1

Combinatorial Discrepancy

Let $(V, \mathcal{S})$ be a set system, where $V = \{v_1, \ldots, v_n\}$ is the ground set and $\mathcal{S} = \{S_1, \ldots, S_m\}$, with $S_i \subseteq V$. (Such a combinatorial structure is often called a *hypergraph*.) We wish to color the elements of V red and blue so that, within each S_i, no color outnumbers the other one by too much. To make this notion precise, we introduce a function χ mapping each $v_j \in V$ to a "color" in $\{-1, 1\}$, and we define the *discrepancy* of the set S_i to be

$$\chi(S_i) = \sum_{v_j \in S_i} \chi(v_j).$$

The maximum value of $|\chi(S_i)|$, over all $S_i \in \mathcal{S}$, is called the discrepancy of the set system (under the given coloring). When no particular coloring is understood, the *discrepancy* of the set system, denoted by $D_\infty(\mathcal{S})$, refers to its minimum discrepancy over all possible colorings.[1]

This type of discrepancy is called *combinatorial* or, more evocatively, *red-blue*. By contrast with some of the discrepancies discussed in subsequent chapters, which involve both continuous and discrete distributions, the red-blue discrepancy compares two discrete distributions. Both types are intimately linked, however, and techniques for red-blue discrepancy often extend effortlessly to the continuous case.

Discrepancy has been defined in the worst-case sense, ie, in the L^∞ norm. This is intuitively appealing but difficult to manipulate algebraically. The L^2 norm provides a friendlier environment, so we define

$$D_2(\mathcal{S}) \stackrel{\text{def}}{=} \min_\chi \sqrt{\chi(S_1)^2 + \cdots + \chi(S_m)^2},$$

[1] For technical convenience, we use absolute values for the discrepancy of set systems but not when referring to the discrepancy of a particular subset.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

over all colorings $\chi : V \mapsto \{-1, 1\}$. This suggests an algebraic characterization of the discrepancy using matrices. Let A be the *incidence matrix* of the set system $(V, \mathcal{S})$; this is the matrix whose n columns are indexed by the elements of V and whose m rows are the characteristic vectors of the sets S_i, so that A_{ij} is 1 if $v_j \in S_i$ and 0 otherwise. The discrepancy of the set system, also denoted by $D_\infty(A)$, can be expressed as the L^∞ norm of a column vector:

$$D_\infty(A) = \min_{x \in \{-1,1\}^n} \|Ax\|_\infty .$$

Similarly,

$$D_2(A) = \min_{x \in \{-1,1\}^n} \|Ax\|_2 .$$

Here is an overview of this chapter:

- In §1.1 we show that, in the absence of any special assumptions on the set system, a random coloring is nearly optimal. It ensures a discrepancy on the order of $\sqrt{n \log(2m)}$. We give several methods for computing such a coloring deterministically and, in the process, introduce a general derandomization technique.
- We show in §1.2 that if the number of sets in $\mathcal{S}$ is small enough, eg, $O(n)$, then the discrepancy can be kept in $O(\sqrt{n})$. (The bound is proven to be optimal in §1.5.) This gives us the opportunity to introduce the powerful *entropy method* of discrepancy theory.
- In §1.3 we establish the classical Beck-Fiala theorem, which says that if no element belongs to more than a constant number of sets, then the discrepancy can be kept constant.
- We discuss the case of *range spaces* in §1.4. These are well-structured set systems of central importance in discrete and computational geometry. We derive several results that form the foundation of our treatment of geometric sampling in Chapter 4.
- In §1.5 we describe several methods for deriving lower bounds on the discrepancy of set systems. All of them have to do with the spectrum of $A^T A$. The simplest one relates the discrepancy to the smallest eigenvalue. We apply this *eigenvalue bound* to derive a classical theorem of Roth on the discrepancy of arithmetic progressions. This result is optimal, but in general the eigenvalue bound is weak because it does not exploit the fact that the coloring x is a vector with ± 1 coordinates. To do that, we introduce the notion of *hereditary discrepancy* and show how determinants

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

can be used to prove lower bounds. We give an application to set systems formed by points and halfplanes. Finally, we derive the powerful *trace bound*, which allows us to avoid determinants and eigenvalues altogether and prove tight lower bounds in a surprisingly simple manner. We give two examples: points in lines, and points in higher-dimensional boxes.

1.1 Greedy Methods

Given a set system $(V, \mathcal{S})$, with $|V| = n$ and $|\mathcal{S}| = m$, pick a random coloring χ, meaning that for each v_j, the "color" $\chi(v_j)$ is chosen randomly, uniformly, and independently, in $\{-1, 1\}$. We say that S_i is *bad* if $|\chi(S_i)| > \sqrt{2|S_i| \ln(2m)}$. By Chernoff's bound,[2] we immediately derive

$$\text{Prob}[\, S_i \text{ is bad }] < \frac{1}{m};$$

therefore, with nonzero probability, no S_i is bad.

Theorem 1.1 *The discrepancy of a set system $(V, \mathcal{S})$ does not exceed $\sqrt{2n \ln(2m)}$, where $|V| = n$ and $|\mathcal{S}| = m$. This is achieved by a random coloring.*

Let us slightly relax the bound and say that S_i is bad if

$$|\chi(S_i)| > \sqrt{3|S_i| \ln(2m)}.$$

Then, by Chernoff's bound, the probability that no S_i is bad exceeds $1 - 1/\sqrt{m}$. Note that if the first coloring we try fails, we should keep on trying. The probability of being still unsuccessful after k attempts is only $O(1/m^{k/2})$.

The Method of Conditional Expectations

We now describe a general technique for derandomizing the probabilistic coloring algorithm, ie, transforming it into one that does the same thing without using random bits.

The idea is to assign $\chi(v_1)$, $\chi(v_2)$, etc, in that order, without ever backtracking. Let $B = \sum_{i=1}^{m} B_i$, where B_i is the indicator variable equal to 1

[2] See Lemma A.5.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

if S_i is bad and 0 otherwise. We know that

$$\mathbf{E}B = \sum_{i=1}^{m} \mathbf{E}B_i = \sum_{i=1}^{m} \text{Prob}[\, S_i \text{ is bad }] < \frac{1}{\sqrt{m}}. \tag{1.1}$$

Let $\varepsilon_1 = \pm 1$ be such that

$$\mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1 \,] \leq \mathbf{E}[\, B \,|\, \chi(v_1) = -\varepsilon_1 \,].$$

We have

$$\mathbf{E}B = \mathbf{E}_{\chi(v_1)}\, \mathbf{E}[\, B \,|\, \chi(v_1) \,] \geq \mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1 \,]. \tag{1.2}$$

In general, let $\varepsilon_k \in \{-1, 1\}$ minimize the function of x,

$$\mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_{k-1}) = \varepsilon_{k-1}, \chi(v_k) = x \,].$$

Note that

$$\begin{aligned} &\mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_{k-1}) = \varepsilon_{k-1} \,] \\ &\quad = \mathbf{E}_{\chi(v_k)}\, \mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_{k-1}) = \varepsilon_{k-1}, \chi(v_k) \,] \\ &\quad \geq \mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_k) = \varepsilon_k \,]. \end{aligned}$$

It follows from (1.2) that

$$\mathbf{E}B \geq \mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_k) = \varepsilon_k \,].$$

At $k = n$, no randomness is left, so from (1.1),

$$\frac{1}{\sqrt{m}} > \mathbf{E}B \geq \mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_n) = \varepsilon_n \,].$$

The right-hand side denotes the number of bad S_i's in the final coloring, which, being less than one, is therefore zero. Thus, the assignment $\chi(v_i) = \varepsilon_i$ $(1 \leq i \leq n)$ guarantees that each S_i satisfies

$$|\chi(S_i)| \leq \sqrt{3|S_i| \ln(2m)}.$$

The entire procedure can be carried out in polynomial time. Indeed, there are n basic coloring steps, and each of them involves the calculation of two conditional expectations of the form

$$\mathbf{E}[\, B \,|\, \chi(v_1) = \varepsilon_1, \ldots, \chi(v_k) = \varepsilon_k \,].$$

Each such conditional expectation is a sum of m terms of the form

$$\text{Prob}[\, |\chi(S_i)| > \sqrt{3|S_i| \ln(2m)} \;\big|\; \chi(v_1) = \varepsilon_1, \ldots, \chi(v_k) = \varepsilon_k \,],$$

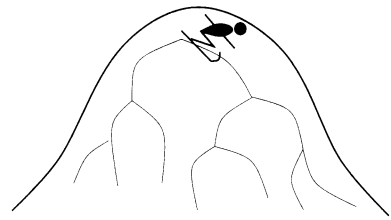
each of which is a sum of at most $2n$ probabilities from the binomial distri-

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

bution[3] $B(|S_i \cap \{v_{k+1}, \ldots, v_n\}|, 1/2)$. Note that the difficulty of computing huge binomial coefficients is easily circumvented. Using the bound of $1/\sqrt{m}$ in (1.1) provides some slack that allows us to perform calculations with relative error $1/(nm)^{O(1)}$. This, in turn, makes a computer word size of $O(\log(n+m))$ sufficient.

The algorithm we just described is an instance of a very general derandomization technique. We will encounter it again in Chapter 7. It is *greedy* in that it follows a locally optimal strategy, but the cost function encodes information about the future. Intuitively, the idea is to keep, at all times, the relative density of good events reachable from the current state bounded from below. As long as these densities—or good enough approximations thereof—can be computed effectively, the method yields a polynomial algorithm for reaching a good event.

A steep trail might be followed by a flat portion. So, to ensure a fast descent, at each fork the skier opts for the trail whose average slope over all descents from that trail is maximum in absolute value. The decision is not a local one.



The Hyperbolic Cosine Algorithm

A somewhat simpler approach is to choose a cost function based on the partial discrepancies incurred up to the current point in time. Suppose that $\chi(v_1), \ldots, \chi(v_k)$ have already been assigned. For each S_i, let $p_{i,k}$ (resp. $m_{i,k}$) be the number of $v_j \in S_i$ $(j \leq k)$, such that $\chi(v_j) = 1$ (resp. $\chi(v_j) = -1$). For $1 \leq k \leq n$, we define $H(k) = \sum_{1 \leq i \leq m} H(i,k)$, where[4]

$$H(i,k) = \cosh(\alpha(p_{i,k} - m_{i,k}))$$

and $\alpha = \sqrt{2\ln(2m)/n}$. Note that $p_{i,k} - m_{i,k}$ is precisely the "current" discrepancy $\chi(S_i)$. The strategy is to choose the assignment of $\chi(v_{k+1}) = \pm 1$ that produces the smaller value of $H(k+1)$. If $v_{k+1} \notin S_i$, then obviously $H(i,k+1) = H(i,k)$. Otherwise, by elementary properties of the hyperbolic cosine, the two possible values of $H(i,k+1)$ average to exactly $H(i,k)\cosh(\alpha)$. It follows that the two values of $H(k+1)$ corresponding

[3] See Appendix A.

[4] Recall that $\cosh x = (e^x + e^{-x})/2$.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

to $\chi(v_{k+1}) = \pm 1$ average to at most $H(k)\cosh(\alpha)$, which, by taking Taylor expansions, is easily shown to be less than $H(k)e^{\alpha^2/2}$. This remains true if we extend $H(k)$ to the case $k = 0$ by setting $H(0) = m$. It follows that $H(k) < me^{k\alpha^2/2}$, for $k > 0$, and hence,

$$\tfrac{1}{2}\, e^{\alpha \max_i |\chi(S_i)|} < \cosh(\alpha \max_i |\chi(S_i)|) \leq H(n) < me^{n\alpha^2/2}.$$

Thus, the discrepancy of the set system is at most $\sqrt{2n\ln(2m)}$.

