The celebrated theorem due to P. Turán (1941) states: if a graph G has n vertices and has no k-clique then it has at most $(1 - 1/(k-1))\, n^2/2$ edges (see Theorem 4.8). Its dual form states (see Exercise 4.8):

If G has n vertices and $nk/2$ edges, then $\alpha(G) \geq n/(k+1)$.

This dual form of Turán's theorem also follows from Theorem 18.4: fixing the total number of edges, the sum $\sum_{i=1}^{n} 1/(d_i + 1)$ is minimized when the d_i's are as nearly equal as possible, and, by Theorem 1.8, $\frac{1}{2}\sum_{i=1}^{n} d_i$ is exactly the number of edges in G.

18.5 Crossings and incidences

Given a set P of n points and a set L of m lines in the plane, the point-line incidence graph is a bipartite $n \times m$ graph with parts P and L, where $p \in P$ and $l \in L$ are adjacent iff the point p lies on the line l (see Fig. 18.1). How many edges can such a graph have?

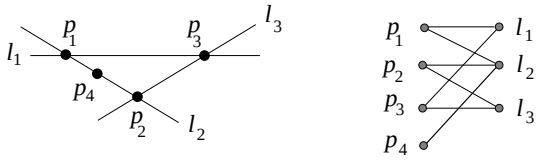

Fig. 18.1 We have four points and three lines. The number of incidences (edges in the point-line incidence graph on the right) is 7.

Since any two points can lie on at most one common line, and two lines intersect in at most one point, each point-line incidence graph is C_4-free, that is, contains no cycles on four vertices. We already know (see Exercise 2.6) that the number of edges in such graphs cannot exceed either $nm^{1/2} + m$ or $mn^{1/2} + n$. For $n = m$ this is about $n^{3/2}$. Szemerédi and Trotter (1983) obtained a much better upper bound which, for $n = m$, is about $n^{4/3} \ll n^{3/2}$. We will derive this theorem from another (seemingly unrelated) result about the number of crossings when a graph is drawn on the plane.

18.5.1 Crossing number

Given a graph G, the crossing number of the graph, denoted $\mathrm{cr}(G)$, is the minimum number of edge-crossings possible amongst all drawings of the graph with edges as straight line segments and vertices as points in the plane. Thus a graph G is planar if and only if $\mathrm{cr}(G) = 0$. A natural question is: given a graph with e edges and n vertices, how large is its crossing number?

The well-known Euler's polyhedron formula states that if a finite, connected, planar graph is drawn in the plane without any edge intersections, and n is the number of vertices, e is the number of edges and f is the number of faces (regions bounded by edges, including the outer, infinitely-large region), then $n - e + f = 2$. If $e \geq 3$ then every face is adjacent to at least three edges, whereas every edge is adjacent to exactly two faces. By double counting the edge-face incidences, we get $3f \leq 2e$. Eliminating f, we conclude that $e \leq 3n - 6$ for all planar graphs.

If a graph G can be drawn with only $\mathrm{cr}(G)$ crossings, then we can delete one of the crossings by removing an edge associated with that crossing, and so we can remove all the crossings by deleting at most $\mathrm{cr}(G)$ edges, leaving at least $e - \mathrm{cr}(G)$ edges (and v vertices). Since the graph obtained is planar, we obtain the following lower bound on the crossing number of any graph G:

$$\mathrm{cr}(G) \geq e - 3n + 6 > e - 3n\,. \tag{18.3}$$

By applying this inequality to *random* induced subgraphs of G, Ajtai, Chvátal, Newborn, and Szemerédi (1982), and Leighton (1984) were able to improve this lower bound.

Theorem 18.5 (The crossing number inequality)**.** *Let G be a graph with n vertices and $e \geq 4n$. Then*

$$\mathrm{cr}(G) \geq \frac{e^3}{64n^2}\,.$$

Proof. Let G be embedded in the plane and suppose the crossing number of the drawing is x. Independently select vertices of G with probability p, and let H be the (induced) subgraph of edges between selected vertices. By the linearity of expectation, H is expected to have pn vertices and p^2e edges. (The events that each edge ends up in H are not quite independent, but the great thing about linearity of expectation is that it works even without assuming any independence.) Observe that each crossing involves two edges and four vertices. Thus, the probability that the crossing survives in this drawing is only p^4. By one last application of linearity of expectation, the expected number of crossings of this drawing that survive for H is p^4x. This particular drawing may not be the optimal one for H, so we end up with an inequality $\mathrm{E}\,[\mathrm{cr}(H)] \leq p^4x$. By (18.3), the number of crossings in any graph H is always at least the number of edges minus three times the number of vertices of H. Consequently

$$p^4x \geq \mathrm{E}\,[\mathrm{cr}(H)] \geq p^2e - 3pn\,.$$

Taking $p := 4n/e$ gives the desired lower bound on $x = \mathrm{cr}(G)$. □

18.5.2 The Szemerédi–Trotter theorem

From the above result on crossing numbers one deduces a short proof of the Szemerédi–Trotter theorem in combinatorial geometry. It gives an almost tight upper bound on the number of incidences, that is, on the number of point-line pairs such that the point lies on the line.

