$$D(s_2) - D(s_1) - (D(s_3) - D(s_2)),$$

[1] 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 invariant area measure for h under the group of motions is $dh = d\rho d\theta$ [18]. Crofton's theorem says that the measure of the set of lines intersecting a convex set $A \subset \mathcal{U}$ is equal to the perimeter of A.

where s_1, s_2, s_3 are appropriate halfplanes. By Cauchy-Schwarz,

$$\Delta^2(h) \leq 6(D^2(s_1) + D^2(s_2) + D^2(s_3)),$$

and therefore, up to within a constant factor, $\mathbf{E}_h\Delta^2$ is a lower bound on $\sup_s D^2(s)$ (over all halfplanes s), and hence, on $\sup_h D^2(h^+)$. (The step from s to h^+ is trivial: indeed, if the worst-case halfplane s is of the form h^-, then either h^+ has high discrepancy or then any halfplane h_0^+ that contains the whole grid must have high discrepancy). Thus, to establish the case $d = 2$ of Theorem 1.1, it suffices to prove that $\mathbf{E}_h\Delta^2 = \Omega(n^{1/2})$.

2 The Intuition

Let h be a random line and suppose, for the sake of contradiction, that $D(h^+)$ is zero everywhere (of course, we know that it must be at least ± 1 somewhere, so we might want to think of zero as just meaning "very small"). Zero discrepancy implies that, within each of the two half-slabs in slab(h), red and blue points match perfectly, i.e., $R_h^+ = B_h^+$ and $R_h^- = B_h^-$. Now, pick two random integers $1 \leq i < j \leq n$ and assume that the points p_i, p_j fall in slab(h) (Fig.1). If p_ip_j crosses h, then obviously p_i and p_j are equally likely to be bichromatic or monochromatic; we write this as $BC = MC$, which stands for: "probability of bichromatic crossing = probability of monochromatic crossing". On the other hand, if p_ip_j does not cross h then $BN > MN$, meaning that the probability that p_ip_j is bichromatic noncrossing exceeds that of being monochromatic noncrossing: $BN > MN$ is the key inequality driving the whole argument so it is crucial to understand it. Without loss of generality, suppose that p_i is blue and lies on the side of the origin O. Then, p_j is picked randomly among $B_h^+ - 1$ blue points (-1 because $i \neq j$) and R_h^+ red points. Since $B_h^+ = R_h^+$, the point p_j is thus a tiny bit more likely to be red than blue.

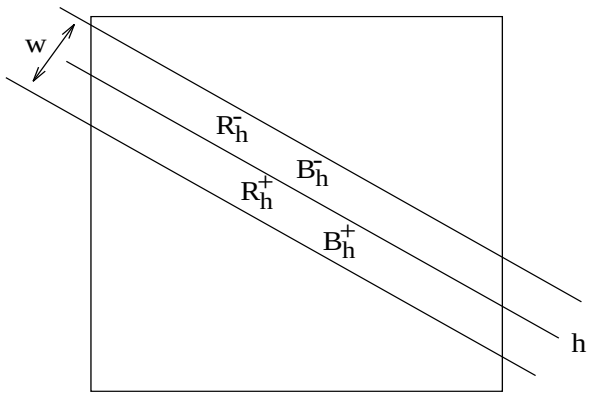

Figure 1: Assuming no discrepancy: $R_h^+ = B_h^+$ and $R_h^- = B_h^-$.

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 will apply to the other case as well). The inequality says that a random bichromatic segment p_ip_j (i.e., with random i, j) 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. But throwing a random slab on top of p_ip_j is similar to throwing a random needle of length $|p_ip_j|$ inside the slab. This is the Buffon needle experiment, 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 p_i, p_j are grid points, a random $p_i p_j$ 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 $p_i p_i$ is forbidden from the probabilistic experiment. But even if we allowed such events they would still be very unlikely to come up. So, how can such a seemingly minor requirement cause all the cancellations necessary to make the discrepancy "bubble up to the surface"? We explain how to achieve this cancellation phenomenon in the next section. Splitting the slab in half is part of a general differencing scheme which we describe in Section 5.