Remark: The weight function $H(k)$ takes into account only the discrepancies of the prefixes of sets in $\mathcal{S}$ examined so far, and not the respective sizes of these prefixes. So, a small prefix may end with a discrepancy similar to that of a large prefix. For example, if all of the sets have linear size except for one of them that is very small and is spread evenly among $v_1, \ldots, v_n$, then $H(k)$ will not be influenced much by the small set, and an adversary can easily drive up its discrepancy as high as linear in its actual size. Our next discussion corrects this undesirable feature.

The Unbiased Greedy Algorithm

In the event that some sets of $\mathcal{S}$ might be small, it is desirable to have the stronger inequality, $|\chi(S_i)| \leq \sqrt{2|S_i|\ln(2m)}$, for each $S_i \in \mathcal{S}$, as was provided by the derandomization method. A simple modification of the cost function achieves just that. We follow the same approach as the one used in the hyperbolic cosine algorithm. Only the definition of the cost function, renamed $G(k)$, is different. Given $S_i \in \mathcal{S}$, we fix a parameter $\varepsilon_i \in (0,1)$, to be specified later, and define, for $1 \leq k \leq n$, $G(k) = \sum_{1 \leq i \leq m} G(i,k)$, where

$$G(i,k) = (1+\varepsilon_i)^{p_{i,k}}(1-\varepsilon_i)^{m_{i,k}} + (1+\varepsilon_i)^{m_{i,k}}(1-\varepsilon_i)^{p_{i,k}}.$$

We can verify that the two possible values of $G(k+1)$ average out to $G(k)$ (hence the term *unbiased*). Thus, always picking the assignment of $\chi(v_{k+1})$ that minimizes $G(k+1)$ implies that $G(k+1) \leq G(k)$. It is natural to define $G(0) = 2m$. Obviously, $G(i,n) \leq G(n) \leq 2m$, for any $1 \leq i \leq m$. Now, observe that

$$G(i,n) = (1-\varepsilon_i^2)^{\frac{|S_i|-|\chi(S_i)|}{2}} \left((1+\varepsilon_i)^{|\chi(S_i)|} + (1-\varepsilon_i)^{|\chi(S_i)|}\right). \tag{1.3}$$

It follows that

$$(1-\varepsilon_i^2)^{\frac{|S_i|-|\chi(S_i)|}{2}} (1+\varepsilon_i)^{|\chi(S_i)|} \leq 2m, \tag{1.4}$$

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

and hence

$$|S_i| \ln(1-\varepsilon_i^2) + |\chi(S_i)| \ln\left(\frac{1+\varepsilon_i}{1-\varepsilon_i}\right) \leq 2\ln(2m).$$

Let $\lambda(x) = \frac{1}{2}\ln((1+x)/(1-x))$ and $f(x) = \lambda(x)^2 + \ln(1-x^2)$. For any $x \in [0,1)$, $f(x) \geq 0$. (This follows trivially from the fact that $f(0) = f'(0) = 0$ and that the derivative of $(1-x^2)f'(x)$ is positive over $(0,1)$.) We derive

$$2|\chi(S_i)|\lambda(\varepsilon_i) - |S_i|\lambda(\varepsilon_i)^2 \leq 2\ln(2m).$$

Since $\lambda(\varepsilon_i)$ varies continuously from 0 to infinity as ε_i goes from 0 to 1, we can choose

$$\lambda(\varepsilon_i) = \sqrt{2\ln(2m)/|S_i|},$$

which gives us $|\chi(S_i)| \leq \sqrt{2|S_i|\ln(2m)}$, as desired.

Theorem 1.2 *In $O(nm)$ time, it is possible to color the elements of a set system $(V,\mathcal{S})$ such that the discrepancy of any $S_i \in \mathcal{S}$ is at most $\sqrt{2|S_i|\ln(2m)}$ in absolute value, where $n = |V|$ and $m = |\mathcal{S}|$.*

Remark: The unbiased greedy approach entails nothing more than replacing the multiplicative factor $e^{\pm\alpha}$ of the hyperbolic cosine algorithm by the first two terms of its Taylor expansion, that is, $1 \pm \alpha$. This modification implies that the assignment of the next $\chi(v)$ corresponds to a fair game, ie, a random choice multiplies $G(k)$ by 1 on average, as opposed to $\cosh(\alpha)$ in the case of the hyperbolic cosine algorithm. By setting a fixed upper bound on $G(k)$, we thus prevent big sets from overinfluencing the assignment process. Indeed, observe that in the expression for $G(i,n)$ given in (1.3), both the discrepancy and the size of S_i are taken into account.

1.2 The Entropy Method

We consider the particular case of a set system $(V,\mathcal{S})$, where $|V| = |\mathcal{S}| = n$ (which is easily generalized to the nonsquare case). The method is based on the use of partial colorings and the pigeonhole principle. The entropy function is used to simplify an otherwise complicated counting argument. The idea is to argue that many colorings have almost the same discrepancy vectors (ie, each $\chi(S_i)$ differs little among the various colorings). Thus, subtracting two such colorings and dividing by two gives a partial coloring, ie, a coloring in $\{-1,0,1\}$, with low discrepancy. If we can show that the

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

new coloring uses few zeros, then an appropriate recursion "completes" the coloring without increasing the discrepancy by too much.

Theorem 1.3 *Any set system $(V, \mathcal{S})$ such that $|V| = |\mathcal{S}| = n$ has $O(\sqrt{n})$ discrepancy. Some set systems have a matching lower bound.*

The lower bound part of the proof is given in §1.5. Note that the upper bound of $O(\sqrt{n})$ represents a constant number of standard deviations from the random coloring of a given set of size $\Theta(n)$. For this reason, this is often referred to as the *standard deviation bound.* We begin by defining a *partial coloring* of V as a map $\chi : V \mapsto \{-1, 0, 1\}$. As usual, the discrepancy of $S_i \in \mathcal{S}$ is defined as

$$\chi(S_i) = \sum_{v \in S_i} \chi(v).$$

Theorem 1.3 is a simple corollary of the following:

Lemma 1.4 *Let $(V, \mathcal{S})$ be a set system, where $|V| = n$ and $|\mathcal{S}| = m \geq n$. There exists a partial coloring χ of V, such that at most $0.9n$ elements of V are colored zero and, for each $S_i \in \mathcal{S}$,*

$$|\chi(S_i)| < c\sqrt{n \ln \frac{2m}{n}},$$

for some constant $c > 0$.

Apply the lemma to $(V, \mathcal{S})$, and let Z be the subset of 0-colored elements in V. Unless Z is empty, apply the lemma to the subsystem $(Z, \mathcal{S}|_Z)$, where $\mathcal{S}|_Z$ is the collection of subsets of Z of the form $S \cap Z$, for any $S \in \mathcal{S}$. Then, iterate in this fashion until every element of V is ± 1-colored. The discrepancy of each $S_i \in \mathcal{S}$ will be at most

$$\sum_{k \geq 0} c\sqrt{(0.9)^k n \ln \frac{2m}{(0.9)^k n}} = O(\sqrt{n}),$$

which establishes Theorem 1.3. □

It is easy to generalize the theorem to the case where $m > n$. We find that it is possible to two-color the set system $(V, \mathcal{S})$ so that its discrepancy is in $O(\sqrt{n \ln(2m/n)})$. This represents an improvement over the random coloring provided that $m = n^{1+o(1)}$.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

It now remains to prove Lemma 1.4. Let χ_0 be a random two-coloring of V. Given $S_i \in \mathcal{S}$, let

$$\chi_0^*(S_i) = \left\lfloor \frac{\chi_0(S_i)}{c\sqrt{n \ln(2m/n)}} \right\rfloor ,$$

for some large enough constant c. By Chernoff's bound,[5] the probability p_k that $\chi_0^*(S_i) = k$ is less than $(2m/n)^{-ck^2}$. Since the function $f(x) = -x \log x$ increases as x goes from 0 to $1/e$, the entropy of $\chi_0^*(S_i)$ satisfies:

$$H(\chi_0^*(S_i)) \stackrel{\text{def}}{=} \sum_{-\infty<k<+\infty} p_k \log \frac{1}{p_k} < p_0 \log \frac{1}{p_0} + \sum_{|k|>0} ck^2 \left(\frac{2m}{n}\right)^{-ck^2} \log \frac{2m}{n} .$$

For c large enough, we easily verify that

$$\sum_{|k|>0} p_k < \frac{n}{cm} ,$$

and therefore $-p_0 \log p_0 < n/10m$. It follows that

$$H(\chi_0^*(S_i)) < \frac{n}{5m} .$$

Let χ_0^* be the vector

$$(\chi_0^*(S_1), \ldots, \chi_0^*(S_m)).$$

Because of the subadditivity of entropy (Appendix A.3),

$$H(\chi_0^*) < \frac{n}{5} .$$

Thus, by Lemma A.8, there exists a vector $(d_1, \ldots, d_m)$ such that

$$\text{Prob}[\, \chi_0^* = (d_1, \ldots, d_m)\,] > 2^{-n/5}.$$

In other words, the set C of two-colorings producing the vector $(d_1, \ldots, d_m)$ is of size greater than $2^{4n/5}$. Pick one coloring χ_1 in C and for each $\chi \in C$ form the partial coloring $\chi' = \frac{1}{2}(\chi - \chi_1)$. Note that

$$|\chi'(S_i)| < \frac{c}{2}\sqrt{n \ln \frac{2m}{n}}, \tag{1.5}$$

for each $S_i \in \mathcal{S}$. The number of partial colorings with at most $n/10$ nonzeros is equal to

$$\sum_{0 \le k \le n/10} \binom{n}{k} 2^k < 2^{4n/5} < |C|.$$

[5] See Lemma A.5.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

Therefore, there exists a partial coloring with more than $n/10$ nonzeros that satisfies (1.5), which proves Lemma 1.4. □

1.3 The Beck-Fiala Theorem

We briefly discuss a result commonly known as the Beck-Fiala theorem: It states that efficient colorings are always possible as long as no element of the ground set appears in too many S_i's. Let t be the *degree* of the set system, ie, the maximum number of sets S_i to which any given $v \in V$ belongs. Initialize each $\chi(v_k)$ to 0 and call v_k *undecided*. The algorithm will make the v_k's decided as it goes. A set S_i is said to be *stable* if it contains at most t undecided elements. Because of the degree condition, the number of nonstable sets is strictly less than the number of undecided elements. Thus, if we regard the sequence $(\chi(v_1), \ldots, \chi(v_n))$ as a vector in $\mathbf{R}^n$, by simple linear algebra we can move the vector along (at least) a line while both changing only undecided coordinates and maintaining the discrepancy $\chi(S_i)$ of *each* nonstable set S_i equal to 0.

Let us stop our continuous motion as soon as one (or several) of the $\chi(v_k)$'s becomes equal to ± 1. At that stage, we pull all such v_k's out of the game by declaring them no longer undecided. Note that this might cause new sets S_i to become stable. Obviously, it is still the case that the undecided elements outnumber the nonstable sets, so we can repeat the same process and move the vector of undecided colors accordingly. Iterating in this fashion as long as some v_k remains undecided will eventually make every S_i stable.

Note that the discrepancy of any S_i is 0 until it becomes stable, and from that point on at most t of its elements can have color updates; none of those $\chi(v_k)$ was equal to ± 1 to begin with (else they would not be undecided). So each was in $(-1, 1)$, and therefore the total change on $\chi(S_i)$ amounts to strictly less than $2t$. Since the final value must be integral, $|\chi(S_i)| \leq 2t - 1$ for all $i \leq m$. We have proven the Beck-Fiala theorem.

