Stasys Jukna

Extremal Combinatorics

With Applications in Computer Science

Second Edition

Springer

Texts in Theoretical Computer Science
An EATCS Series

For further volumes:
www.springer.com/series/3214

Stasys Jukna

Extremal Combinatorics

With Applications in Computer Science

Second Edition

Springer

Prof. Dr. Stasys Jukna
Goethe Universität Frankfurt
Institut für Informatik
Robert-Mayer Str. 11-15
60054 Frankfurt am Main
Germany
jukna@thi.informatik.uni-frankfurt.de

Vilnius University
Institute of Mathematics and Informatics
Akademijos 4
08663 Vilnius
Lithuania

ISSN 1862-4499 Texts in Theoretical Computer Science. An EATCS Series
ISBN 978-3-642-17363-9 e-ISBN 978-3-642-17364-6
DOI 10.1007/978-3-642-17364-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011937551

ACM Codes: G.2, G.3, F.1, F.2, F.4.1

Cover design: KünkelLopka GmbH, Heidelberg

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Indrė

Preface

Preface to the First Edition

Combinatorial mathematics has been pursued since time immemorial, and at a reasonable scientific level at least since Leonhard Euler (1707–1783). It rendered many services to both pure and applied mathematics. Then along came the prince of computer science with its many mathematical problems and needs – and it was combinatorics that best fitted the glass slipper held out. Moreover, it has been gradually more and more realized that combinatorics has all sorts of deep connections with "mainstream areas" of mathematics, such as algebra, geometry and probability. This is why combinatorics is now a part of the standard mathematics and computer science curriculum.

This book is as an introduction to *extremal combinatorics* – a field of combinatorial mathematics which has undergone a period of spectacular growth in recent decades. The word "extremal" comes from the nature of problems this field deals with: if a collection of finite objects (numbers, graphs, vectors, sets, etc.) satisfies certain restrictions, how large or how small can it be?

For example, how many people can we invite to a party where among each three people there are two who know each other and two who don't know each other? An easy Ramsey-type argument shows that at most five persons can attend such a party. Or, suppose we are given a finite set of nonzero integers, and are asked to mark an as large as possible subset of them under the restriction that the sum of any two marked integers cannot be marked. It turns out that (independent of what the given integers actually are!) we can always mark at least one-third of them.

Besides classical tools, like the pigeonhole principle, the inclusion-exclusion principle, the double counting argument, induction, Ramsey argument, etc., some recent weapons – the probabilistic method and the linear algebra method – have shown their surprising power in solving such problems. With a mere knowledge of the concepts of linear independence and discrete probability, completely unexpected connections can be made between algebra, probability, and combinatorics. These techniques have also found striking applications in other areas of discrete mathematics and, in particular, in the theory of computing.

Nowadays we have comprehensive monographs covering different parts of extremal combinatorics. These books provide an invaluable source for students and researchers in combinatorics. Still, I feel that, despite its great potential and surprising applications, this fascinating field is not so well known for students and researchers in computer science. One reason could be that, being comprehensive and in-depth, these monographs are somewhat too difficult to start with for the beginner. I have therefore tried to write a "guide tour" to this field – an introductory text which should

- be self-contained,
- be more or less up-to-date,
- present a wide spectrum of basic ideas of extremal combinatorics,
- show how these ideas work in the theory of computing, and
- be accessible to graduate and motivated undergraduate students in mathematics and computer science.

Even if not all of these goals were achieved, I hope that the book will at least give a first impression about the power of extremal combinatorics, the type of problems this field deals with, and what its methods could be good for. This should help students in computer science to become more familiar with combinatorial reasoning and so be encouraged to open one of these monographs for more advanced study.

Intended for use as an introductory course, the text is, therefore, far from being all-inclusive. Emphasis has been given to theorems with elegant and beautiful proofs: those which may be called the gems of the theory and may be relatively easy to grasp by non-specialists. Some of the selected arguments are possible candidates for *The Book*, in which, according to Paul Erdős, God collects the perfect mathematical proofs.* I hope that the reader will enjoy them despite the imperfections of the presentation.