3 Quasi-Orthogonality via Finite-Differencing

Orthogonal functions are often used in discrepancy theory to "simulate" probabilistic independence. Suppose that h were fixed and that the two-coloring were random (of course, this is the opposite of what we have: h is random and the coloring is fixed). Let $q_1, \ldots, q_k$ be the points of P inside slab(h), and let x_i be the contribution of q_i to $\Delta(h)$, i.e., x_i is $+1$ if q_i is red and lies in the half-slab closer to the origin or it is blue and lies in the other half-slab, and $x_i = -1$ in the other cases. Then, the differential discrepancy $\Delta(h)$ is of the form $\sum_{1 \le i \le k} x_i$, and it satisfies

$$\mathbf{E}\,\Delta^2(h) = \sum_i \mathbf{E}\,x_i^2 + \sum_{i \neq j} \mathbf{E}\,x_i x_j.$$

Since the points are colored randomly independently, each "covariance" term $\mathbf{E}\,x_i x_j$ ($i \neq j$) vanishes (expectation and multiplication commute), and we find that the average square differential discrepancy is $\sum_i \mathbf{E}\,x_i^2 = k$, which is about $\sqrt{n}$. Of course, this has no bearing on our problem since the coloring is fixed and h is random. By analogy, however, we think of $\mathbf{E}\,x_i x_j = 0$ as an orthogonality condition, and we attempt to write $\Delta(h)$ as a linear combination of orthogonal functions,

$$\Delta(h) = \sum_i \lambda_i g_i(h).$$

(Note that the sum is an integral in the case of Fourier transforms.) Then,

$$\int \Delta^2(h)\,dh = \sum_i \lambda_i^2 \int g_i^2(h)\,dh + \sum_{i \neq j} \lambda_i \lambda_j \int g_i(h) g_j(h)\,dh = \sum_i \lambda_i^2 \int g_i^2(h)\,dh.$$

This is the analog of the Parseval-Plancherel identity of Fourier analysis. The hope, of course, is that the coefficients λ_i^2 can be estimated from below with reasonable accuracy. This is the main idea behind many lower bound proofs in discrepancy theory.

In the case of the Fourier transform approach, there is the additional advantage that the discrepancy can be expressed as the convolution of two functions: one is the characteristic function of a fixed halfspace while the other is the discrepancy measure induced by the point set. Besides Parseval-Plancherel, the advantage of switching over to the frequency domain is that the convolution now becomes a product, so the Fourier transforms of the two convolved functions are in effect "separated," and thus can be analyzed in isolation —albeit at the price of formidable technical difficulties.

Instead of trigonometric functions, we use a simpler system of *quasi-orthogonal* functions, with still the same purpose of "simulating" probabilistic independence. We say that a system of functions $f_1, \ldots, f_n : \mathbb{R}^2 \mapsto \mathbb{R}$ with bounded L_2-norm is *quasi-orthogonal* if

$$\sum_{i \neq j} \left| \int_{\mathbb{R}^2} f_i(q) f_j(q)\, dq \right| < c \cdot \sum_i \int_{\mathbb{R}^2} f_i^2(q)\, dq,$$

for some constant $0 < c < 1$. We wish to express the discrepancy as a linear combination of quasi-orthogonal functions. We have a slightly weakened Parseval-Plancherel identity, which we can still use to derive a bound on the L_2-norm of the discrepancy.

Finite-differencing provides a simple scheme to create quasi-orthogonal functions that relate naturally to the discrepancy. Given h crossing $\mathcal{U}$, let $\text{slab}^+(h)$ be the half-slab of $\text{slab}(h)$ closer to the origin and let $\text{slab}^-(h)$ be the other one. Given a point $p_i \in P$, we define the function $f_i(h)$ as follows:

$$f_i(h) = \begin{cases} 1 & \text{if } p_i \in \text{slab}^+(h) \\ -1 & \text{if } p_i \in \text{slab}^-(h) \\ 0 & \text{if } p_i \notin \text{slab}(h). \end{cases}$$

Let c_i denote the color of p_i: 1 if red and -1 if blue. We can express the differential discrepancy of h as

$$\Delta(h) = \sum_{1 \leq i \leq n} c_i f_i(h).$$

By the Buffon needle experiment mentioned earlier, it is easily shown that the family of functions $\{f_i\}$ is quasi-orthogonal. Note that since each c_i is ± 1, we immediately have a (slightly weakened) version of Parseval-Plancherel. In connection to the previous section, observe that the term

$$\sum_{i \neq j} \left| \int_{\mathbb{R}^2} f_i(h) f_j(h)\, dh \right|$$

corresponds to the contribution of the nontrivial segments $p_i p_j$ $(i \neq j)$ to the discrepancy.