Theorem 1.5 *The discrepancy of a set system of degree t is less than $2t$.*

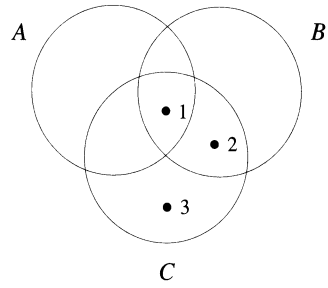
1.4 Discrepancy and the VC-Dimension

Range spaces denote particular (finite or infinite) set systems that arise naturally in geometry. Despite their strong geometric connection, range spaces are defined purely in combinatorial terms. In keeping with common

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

usage, we use the notation $\Sigma = (X, \mathcal{R})$, instead of $(V, \mathcal{S})$, to refer to a range space. For example, X might be a set of n points in $\mathbf{R}^2$, while $\mathcal{R}$ is the collection of sets of the form $X \cap D$, where D is a disk. Note that Σ is a subsystem of the infinite geometric set system $(\mathbf{R}^2, \mathcal{R}^o)$, where $\mathcal{R}^o$ denotes the set of all disks.

The range space consists of three subsets: $A = \{1\}$, $B = \{1, 2\}$, *and* $C = \{1, 2, 3\}$.



It is common to consider infinite range spaces of the form $(\mathbf{R}^d, \mathcal{R}^o)$, where $\mathcal{R}^o$ is obtained by letting a group of transformations act on a fixed subset of $\mathbf{R}^d$; the elements of X are called *points* and the subsets in $\mathcal{R}$ are called *ranges*. In practice, the set systems considered in a geometric context are often subsystems of infinite range spaces. A single parameter, called the Vapnik-Chervonenkis dimension (VC-dimension), characterizes the ability of sampling effectively from such range spaces. In any reference to a "range space" it is usually implicit that the VC-dimension is finite. We shall show that the discrepancy of range spaces is smaller than that of more general set systems. If $|X| = n$, then the discrepancy is within $O(n^{1/2-\varepsilon})$, for some small constant $\varepsilon > 0$, which beats the standard deviation bound of $O(\sqrt{n})$.

Primal and Dual Shatter Functions

Let $\Sigma = (X, \mathcal{R})$ be a (finite or infinite) set system. Given $Y \subseteq X$, let $(Y, \mathcal{R}|_Y)$ denote the set system *induced* by Y, ie, $\{ Y \cap R \mid R \in \mathcal{R} \}$. Note that although the same set $Y \cap R$ may be obtained for several R's, only one copy appears in $\mathcal{R}|_Y$; in other words, $\mathcal{R}|_Y$ is not a multiset. A subset Y of X is said to be *shattered* (by $\mathcal{R}$) if $\mathcal{R}|_Y = 2^Y$, meaning that every subset of Y (including the empty set) is of the form $Y \cap R$, for some $R \in \mathcal{R}$. The supremum of all sizes of finite shattered subsets of X is called the *Vapnik-Chervonenkis dimension* of Σ, or *VC-dimension* for short. For example, it is easy to see that $d + 1$ is the VC-dimension of the range space formed by points and halfspaces in d-space (Fig. 1.1). One should not be fooled, however; in general, evaluating the VC-dimension is no simple matter.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

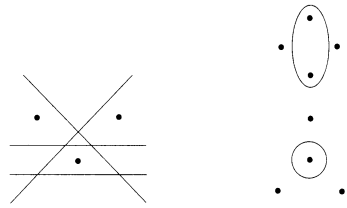


Fig. 1.1. The VC-dimension of halfplanes and points is three, not four.

If X contains arbitrarily large shattered subsets, then we say that the VC-dimension is infinite. In a Euclidean space of fixed dimension, a rule of thumb is that if Σ is a range space defined by some shape of constant description size (eg, a simplex, an ellipsoid, but not, say, an arbitrary convex set), then the VC-dimension is bounded. We define the *shatter function* $\pi_{\mathcal{R}}$ of a range space Σ as follows: $\pi_{\mathcal{R}}(m)$ is the maximum number of subsets in any subsystem $(Y, \mathcal{R}|_Y)$ induced by some $Y \subseteq X$ of size m.

Lemma 1.6 *If the shatter function of an infinite range space is bounded by a fixed-degree polynomial, then its VC-dimension is bounded by a constant. Conversely, if the VC-dimension is $d \geq 1$ then, for any $m \geq d$,*

$$\pi_{\mathcal{R}}(m) \leq \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{d} < \left(\frac{em}{d}\right)^d.$$

Proof: The first part of the lemma is obvious. We prove the inequality on $\pi_{\mathcal{R}}(m)$ by induction on d and m. Let $f(d, m)$ be the maximum value at m of the shatter function of any range space of VC-dimension at most d. Trivially, $f(0, m)$ and $f(d, 0)$ are at most 1, so assume that $d, m > 0$. Fix $Y \subseteq X$ of size m and let $y \in Y$. Let C be the number of distinct sets of the form $Y \cap R$, where $R \in \mathcal{R}$. It might be tempting to say that C is equal to the number A of sets of the form $(Y \setminus \{y\}) \cap R$. But this might fall slightly short of the count, because two distinct R and R' can produce the same set $(Y \setminus \{y\}) \cap R$. For this to happen their restriction to Y must differ by exactly y. Thus, if we define B to be the number of sets $Y \cap R$ that can be expressed as the disjoint union of $Y \cap R'$ $(R' \in \mathcal{R})$ and $\{y\}$, we can then safely write $C = A + B$.

Note that in the definition of B the sets $Y \cap R'$ cannot shatter any

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

subset of $Y \setminus \{y\}$ of size d; otherwise, we could form a subset of X of size $d+1$ that is shattered by $\mathcal{R}$, which would give a contradiction. Therefore, $B \leq f(d-1, m-1)$. Since, obviously $A \leq f(d, m-1)$, we have the recurrence

$$f(d,m) \leq f(d,m-1) + f(d-1,m-1).$$

Checking that $f(d,m)$ satisfies the inequality of the lemma is routine. We can also solve the recurrence visually by counting the number of paths connecting an integral point on the x-axis between 0 and d to the point (d,m), using only vertical edges and edges oriented at 45 degrees. □

The estimate given by the lemma is optimal. It is not too hard to see that infinite range spaces cannot have sublinear shatter functions. In other words, if $\pi_{\mathcal{R}}(m) = o(m)$, then $\mathcal{R}$ is finite, and hence $\pi_{\mathcal{R}}(m) = O(1)$. Also, by keeping track of the growth of shatter functions, it is quite easy to show that the class of range spaces of bounded VC-dimension is closed under union, intersection, and complementation. More precisely, if $(X, \mathcal{R})$ is of bounded VC-dimension, then so is $(X, \mathcal{S})$, where any $\mathcal{S}$ is a finite combination of unions, intersections, and complementations of subsets of $\mathcal{R}$.

If we represent a (finite) set system by its incidence matrix, transposition has an obvious interpretation: We no longer look at which elements lie in a given set but at which sets contain a given element. In other words, we switch the roles of elements and subsets (or points and ranges). Of course, we can do the same even when the set system is an infinite range space. Given a range space $\Sigma = (X, \mathcal{R})$, this suggests introducing the set $\mathcal{R}^* = \{ R_x \,|\, x \in X \}$, where $R_x = \{ R \in \mathcal{R} \,|\, x \in R \}$. The range space $\Sigma^* = (X^*, \mathcal{R}^*)$, where $X^* = \mathcal{R}$, is called the *dual* of Σ.

If Σ is *separable*, meaning that for every $x, y \in X$ there exists $R \in \mathcal{R}$ that contains x but not y (ie, no column appears twice), then duality is involutory; in other words, the dual of $(X^*, \mathcal{R}^*)$ is isomorphic to $(X, \mathcal{R})$. The shatter function of Σ^*, denoted by $\pi^*_{\mathcal{R}}$, is called the *dual shatter function* of $(X, \mathcal{R})$. Although the VC-dimension of a range space might be sometimes quite difficult to evaluate, its dual shatter function is often easier to estimate. For example, in the case of the range space defined by points and balls in d-space, the dual shatter function corresponds to the number of regions into which m balls cut up $\mathbf{R}^d$, which can be shown without difficulty to be $O(m^d)$. To distinguish $(X, \mathcal{R})$ from its dual, we sometimes refer to it as the *primal* range space and we call $\pi_{\mathcal{R}}$ the primal shatter function.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

Lemma 1.7 *If a range space has VC-dimension d, then its dual has VC-dimension less than 2^{d+1}.*

Proof: Arguing by contradiction, suppose that the dual range space has VC-dimension at least 2^{d+1}. This implies the existence of 2^{d+1} ranges of $\mathcal{R}$ that are shattered in the dual range space. In other words, there exist $2^{2^{d+1}}$ points of X such that the 2^{d+1}-by-$2^{2^{d+1}}$ incidence matrix A contains all possible column patterns (this is the matrix whose rows are the characteristic vectors of the chosen ranges with respect to the $2^{2^{d+1}}$ points). Let $u_0, \ldots, u_{2^{d+1}-1}$ denote the $(d+1)$-bit binary representations of $0, 1, \ldots, 2^{d+1}-1$, respectively. Form a 2^{d+1}-by-$(d+1)$ matrix B by making u_0 the first row, u_1 the second row, etc. Each column of B must appear somewhere as a column of A. This shows that the subset of X associated with the columns of A corresponding to those of B is indeed shattered in the primal range space. This subset is of size $d+1$, so we have a contradiction. □

Beating the Standard Deviation Bound

As usual, let $(X, \mathcal{R})$ be a range space of VC-dimension d, with $|X| = n$. A random two-coloring of X ensures that, with reasonably high probability, no color outnumbers the other by more than $O(\sqrt{n \log n})$. By using structural properties of range spaces, it is possible to reduce this upper bound to $o(\sqrt{n})$. Specifically, we show that the discrepancy of $(X, \mathcal{R})$ is within a polylogarithmic factor of $O(n^{1/2-1/2d})$, for $d > 1$. Unfortunately, the proof is inherently existential (using the pigeonhole principle in a manner similar to the treatment of square matrices in §1.2), and it does not yield an efficient coloring algorithm. Recall that by Lemma 1.6 the primal shatter function of the range space is in $O(m^d)$. We prove the stronger result:

Theorem 1.8 *The discrepancy of a range space whose primal shatter function is bounded by cm^d, for some constants $c > 0, d > 1$, is*

$$O(n)^{1/2-1/2d}(\log n)^{1+1/2d}.$$

Note that the "big-oh" notation hides a constant that depends only on c and d. We begin with a simple technical lemma demonstrating once again the usefulness of partial colorings. Recall that a partial coloring is a map $\chi : X \mapsto \{-1, 0, 1\}$.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

Lemma 1.9 *Let $(X, \mathcal{R}_0)$ and $(X, \mathcal{R}_1)$ be two set systems defined on the same ground set X of size n. Assume that*

$$\prod_{R \in \mathcal{R}_0} (|R| + 1) \leq 2^{(n-1)/5},$$

and that $|R| \leq r$, for each $R \in \mathcal{R}_1$. Then there exists a partial coloring $\chi : X \mapsto \{-1, 0, 1\}$ such that

(i) *χ is nonzero over at least one-tenth of X;*
(ii) *$\chi(R) = 0$, for each $R \in \mathcal{R}_0$;*
(iii) *$|\chi(R)| \leq \sqrt{2r \ln(4|\mathcal{R}_1|)}$, for each $R \in \mathcal{R}_1$.*

Proof: Let C be the set of two-colorings of X such that (iii) holds. The argument leading to Theorem 1.1 also shows that $|C| \geq 2^{n-1}$. Order $\mathcal{R}_0$ in arbitrary fashion and, given a two-coloring χ of X, consider the sequence $(\chi(R) : R \in \mathcal{R}_0)$. If $\chi \in C$, then there are at most

$$\prod_{R \in \mathcal{R}_0} (|R| + 1) \leq 2^{(n-1)/5}$$

distinct sequences. (The factor is not $2|R|+1$ because $|R|$ and $\chi(R)$ always have the same parity.) By the pigeonhole principle, there must be a collection C_1 of at least $2^{n-1}/2^{(n-1)/5}$ colorings of C with the same sequence. Choose some $\chi_0 \in C_1$, and for each $\chi \in C_1$ define the partial coloring

$$\chi'(x) = \frac{\chi(x) - \chi_0(x)}{2}.$$

Notice that each χ' satisfies (ii) and (iii). It remains to show that one of them is nonzero over at least one-tenth of X. The number of partial colorings with at most $n/10$ nonzeros is equal to

$$\sum_{0 \leq k \leq n/10} \binom{n}{k} 2^k < 2^{4(n-1)/5} \leq |C_1|.$$

Therefore, there exists a partial coloring of C_1 satisfying (i).