A possible feature and main departure from traditional books in combinatorics is the choice of topics and results, influenced by the author's twenty years of research experience in the theory of computing. Another departure is the inclusion of combinatorial results that originally appeared in computer science literature. To some extent, this feature may also be interesting for students and researchers in combinatorics. In particular, some impressive applications of combinatorial methods in the theory of computing are discussed.

Teaching. The text is *self-contained.* It assumes a certain mathematical maturity but *no* special knowledge in combinatorics, linear algebra, prob-

* "You don't have to believe in God but, as a mathematician, you should believe in The Book." (Paul Erdős)

For the first approximation see M. Aigner and G.M. Ziegler, *Proofs from THE BOOK.* Second Edition, Springer, 2000.

ability theory, or in the theory of computing — a standard mathematical background at undergraduate level should be enough to enjoy the proofs. All necessary concepts are introduced and, with very few exceptions, all results are proved before they are used, even if they are indeed "well-known." Fortunately, the problems and results of combinatorics are usually quite easy to state and explain, even for the layman. Its accessibility is one of its many appealing aspects.

The book contains much more material than is necessary for getting acquainted with the field. I have split it into relatively short chapters, each devoted to a particular proof technique. I have tried to make the chapters almost *independent*, so that the reader can choose his/her own order to follow the book. The (linear) order, in which the chapters appear, is just an extension of a (partial) order, "core facts first, applications and recent developments later." Combinatorics is broad rather than deep, it appears in different (often unrelated) corners of mathematics and computer science, and it is about techniques rather than results – this is where the independence of chapters comes from.

Each chapter starts with results demonstrating the particular technique in the simplest (or most illustrative) way. The relative importance of the topics discussed in separate chapters is not reflected in their length – only the topics which appear for the first time in the book are dealt with in greater detail. To facilitate the understanding of the material, over 300 exercises of varying difficulty, together with hints to their solution, are included. This is a vital part of the book – many of the examples were chosen to complement the main narrative of the text. Some of the hints are quite detailed so that they actually sketch the entire solution; in these cases the reader should try to fill out all missing details.

Acknowledgments. I would like to thank everybody who was directly or indirectly involved in the process of writing this book. First of all, I am grateful to Alessandra Capretti, Anna Gál, Thomas Hofmeister, Daniel Kral, G. Murali Krishnan, Martin Mundhenk, Gurumurthi V. Ramanan, Martin Sauerhoff and P.R. Subramania for comments and corrections.

Although not always directly reflected in the text, numerous earlier discussions with Anna Gál, Pavel Pudlák, and Sasha Razborov on various combinatorial problems in computational complexity, as well as short communications with Noga Alon, Aart Blokhuis, Armin Haken, Johan Håstad, Zoltan Füredi, Hanno Lefmann, Ran Raz, Mike Sipser, Mario Szegedy, and Avi Wigderson, have broadened my understanding of things. I especially benefited from the comments of Aleksandar Pekec and Jaikumar Radhakrishnan after they tested parts of the draft version in their courses in the BRICS International Ph.D. school (University of Aarhus, Denmark) and Tata Institute (Bombay, India), and from valuable comments of László Babai on the part devoted to the linear algebra method.

I would like to thank the Alexander von Humboldt Foundation and the German Research Foundation (Deutsche Forschungsgemeinschaft) for sup-

porting my research in Germany since 1992. Last but not least, I would like to acknowledge the hospitality of the University of Dortmund, the University of Trier and the University of Frankfurt; many thanks, in particular, to Ingo Wegener, Christoph Meinel and Georg Schnitger, respectively, for their help during my stay in Germany. This was the time when the idea of this book was born and realized. I am indebted to Hans Wössner and Ingeborg Mayer of Springer-Verlag for their editorial help, comments and suggestions which essentially contributed to the quality of the presentation in the book.