Theorem 18.6 (Szemerédi–Trotter 1983)**.** *Let P be a set of n distinct points in the plane, and let L be a set of m distinct lines. Then the number of incidences between P and the lines in L is at most $4(mn)^{2/3} + m + 4n$.*

The original proof of this theorem was somewhat complicated, using a combinatorial technique known as cell decomposition. Later, Székely (1997) discovered a much simpler proof using crossing numbers of graphs.

Proof (due to Székely 1997). Let $x = |\{(p, l) \in P \times L : \ p \in l\}|$ be the number of incidences. Let G be the graph whose vertex set is P and whose vertices are adjacent if they are consecutive on some line in L. A line $l \in L$ which is incident to k_l points in P will thus contain $k_l - 1$ line segments between points in P. Since the sum of all the k_l over all lines $l \in L$ is exactly the total number x of incidences, the graph G has $x - m$ edges. Clearly $\mathrm{cr}(G) < m^2$ since two lines cross at no more than one point. By the result on crossing numbers, we deduce

$$m^2 > \frac{(x-m)^3}{64n^2} - n$$

(we put "$-n$" just to eliminate the condition $e \geq 4n$) and therefore $x \leq 4(mn)^{2/3} + m + 4n$. □

To see that the theorem is tight up to a constant factor, take the grid $P = [k] \times [4k^2]$ together with the set L of all straight lines $y = ax + b$ with slope $a \in [k]$ and intercept $b \in [2k^2]$. Then for $x \in [k]$ one has $ax + b \leq ak + b \leq k^2 + 2k^2 < 4k^2$. So, for each $x = 1, \ldots, k$ each line contains a point (x, y) of P. We get a total of roughly $2k^4$ incidences, as compared to the upper bound of roughly $4k^4$.

In applications the following corollary of this theorem is often used (we will also use it in Sect. 25.4). We will say that a function f "is at most about" another function g if $f = O(g)$.

Theorem 18.7. *For n points in the plane, the number of lines, each containing at least k of them, is at most about $n^2/k^3 + n/k$.*

Proof. Let P be a set of n points, and L a set of m lines, each of which contains at least k points of P. Then these lines generate at least mk incidences and so, by Theorem 18.6, we have that $m(k-1) \leq 4(mn)^{2/3} + 4n$. If $n \leq (nm)^{2/3}$ then the right-hand side is at most $8(mn)^{2/3}$, from which $m = O(n^2/k^3)$ follows. If $n \geq (nm)^{2/3}$ then the right hand side is at most $8n$, from which $m = O(n/k)$ follows. □

The importance of Theorem 18.7 lies in the fact that the exponent of k in the denominator is *strictly* larger than 2. A bound of $m \leq \binom{n}{2}/\binom{k}{2}$, which is about n^2/k^2, is trivial by just double-counting the pairs of points. (Prove this!)

The so-called *Two Extremities Theorem* says that finite collections of points in the plane fall into one of two extremes: one where a large fraction of points lie on a single line, and one where a large number of lines are needed to connect all the points.

Theorem 18.8 (Beck 1983)**.** *Given any n points in the plane, at least one of the following statements is true:*

1. *There is a line which contains at least $\Omega(n)$ of the points.*
2. *There exist at least $\Omega(n^2)$ lines, each of which contains at least two of the points.*

Proof. Consider a set P of n points in the plane. Let t be a positive integer. Let us say that a pair of points x, y in the set P is t-connected if the (unique) line connecting x and y contains between 2^t and $2^{t+1}-1$ points of P (including x and y). By Theorem 18.7, the number of such lines is at most about $n^2/2^{3t} + n/2^t$. Since each such line connects together at most about 2^{2t} pairs of points of P, we thus see that at most about $n^2/2^t + n2^t$ pairs of points can be t-connected.

Now, let C be a large constant. By summing the geometric series, we see that the number of pairs of points which are t-connected for some t satisfying $C \leq 2^t \leq n/C$ is at most about n^2/C. On the other hand, the total number of pairs is $\binom{n}{2}$.

Thus if we choose constant C to be large enough, we can find at least, say, $n^2/4$ pairs of points which are not t-connected for any $C \leq 2^t \leq n/C$. The lines that connect these pairs either pass through fewer than C points, or pass through more than n/C points. If the latter case holds for even one of these pairs, then we have the first conclusion of Beck's theorem. Thus we may assume that all of the $n^2/4$ pairs are connected by lines which pass through fewer than C points. But each such line can connect at most C^2 pairs of points. Thus there must be at least $n^2/4C^2$ lines connecting at least two points of P. □

More about combinatorial problems in geometry as well as their cute solutions can be found in a beautiful book by Matoušek (2002).