The reader will appreciate the family resemblance between this proof and the entropy method: two different ways of counting essentially the same things. □

We are now ready to prove Theorem 1.8. As we noticed earlier, the class of range spaces of bounded VC-dimension is closed under union, intersection, and complementation. In particular, the range space $(X, \mathcal{S})$, where $\mathcal{S}$ consists of the sets of the form $R \setminus R'$, for $R, R' \in \mathcal{R}$, has bounded VC-dimension. This is best seen by the fact that its primal shatter function is

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

in $O(m^{2d})$. We need to use a result about range spaces that is proven in Chapter 4. Given some parameter $0 < \varepsilon < 1$ to be determined later, a set $N \subseteq X$ that intersects every $S \in \mathcal{S}$ of size greater than $\varepsilon|X|$ is called an ε-net for $(X, \mathcal{S})$. By Theorem 4.3 (page 172), there exists such a set N of size $O(\varepsilon^{-1} \log n)$. (Better bounds can be found, but they are not needed here.)

We "factor" the range space $(X, \mathcal{R})$ by grouping into the same equivalence class all the sets of $\mathcal{R}$ that have the same restriction to N: Two sets R, R' are in the same class if and only if $N \cap R = N \cap R'$. Let $\mathcal{R}_0$ be the subset of $\mathcal{R}$ obtained by taking one representative from each class. For each $R \in \mathcal{R}$, form the sets $R \setminus R_0$ and $R_0 \setminus R$, where R_0 is the representative in the class of R. Let $\mathcal{R}_1$ denote the collection of all such sets (for each $R \in \mathcal{R}$). Note that no $R_1 \in \mathcal{R}_1$ intersects N. Because N is an ε-net for $(X, \mathcal{S})$, it follows that the size of R_1 cannot exceed εn. We verify that by choosing

$$\varepsilon = \frac{c(\log n)^{1+1/d}}{n^{1/d}},$$

for a large enough constant c and setting $r = \varepsilon n$, the conditions of Lemma 1.9 are satisfied. Given any range $R \in \mathcal{R}$, let R_0 be its representative in $\mathcal{R}_0$. Because $R \cap R_0 = R_0 \setminus (R_0 \setminus R)$, we can express R as the disjoint union:

$$R = (R \setminus R_0) \cup (R_0 \setminus (R_0 \setminus R)).$$

Noting that if $B \subseteq A$,

$$|\chi(A \setminus B)| = |\chi(A) - \chi(B)| \leq |\chi(A)| + |\chi(B)|,$$

it follows that

$$\begin{aligned} |\chi(R)| &\leq |\chi(R \setminus R_0)| + |\chi(R_0)| + |\chi(R_0 \setminus R)| \\ &\leq 2\sqrt{2r \ln(4|\mathcal{R}_1|)} \\ &= O(n^{1/2-1/2d})(\log n)^{1+1/2d}. \end{aligned}$$

Let $Y \subseteq X$ be the set of 0-colored points. If Y is nonempty, we repeat the same argument with respect to $(Y, \mathcal{R}|_Y)$ and iterate in this fashion until all the points of X are colored. In the end, the discrepancy of any subset follows (at worst) a geometric progression summing up to $O(n^{1/2-1/2d})(\log n)^{1+1/2d}$. This completes the proof of Theorem 1.8. □

The bound can be reduced to $O(n^{1/2-1/2d})$ by a more complicated argument. It cannot be improved further. Indeed, by Theorem 3.9 (page 156), the red-blue discrepancy of the range space $(\mathbf{R}^d, \mathcal{R})$, where $\mathcal{R}$ is the set

of halfspaces, is $\Omega(n^{1/2-1/2d})$. It is easy to see that the primal shatter function of that range space is in $O(m^d)$.

By using a spanning path argument quite similar to the one given in the next chapter (§2.8), one can prove the theorem below. Again, we must mention that the big-oh notation hides a constant that depends on only c and d. Surprisingly, the upper bound is known to be optimal for all $d > 1$. We omit the proof, which is simply a combinatorial version of the geometric proof of Theorem 2.19 (page 124).

Theorem 1.10 *The discrepancy of a range space whose dual shatter function is bounded by cm^d, for arbitrary constants $c > 0, d > 1$, is $O(n^{1/2-1/2d}\sqrt{\log n})$.*

1.5 Lower Bounds

We discuss spectral techniques for deriving lower bounds on the discrepancy of set systems. This common algebraic thread will persist in our treatment of geometric discrepancy in Chapter 3. There, of course, additional tools, mostly geometric and analytical, will be brought to bear. Although discrepancy questions can be stated purely combinatorially, we are faced here with a situation—unfortunately all too frequent—where counting arguments alone are by and large useless; one notable exception is the matching lower bound of Theorem 1.3 (page 8), which can be derived probabilistically.

All of our techniques rely on asymptotic estimations of the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ of $A^T A$, where A is the incidence matrix of the set system. We show successively how the discrepancy can be bounded from below in terms of (i) the smallest eigenvalue λ_n, (ii) the determinant $\prod \lambda_i$, and (iii) the traces $\sum \lambda_i$ and $\sum \lambda_i^2$. We begin our discussion with a rare case: a set system whose discrepancy can be bounded directly. This warmup exercise nevertheless brings out the spectral flavor that permeates most of this section.

The Hadamard Matrix Bound

It is not hard to exhibit a set system whose discrepancy is $\Omega(\sqrt{n})$. As we just said, this can be established by an elementary, but tedious, counting argument. A more elegant, algebraic proof is given next. Let $H = (h_{ij})$

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

be a Hadamard matrix[6] of order n. The matrix H is orthogonal, and its elements are all ± 1. Here is a Hadamard matrix of order 8:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}.$$

The matrix $A = \frac{1}{2}(H + J)$, where J denotes the matrix full of ones, is the incidence matrix of a set system $(V, \mathcal{S})$, ie, each row of A is the n-bit characteristic vector of a distinct set of the system. We show that its discrepancy in the L^∞ norm is at least $\sqrt{n}/2$. Let H_i (resp. J_i) be the i-th column of H (resp. J). Given a coloring $x \in \{-1, 1\}^n$,

$$\|Ax\|_2^2 = (Ax)^T(Ax) = \frac{1}{4}\sum_{i,j} x_i x_j (H_i + J_i)^T (H_j + J_j).$$

Expanding the sum, we find that:

1. By orthogonality, the term $\sum_{i,j} x_i x_j H_i^T H_j$ is equal to $\sum_i x_i^2 H_i^T H_i$, which is $\sum_i x_i^2 n$.
2. Because $J_j = H_1$, we can write $\sum_{i,j} x_i x_j H_i^T J_j$ as

$$\left(\sum_j x_j\right)\sum_i x_i H_i^T H_1.$$

   By orthogonality, this is $(\sum_j x_j)x_1 n$. Obviously, we find the same value for $\sum_{i,j} x_i x_j J_i^T H_j$ as well.
3. The term $\sum_{i,j} x_i x_j J_i^T J_j$ is equal to $(\sum_i x_i)^2 n$.

Putting everything together, we obtain a lower bound on the L^2 norm of Ax.

[6] See Appendix B.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

$$\begin{aligned} 4\|Ax\|_2^2 &= n\sum_i x_i^2 + 2n\Big(\sum_i x_i\Big)x_1 + n\Big(\sum_i x_i\Big)^2 \\ &= n\Big(x_1 + \sum_i x_i\Big)^2 + n\sum_{i>1} x_i^2 \\ &\geq n\sum_{i>1} x_i^2 = n(n-1). \end{aligned}$$

It follows that at least one coordinate of Ax must exceed $\sqrt{n-1}/2$ in absolute value. This establishes the lower bound of Theorem 1.3 (page 8), for the case where n is a power of 2. The other cases are handled by a standard padding argument.

Although the proof may seem too ad hoc to lend itself to grand statements about lower bounds, it does point the way to the spectral route which we are about to explore now. To minimize $\|Ax\|_2$ for a fixed-length x is a straightforward eigenvalue problem. We formalize this idea below and apply it to the discrepancy of arithmetic progressions.

The Eigenvalue Bound

Let A be the incidence matrix of a set system on n elements; we do not assume that the matrix is square. We consider the mean-square discrepancy $D_2(A)^2$, defined as

$$\min\Big\{\, \|Ax\|_2^2 \,:\, x \in \{-1,1\}^n \,\Big\}.$$

The matrix A^TA is positive semidefinite, and therefore it is diagonalizable and its eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ are nonnegative reals. Suppose that $x = x_1v_1 + \cdots + x_nv_n$, where $\{v_i\}$ is an orthonormal eigenbasis, with v_i associated with λ_i. We have

$$\begin{aligned} \|Ax\|_2^2 = x^TA^TAx &= \Big(\sum_i x_iv_i\Big)^T\Big(\sum_i \lambda_i x_i v_i\Big) \\ &= \sum_{i=1}^n \lambda_i x_i^2 \geq \lambda_n\|x\|_2^2, \end{aligned} \tag{1.6}$$

and thus

$$D_2(A) \geq \sqrt{n\lambda_n}. \tag{1.7}$$

The set $\{-1,1\}^n$ of all "colorings" is contained in the Euclidean sphere of radius $\sqrt{n}$ centered at the origin. Geometrically, A^TA transforms the

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

corresponding ball into an ellipsoid. Indeed, expressed over the eigenbasis, the image of a coloring x under the linear transformation A^TA is a vector whose coordinates in the eigenbasis are $(y_1, \ldots, y_n)$, where

$$\left(\frac{y_1}{\lambda_1}\right)^2 + \cdots + \left(\frac{y_n}{\lambda_n}\right)^2 = \|x\|_2^2 = n.$$

Note that by (1.7) the minimum distance from the origin to the ellipsoid's boundary, which is $\lambda_n\sqrt{n}$, is a lower bound (up to a factor of $\sqrt{n}$) on the mean-square discrepancy. To derive a high lower bound on the discrepancy, we therefore must be able to show that the ellipsoid in question is not too "flat."

What we have just done is to relax the constraints $x_i = \pm 1$ into $x^Tx = n$. This gives rise to the standard quadratic programming problem:

$$\textit{minimize} \quad x^TA^TAx\,,$$

subject to $x^Tx = n$. This leads to minimizing the Rayleigh quotient $\|Ax\|_2^2/\|x\|_2^2$, which by the Courant-Fischer characterization of eigenvalues gives precisely the smallest eigenvalue λ_n. This shows that in the relaxation problem the inequality (1.6) cannot be improved.

Roth's $\frac{1}{4}$-Theorem

We use (1.7) to prove a beautiful result on the discrepancy of arithmetic progressions. Van der Waerden's theorem is a classical result in Ramsey theory, which says that any two-coloring of the integers contains an arbitrarily long monochromatic arithmetic progression. Roth established a complementary result by proving that not all arithmetic progressions can be evenly bicolored. This is known as *Roth's $\frac{1}{4}$-theorem.* It is easily derived from the spectral bound of (1.7).

Theorem 1.11 *Any two-coloring of the integers $\{1, \ldots, n\}$ contains an arithmetic progression whose discrepancy is $\Omega(n^{1/4})$.*

Put differently, there is a constant $c > 0$ such that, no matter how we color the first n integers red or blue, there exists an arithmetic progression over which the numbers of red and blue integers differ by at least $cn^{1/4}$ (Fig. 1.2). The bound of $\Omega(n^{1/4})$ is tight.