My deepest thanks to my wife, Daiva, and my daughter, Indrė, for being there.

Frankfurt/Vilnius March 2001 *Stasys Jukna*

Preface to the Second Edition

This second edition has been extended with substantial new material, and has been revised and updated throughout. In particular, it offers three new chapters about expander graphs and eigenvalues, the polynomial method and error-correcting codes. Most of the remaining chapters also include new material such as the Kruskal–Katona theorem about shadows, the Lovász–Stein theorem about coverings, large cliques in dense graphs without induced 4-cycles, a new lower bounds argument for monotone formulas, Dvir's solution of finite field Kakeya's conjecture, Moser's algorithmic version of the Lovász Local Lemma, Schöning's algorithm for 3-SAT, the Szemerédi–Trotter theorem about the number of point-line incidences, applications of expander graphs in extremal number theory, and some other results. Also, some proofs are made shorter and new exercises are added. And, of course, all errors and typos observed by the readers in the first edition are corrected.

I received a lot of letters from many readers pointing to omissions, errors or typos as well as suggestions for alternative proofs – such an enthusiastic reception of the first edition came as a great surprise. The second edition gives me an opportunity to incorporate all the suggestions and corrections in a new version. I am therefore thankful to all who wrote me, and in particular to: S. Akbari, S. Bova, E. Dekel, T. van Erven, D. Gavinsky, Qi Ge, D. Gunderson, S. Hada, H. Hennings, T. Hofmeister, Chien-Chung Huang, J. Hünten, H. Klauck, W. Koolen-Wijkstra, D. Krämer, U. Leck, Ben Pak Ching Li, D. McLaury, T. Mielikäinen, G. Mota, G. Nyul, V. Petrovic, H. Prothmann, P. Rastas, A. Razen, C. J. Renteria, M. Scheel, N. Schmitt, D. Sieling, T. Tassa, A. Utturwar, J. Volec, F. Voloch, E. Weinreb, A. Windsor, R. de Wolf, Qiqi Yan, A. Zilberstein, and P. Zumstein.

I thank everyone whose input has made a difference for this new edition. I am especially thankful to Thomas Hofmeister, Detlef Sieling and Ronald de Wolf who supplied me with the reaction of their students. The “error-probability” in the 2nd edition was reduced by Ronald de Wolf and Philipp Zumstein who gave me a lot of corrections for the new stuff included in this edition. I am especially thankful to Ronald for many discussions—his help was extremely useful during the whole preparation of this edition. All remaining errors are entirely my fault.

Finally, I would like to acknowledge the German Research Foundation (Deutsche Forschungsgemeinschaft) for giving an opportunity to finish the 2nd edition while working within the grant SCHN 503/5-1.

Frankfurt/Vilnius August 2011 *S. J.*

Notation

In this section we give the notation that shall be standard throughout the book.

Sets

We deal exclusively with finite objects. We use the standard set-theoretical notation:

- $|X|$ denotes the *size* (the *cardinality*) of a set X.
- A *k-set* or *k-element set* is a set of k elements.
- $[n] = \{1, 2, \dots, n\}$ is often used as a "standard" n-element set.
- $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$.
- $\overline{A} = X \setminus A$ is the complement of A.
- $A \oplus B = (A \setminus B) \cup (B \setminus A)$ (symmetric difference).
- $A \times B = \{(a, b) : a \in A, b \in B\}$ (Cartesian product).
- $A \subseteq B$ if B contains all the elements of A.
- $A \subset B$ if $A \subseteq B$ and $A \neq B$.
- 2^X is the set of all subsets of the set X. If $|X| = n$ then $|2^X| = 2^n$.
- A *permutation* of X is a one-to-one mapping (a bijection) $f\colon X \to X$.
- $\{0,1\}^n = \{(v_1, \dots, v_n) : v_i \in \{0,1\}\}$ is the (binary) n-cube.
- 0-1 vector (matrix) is a vector (matrix) with entries 0 and 1.
- A *unit vector* e_i is a 0-1 vector with exactly one 1 in the i-th position.
- An $m \times n$ matrix is a matrix with m rows and n columns.
- The *incidence vector* of a set $A \subseteq \{x_1, \dots, x_n\}$ is a 0-1 vector $v = (v_1, \dots, v_n)$, where $v_i = 1$ if $x_i \in A$, and $v_i = 0$ if $x_i \notin A$.
- The *characteristic function* of a subset $A \subseteq X$ is the function $f : X \to \{0,1\}$ such that $f(x) = 1$ if and only if $x \in A$.