18.6 Far away strings

The Hamming distance between two binary strings is the number $\mathrm{dist}(x, y)$ of positions in which these strings differ. How many binary strings can we find such that each two of them lie at Hamming distance at least $n/2$? In

Sect. 14.3 we used Hadamard matrices to construct such a set consisting of $2n$ strings (see Theorem 14.10). But what if we relax the condition and only require the pairwise distance be at least, say, $n/4$? It turns out that then much larger sets exist.

To show this, we will use the following Chernoff's inequality: If X is the sum of n independent and uniformly distributed 0-1 variables, then $\Pr[X \leq n/2 - a] \leq \mathrm{e}^{-2a^2/n}$.

Theorem 18.9. *There exists a set of* $\mathrm{e}^{n/16}$ *binary strings of length* n *such that any pair is at Hamming distance at least* $n/4$ *from each other.*

Proof. Consider a random string in $\{0,1\}^n$ generated by picking each bit randomly and independently. For any two such strings x and y, let X_i be the indicator random variable for the event that $x_i \neq y_i$. Then $\mathrm{E}[X_i] = 1/2$, and $\mathrm{dist}(x,y) = X_1 + \cdots + X_n$. By the linearity of expectation, $\mathrm{E}[\mathrm{dist}(x,y)] = n/2$. Using Chernoff's inequality, we have that

$$\Pr[\mathrm{dist}(x,y) \leq n/2 - a] \leq \mathrm{e}^{-2a^2/n}.$$

Now generate $M := \mathrm{e}^{n/16}$ strings at random and independently. Set $a := n/4$. By the union bound, the probability that any pair of these strings lies at distance at most $n/4$, is at most $\binom{M}{2}\mathrm{e}^{-2a^2/n} < M^2\mathrm{e}^{-n/8} = 1$, implying that the desired set of strings exists. □

This result has an interesting interpretation in the Euclidean setting. Recall that a *unit vector* is a vector $x \in \mathbb{R}^n$ such that $\|x\| = 1$, where $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$ is the norm of x. The set of all unit vectors forms the *unit sphere*. The Euclidean distance between two vectors $x, y \in \mathbb{R}^n$ is the norm $\|x - y\|$ of their difference.

Corollary 18.10. *The unit sphere in* $\mathbb{R}^n$ *contains a set of* $\mathrm{e}^{n/16}$ *points, each two of which are at Euclidean distance at least one from each other.*

Proof. Let $P \subseteq \{0,1\}^n$ be the set of binary strings guaranteed by Theorem 18.9. Associate with each binary string $u = (u_1, \ldots, u_n)$ a unit vector $x_u \in \mathbb{R}^n$ whose i-th coordinate is defined by $x_u(i) := \frac{1}{\sqrt{n}}(-1)^{u_i}$. Then, for any two vectors $u, v \in P$ and for any coordinate i, we have that

$$\Big(x_u(i) - x_v(i)\Big)^2 = \frac{1}{n}\Big((-1)^{u_i} - (-1)^{v_i}\Big)^2 = \begin{cases} 0 & \text{if } u_i = v_i, \\ \frac{4}{n} & \text{if } u_i \neq v_i. \end{cases}$$

Hence,

$$\|x_u - x_v\|^2 = \sum_{i=1}^{n} \Big(x_u(i) - x_v(i)\Big)^2 = \frac{4}{n} \cdot \mathrm{dist}(x,y) \geq 1,$$

as desired. □

18.7 Low degree polynomials

In this section we consider polynomials $f(x_1, \ldots, x_n)$ on n variables over the field $\mathbb{F}_2$. Such a polynomial has degree at most d if it can be written in the form

$$f(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^{m} \prod_{j \in S_i} x_j \,,$$

where $a_0 \in \{0, 1\}$ and $S_1, \ldots, S_m$ are subsets of $\{1, \ldots, n\}$ of size at most d; here and throughout the section the sum is modulo 2.

If $f_1, \ldots, f_m$ are polynomials of degree at most d, then their product can have degree up to dm. The following result says that the product can still be approximated quite well by a polynomial of relatively small degree.