Note that, to prove any meaningful lower bound, it is crucial to consider arithmetic progressions of different step sizes (ie, distinct differences between consecutive elements). Indeed, any arithmetic progression of step size 10 can be made of low discrepancy by coloring the first 10 elements

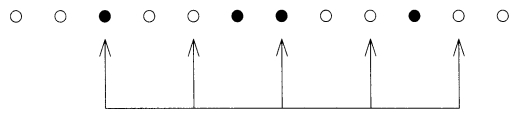


Fig. 1.2. The discrepancy of this arithmetic progression is one.

red, the next 10 elements blue, the following 10 red, etc. The theorem says that neither this coloring scheme nor, for that matter, any other one can be made to handle all step sizes at once.

Why $n^{1/4}$? We know that we need to consider many step sizes. But, of course, we must also deal with long arithmetic progressions, since sparse sets have low discrepancy. We shall occupy the middle ground in trading off step sizes and lengths by choosing progressions of lengths and step sizes roughly $\sqrt{n}$. Notice then that a random coloring guarantees discrepancy in $O(n^{1/4}\sqrt{\log n})$ for those progressions (Theorem 1.1), which is about the bound of the theorem.

We now prove Theorem 1.11. We consider only arithmetic progressions in $\{1,\ldots,n\}$ of length $s = \lfloor\sqrt{n/6}\rfloor$. Such a progression, denoted by $S(p,q)$, is characterized by a starting point p ($1 \le p \le n$) and a step size q ($1 \le q \le 6s$). We construct $S(p,q)$ as follows: Starting at p, we jump by steps of length q and iterate $s-1$ times; so

$$S(p,q) = \{\, p, p+q, p+2q, \ldots, p+(s-1)q \,\}.$$

If we should land past n, we simply wrap around (by performing the jumps modulo n). Note that because $6s(s-1) < n$, there is no risk of reaching p again. This means that $S(p,q)$ is the disjoint union of two "standard" arithmetic progressions.

Given $1 \le q \le 6s$, let A_q be the n-by-n matrix whose p-th row is the characteristic vector of the set $S(p,q)$. The matrix A_q is a circulant matrix[7] obtained by permuting cyclically the characteristic vector of $S(1,q)$. Because it is circulant, the inner product of two column vectors $A_q^{(i)}$ and $A_q^{(j)}$ is equal to the inner product of $A_q^{(i+1)}$ and $A_q^{(j+1)}$ (superscripts are understood to be modulo n). Therefore, $A_q^T A_q$ is also circulant.

[7] Recall that an n-by-n matrix $M = (m_{ij})$ is called circulant if each row past the first one derives from the previous row by shifting each element to the right by one position, ie, $m_{i+1,j+1} = m_{i,j}$ (indices modulo n).

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

We form the incidence matrix A of our set system by stacking up the matrices $A_1, \ldots, A_{6s}$ vertically, one on top of the other:

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{6s} \end{bmatrix}.$$

Note that although A is not circulant, the matrix $M = A^T A$ is equal to $\sum_{q=1}^{6s} A_q^T A_q$ and therefore is itself circulant.

Let $\zeta = e^{2\pi ik/n}$ be an arbitrary n-th root of unity. It is easily verified that $z = (1, \zeta, \ldots, \zeta^{n-1})^T$ is an eigenvector of M. Indeed, observe that the second coordinate of Mz is equal to the first one multiplied by ζ, the third one is the second coordinate multiplied by ζ, and so on. Thus, the vector Mz is obtained by multiplying $(1, \zeta, \ldots, \zeta^{n-1})$ by the first coordinate of Mz. So, clearly z is an eigenvector for M. Let $\lambda(z)$ denote its associated eigenvalue. There are n roots of unity, and the corresponding vectors are orthogonal to each other (see Appendix B); therefore, we have a complete basis of eigenvectors. Also,

$$z^* M z = \lambda(z) z^* z = n\lambda(z),$$

where $*$ denotes the Hermitian transpose. On the other hand,

$$z^* M z = (Az)^* Az = \sum_{q=1}^{6s} \sum_{i=1}^{n} \left| \sum_{j=1}^{n} A_q(i,j) \zeta^{j-1} \right|^2.$$

Note that because $|\zeta| = 1$,

$$\left| \sum_j a_{ij} \zeta^{j-1} \right|^2 = \left| \sum_j a_{ij} \zeta^{j+k} \right|^2,$$

for any k; therefore, all the rows of A_q, for a fixed q, have the same contribution. This yields

$$n\lambda(z) = \sum_{q=1}^{6s} n \left| \sum_{k=0}^{s-1} \zeta^{qk} \right|^2.$$

By the pigeonhole principle, for at least two distinct $1 \leq q_1 < q_2 \leq 6s$, the angles $\arg(\zeta^{q_1})$ and $\arg(\zeta^{q_2})$ fall in an interval of length $2\pi/6s$ (around the unit circle in the complex plane). If this interval contains the angle zero, then for one of the q_i's we have $0 \leq \arg(\zeta^{q_i}) \leq \pi/3s$; otherwise, we have $|\arg(\zeta^{q_0})| \leq \pi/3s$, for $q_0 = q_2 - q_1$. Thus, in general, there exists

$1 \le q_0 \le 6s$ such that

$$\left| \arg(\zeta^{q_0 k}) \right| \le \frac{k\pi}{3s},$$

for each $0 \le k < s$. This shows that the real part of each $\arg(\zeta^{q_0 k})$ is at least $1/2$. Thus,

$$n\lambda(z) \ge n \left| \sum_{k=0}^{s-1} \zeta^{q_0 k} \right|^2 \ge \frac{ns^2}{4}.$$

Since this is true for any eigenvalue of M, it follows from (1.7) that[8]

$$D_\infty(A)^2 \ge \frac{D_2(A)^2}{6sn} \gg \sqrt{n}.$$

Because of the wrap-around it can be argued that A is not the incidence matrix of a set of arithmetic progressions. As we remarked earlier, however, each set represented by A can be partitioned into two valid arithmetic progressions. If the set has high discrepancy, then so must at least one of its two constituent progressions. Thus, the lower bound on $D_\infty(A)$ completes the proof of Theorem 1.11. □

The View from Harmonic Analysis

We give another proof of Theorem 1.11 (page 20), this time using Fourier transforms as our main tool. As it turns out, the proof is really the same as the previous one, even though it looks quite different on the surface. It is instructive to see why, because it brings together two key tools in discrepancy theory, eigenvalues and Fourier transforms, and their common link, the convolution operator. We will discuss this connection in depth in the next two chapters. Our discussion here is to serve as a kinder, gentler introduction to this material.

Recall that our aim is to show that any two-coloring of the integers $\{1, \ldots, n\}$ contains an arithmetic progression whose discrepancy is $\Omega(n^{1/4})$. Fix a coloring χ:

$$\chi(m) = \begin{cases} 1 & \text{if } m \text{ is red and } 1 \le m \le n, \\ -1 & \text{if } m \text{ is blue and } 1 \le m \le n, \\ 0 & \text{else.} \end{cases}$$

[8] Recall that $x \gg y$ means $x = \Omega(y)$.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

As in the previous proof, we are interested in only arithmetic progressions of length within $\sqrt{n}$; put $s = \lfloor \sqrt{n} \rfloor$. Given a step size q, we define the characteristic function c_q of the corresponding arithmetic progression of length $2s + 1$:

$$c_q(m) = \begin{cases} 1 & \text{if } m \text{ is a multiple of } q \text{ and } |m| \leq sq, \\ 0 & \text{else.} \end{cases}$$

Regard $c_q(m)$ as a "comb" centered at 0. Slide it over so that its center coincides with p. The portion of the comb within $[1, n]$ defines an arithmetic progression whose discrepancy we denote by $\Delta_q(p)$. It is immediate that

$$\Delta_q(p) = \sum_{k=1}^{n} \chi(k) c_q(k - p).$$

Since $\chi(k)$ is 0 outside $[1, n]$ and c_q is even, we have

$$\Delta_q(p) = \sum_{k \in \mathbf{Z}} \chi(k) c_q(p - k);$$

in other words, $\Delta_q = \chi \star c_q$. Taking Fourier transforms on the group $\mathbf{Z}$ (see Appendix B), we find that

$$\widehat{c}_q(t) = \sum_{m \in \mathbf{Z}} c_q(m) e^{-2\pi i m t} = \sum_{|k| \leq s} e^{-2\pi i k q t}.$$

By the same pigeonhole argument used in the previous proof, there exists some $1 \leq q(t) \leq bs$ for some fixed b large enough, such that in the sum $\sum_{|k| \leq s} e^{-2\pi i k q(t) t}$ the real part of each summand exceeds some fixed positive constant. Therefore,

$$|\, \widehat{c}_{q(t)}(t) \,| \gg s.$$

By the Parseval-Plancherel identity and the convolution theorem,

$$\begin{aligned} \sum_{q=1}^{bs} \sum_{p \in \mathbf{Z}} \Delta_q(p)^2 &= \sum_{q=1}^{bs} \int_0^1 |\widehat{\Delta}_q(t)|^2 \, dt \\ &= \sum_{q=1}^{bs} \int_0^1 |\widehat{\chi}(t)|^2 \, |\widehat{c}_q(t)|^2 \, dt \geq \int_0^1 |\widehat{\chi}(t)|^2 \, |\widehat{c}_{q(t)}(t)|^2 \, dt \\ &\gg s^2 \int_0^1 |\widehat{\chi}(t)|^2 \, dt = s^2 \sum_{p \in \mathbf{Z}} |\chi(p)|^2 = ns^2. \end{aligned}$$

So, for some step size $q_0 \leq b\sqrt{n}$, we have

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

$$\sum_{p \in \mathbf{Z}} \Delta_{q_0}(p)^2 \gg ns.$$

Since $\Delta_{q_0}(p)$ is zero for all values of x outside an interval of length $O(n)$, we find that $\Delta_{q_0}(p_0)^2 \gg \sqrt{n}$ for some p_0, which proves Theorem 1.11. □

DISCUSSION

Why are the two proofs of Theorem 1.11 (page 20) really the same in disguise? Recall that the matrix $M = A^T A$ (in the first proof) is circulant. Let F be the Fourier matrix[9] of order n. Because M is circulant, given any coloring $x = (x_1, \ldots, x_n)^T \in \{-1, 1\}^n$, the vector Mx is the convolution of x with a certain vector v. Its Fourier transform FMx is therefore the coordinate-wise product of Fv and Fx. If $Fv = (\lambda_1, \ldots, \lambda_n)^T$, it then follows that $FMx = \Lambda Fx$, where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

By premultiplying by the inverse matrix F^{-1}, we find that $M = F^{-1}\Lambda F$. Taking the Fourier transform diagonalizes our matrix. This is no big surprise since it is a "convolution" matrix.

The moral of the story is this: The eigenvalue method is the most general and direct line of attack on the L^2 norm of the discrepancy. In practice, however, getting a handle on the eigenvalues is no simple matter. But whenever the set system is defined by some form of convolution (a "comb" in the case of arithmetic progressions), the Fourier transform method brings those eigenvalues to the fore (via diagonalization). Geometric discrepancy with respect to boxes or disks is defined by translating (and sometimes rotating or scaling) some fixed object across space and defining one subset of the set system for each position: Translating is just like sliding a comb and acts as a convolution operator in defining the set system. Thus, it is little surprise that Fourier transforms should play a major role.

[9] See the discussion of the discrete Fourier transform in Appendix B.

Hereditary Discrepancy and Determinants

The eigenvalue bound is often too weak to be useful because it makes no attempt to exploit the fact that the coloring is a vector with ± 1 coordinates. By introducing the notion of hereditary discrepancy, we are able to use that fact and relate the discrepancy, not only to the minimum eigenvalue, but to the entire spectrum of the incidence matrix.