Arithmetic

Some of the results are asymptotic, and we use the standard asymptotic notation: for two functions f and g, we write $f = O(g)$ if $f \leq c_1 g + c_2$ for all possible values of the two functions, where c_1, c_2 are absolute constants. We write $f = \Omega(g)$ if $g = O(f)$, and $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$. If the limit of the ratio f/g tends to 0 as the variables of the functions tend to infinity, we write $f = o(g)$. Finally, $f \lesssim g$ means that $f \leq (1 + o(1))g$, and $f \sim g$ denotes that $f = (1 + o(1))g$, i.e., that f/g tends to 1 when the variables tend to infinity. If x is a real number, then $\lceil x \rceil$ denotes the smallest integer not less than x, and $\lfloor x \rfloor$ denotes the greatest integer not exceeding x. As customary, $\mathbb{Z}$ denotes the set of integers, $\mathbb{R}$ the set of reals, $\mathbb{Z}_n$ an additive group of integers modulo n, and $\mathrm{GF}(q)$ (or $\mathbb{F}_q$) a finite Galois field with q elements. Such a field exists as long as q is a prime power. If $q = p$ is a prime then $\mathbb{F}_p$ can be viewed as the set $\{0, 1, \ldots, p-1\}$ with addition and multiplication performed modulo p. The sum in $\mathbb{F}_2$ is often denoted by $\oplus$, that is, $x \oplus y$ stands for $x + y \bmod 2$. We will often use the so-called *Cauchy–Schwarz inequality* (see Proposition 13.4 for a proof): if $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are real numbers then

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right).$$

If not stated otherwise, e = 2.718... will always denote the base of the natural logarithm.

Graphs

A *graph* is a pair $G = (V, E)$ consisting of a set V, whose members are called *vertices* (or *nodes*), and a family E of 2-element subsets of V, whose members are called *edges.* A vertex v is *incident* with an edge e if $v \in e$. The two vertices incident with an edge are its *endvertices* or *endpoints*, and the edge *joins* its ends. Two vertices u, v of G are *adjacent*, or *neighbors*, if $\{u, v\}$ is an edge of G. The number $d(u)$ of neighbors of a vertex u is its *degree.* A *walk* of length k in G is a sequence $v_0, e_1, v_1 \ldots, e_k, v_k$ of vertices and edges such that $e_i = \{v_{i-1}, v_i\}$. A walk without repeated vertices is a *path.* A walk without repeated edges is a *trail.* A *cycle* of length k is a path $v_0, \ldots, v_k$ with $v_0 = v_k$. A (connected) *component* in a graph is a set of its vertices such that there is a path between any two of them. A graph is *connected* if it consists of one component. A *tree* is a connected graph without cycles. A *subgraph* is obtained by deleting edges and vertices. A *spanning subgraph* is obtained by deleting edges only. An *induced subgraph* is obtained by deleting vertices (together with all the edges incident to them).

A *complete graph* or *clique* is a graph in which every pair is adjacent. An *independent set* in a graph is a set of vertices with no edges between them. The greatest integer r such that G contains an independent set of size r is the *independence number* of G, and is denoted by $\alpha(G)$. A graph is *bipartite* if its vertex set can be partitioned into two independent sets.