Lemma 18.11 (Razborov 1987)**.** *Let $f = \prod_{i=1}^{m} f_i$, where $f_1, \ldots, f_m$ are polynomials of degree at most d over $\mathbb{F}_2$. Then, for any $r \geq 1$, there exists a polynomial g of degree at most dr such that g differs from f on at most 2^{n-r} inputs.*

Proof. Let $\boldsymbol{S}$ be a random subset of $\{1, \ldots, m\}$, that is, we choose $\boldsymbol{S}$ randomly from the family of all 2^m subsets with probability 2^{-m}. Let $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_r$ be independent copies of $\boldsymbol{S}$. Consider a (random) function of the form

$$\boldsymbol{g} = \prod_{j=1}^{r} \boldsymbol{h}_j \,, \text{ where } \boldsymbol{h}_j = 1 - \sum_{i \in \boldsymbol{S}_j} (1 - f_i) \,. \tag{18.4}$$

We claim that, for every (fixed) input $a \in \{0, 1\}^n$,

$$\Pr[\boldsymbol{g}(a) \neq f(a)] \leq 2^{-r} \,. \tag{18.5}$$

Indeed, if $f(a) = 1$ then all $f_i(a) = 1$, and hence, $\boldsymbol{g}(a) = 1$ with probability 1. Suppose now that $f(a) = 0$. Then $f_{i_0}(a) = 0$ for at least one i_0. Since each of the sets $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_r$ contains i_0 with probability $1/2$, we have that $\Pr[\boldsymbol{h}_j(a) = 1] \leq 1/2$ for all $j = 1, \ldots, r$ (consult Exercise 18.11 for this conclusion). Hence,

$$\Pr[\boldsymbol{g}(a) = 0] = 1 - \Pr[\boldsymbol{h}_1(a) = \ldots = \boldsymbol{h}_r(a) = 1] \geq 1 - 2^{-r} \,,$$

as claimed.

For an input vector $a \in \{0, 1\}^n$, let X_a denote the indicator random variable for the event that $\boldsymbol{g}(a) \neq f(a)$, and let X be the sum of X_a over all a. By (18.5) and the linearity of expectation, the expected number of inputs on which $\boldsymbol{g}$ differs from f is

$$\mathrm{E}[X] = \sum_a \mathrm{E}[X_a] = \sum_a \Pr[X_a = 1] \leq 2^{n-r} \,.$$

By the pigeonhole principle of expectation, there must be a point in the probability space for which this holds. This point is a polynomial of the form (18.4); it has degree at most dr and differs from f on at most 2^{n-r} inputs. □

Razborov used this lemma to prove that the majority function cannot be computed by constant-depth polynomial-size circuits with unbounded fanin And, Or and Parity gates. The majority function is a boolean function $\mathrm{Maj}_n(x_1, \ldots, x_n)$ which outputs 1 if and only if $x_1 + \cdots + x_n \geq n/2$.

Theorem 18.12 (Razborov 1987)**.** *Every unbounded fanin depth-c circuit with And, Or and Parity gates computing* Maj_n *requires* $2^{\Omega(n^{1/2c})}$ *gates.*

The idea is as follows. If f can be computed by a depth-c circuit of size ℓ then, by Lemma 18.11, there exists a polynomial g of degree at most r^c such that g differs from f on at most $\ell \cdot 2^{n-r}$ inputs. The desired lower bound is then obtained by showing that the majority function cannot be approximated sufficiently well by such polynomials (see Lemma 13.8). Taking r to be about $n^{1/(2c)}$ and making necessary computations this leads to a lower bound $\ell \geq 2^{\Omega(n^{1/(2c)})}$. This final step requires some routine calculations, and we omit it.

18.8 Maximum satisfiability

In most of the above applications it was enough to take a uniform distribution, that is, every object had the same probability of appearing. In this section we will consider the situation where the distribution essentially depends on the specific properties of a given family of objects.

An *And-Or formula* or a *CNF* (or simply, a *formula*) over a set of variables $x_1, \ldots, x_n$ is an And of an arbitrary number of *clauses*, where a clause is an Or of an arbitrary number of *literals*, each literal being either a variable x_i or a negated variable $\overline{x}_i$. For example:

$$F = (x_1 \vee \overline{x}_3)(\overline{x}_1 \vee x_2 \vee \overline{x}_3)(\overline{x}_2)(\overline{x}_1 \vee \overline{x}_2)\,.$$

An *assignment* is a mapping which assigns each variable one of the values 0 or 1. We can look at such assignments as binary vectors $v = (v_1, \ldots, v_n) \in \{0,1\}^n$, where v_i is the value assigned to x_i. If y is a literal, then we say that v *satisfies* y if either $y = x_i$ and $v_i = 1$, or $y = \overline{x}_i$ and $v_i = 0$. An assignment satisfies a clause if it satisfies at least one of its literals. An assignment satisfies a formula if it satisfies each of its clauses. For the formula above, the assignment $v = (1, 0, 0)$ is satisfying. A formula is *satisfiable* if at least one assignment satisfies it. A formula F is *k-satisfiable* if any subset of k clauses of F is satisfiable.