Let $(V, \mathcal{S})$ be a set system. Given $W \subseteq V$, recall that $\mathcal{S}|_W$ is the collection of subsets of W of the form $S \cap W$, where $S \in \mathcal{S}$. The *hereditary discrepancy* of $(V, \mathcal{S})$, denoted by $\operatorname{herdisc}(\mathcal{S})$, is defined to be the maximum value of $D_\infty(\mathcal{S}|_W)$ over all $W \subseteq V$. The motivation behind this notion is that even though some subsystem might have a huge discrepancy, $D_\infty(\mathcal{S})$ itself might be very small by some "fluke." Indeed, by adding only $O(D_\infty(\mathcal{S}))$ new elements to V and the sets of $\mathcal{S}$, we can easily make the discrepancy vanish entirely. Besides its built-in robustness, the hereditary view has an unintended benefit: It allows us to bound the discrepancy in terms of the full spectrum of eigenvalues and not only the smallest one. We consider the product of the eigenvalues (the determinant) in this section and their sum (the trace) later in this chapter.

How much are we giving up by adopting the hereditary viewpoint? In geometry, not much. Indeed, the hereditary discrepancy is particularly well suited for geometric applications, because geometric set systems typically are hereditary themselves: remove points from a set system of points and disks, and you still get a set system of points and disks. And so, in geometry at least, adding the adjective hereditary before the word discrepancy does not narrow the view.

Theorem 1.12 (THE DETERMINANT BOUND) *If A is an n-by-n incidence matrix of a set system, then*

$$\operatorname{herdisc}(A) \geq \tfrac{1}{2} |\det A|^{1/n}.$$

Corollary 1.13 $\operatorname{herdisc}(A) \geq \frac{1}{2} \max_{k,B} |\det B|^{1/k}$, *where B ranges over all k-by-k submatrices of A.*

To minimize the quadratic form $\|Ax\|_2^2$ is the stuff of linear algebra textbooks. The hereditary discrepancy adds three twists: The vector x has ± 1 coordinates; the norm is not Euclidean; and, if that were not bad enough, all submatrices of A come into play. Our first objective is to find our way back to linear algebra. Once we have done that, we will see that all three twists in fact help produce stronger results than the eigenvalue bound. To prove Theorem 1.12 we introduce a weighted version of the discrepancy.

Recall that

$$D_\infty(A) = \min\Big\{ \|Ax\|_\infty \,:\, x \in U \Big\},$$

where $U = \{-1,1\}^n$. A more general definition might allow the range of x to vary, as in

$$D^c_\infty(A) = \min\Big\{ \|Ax\|_\infty \,:\, x \in c + U \Big\},$$

where $c \in \mathbf{R}^n$. To establish a lower bound on herdisc (A), a reasonable approach is to bound

$$\text{intdisc}\,(A) \stackrel{\text{def}}{=} \max\Big\{ D^c_\infty(A) \,:\, c \in \{-1,0,1\}^n \Big\},$$

in view of the fact that

$$\text{herdisc}\,(A) \geq \text{intdisc}\,(A). \tag{1.8}$$

This inequality is quite obvious. Think of the hereditary discrepancy in the context of a game: First, your adversary sets any number of x_i's to 0; then you complete the coloring x so as to minimize the (regular) discrepancy. To see the connection with $D^c_\infty(A)$, consider any $c \in \{-1,0,1\}^n$. Form a ± 1-coloring by setting $x_i = -c_i$ for each $c_i \neq 0$ and completing the assignment of x by minimizing the (regular) discrepancy. By definition, the resulting discrepancy Δ is at least $D^c_\infty(A)$. On the other hand, pursuing the game analogy, the assignments $x_i = -c_i$ correspond to the adversary's annulling some of the columns. This shows that $\Delta \leq$ herdisc (A) and establishes (1.8). Note that this reasoning should give no reason to think that the inequality is actually an equality; for example, with the set system $\{a\}$, $\{a,b\}$, $\{b,c\}$, $\{a,c,d\}$, D^c_∞ is at most 1 while the hereditary discrepancy is 2.

Relaxing c in the definition of intdisc leads to the *linear discrepancy* of A:

$$\text{lindisc}\,(A) \stackrel{\text{def}}{=} \max\Big\{ D^c_\infty(A) \,:\, c \in [-1,1]^n \Big\}.$$

As one might expect, the benefit of such a relaxation is to make it amenable to linear algebra. Fortunately, relaxing c does not have drastic effects on the discrepancy.

Lemma 1.14

$$\text{lindisc}\,(A) \leq 2\,\text{intdisc}\,(A).$$

We can now finish the proof of the theorem. Given any $c \in [-1,1]^n$,

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

there exists some $x_c \in U$ such that

$$\|A(x_c + c)\|_\infty \leq \text{lindisc}\,(A).$$

Thus, if Y denotes the set of points $y \in \mathbf{R}^n$ satisfying $\|Ay\|_\infty \leq \text{lindisc}\,(A)$, the set $Y + U$ covers the entire cube $[-1,1]^n$ (Fig. 1.3). But the pieces $Y + x$ $(x \in U)$ can be obtained by cutting up a single copy of Y. Formally speaking, Y encloses $[-1,1]^n$ in $(\mathbf{R}/2\mathbf{Z})^n$. (Topologically, we are identifying the opposite facets of U.) It follows that the volume of Y is at least that of U, ie, 2^n. Obviously, we can assume without loss of generality that A is nonsingular (else, $\det A = 0$ and the theorem is trivial). Then, $Y = A^{-1}[-\text{lindisc}\,(A), \text{lindisc}\,(A)]^n$, from which we derive that $\text{vol}(Y) = (2\,\text{lindisc}\,(A))^n |\det A^{-1}|$, and hence $\text{lindisc}\,(A) \geq |\det A|^{1/n}$. In view of (1.8) and Lemma 1.14, the proof of Theorem 1.12 is now complete. □

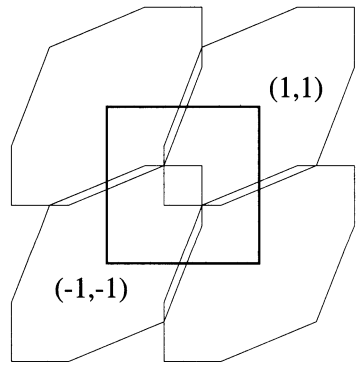


Fig. 1.3. The set $Y + \{-1,1\}^n$ covers the cube $[-1,1]^n$.

It remains for us to prove Lemma 1.14. Put

$$t = \max_{c \in \{-1,0,1\}^n} D_\infty^c(A).$$

It suffices to show that, given any point $c \in [-1,1]^n$, there exists $x_0 \in U + c$ such that $\|Ax_0\|_\infty \leq 2\,\text{intdisc}\,(A)$. This is obviously true if the coordinates of c are integers (ie, $-1, 0, 1$). Let us say that c is *k-good* if its coordinates are rationals in $[-1,1]$ whose binary expansions do not extend beyond the k-th bit (ie, all of the lower order bits are 0 past position k). Proceeding by induction, assume that the proposition is true for any k-good point. Now suppose that c is $(k+1)$-good. It is elementary to see that there always exists a translation vector $a \in U$ such that $b = 2c + a$ falls in

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

the box $[-1,1]^n$. It follows that b is k-good. By induction, there exists $x_1 \in U + b$ such that $\|Ax_1\|_\infty \leq 2\,\mathrm{intdisc}\,(A)$. Thus, for some $x_2 \in U$, we have (dividing by 2)

$$\|A(c+\gamma)\|_\infty \leq \mathrm{intdisc}\,(A),$$

where $\gamma = (x_2+a)/2$. Since $\gamma \in \{-1,0,1\}^n$, we have $D^{[\gamma]}(A) \leq \mathrm{intdisc}\,(A)$, and therefore $\|A(x_3+\gamma)\|_\infty \leq \mathrm{intdisc}\,(A)$, for some $x_3 \in U$. Subtracting the last two inequalities yields $\|A(c-x_3)\|_\infty \leq 2\,\mathrm{intdisc}\,(A)$, which completes the induction. A standard compactness argument finishes the proof of Lemma 1.14. □

Application: Points and Halfplanes

We show how the notion of hereditary discrepancy can be used to prove a tight lower bound for the standard L^∞ (red-blue) discrepancy formed by points and halfplanes. The discrepancy of a square matrix like Hadamard was easily bounded, but in general specific matrices are hopelessly difficult to tackle. Chapter 3 treats the case of several geometric incidence matrices, but these are not square. The determinant bound for the hereditary discrepancy allows us to derive a tight bound on the discrepancy of an important class of square matrices.

Let $A = (a_{ij})$ be the n-by-n incidence matrix of a set system formed by n points in the plane and n closed halfplanes: $a_{ij} = 1$ if the i-th halfplane contains the j-th point, else $a_{ij} = 0$. We prove the following bound, which is optimal:

Theorem 1.15 *There exist n points and n halfplanes in $\mathbf{R}^2$, such that the n-by-n incidence matrix $A = (a_{ij})$ has discrepancy $D_\infty(A) = \Omega(n^{1/4})$.*

The point set $\{p_i\}$ consists of the n integer points in $[1, \sqrt{n}\,]^2$; we assume that n is a large square. The discrepancy vector x is formed by associating its i-th coordinate x_i with the ± 1-color of point p_i. Let us relax the assumption that $x_i = \pm 1$ and instead regard x as any vector in $\mathbf{R}^n$. Given a closed halfplane h bounded above by a nonvertical line, let $f(h)$ denote the sum $\sum_{p_i \in h} x_i$. We define ω to be the unique motion-invariant measure for lines that provides a probability measure for the lines crossing the square $[1, \sqrt{n}\,]^2$; see [265] for details.[10] Alexander [10] has proven that if

[10] Intuitively, the probability that a random line hits an object should not depend on

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

$x_1 + \cdots + x_n = 0$, then

$$\int f(h)^2 \, d\omega(h) \gg \frac{\|x\|_2^2}{\sqrt{n}} . \tag{1.9}$$

This bound is an easy consequence of the finite differencing method developed in Chapter 3, so we do not reproduce it here.

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 large enough, say, $N = 2^n$, we can ensure that the ω-area σ of N of these regions is exactly the same, say around $1/N$, while the other $O(n^2)$ regions have smaller areas. Computing $\int f(h)^2 \, d\omega(h)$ by integrating f^2 over only the equal-area regions produces an absolute error of $O(n^2/N) \sup f^2$. Obviously, $|f(h)|$ cannot exceed

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

by Cauchy-Schwarz, so the error is bounded by $O(n^3 \|x\|_2^2/N)$. We define B to be the N-by-n matrix whose rows are indexed by the N equal-area regions $\widehat{\sigma}$ and are the characteristic vectors of the set of x_i's appearing in (the unique linear form) $f(h)$, given $h \in \widehat{\sigma}$. It follows that

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

Because $\sigma = 1/N \pm O(n^2/N^2)$, we have

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

Lemma 1.16

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

Proof: Let $\mu_1 \ge \cdots \ge \mu_n \ge 0$ 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 rank of the linear system

$$\begin{cases} x_1 + \cdots + x_n = 0 \\ \xi_j = 0 \quad (j < n-1) \end{cases}$$

is at most $n - 1$. Feasible solutions lie in the (ξ_{n-1}, ξ_n)-plane, so they intersect the cylinder $\xi_{n-1}^2 + \xi_n^2 = 1$. A solution x at the intersection is of

its particular placement but only on its shape. In the plane, the measure for a line $h : ax + by = 1$ has density $d\omega(h) = c(a^2 + b^2)^{-3/2} \, da \, db$, for some normalizing constant $c > 0$ adjusted to make the probabilities sum up to one.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

unit length, so

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

and

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

hence,

$$\mu_{n-1} \geq \Omega(N/\sqrt{n}). \tag{1.10}$$

Next, we derive a lower bound on the smallest eigenvalue. With N large enough, we can easily 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 $x_i^2/2$. We conclude that $\|Bx\|_2^2 \geq \frac{1}{2}\|x\|_2^2$, and so $\mu_n \geq 1/2$. Since $\det B^T B$ is the product of the eigenvalues, the lemma follows from (1.10). □

The Binet-Cauchy formula says that[11]

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

This implies the existence of an n-by-n submatrix A of B such that

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

hence,

$$\det A^T A = \Omega(n)^{n/2}. \tag{1.11}$$

Bringing in the hereditary discrepancy herdisc (A), it follows from Theorem 1.12 (page 26) that

$$\text{herdisc}\,(A) = \Omega(n^{1/4}).$$

Let A' be the (or any) submatrix of A whose discrepancy is this hereditary

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

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

discrepancy. The matrix M derived from A by zeroing out the columns not in the submatrix A' is the incidence matrix of a set system of n halfplanes and at most n points. We can make it n-by-n by adding artificial points outside all of the halfplanes (which is possible since they all face down). This completes the proof of Theorem 1.15. □

The Trace Bound

The determinant bound of Theorem 1.12 has two weaknesses: One is that the matrix might have high discrepancy and null determinant (say, one row is duplicated); the other one is that determinants for set systems can be very difficult to estimate asymptotically.[12] We establish a connection between the hereditary discrepancy and the traces of $A^T A$ and its square.