Splitting $\text{slab}(h)$ into two half-slabs is part of a general finite-differencing scheme which allows us to generalize the technique to higher dimensions. Indeed, it will then be necessary to subdivide the slab into more than two parts and assign appropriate weights to each half-slab.

4 The Two-Dimensional Case

Recall that it suffices to show that $\mathbf{E}_h \Delta^2 = \Omega(n^{1/2})$. Writing $\Delta(h)$ in unary, we have

$$\Delta^2(h) = \Big(\sum_{1 \leq i \leq R_h^+} 1 + \sum_{1 \leq i \leq B_h^-} 1 - \sum_{1 \leq i \leq R_h^-} 1 - \sum_{1 \leq i \leq B_h^+} 1 \Big)^2.$$

If we expand that expression we find that: (i) companion pairs amount to $R_h^+ + B_h^- + R_h^- + B_h^+ = |P \cap \text{slab}(h)|$; (ii) mixed pairs contribute twice the number of segments that are either monochromatic noncrossing or bichromatic crossing minus twice the number of segments that are bichromatic noncrossing or monochromatic crossing. Thus, $\Delta^2(h)$ is equal to

$$|P \cap \text{slab}(h)| + 2\binom{n}{2} \Big\{ \text{Prob}_s[\, s \in MN\,] - \text{Prob}_s[\, s \in MC\,] + \text{Prob}_s[\, s \in BC\,] - \text{Prob}_s[\, s \in BN\,] \Big\},$$

where $s = p_i p_j$ and i, j are picked randomly with $1 \le i < j \le n$. The symbols MN, MC, BC, BN now denote the sets corresponding to the probabilities of Section 2: for example, MN denotes the set of segments $p_i p_j \subset \mathrm{slab}(h)$ that avoid h and have monochromatic endpoints. By integrating over all h crossing $\mathcal{U}$ we find that

$$\mathbf{E}_h\, \Delta^2 \ge \sum_{1 \le i \le n} \mathrm{Prob}_h[\, p_i \in \mathrm{slab}(h)\,] - 2 \sum_{1 \le i < j \le n} P(i,j),$$

where

$$P(i,j) = \Big|\mathrm{Prob}_h[\, p_i p_j \text{ lies in slab}(h) \text{ \& crosses } h\,] - \mathrm{Prob}_h[\, p_i p_j \text{ lies in slab}(h) \text{ \& avoids } h\,]\Big|.$$

By Crofton's theorem, a random line h is at a distance at most $w/2$ from p_i with probability πw, therefore

$$\mathbf{E}_h\, \Delta^2 \ge \pi w n - 2 \sum_{1 \le i < j \le n} \Big|\pi(w, |p_i p_j|) - 4\pi(w/2, |p_i p_j|)\Big|, \tag{1}$$

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

Lemma 4.1

$$\Big|\pi(w,y) - 4\pi(w/2,y)\Big| \le \frac{w^4}{2y^3}.$$

Proof: The inequality can be derived very elegantly from Crofton's theorem, but a direct analytical proof is even shorter. Since $x \le w \le y$, the maximum angle between a segment of length y and the axis of an enclosing slab of width x is $\alpha = \arcsin(x/y)$, so we have

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

therefore

$$\pi(w,y) - 4\pi(w/2,y) = 2w\Big(\arcsin\frac{w}{y} - 2\arcsin\frac{w}{2y}\Big) + 2y\Big(3 + \sqrt{1 - \Big(\frac{w}{y}\Big)^2} - 4\sqrt{1 - \Big(\frac{w}{2y}\Big)^2}\Big).$$

Since $y \ge 1/4(\sqrt{n}+1)$ and $w\sqrt{n} = c_1$ is arbitrarily small, a Taylor expansion around $w = 0$ gives the power series

$$\frac{w^4}{16y^3} + \frac{3w^6}{128y^5} + O(w^8/y^7),$$

which completes the proof. □

It follows from (1) and Lemma 4.1 that

$$\mathbf{E}_h \Delta^2 \ge \pi w n - \sum_{1 \le i < j \le n} \frac{w^4}{|p_i p_j|^3}. \tag{2}$$