It is an interesting "Helly-type" phenomenon, first established by Lieberher and Specker (1981), which says that if a formula is 3-satisfiable then at least

2/3 of its clauses are simultaneously satisfiable. For 2-satisfiable formulas this fraction is $2/(1+\sqrt{5}) > 0.618$ (the inverse of the golden ratio). The original proof of these facts was rather involved. Yannakakis (1994) has found a very simple proof of these bounds using the probabilistic method.

Theorem 18.13 (Yannakakis 1994)**.** *If F is a 3-satisfiable formula then at least a 2/3 fraction of its clauses are simultaneously satisfiable.*

Proof. Given a 3-satisfiable formula F, define a random assignment $\boldsymbol{v} = (v_1, \ldots, v_n)$, where each bit v_i takes its value independently from other bits and with probability

$$\Pr[v_i = 1] = \begin{cases} 2/3 & \text{if } F \text{ contains a unary clause } (x_i); \\ 1/3 & \text{if } F \text{ contains a unary clause } (\overline{x}_i); \\ 1/2 & \text{otherwise.} \end{cases}$$

Note that this definition is consistent since it is impossible to have the unary clauses (x_i) and $(\overline{x}_i)$ in the same 3-satisfiable formula. Simple (but crucial) observation is that each singular literal $y \in \{x_i, \overline{x}_i\}$, which appears in the formula F, is falsified with probability $\leq 2/3$ (independent of whether this literal forms a unary clause or not). To see this, let $y = x_i$ and $p = \Pr[v_i = 0]$. We have three possibilities:

- either (x_i) is a unary clause of F, and in this case $p = 1 - 2/3 = 1/3$;
- or F contains a unary clause $(\overline{x}_i)$, and in this case $p = 1 - 1/3 = 2/3$;
- or neither x_i nor $\overline{x}_i$ appears in a unary clause, in which case $p = 1/2$.

Using this observation, we can prove the following fact.

Claim 18.14. Every clause is satisfied by $\boldsymbol{v}$ with probability at least 2/3.

For unary clauses the claim is trivial. On the other hand, if C contains three or more literals, then, by the above observation, each of these literals can be falsified with probability at most 2/3, and hence, the clause is satisfied with probability at least $1 - (2/3)^3 = 0.7037... > 2/3$; for longer clauses the probabilities are even better.

It remains to consider binary clauses. Assume w.l.o.g. that $C = (x_1 \vee x_2)$. If at least one of x_1 and x_2 is satisfied with probability 1/2 then the clause C is satisfied with probability $1 - \Pr[v_1 = 0] \cdot \Pr[v_2 = 0] \geq 1 - \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{3}$. Thus, the only bad case would be when both literals x_1 and x_2 are satisfied only with probability 1/3. But this is impossible because it would mean that the formula F contains the clauses $(x_1 \vee x_2), (\overline{x}_1), (\overline{x}_2)$, which contradicts the fact that F is 3-satisfiable.

We now conclude the proof of the theorem in a standard manner. Suppose that F consists of the clauses $C_1, \ldots, C_m$. Let X_i denote the indicator random variable for the event "the i-th clause C_i is satisfied by $\boldsymbol{v}$". Then $X = \sum_{i=1}^{m} X_i$ is the total number of satisfied clauses of F. By Claim 18.14, $\Pr[X_i = 1] \geq 2/3$ for each i, and by the linearity of expectation, $\mathrm{E}[X] = \sum_{i=1}^{m} \mathrm{E}[X_i] \geq \frac{2m}{3}$.

By the pigeonhole property of the expectation, at least one assignment $\boldsymbol{v}$ must satisfy so many clauses of F, as desired. □

It is worth mentioning that, for large values of k, the right fraction for all k-satisfiable formulas is $3/4$. Namely, Trevisan (2004) has proved that, if r_k stands for the largest real such that in any k-satisfiable formula at least an r_k-th fraction of its clauses are satisfied simultaneously, then $\lim_{k\to\infty} r_k = 3/4$.

18.9 Hash functions

A set V of vectors of length t over an alphabet $A = \{1, \ldots, n\}$ is called *k-separated* if for every k distinct vectors there is a coordinate in which they are all distinct. How many vectors can such a set have?

This question is equivalent to the question about the maximum size $N = N(n, k, t)$ of a domain for which there exists a family of (n, k) *hash functions* with t members, that is, a family of t partial functions $f_1, \ldots, f_t$ mapping a domain of size N into a set of size n so that every subset of k elements of the domain is mapped in a one-to-one fashion by at least one of the functions. To see this equivalence, it is enough to consider the set of vectors $(f_1(x), \ldots, f_t(x))$ for each point x of the domain.