Theorem 1.17 (THE TRACE BOUND) *If A is an n-by-n incidence matrix and $M = A^T A$, then*

$$\operatorname{herdisc}(A) \geq \frac{1}{4}\, c^{n \operatorname{tr} M^2 / \operatorname{tr}^2 M} \sqrt{\frac{\operatorname{tr} M}{n}},$$

for some constant $0 < c < 1$.

How does this compare with the determinant bound? The latter can be rewritten as roughly $\sqrt{(\det M)^{1/n}}$, and so inside the square roots we find the arithmetic mean of the eigenvalues of M being compared against the (never bigger) geometric mean: a sign of progress. There is a correction factor, however. Note that it is inevitable since $\sqrt{\operatorname{tr} M/n}$ alone cannot bound the discrepancy: Try the matrix A full of ones to see why. For the trace bound to be of any use, however, it is crucial that the exponent $n \operatorname{tr} M^2 / \operatorname{tr}^2 M$ in the correction factor be essentially constant. One easily verifies that if θ is the angle between the vectors $(1, \ldots, 1)$ and $(\lambda_1, \ldots, \lambda_n)$ then, by projection,

$$\frac{n \operatorname{tr} M^2}{\operatorname{tr}^2 M} = \frac{1}{\cos^2 \theta}.$$

[12] For example, the Riemann hypothesis can be expressed as an upper bound on a very simple determinant. The Redheffer matrix [28] has 1's in the leftmost column and at entry (i, j) if i divides j, and 0's elsewhere. Its determinant is the Mertens function $\sum_{k=1}^{n} \mu(k)$, where $\mu(n)$ is the Möbius function. It is $O(n^{1/2+\varepsilon})$, for any $\varepsilon > 0$, if and only if the Riemann hypothesis is true.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

So, to say that the exponent should not be large is another way of requiring that the eigenvalue distribution be fairly uniform. Fortunately, the correction factor is constant in many applications.

Be that as it may, what gives the trace bound its power is that both $\operatorname{tr} M$ and $\operatorname{tr} M^2$ have simple combinatorial meanings. For example, the trace of M is the number of ones in A, ie, the number of incidences in the set system. Similarly, the trace of M^2 is the number of rectangles of ones in A or, equivalently, the number of closed paths (simple and non simple) of length 4 in the bipartite graph corresponding to A. The trace bound follows easily from the lemma below.

Lemma 1.18 *For any* $1 \leq k \leq n$,

$$\operatorname{lindisc}(A) \geq 18^{-n/k}\sqrt{\lambda_k},$$

where $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ *are the eigenvalues of* $M = A^T A$.

To see why the lemma implies Theorem 1.17, we use a second-moment probabilistic argument. For $x \geq 0$, let $\mathcal{E}_x$ be the event, $\lambda \geq \mathbf{E}\lambda - x$, and let p be its probability. The sequence of derivations,

$$\begin{aligned} 0 &= \mathbf{E}[\lambda \,|\, \mathcal{E}_x]\, p + \mathbf{E}[\lambda \,|\, \overline{\mathcal{E}}_x]\,(1-p) - \mathbf{E}\lambda \\ &= (\mathbf{E}[\lambda \,|\, \mathcal{E}_x] - \mathbf{E}\lambda)\, p + (\mathbf{E}[\lambda \,|\, \overline{\mathcal{E}}_x] - \mathbf{E}\lambda)\,(1-p) \\ &\leq (\mathbf{E}[\lambda \,|\, \mathcal{E}_x] - \mathbf{E}\lambda)\, p - x(1-p), \end{aligned}$$

leads to

$$\mathbf{E}[\lambda \,|\, \mathcal{E}_x] \geq \mathbf{E}\lambda + x(1/p - 1).$$

Consider the random variable $\mathbf{E}[\lambda \,|\, Y]$. Let Y be the event $\mathcal{E}_x$ with probability p and $\overline{\mathcal{E}}_x$ with probability $1-p$. The conditional variance $\mathbf{var}[\lambda \,|Y]$, defined as the variance of the random variable $\mathbf{E}[\lambda \,|\, Y]$, cannot exceed the (unconditional) variance and therefore,

$$\begin{aligned} \mathbf{var}\,\lambda &\geq \mathbf{var}\,\mathbf{E}[\lambda \,|\, Y] = \mathbf{E}\,(\mathbf{E}[\lambda \,|\, Y] - \mathbf{E}\,\lambda)^2 \\ &\geq (\mathbf{E}[\lambda \,|\, \mathcal{E}_x] - \mathbf{E}\lambda)^2 p + (\mathbf{E}[\lambda \,|\, \overline{\mathcal{E}}_x] - \mathbf{E}\lambda)^2 (1-p) \\ &\geq x^2(1/p-1)^2 p + x^2(1-p) \geq x^2(1/p-1), \end{aligned}$$

which shows that

$$p \geq \frac{1}{1 + x^{-2}\,\mathbf{var}\lambda}\,. \tag{1.12}$$

By setting $x = 3\mathbf{E}\lambda/4$ and k to be about $n/(2\mathbf{var}\lambda/\mathbf{E}^2\lambda + 1)$, we find that $\lambda_k \geq \mathbf{E}\lambda/4$. Since $\mathbf{E}\lambda = \operatorname{tr} M/n$ and $\mathbf{var}\lambda = \operatorname{tr} M^2/n - (\operatorname{tr} M)^2/n^2$, it

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

follows from (1.8) and Lemmas 1.14 and 1.18 that

$$\text{herdisc}\,(A) \geq \frac{1}{2}\,\text{lindisc}\,(A) \geq \frac{1}{2}\,18^{-n/k}\sqrt{\lambda_k} \geq \frac{1}{4}\,18^{-2n\,\text{tr}\,M^2/\text{tr}^2\,M}\sqrt{\frac{\text{tr}\,M}{n}}\,,$$

which proves Theorem 1.17. □

Another interesting expression for the tail of the eigenvalue distribution is obtained by setting $x = \varepsilon\sqrt{\text{tr}\,M^2/n}$ in (1.12). Then,

$$\text{Prob}\left\{\lambda \geq \text{tr}\,M/n - \varepsilon\sqrt{\text{tr}\,M^2/n}\right\} \quad \geq \quad \frac{1}{1 + n\varepsilon^{-2}\,\textbf{var}\lambda/\text{tr}\,M^2} \geq \frac{1}{1+\varepsilon^{-2}},$$

which is independent of n. We conclude:

Lemma 1.19 *Let A be an n-by-n 0/1 matrix, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $M = A^TA$. Then, for any fixed $\varepsilon > 0$,*

$$\lambda_k \geq \text{tr}\,M/n - \varepsilon\sqrt{\text{tr}\,M^2/n}\,,$$

for some $k = \Omega_\varepsilon(n)$.

We now prove Lemma 1.18. A singular-value decomposition of the matrix A allows us to rewrite it as UDV^T, where U (resp. V) is the orthogonal matrix whose columns are the eigenvectors of AA^T (resp. A^TA) and D is the n-by-n diagonal matrix whose only nonzero entries are $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots$ (the singular values of A, which are the square roots of the eigenvalues of A^TA or, equivalently, AA^T). Let L be the subspace spanned by the k eigenvectors of A^TA corresponding to $\lambda_1, \ldots, \lambda_k$. The projection of a unit cube in $\mathbf{R}^n$ to a k-dimensional subspace is a convex polytope of volume between c^{-n} and c^n, for some constant $c > 0$. A simple argument that adds the contribution of each of the $\binom{n}{k}2^{n-k}$ k-faces of the cube shows that $c \leq 3$. It follows that

$$\text{vol}\,(A\,\text{proj}_L[-1,1]^n) = \sqrt{\lambda_1 \cdots \lambda_k}\,\text{vol}\,(\text{proj}_L[-1,1]^n) \geq 2^k 3^{-n}\lambda_k^{k/2}. \tag{1.13}$$

Given any $x \in L$ and $y \in L^\perp$, A^TAx lies in L and so $(Ax)^T(Ay) = (A^TAx)^Ty = 0$. In fact, not only are AL and $A(L^\perp)$ orthogonal, but they span all of $A\mathbf{R}^n$ and therefore $(AL)^\perp = A(L^\perp)$. It easily follows that

$$A\text{proj}_L[-1,1]^n = \text{proj}_{AL}A[-1,1]^n$$

and by (1.13)

$$\text{vol}\,(\text{proj}_{AL}A[-1,1]^n) \geq 2^k 3^{-n}\lambda_k^{k/2}.$$

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

By definition (page 27), for any $c \in [-1,1]^n$, there exists some $x \in \{-1,1\}^n$ such that $Ac = Ax + y$, where $y \in [-\text{lindisc}\,(A), \text{lindisc}\,(A)]^n$. The image of the cube $[-1,1]^n$ under the transformation given by A is a polyhedron in $\mathbf{R}^n$ whose vertices belong to $A\{-1,1\}^n$. It follows that the polyhedron $A[-1,1]^n$ is covered by the $\leq 2^n$ n-dimensional cubes of side length $2\,\text{lindisc}\,(A)$ centered at the vertices of $A[-1,1]^n$. Projecting onto AL and accounting for the dilation factor of 3^n, we find that

$$\text{vol}\,(\text{proj}_{AL} A[-1,1]^n) \leq 6^n (2\,\text{lindisc}\,(A))^k,$$

which proves Lemma 1.18, and hence the trace bound. □

Application I: Points and Lines

In Chapter 6, we prove the existence of n points and n lines in the plane, all of them distinct, such that each point belongs to $\Theta(n^{1/3})$ lines and each line contains $\Theta(n^{1/3})$ points (Lemma 6.25, page 263). The trace of M is the number of incidences, ie, $\Theta(n^{4/3})$. The trace of M^2 is equal to the number of rectangles of ones in A. Since no proper rectangle can occur (two lines intersect in at most one point), we are left with degenerate rectangles formed by two ones along the same row (or the same column). There are $\Theta(n(n^{1/3})^2) = \Theta(n^{5/3})$ of those. It follows that the trace of M^2 is $O(n^{5/3})$. By the trace bound, the hereditary discrepancy of the set system is at least $\frac{1}{4} c^{n\,\text{tr}\,M^2/\text{tr}^2\,M} \sqrt{\frac{\text{tr}\,M}{n}}$, which is $\Omega(n^{1/6})$. Notice how the exponent miraculously reduces to a constant!