Each square of the chessboard has length $\lambda_0 = 1/4(\sqrt{n}+1)$, and there are at most $8k$ grid points at L_∞-distance $\lambda_0 k$ from a given grid point, therefore

$$\sum_{1 \le i < j \le n} \frac{1}{|p_i p_j|^3} \le \sum_{1 \le k \le m-1} \frac{8n}{\lambda_0^3 k^2} = O(n^{5/2}),$$

and

$$\mathbf{E}_h \Delta^2 \geq \pi w n - O(w^4 n^{5/2}).$$

Setting $c_1 = w\sqrt{n}$ small enough yields $\mathbf{E}_h \Delta^2 = \Omega(n^{1/2})$, which proves Theorem 1.1 in the case $d = 2$. □

The next section strengthens the proof technique in anticipation of the higher-dimensional case. In particular, it describes a way to circumvent the expansion in Taylor series in Lemma 4.1 and basically avoid all calculations.

5 Difference Operators

In the two-dimensional case we argued that there are only $O(k)$ grid points at L_∞-distance $\lambda_0 k$ from a given grid point, and therefore the sum $\sum |p_i p_j|^{-3}$ yields the convergent series $\sum k^{-2}$. Proceeding backwards, we know that in d-space the analog number of grid points will be $O(k^{d-1})$, so instead of the sum $\sum |p_i p_j|^{-3}$ we need a more rapidly convergent series, e.g., $\sum |p_i p_j|^{-d-1}$. To achieve this, we generalize the notion of differential discrepancy, so as to strengthen the upper bound in Lemma 4.1 to w^{c+1}/y^c, for *any* constant $c > 0$. The idea is to apply differencing techniques to bootstrap our original method and thus produce more cancellations. For convenience, let us remain in two dimensions for the time being.

Let σ denote the operator "shift variable i back by one", e.g., $\sigma[(i+1)^2 x + \sin(iy)] = i^2 x + \sin(iy - y)$. We begin with a simple technical lemma, which illustrates our basic approach.

Lemma 5.1 *Let f be a real function locally smooth at 0. The function $(1-\sigma)^{m+1}[f(ix)]$ is locally smooth and vanishes at 0 and so do its first m derivatives.*

Proof: The smoothness part is obvious. Now, observe that the k-th order term in the Taylor expansion of $(1-\sigma)^{m+1}[f(ix)]$ at $x = 0$ is obtained by multiplying the corresponding term in the expansion of $f(x)$ by $(1-\sigma)^{m+1}[i^k]$, which is 0 for any $k \leq m$. □

We rename $\Delta(h)$ the *first-order* differential discrepancy of h, and denote it $\Delta_1(h)$.[2] To define the *k-th order differential discrepancy*, we subdivide $\mathrm{slab}(h)$ into $k+1$ congruent slabs of width $w_k = w/(k+1)$, denoted $\mathrm{slab}_0, \ldots, \mathrm{slab}_k$ in the direction from O to h. (We omit the argument h, which is understood.) Let $R_i = |R \cap \mathrm{slab}_i|$ (resp. $B_i = |B \cap \mathrm{slab}_i|$) be the number of red (resp. blue) points in slab_i. Following the rules of forward-differencing, we define

$$\Delta_k(h) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} (R_i - B_i).$$

As before, by Cauchy-Schwarz, any lower bound on $\mathbf{E}_h \Delta_k^2(h)$ carries over to $\sup_h D^2(h^+)$ (for k constant). We expand $\Delta_k^2(h)$ and classify the matching pairs according to their types: companion, mixed, noncrossing, crossing 1 line, 2 lines, etc.

$$\Delta_k^2(h) = \sum_{0 \leq i,j \leq k} (-1)^{i+j} \binom{k}{i} \binom{k}{j} (R_i - B_i)(R_j - B_j).$$

[2] It might be argued that $\Delta(h)$ should be called the second-order differential discrepancy, but for notational convenience it is preferable to let the number refer to the number of splitting lines.