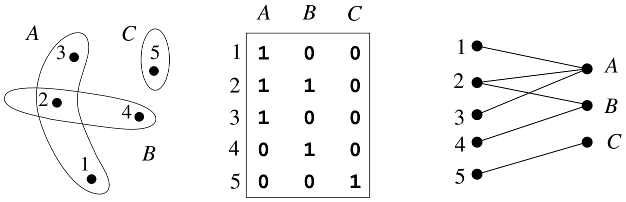
A *legal coloring* of $G = (V, E)$ is an assignment of colors to each vertex so that adjacent vertices receive different colors. In other words, this is a partition of the vertex set V into independent sets. The minimum number of colors required for that is the *chromatic number* $\chi(G)$ of G.

Set systems

A *set system* or *family of sets* $\mathcal{F}$ is a collection of sets. Because of their intimate conceptual relation to graphs, a set system is often called a *hypergraph.* A family is k-*uniform* if all its members are k-element sets. Thus, graphs are k-uniform families with $k = 2$.

In order to prove something about families of sets (as well as to interpret the results) it is often useful to keep in mind that any family can be looked at either as a 0-1 matrix or as a bipartite graph.

Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a family of subsets of a set $X = \{x_1, \ldots, x_n\}$. The *incidence matrix* of $\mathcal{F}$ is an $n \times m$ 0-1 matrix $M = (m_{i,j})$ such that $m_{i,j} = 1$ if and only if $x_i \in A_j$. Hence, the j-th column of M is the incidence vector of the set A_j. The *incidence graph* of $\mathcal{F}$ is a bipartite graph with parts X and $\mathcal{F}$, where x_i and A_j are joined by an edge if and only if $x_i \in A_j$.



$$\begin{array}{c|ccc|} & A & B & C \\ \hline 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 4 & 0 & 1 & 0 \\ 5 & 0 & 0 & 1 \\ \hline \end{array}$$

Fig. 0.1 Three representations of the family $\mathcal{F} = \{A, B, C\}$ over the set of points $X = \{1, 2, 3, 4, 5\}$ with $A = \{1, 2, 3\}$, $B = \{2, 4\}$ and $C = \{5\}$.

For small n, the system of all subsets of an n-element set, ordered by set-inclusion, can be represented by a so-called *Hasse diagram.* The k-th level here contains all k-element subsets, $k = 0, 1, \ldots, n$.

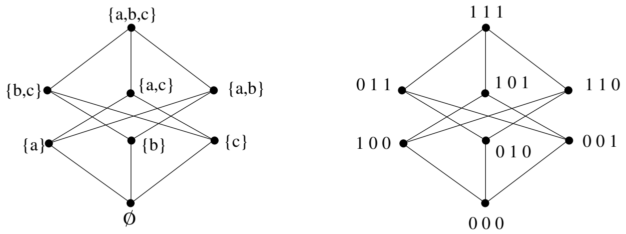


Fig. 0.2 A Hasse diagram of the family of all subsets of {a,b,c} ordered by set-inclusion, and the set of all binary strings of length three; there is an edge between two strings if and only if they differ in exactly one position.

Contents

Part II Extremal Set Theory

Part III The Linear Algebra Method

Part IV The Probabilistic Method

Part V Fragments of Ramsey Theory

Part I
The Classics

1. Counting

We start with the oldest combinatorial tool — *counting*.

1.1 The binomial theorem

Given a set of n elements, how many of its subsets have exactly k elements? This number (of k-element subsets of an n-element set) is usually denoted by $\binom{n}{k}$ and is called the *binomial coefficient*. Put otherwise, $\binom{n}{k}$ is the number of possibilities to choose k distinct objects from a collection on n distinct objects.

The following identity was proved by Sir Isaac Newton in about 1666, and is known as the *Binomial theorem*.