The problem of estimating $N(n, k, t)$, which is motivated by the numerous applications of perfect hashing in theoretical computer science, has received a considerable amount of attention. The interesting case is when the number t of hash functions is much bigger than the size n of the target set (and, of course, $n \geq k$). The following are the best known estimates for $N(n, k, t)$:

$$\frac{1}{k-1} \log \frac{1}{1 - g(n,k)} \lesssim \frac{1}{t} \log N(n, k, t) \tag{18.6}$$

and

$$\frac{1}{t} \log N(n, k, t) \lesssim \min_{1 \leq r \leq k-1} g(n, r) \log \frac{n - r + 1}{k - r}, \tag{18.7}$$

where

$$g(n, k) := \frac{(n)_k}{n^k} = \frac{n(n-1)\cdots(n-k+1)}{n^k}.$$

In particular, (18.7) implies that

$$N(n, k, t) \leq \left(\frac{n}{k}\right)^t.$$

The lower bound (18.6), proved by Fredman and Komlós (1984), can be derived using a probabilistic argument (the *deletion method*) discussed in Chap. 20: one chooses an appropriate number of vectors randomly, shows

that the expected number of non-separated k-tuples is small, and omits a vector from each such "bad" k-tuple. The proof of the upper bound (18.7) was much more difficult. For $r = k-1$, a slightly weaker version of this bound was proved in Fredman and Komlós (1984), and then extended to (18.7) by Körner and Marton (1988). All these proofs rely on certain techniques from information theory.

A short and simple probabilistic proof of (18.7), which requires no information-theoretic tools, was found by Nilli (1994) (c/o Noga Alon). We only present the key lemma of this proof.

Lemma 18.15. *Let U be a set of m vectors of length t over the alphabet $B \cup \{*\}$, where $B = \{1, \dots, b\}$, and let x_v denote the number of non-$*$ coordinates of $v \in U$. Let $\overline{x} = \sum x_v/m$ be the average value of x_v. If for every d distinct vectors in U there is a coordinate in which they all are different from $*$ and are all distinct, then*

$$m \le (d-1)\left(\frac{b}{d-1}\right)^{\overline{x}}.$$

Proof. For every coordinate i, choose randomly and independently a subset $\boldsymbol{D}_i$ of cardinality $d-1$ of B. Call a vector $v \in U$ *consistent* if for every i, $v_i \in \boldsymbol{D}_i \cup \{*\}$. Since each set $\boldsymbol{D}_i$ has size $d-1$, the assumption clearly implies that for any choice of the sets D_i there are no more than $d-1$ consistent vectors. On the other hand, for a fixed vector v and its coordinate i, $\Pr[v_i \in \boldsymbol{D}_i] = (d-1)/b$. So, each vector v is consistent with probability $\big((d-1)/b\big)^{x_v}$ and, by the linearity of expectation, the expected number of consistent vectors in U is

$$\sum_{v \in U} \left(\frac{d-1}{b}\right)^{x_v} \ge m \left(\frac{d-1}{b}\right)^{\overline{x}},$$

where the inequality follows from Jensen's inequality (see Proposition 1.12), since the function $g(z) = \big((d-1)/b\big)^z$ is convex. □

18.10 Discrepancy

Let $X_1, \dots, X_k$ be n-element sets, and $X = X_1 \times \cdots \times X_k$. A subset T_i of X is called a *cylinder* in the i-th dimension if membership in T_i does not depend on the i-th coordinate. That is, $(x_1, \dots, x_i, \dots, x_k) \in T_i$ implies that $(x_1, \dots, x'_i, \dots, x_k) \in T_i$ for all $x'_i \in X_i$. A subset $T \subseteq X$ is a *cylinder intersection* if it is an intersection $T = T_1 \cap T_2 \cap \cdots \cap T_k$, where T_i is a cylinder in the i-th dimension. The *discrepancy* of a function $f : X \to \{-1, 1\}$ on a set T is the absolute value of the sum of the values of f on points in T, divided by the total number $|X|$ of points:

$$\mathrm{disc}_T(f) = \frac{1}{|X|}\left|\sum_{x\in T} f(x)\right| .$$

The *discrepancy* of f is the maximum $\mathrm{disc}(f) = \max_T \mathrm{disc}_T(f)$ over all cylinder intersections $T \subseteq X$.

The importance of this measure stems from the fact that functions with small discrepancy have large *multi-party communication complexity*. (We will discuss this in Sect. 27.4 devoted to multi-party games.) However, this fact alone does not give immediate lower bounds for the multi-party communication complexity, because $\mathrm{disc}(f)$ is very hard to estimate. Fortunately, the discrepancy can be bounded from above using the following more tractable measure.