The companion pairs amount to

$$\sum_{0\le i\le k}\binom{k}{i}^2(R_i+B_i)+2\sum_{0\le i\le k}\binom{k}{i}^2(M_{i,i}-B_{i,i}),$$

while the mixed pairs contribute

$$2\sum_{0\le i<j\le k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}(M_{i,j}-B_{i,j}),$$

where $M_{i,j}$ (resp. $B_{i,j}$) denotes the number of monochromatic (resp. bichromatic) segments p_up_v $(u<v)$ with one endpoint in slab_i and the other one in slab_j. Thus,

$$\Delta_k^2(h)=\sum_{0\le i\le k}\binom{k}{i}^2|P\cap\mathrm{slab}_i|+2\sum_{0\le i\le j\le k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}(M_{i,j}-B_{i,j}).$$

Taking the expectation for a random line h immediately yields

$$\mathbf{E}_h\,\Delta_k^2\ge\pi wn-2\sum_{1\le u<v\le n}|C_{u,v}|, \tag{3}$$

where

$$C_{u,v}=\sum_{0\le i\le j\le k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}\mathrm{Prob}_h[\,p_up_v\in M_{i,j}\cup B_{i,j}\,]. \tag{4}$$

(By abuse of notation, we use the same symbols for the sets $M_{i,j},B_{i,j}$ and their cardinalities.) By symmetry, the probability that a fixed p_up_v lies in $M_{i,j}\cup B_{i,j}$ (for a random $\mathrm{slab}(h)$) is a function only of $|p_up_v|$ and $|j-i|$, and thus can be written as $\pi_{|j-i|}(|p_up_v|)$.

Lemma 5.2 *There exists an even function C smooth at 0 such that*

1. *For $i>0$, $\pi_i(y)=y(1-\sigma)^2[C((i+1)w_k/y)]$.*
2. *Setting $i=0$ in the expression above gives $2\pi_0(y)$.*

Proof: Setting $\alpha_i=\arcsin(iw_k/y)$, we derive

$$\pi_0(y)=2\int_0^{\alpha_1}(w_k-y\sin\theta)\,d\theta$$

and for $i>0$,

$$\pi_i(y)=2\int_{\alpha_{i-1}}^{\alpha_i}(y\sin\theta-(i-1)w_k)\,d\theta+2\int_{\alpha_i}^{\alpha_{i+1}}((i+1)w_k-y\sin\theta)\,d\theta.$$

We can check without evaluating these integrals that setting $i=0$ in the second expression (the one for $i>0$) gives twice the expression for $\pi_0(y)$, so (2) follows from (1). For $i>0$,

$$\pi_i(y)=y\Big\{\int_{\alpha_{i-1}}^{\alpha_i}\diamond-\int_{\alpha_i}^{\alpha_{i+1}}\diamond\Big\}+w_k\Big\{(i-1)\int_{\alpha_{i-1}}^{\alpha_i}\square-(i+1)\int_{\alpha_i}^{\alpha_{i+1}}\square\Big\},$$

which can be expressed as

$$y\Big\{A(\alpha_{i-1}) - 2A(\alpha_i) \quad + \quad A(\alpha_{i+1})\Big\}$$
$$+ \quad y\Big\{\frac{(i-1)w_k}{y}B(\alpha_{i-1}) - \frac{2iw_k}{y}B(\alpha_i) + \frac{(i+1)w_k}{y}B(\alpha_{i+1})\Big\}.$$

It follows that

$$\pi_i(y) = y(1-\sigma)^2\Big[C((i+1)w_k/y))\Big],$$

where

$$C(ix) = A(\arcsin ix) + (ix)B(\arcsin ix).$$

It is immediate to verify that C is even and smooth at 0. □

The characterization of $C_{u,v}$ in terms of iterated differences given below produces the cancellations needed to make the discrepancy "bubble up" to the surface.

Lemma 5.3 *We have $C_{u,v} = L(|p_u p_v|)$, where (at $i = k$)*

$$L(y) = \frac{(-1)^k y}{2}(1-\sigma)^{2(k+1)}\Big[C((i+1)w_k/y)\Big].$$

Proof: Let $y = |p_u p_v|$; recall (4) that $C_{u,v} = L(y)$, where

$$L(y) = \sum_{0\le i\le j\le k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}\pi_{j-i}(y) = \sum_{0\le i\le k}(-1)^i\pi_i(y)\sum_{0\le j\le k-i}\binom{k}{j}\binom{k}{k-j-i}.$$