Binomial Theorem. *Let n be a positive integer. Then for all x and y,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\,.$$

Proof. If we multiply the terms

$$(x+y)^n = \underbrace{(x+y)\cdot(x+y)\cdot\ldots\cdot(x+y)}_{n-\text{times}},$$

then, for every $k = 0, 1, \ldots, n$, there are exactly $\binom{n}{k}$ possibilities to obtain the term $x^k y^{n-k}$. Why? We obtain the term $x^k y^{n-k}$ precisely if from n possibilities (terms $x+y$) we choose the first number x exactly k times. □

Note that this theorem just generalizes the known equality:

$$(x+y)^2 = \binom{2}{0}x^0y^2 + \binom{2}{1}x^1y^1 + \binom{2}{2}x^2y^0 = x^2 + 2xy + y^2\,.$$

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_1,
© Springer-Verlag Berlin Heidelberg 2011

Be it so simple, the binomial theorem has many applications.

Example 1.1 (Parity of powers). To give a typical example, let us show the following property of integers: If n, k are natural numbers, then n^k is odd iff n is odd.

One direction ($\Rightarrow$) is trivial: If $n = 2m$ is even, then $n^k = 2^k(m^k)$ must be also even. To show the other direction ($\Leftarrow$), assume that n is odd, that is, has the form $n = 2m + 1$ for a natural number m. The binomial theorem with $x = 2m$ and $y = 1$ yields:

$$n^k = (2m+1)^k = 1 + (2m)^1\binom{k}{1} + (2m)^2\binom{k}{2} + \cdots + (2m)^k\binom{k}{k}.$$

That is, the number n^k has the form "1 plus an even number", and must be odd.

The *factorial* of n is the product $n! := n(n-1)\cdots 2 \cdot 1$. This is extended to all non-negative integers by letting $0! = 1$. The *k-th factorial* of n is the product of the first k terms:

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

Note that $\binom{n}{0} = 1$ (the empty set) and $\binom{n}{n} = 1$ (the whole set). In general, binomial coefficients can be written as quotients of factorials:

Proposition 1.2.

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}.$$

Proof. Observe that $(n)_k$ is the number of (ordered!) strings $(x_1, x_2, \ldots, x_k)$ consisting of k different elements of a fixed n-element set: there are n possibilities to choose the first element x_1; after that there are still $n-1$ possibilities to choose the next element x_2, etc. Another way to produce such strings is to choose a k-element set and then arrange its elements in an arbitrary order. Since each of $\binom{n}{k}$ k-element subsets produces exactly $(k)_k = k!$ such strings, we conclude that $(n)_k = \binom{n}{k}k!$. □

There are a lot of useful equalities concerning binomial coefficients. In most situations, using their *combinatorial* nature (instead of algebraic, as given by the previous proposition) we obtain the desired result fairly easily. For example, if we observe that each subset is uniquely determined by its complement, then we immediately obtain the equality

$$\binom{n}{n-k} = \binom{n}{k}. \tag{1.1}$$

By this equality, for every fixed n, the value of the binomial coefficient $\binom{n}{k}$ increases till the middle and then decreases. By the binomial theorem, the sum of all these $n+1$ coefficients is equal to the total number 2^n of all subsets of an n-element set:

$$\sum_{k=0}^{n}\binom{n}{k}=\sum_{k=0}^{n}\binom{n}{k}1^k1^{n-k}=(1+1)^n=2^n\,.$$

In a similar (combinatorial) way other useful identities can be established (see Exercises for more examples).

Proposition 1.3 (Pascal Triangle)**.** *For every integers $n \geq k \geq 1$, we have*

$$\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}.$$

Proof. The first term $\binom{n-1}{k-1}$ is the number of k-sets containing a fixed element, and the second term $\binom{n-1}{k}$ is the number of k-sets avoiding this element; their sum is the whole number $\binom{n}{k}$ of k-sets. □

For growing n and k, exact values of binomial coefficients $\binom{n}{k}$ are hard to compute. In applications, however, we are often interested only in their rate of growth, so that (even rough) estimates suffice. Such estimates can be obtained, using the Taylor series of the exponential and logarithmic functions:

$$\mathrm{e}^t=1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\cdots \qquad \text{for all } t\in\mathbb{R} \tag{1.2}$$

and

$$\ln(1+t)=t-\frac{t^2}{2}+\frac{t^3}{3}-\frac{t^4}{4}+\cdots \qquad \text{for } -1<t\leq 1. \tag{1.3}$$