Theorem 1.20 *There exist n points and n lines in $\mathbf{R}^2$, such that the n-by-n incidence matrix $A = (a_{ij})$ has discrepancy $D_\infty(A) = \Omega(n^{1/6})$.*

One can use the method of partial colorings (page 15) to show that the lower bound is optimal up to within a logarithmic factor. It is interesting to contrast the exponent of $1/6$ for points and lines vs. $1/4$ for points and halfplanes.

Application II: Boxes in Higher Dimension

In fixed dimension, it is a rule of thumb that discrepancies for boxes are $(\log n)^{\Theta(1)}$ if the orientation is fixed and $n^{\Theta(1)}$ if rotations are allowed. If we let the dimension increase, however, we expect this gap to be eventually bridged since in dimension high enough any $0/1$ matrix is an incidence

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

matrix for points and boxes. An interesting question thus is: At which dimension do we switch from logarithmic to polynomial? The trace bound indicates that the transition to polynomial discrepancy occurs at dimension as low as $O(\log n)$.

Theorem 1.21 *There exist n points and n axis-parallel boxes in $\mathbf{R}^d$, for any $d = O(\log n)$, such that the n-by-n incidence matrix $A = (a_{ij})$ has discrepancy $D_\infty(A) = 2^{\Omega(d)}$.*

The lower bound is actually more general than stated, as it holds for points and boxes in the Hamming cube $\{0,1\}^d$. Theorem 1.21 follows easily from the lemma below.

Lemma 1.22 *For any n large enough, there exists a set system of n points and n boxes in $\{0,1\}^d$, where $d = \Theta(\log n)$, such that the n-by-n incidence matrix $A = (a_{ij})$ has discrepancy $D_\infty(A) = \Omega(n^{0.0477})$.*

The theorem is essentially a restatement of Lemma 1.22 if $d \geq b \log n$, for some constant $b > 0$. So, assume that $d < b \log n$. Set n_0 to be about $2^{d/b}$ so that we can apply the lemma with respect to n_0 and d. Now, pad the set system to be n-by-n by adding $n - n_0$ points and boxes with no new incidences. The lower bound of $\Omega(n_0^{0.0477})$ is also $\Omega(2^{\Omega(d)})$, which proves Theorem 1.21. □

In view of the trace bound (page 32), Lemma 1.22 follows directly from the existence of m points and n boxes in $\mathbf{R}^{O(\log n)}$ such that: for some constant $c \approx 1.0955$,

(i) $m = \Theta(n)$ and $\operatorname{tr} M = \Theta(n^c)$ with probability at least $1/2$;
(ii) $\mathbf{E} \operatorname{tr} M^2 = O(n^{2c-1})$.

For convenience, we introduce a few parameters:

$$\begin{cases} w &= \frac{1-2p+p^9}{1-2p-(1+2p)p^2 \log e}, \text{ where } p = 0.153\,, \\ c &= 2-(1-p)w\,, \\ g &= n^{c-1}. \end{cases}$$

The dimension d is defined as $w \log n$. The m points are chosen by picking each element of the Hamming cube $\{0,1\}^d$ independently with probability n^{1-w}. (Note that $w \approx 1.067867$, so $n^{1-w} < 1$.) The expected number of points is n and, by Chebyshev's inequality,

Lemma 1.23 *With probability $> 1/2$, the number m of points is $\Theta(n)$.*

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

A box is specified by a word of length d, over the alphabet $\{0, 1, *\}$, containing exactly pd stars. For example, in dimension 5, the word $0*1**$ denotes the three-dimensional box $x_1 = 0$, $x_3 = 1$. We construct the n boxes by specifying g groups of parallel boxes.[13] Each group is defined by selecting the location of the stars first and then taking all the corresponding boxes. To select the stars, we pick pd coordinates uniformly at random (without replacement) and make them stars. In our previous example, the group of parallel boxes consists of $0*0**$, $0*1**$, $1*0**$, and $1*1**$. We have precisely $2^{(1-p)d}g = n$ boxes. Each point in the set system belongs to exactly one box in each of the g groups, so that $\operatorname{tr} M = mg$. By Lemma 1.23, we have the following:

Lemma 1.24 *With probability* $> 1/2$, *the trace of* M *is* $\Theta(n^c)$ *and (i) holds.*

To find an upper bound on the trace of M^2, we express it as a sum of four terms:

$$\operatorname{tr} M^2 = O(\sigma_{1,1} + \sigma_{1,2} + \sigma_{2,1} + \sigma_{2,2}),$$

where $\sigma_{i,j}$ is the number of pairs (I, J) such that $I \supseteq J$, where I is the intersection of i distinct boxes and J is a set of j distinct points. To bound these numbers is easy. Any one of the 2^{pd} Hamming cube vertices lying in a given box belongs to the set system with probability n^{1-w}. There are n boxes, so

$$\mathbf{E}\,\sigma_{1,2} = O\Big(n(2^{pd}n^{1-w})^2\Big) = O(n^{3-2(1-p)w}).$$

Regarding $\sigma_{2,1}$, note that boxes within the same group are disjoint, so only pairs in distinct groups can contribute to $\sigma_{2,1}$. Fix two such groups. Any one of the 2^d points of the Hamming cube belongs to exactly one pair of boxes. Since such a point is picked with probability n^{1-w}, we have $\mathbf{E}\,\sigma_{2,1} = O(g^2 2^d n^{1-w}) = O(n^{2c-1})$. To summarize,

$$\mathbf{E}\,\sigma_{1,1} = \mathbf{E}\operatorname{tr} M = n^c, \quad \mathbf{E}\,\sigma_{1,2} = O(n^{2c-1}), \quad \mathbf{E}\,\sigma_{2,1} = O(n^{2c-1}). \tag{1.14}$$

Finally, we turn to the expectation of $\sigma_{2,2}$: Again, fix two groups of parallel boxes, and let x be the number of stars common to both star patterns. As we just saw, any point of the Hamming cube belongs to exactly one pair of boxes, and this point can be paired with exactly $2^x - 1$

[13] By rounding off, if necessary, we can assume that g, d, and pd are all integral.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

other points. Each point being picked with probability n^{1-w}, it follows that

$$\sigma_{2,2} = O(g^2 2^{d+x} n^{2-2w})$$

and, hence,

$$\mathbf{E}\,\sigma_{2,2} = O(n^{2c-w})\,\mathbf{E}\,2^x.$$

What about the expectation of 2^x? Writing

$$N_k \stackrel{\text{def}}{=} N(N-1)\cdots(N-k+1),$$

we use the lower bound $k! > (k/e)^k$ to derive

$$\begin{aligned}\mathbf{E}\,2^x &= \sum_{k=0}^{pd} 2^k \binom{pd}{k}\binom{d-pd}{pd-k}\Big/\binom{d}{pd} = \sum_{k=0}^{pd} \frac{2^k (pd)_k (d-pd)_{pd-k}}{k!(pd-k)!}\Big/\binom{d}{pd} \\ &\le \sum_{k=0}^{pd} \frac{(2ep^2d^2)^k (d-pd)_{pd}}{(kd)^k (1-2p)^k (pd)!}\Big/\binom{d}{pd} \le \sum_{k=0}^{pd} (1-p)^{pd}\left(\frac{2ep^2 d}{(1-2p)k}\right)^k.\end{aligned}$$

The function $(A/x)^x$ reaches its maximum value at $x = A/e$, and so

$$\mathbf{E}\,2^x = O(n^{(\log e)p^2 w(1+2p)/(1-2p)} \log n),$$

which implies that

$$\mathbf{E}\,\sigma_{2,2} = O(n^{2c-w+\frac{1+2p}{1-2p}p^2 w \log e} \log n).$$

In view of (1.14),

$$\begin{aligned}\mathbf{E}\,\mathrm{tr}\,M^2 &= O\Big(n^c + n^{2c-1} + n^{2c-w+\frac{1+2p}{1-2p}p^2 w \log e} \log n\Big) \\ &= O\Big(n^{2c-1} + n^{2c-1-\frac{p^9}{1-2p}} \log n\Big) = O(n^{2c-1}),\end{aligned}$$

which satisfies condition (ii). Lemma 1.22 and Theorem 1.21 follow. □

1.6 Bibliographical Notes

Section 1.1: The method of conditional expectations was developed by Raghavan [254] and Spencer [294]. A similar idea is implicit in an earlier work of Erdős and Selfridge [124]. The hyperbolic cosine algorithm is due to Spencer [292, 294]. The unbiased greedy algorithm was first proposed by Beck and Fiala [38] and Beck [32]; it was rediscovered by the author [68]. Not surprisingly, a similar technique can be used to rederive Chernoff-type bounds and prove tail estimates for martingales.

https://doi.org/10.1017/CBO9780511626371.002 Published online by Cambridge University Press

Section 1.2: The fact that n-by-n matrices have discrepancy $O(\sqrt{n})$ (Theorem 1.3, page 8) was established by Spencer [293]. Using the pigeonhole principle on the discrepancy vector is an idea going back to Beck [31]. The use of entropy in the proof follows a suggestion of Boppana (see Alon and Spencer [20]). Spencer's original proof shows that the constant hiding behind the bound $O(\sqrt{n})$ is less than 6.

Section 1.3: Theorem 1.5 (page 10) is due to Beck and Fiala [38]. The bound was (ever so slightly) improved to $2t - 3$ by Bednarchak and Helm [40]. It is conjectured that $O(\sqrt{t})$ is the correct bound. A bound of $O(\sqrt{t}\log n)$ was established by Srinivasan [295], where n is the number of elements in the set system; an earlier bound of $O(\sqrt{t\log t}\log n)$ was obtained by Beck and Spencer; see also [294].

Section 1.4: The notion of VC-dimension was introduced by Vapnik and Chervonenkis [317]. We chose to open the section with it because of its sheer elegance and its historical importance. In most applications, however, bounds on the primal and shatter functions seem more important than the VC-dimension (which, typically, is difficult to compute). The bound on the primal shatter function (Lemma 1.6) was established independently by Sauer [269], Shelah [284], and Vapnik and Chervonenkis [317]—see also [51]. The fact that infinite range spaces cannot have sublinear primal shatter functions appears in Assouad [25].

Dudley [112] observed that any finite number of set-theoretical operations on range spaces keep the VC-dimension bounded. This is useful to prove that certain geometric range spaces are of bounded VC-dimension. The bound on the dual VC-dimension in Lemma 1.7 comes from Assouad [25]. The discrepancy estimates in Theorem 1.8 (page 14) and Theorem 1.10 (page 17) are due to Matoušek, Welzl, and Wernisch [222]. Lemma 1.9 is adapted from Beck [31]. The bound in Theorem 1.8 has been improved to $O(n^{1/2-1/2d})$ by Matoušek [215]. The optimality of the "dual" bound in Theorem 1.10, for $d = 2, 3$, was established by Matoušek [218] and extended to any dimension by Alon, Rónyai, and Szabó [19].

Section 1.5: The lower bound on the discrepancy of the Hadamard matrix comes from Spencer [293]. The case of arithmetic progressions (Theorem 1.11, page 20) was solved by Roth [262], who used the Fourier transform method. The proof based on eigenvalues is due to Lovász and Sós and appears in Beck and Sós' survey [39]. A matching upper bound of $O(n^{1/4})$ was proven by Matoušek and Spencer [221]. An earlier, breakthrough result (weaker by only a polylogarithmic factor) was obtained by Beck [31],

who also introduced the partial coloring technique. For an excellent introduction to Ramsey theory, see [149].

The notion of hereditary discrepancy and the determinant bound (Theorem 1.12, page 26) were introduced by Lovász, Spencer, and Vesztergombi [199]. The lower bound for halfplane discrepancy (Theorem 1.15, page 29) was established by Chazelle [75]. A proof of optimality was provided by Matoušek [215]. The lower bound (1.9) was proven by Alexander [10].

The trace bound (Theorem 1.17, page 32) is due to Chazelle and Lvov [79], as are the applications to set systems of lines and boxes in higher dimension [80] (Theorems 1.20 and 1.21).