By Lemma 5.2, we find that

$$L(y) = \sum_{0\le i\le k}(-1)^i\binom{2k}{k-i}\pi_i(y) = \sum_{0\le i\le k}(-1)^i\binom{2k}{k-i}\chi_i y(1-\sigma)^2\Big[C((i+1)w_k/y)\Big],$$

where χ_i is $1/2$ if $i = 0$ and 1 otherwise. Because C is even, $(1-\sigma)^2[C((i+1)w_k/y)]$ is the same at $i = j$ and $i = -j$, and hence,

$$\begin{aligned} L(y) &= \frac{y}{2}\sum_{-k\le i\le k}(-1)^i\binom{2k}{k-i}(1-\sigma)^2\Big[C((i+1)w_k/y)\Big] \\ &= \frac{(-1)^k y}{2}(1-\sigma)^{2(k+1)}\Big[C((i+1)w_k/y)\Big], \end{aligned}$$

at $i = k$. □

We are now in a position to generalize Lemma 4.1. By Lemmas 5.1 and 5.3, $L(y)/y$ is a function of $x = w_k/y$ which vanishes at $x = 0$, as do its first $2k+1$ derivatives in x. Therefore, $|L(y)| = O(yx^{2k+2})$, and

$$|C_{u,v}| = O\Big(w^{2k+2}/|p_u p_v|^{2k+1}\Big).$$

This result says that by repeated differencing we can decrease the importance of mixed pairs (i.e., nontrivial segments) arbitrarily. Specifically, for some large enough constant $c_k > 0$ we find from (3) that as long as $w < 1/c_k\sqrt{n}$, the inequality (2) becomes

$$\mathbf{E}_h\,\Delta_k^2 \ge \pi w n - c_k\sum_{1\le i<j\le n}\frac{w^{2k+2}}{|p_i p_j|^{2k+1}}. \tag{5}$$

6 The Discrepancy of Halfspaces

Inequality (5) gives us the main tool to lift the proof for the planar case to d-space. First, we form an n-point grid inside a d-cube $\mathcal{U}$ normalized so that its set of intersecting hyperplanes has measure 1. Then, we consider a random d-dimensional slab of width w subdivided into $k+1$ congruent sub-slabs, for $k = \lceil d/2 \rceil$. The calculations are identical to the two-dimensional case, except for a slight difference in the evaluation of the probability $\pi_i(|p_u p_v|)$.

In d dimensions the volume element for slab(h) is $dh = dr \wedge dS$, where r is the distance of h to O and dS is the area element of the $(d-1)$-unit sphere U^{d-1} centered at O [18]. The points of U^{d-1} whose tangent hyperplanes form an angle θ with $p_u p_v$ form a $(d-2)$-sphere $U(\theta)$ of radius $\cos\theta$: the $(d-2)$-volume of $U(\theta)$ is therefore $(\cos\theta)^{d-2}$ (up to normalizing by the appropriate scaling factor). Thus, because of rotational symmetry, we can evaluate $\pi_i(|p_u p_v|)$ by integrating the displacement in direction θ over all $\theta \in [-\pi/2, \pi/2)$, using the differential element $(\cos\theta)^{d-2} d\theta$. Besides adjusting normalizing factors, the only difference from what we did in the previous section is in the proof of Lemma 5.2: We now have (up to within constant factors),

$$\pi_0(y) = 2 \int_0^{\alpha_1} (\cos\theta)^{d-2} (w_k - y \sin\theta)\, d\theta$$

and for $i > 0$,

$$\pi_i(y) = 2 \int_{\alpha_{i-1}}^{\alpha_i} (\cos\theta)^{d-2} (y \sin\theta - (i-1) w_k)\, d\theta + 2 \int_{\alpha_i}^{\alpha_{i+1}} (\cos\theta)^{d-2} ((i+1) w_k - y \sin\theta)\, d\theta.$$

The multiplicative factor is even and smooth, so everything else in the proof is strictly identical — in particular, the function C is still even —, and therefore the lemma still holds. (One should appreciate the fact that no calculations whatsoever are required.) Since the probability that a random slab contains a fixed grid point is proportional to w, we still have (5), up to within constant factors. There are $O(t^{d-1})$ grid points at L_∞-distance $\lambda_0 t$ from any given point, where $\lambda_0 = \Omega(n^{-1/d})$ is the minimum distance between two grid points. It easily follows that

$$\mathbf{E}_h\, \Delta_k^2 = \Omega(wn) - O(w^{2k+2} n^{1+(2k+1)/d}),$$

and hence, $\mathbf{E}_h\, \Delta_k^2 = \Omega(n^{1-1/d})$. This establishes Theorem 1.1 in all dimensions. Note that trivially any set of points whose max/min pairwise distance ratio is $O(n^{1/d})$ satisfies the same lower bound. We should observe that requiring a lower bound on the smallest interdistance is not a weakness of the argument: indeed, if we place the points along a circle and color the points alternatively, the discrepancy is either 0 or ± 1. Another low-discrepancy construction consists of placing half the points anywhere in general position and placing the remainder by closely pairing the first set.