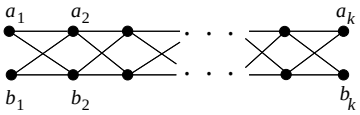

Fig. 18.2 A cube

A k-dimensional *cube* is defined to be a multi-set $D = \{a_1, b_1\} \times \cdots \times \{a_k, b_k\}$, where $a_i, b_i \in X_i$ (not necessarily distinct) for all i. Being a multi-set means that one element can occur several times. Thus, for example, the cube $D = \{a_1, a_1\} \times \cdots \times \{a_k, a_k\}$ has 2^k elements.

Given a function $f : X \to \{-1, 1\}$ and a cube $D \subseteq X$, define the *sign* of f on D to be the value

$$f(D) = \prod_{x\in D} f(x) .$$

Hence, $f(D) = 1$ if and only if $f(x) = -1$ for an even number of vectors $x \in D$. We choose a cube D at random according to the uniform distribution. This can be done by choosing $a_i, b_i \in X_i$ for each i according to the uniform distribution. Let

$$\mathcal{E}(f) := \mathrm{E}\,[f(D)] = \mathrm{E}\left[\prod_{x\in D} f(x)\right]$$

be the expected value of the sign of a random cube D. To stress the fact that the expectation is taken over a particular random object (this time, over D) we will also write $\mathrm{E}_D\,[f(D)]$ instead of $\mathrm{E}\,[f(D)]$.

Example 18.16. The difference between the measures $\mathrm{disc}(f)$ and $\mathcal{E}(f)$ can best be seen in the case when $k = 2$. In this case $X = X_1 \times X_2$ is just a grid, and each function $f : X \to \{-1, 1\}$ is just a ± 1 matrix M_f. Cylinder intersections $T \subseteq X$ in this case correspond to submatrices of M_f, and $\mathrm{disc}_T(f)$ is just the sum of all entries in T divided by $|X|$. Thus, to determine $\mathrm{disc}(f)$

we must consider *all* submatrices of M_f. In contrast, to determine $\mathcal{E}(f)$ it is enough to only consider all $s \times t$ submatrices with $1 \le s, t \le 2$.

The following result was proved in Chung (1990) and generalizes a similar result from Babai et al. (1992).

Theorem 18.17. *For every $f : X \to \{-1, 1\}$,*

$$\operatorname{disc}(f) \le \mathcal{E}(f)^{1/2^k}.$$

The theorem is very useful because $\mathcal{E}(f)$ is a much simpler object than $\operatorname{disc}(f)$. For many functions f, it is relatively easy to compute $\mathcal{E}(f)$ exactly (we will show this in the next section). In Chung and Tetali (1993), $\mathcal{E}(f)$ was computed for some explicit functions, resulting in the highest known lower bounds for the multi-party communication complexity of these functions.

Proof (due to Raz 2000). We will only prove the theorem for $k = 2$; the general case is similar. So let $X = X_1 \times X_2$ and $f : X \to \{-1, 1\}$ be a given function. Our goal is to show that $\operatorname{disc}(f) \le \mathcal{E}(f)^{1/4}$. To do this, pick at random (uniformly and independently) an element $\boldsymbol{x} \in X$. The proof consists of showing two claims.

Claim 18.18. For all functions $h : X \to \{-1, 1\}$, $\mathcal{E}(h) \ge (\mathrm{E}_{\boldsymbol{x}}[h(\boldsymbol{x})])^4$.

Claim 18.19. There exists h such that $\left|\mathrm{E}_{\boldsymbol{x}}[h(\boldsymbol{x})]\right| \ge \operatorname{disc}(f)$ and $\mathcal{E}(h) = \mathcal{E}(f)$.

Together, these two claims imply the theorem (for $k = 2$):

$$\mathcal{E}(f) = \mathcal{E}(h) \ge (\mathrm{E}_{\boldsymbol{x}}[h(\boldsymbol{x})])^4 = \left|\mathrm{E}_{\boldsymbol{x}}[h(\boldsymbol{x})]\right|^4 \ge \operatorname{disc}(f)^4.$$

In the proof of these two claims we will use two known facts about the mean value of random variables:

$$\mathrm{E}\left[\xi^2\right] \ge \mathrm{E}[\xi]^2 \quad \text{for any random variable } \xi; \tag{18.8}$$

and

$$\mathrm{E}[\xi \cdot \xi'] = \mathrm{E}[\xi] \cdot \mathrm{E}[\xi'] \quad \text{if } \xi \text{ and } \xi' \text{ are independent.} \tag{18.9}$$

The first one is a consequence of the Cauchy–Schwarz inequality, and the second is a basic property of expectation.