7 The Discrepancy of Fixed-Size Cubes

Let s be a halfspace achieving the high-discrepancy lower bound of Theorem 1.1. Obviously, we can replace the halfspace by a cube and keep the same discrepancy, provided that the cube is big enough that its intersection with the grid is the same. Given an arbitrarily small positive constant v_0, we wish to exhibit a cube of volume v_0 with a similarly high discrepancy. We do this most simply by covering the intersection $\mathcal{U} \cap s$ by disjoint cubes of volume v_0. Since $O(1/v_0)$ cubes are enough and discrepancy is additive, one of these cubes must have the same

high discrepancy, up to within a factor $\Omega(v_0)$, which proves Theorem 1.2. Note that the same argument works for any regular tesselation of space. However, this does not suggest any way to convert our bound to, say, one on the discrepancy of balls.

Acknowledgments

I wish to thank József Beck and Joel Friedman for helpful discussions.

References

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

[2] Alon, N., Babai, L., Itai, A. *A fast and simple randomized algorithm for the maximal independent set problem*, J. Alg., 7 (1986), 567–583.

[3] Alon, N., Goldreich, O., Hastad, J., Peralta, R. *Simple constructions of almost k-wise independent random variables,* Proc. 31st Ann. IEEE Symp. Found. Comput. Sci. (1990), 544–553.

[4] Beck, J. *On a problem of K.F. Roth concerning irregularities of distribution*, Invent. Math. 74 (1983), 477–87.

[5] Beck, J. *Quasi-random 2-colorings of point sets*, DIMACS Tech. Report 91–20, March 1991.

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

[7] Buffon, G. *Essai d'arithmétique morale*, Supplément à l'Histoire Naturelle, Vol.4, 1777.

[8] Chazelle, B. *An optimal convex hull algorithm in any fixed dimension*, Disc. Comput. Geom., to appear. Prelim. version in Proc. 32nd FOCS, 1991.

[9] Chazelle, B., Matoušek, J. *On linear-time deterministic algorithms for optimization problems in fixed dimension*, Proc. 4th Annu. ACM-SIAM Symp. Disc. Algor. (1993), 281–290.

[10] Chazelle, B., Welzl, E. *Quasi-optimal range searching in spaces of finite VC-dimension*, Disc. Comput. Geom., 4 (1989), 467–489.

[11] Even, G., Goldreich, O., Luby, M., Nisan, N., Velickovic, B. *Approximations of general independent distributions*, Proc. 24th STOC (1992), 10–16.

[12] Luby, M. *A simple parallel algorithm for the maximal independent set problem,* Proc. 17th Ann. ACM Symp. Theory of Comput. (1985), 1–10.

[13] Matoušek, J., Welzl, E., Wernisch, L. *Discrepancy and ε-approximations for bounded VC-dimension*, Combinatorica, to appear; also in Proc. 32nd Ann. IEEE Symp. Foundat. Computer Science (1991), 424–430.

[14] Naor, J., Naor, M. *Small-bias probability spaces: efficient constructions and applications,* Proc. 22nd Ann. ACM Symp. Theory of Comput. (1990), 213–223.

[15] Niederreiter, H. *Random number generation and quasi-Monte Carlo methods*, CBMS-NSF, 1992.

[16] Raghavan, P. *Probabilistic construction of deterministic algorithms: approximating packing integer programs*, J. Comput. System Sci. 37 (1988), 130–143.

[17] Roth, K.F. *On irregularities of distribution*, Mathematika, 1 (1954), 73–79.

[18] Santaló, L.A. *Integral geometry and geometric probability*, Encyclopedia of Math. and its Applicat., Vol. 1, Ed. Gian-Carlo Rota, Addison-Wesley, Reading, Mass., 1976.

[19] Spencer, J. *Ten Lectures on the Probabilistic Method*, CBMS-NSF, SIAM, 1987.

[20] Stolarsky, K.B. *Sums of distances between points on a sphere. II*, Proc. Amer. Math. Soc., 41 (1973), 575–582.