Proof of Claim 18.18. Take a random 2-dimensional cube $D = \{a, a'\} \times \{b, b'\}$. Then

$$\begin{aligned}
\mathcal{E}(h) &= \mathrm{E}_D\left[h(D)\right] = \mathrm{E}_D\Big[\prod_{x\in D} h(x)\Big] \\
&= \mathrm{E}_{a,a'}\mathrm{E}_{b,b'}\left[h(a,b)\cdot h(a,b')\cdot h(a',b)\cdot h(a',b')\right] \\
&= \mathrm{E}_{a,a'}\left[\left(\mathrm{E}_b\left[h(a,b)\cdot h(a',b)\right]\right)^2\right] && \text{by (18.9)} \\
&\geq \left(\mathrm{E}_{a,a'}\mathrm{E}_b\left[h(a,b)\cdot h(a',b)\right]\right)^2 && \text{by (18.8)} \\
&= \left(\mathrm{E}_a\mathrm{E}_b\left[h(a,b)^2\right]\right)^2 && \Pr\left[a'\right] = \Pr\left[a\right] \\
&= \left(\mathrm{E}_a\left(\mathrm{E}_b\left[h(a,b)\right]\right)^2\right)^2 && \text{by (18.9)} \\
&\geq \left(\mathrm{E}_{a,b}\left[h(a,b)\right]\right)^4 && \text{by (18.8).} \quad \square
\end{aligned}$$

Proof of Claim 18.19. Let $T = A\times B$ be a cylinder intersection (a submatrix of X, since $k = 2$) for which $\mathrm{disc}(f)$ is attained. We prove the existence of h by the probabilistic method. The idea is to define a random function $\boldsymbol{g} : X_1 \times X_2 \to \{-1,1\}$ such that the expected value $\mathrm{E}\left[\boldsymbol{g}(x)\right] = \mathrm{E}_{\boldsymbol{g}}\left[\boldsymbol{g}(x)\right]$ is the characteristic function of T. For this, define $\boldsymbol{g}$ to be the product $\boldsymbol{g}(x) = \boldsymbol{g}_1(x)\cdot\boldsymbol{g}_2(x)$ of two random functions, whose values are defined on the points $x = (a,b) \in X_1 \times X_2$ by:

$$\boldsymbol{g}_1(a,b) = \begin{cases} 1 & \text{if } a \in A; \\ \text{set randomly to } \pm 1 & \text{otherwise} \end{cases}$$

and

$$\boldsymbol{g}_2(a,b) = \begin{cases} 1 & \text{if } b \in B; \\ \text{set randomly to } \pm 1 & \text{otherwise.} \end{cases}$$

These function have the property that $\boldsymbol{g}_1$ depends only on the rows and $\boldsymbol{g}_2$ only on the columns of the grid $X_1 \times X_2$. That is, $\boldsymbol{g}_1(a,b) = \boldsymbol{g}_1(a,b')$ and $\boldsymbol{g}_2(a,b) = \boldsymbol{g}_2(a',b)$ for all $a,a' \in X_1$ and $b,b' \in X_2$. Hence, for $x \in T$, $\boldsymbol{g}(x) = 1$ with probability 1, while for $x \notin T$, $\boldsymbol{g}(x) = 1$ with probability $1/2$ and $\boldsymbol{g}(x) = -1$ with probability $1/2$; this is so because the functions $\boldsymbol{g}_1, \boldsymbol{g}_2$ are independent of each other, and $x \notin T$ iff $x \notin A\times X_2$ or $x \notin X_1 \times B$. Thus, the expectation $\mathrm{E}\left[\boldsymbol{g}(x)\right]$ takes the value 1 on all $x \in T$, and takes the value $\frac{1}{2} + (-\frac{1}{2}) = 0$ on all $x \notin T$, i.e., $\mathrm{E}\left[\boldsymbol{g}(x)\right]$ is the characteristic function of the set T:

$$\mathrm{E}\left[\boldsymbol{g}(x)\right] = \begin{cases} 1 & \text{if } x \in T; \\ 0 & \text{if } x \notin T. \end{cases}$$

Now let $\boldsymbol{x}$ be a random vector uniformly distributed in $X = X_1 \times X_2$. Then

$$\begin{aligned}
\mathrm{disc}_T(f) &= \left|\mathrm{E}_{\boldsymbol{x}}\left[f(\boldsymbol{x})\cdot\mathrm{E}_{\boldsymbol{g}}\left[\boldsymbol{g}(\boldsymbol{x})\right]\right]\right| = \left|\mathrm{E}_{\boldsymbol{x}}\mathrm{E}_{\boldsymbol{g}}\left[f(\boldsymbol{x})\cdot\boldsymbol{g}(\boldsymbol{x})\right]\right| \\
&= \left|\mathrm{E}_{\boldsymbol{g}}\mathrm{E}_{\boldsymbol{x}}\left[f(\boldsymbol{x})\cdot\boldsymbol{g}(\boldsymbol{x})\right]\right|.
\end{aligned}$$

So there exists some choice of $g = g_1 \cdot g_2$ such that