This, in particular, implies some useful estimates:

$$1+t<\mathrm{e}^t \qquad \text{for all } t\neq 0, \tag{1.4}$$

$$1-t>\mathrm{e}^{-t-t^2/2} \qquad \text{for all } 0<t<1. \tag{1.5}$$

Proposition 1.4.

$$\left(\frac{n}{k}\right)^k\leq\binom{n}{k} \quad \textit{and} \quad \sum_{i=0}^{k}\binom{n}{i}\leq\left(\frac{\mathrm{e}n}{k}\right)^k\,. \tag{1.6}$$

Proof. Lower bound:

$$\left(\frac{n}{k}\right)^k=\frac{n}{k}\cdot\frac{n}{k}\cdots\frac{n}{k}\leq\frac{n}{k}\cdot\frac{n-1}{k-1}\cdots\frac{n-k+1}{1}=\binom{n}{k}.$$

Upper bound: for $0<t\leq 1$ the inequality

$$\sum_{i=0}^{k}\binom{n}{i} \le \sum_{i=0}^{k}\binom{n}{i}\frac{t^i}{t^k} = \frac{(1+t)^n}{t^k}$$

follows from the binomial theorem. Now substitute $t = k/n$ and use (1.4). □

Tighter (asymptotic) estimates can be obtained using the famous *Stirling formula* for the factorial:

$$n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\,\mathrm{e}^{\alpha_n}\,, \tag{1.7}$$

where $1/(12n+1) < \alpha_n < 1/12n$. This leads, for example, to the following elementary but very useful asymptotic formula for the k-th factorial:

$$(n)_k = n^k \mathrm{e}^{-\frac{k^2}{2n}-\frac{k^3}{6n^2}+o(1)} \quad \text{valid for } k = o(n^{3/4}), \tag{1.8}$$

and hence, for binomial coefficients:

$$\binom{n}{k} = \frac{n^k \mathrm{e}^{-\frac{k^2}{2n}-\frac{k^3}{6n^2}}}{k!}(1+o(1))\,. \tag{1.9}$$

1.2 Selection with repetitions

In the previous section we considered the number of ways to choose r *distinct* elements from an n-element set. It is natural to ask what happens if we can choose the same element repeatedly. In other words, we may ask how many integer solutions does the equation $x_1+\cdots+x_n = r$ have under the condition that $x_i \ge 0$ for all $i = 1,\ldots,n$. (Look at x_i as the number of times the i-th element was chosen.) The following more entertaining formulation of this problem was suggested by Lovász, Pelikán, and Vesztergombi (1977).

Suppose we have r sweets (of the same sort), which we want to distribute to n children. In how many ways can we do this? Letting x_i denote the number of sweets we give to the i-th child, this question is equivalent to that stated above.

The answer depends on how many sweets we have and how fair we are. If we are fair but have only $r \le n$ sweets, then it is natural to allow no repetitions and give each child no more than one sweet (each x_i is 0 or 1). In this case the answer is easy: we just choose those r (out of n) children who will get a sweet, and we already know that this can be done in $\binom{n}{r}$ ways.

Suppose now that we have enough sweets, i.e., that $r \ge n$. Let us first be fair, that is, we want every child gets at least one sweet. We lay out the sweets in a single row of length r (it does not matter in which order, they all are alike), and let the first child pick them up from the left to right. After a while we stop him/her and let the second child pick up sweets, etc. The distribution of sweets is determined by specifying the place (between consecutive sweets) of where to start with a new child. There are $r-1$ such places, and we have to select $n-1$ of them (the first child always starts at the beginning, so we have no choice here). For example, if we have $r = 9$ sweets and $n = 6$ children, a typical situation looks like this:

$$\begin{array}{cccccccccccccc} \square & \curlywedge & \square & \square & \square & \curlywedge & \square & \curlywedge & \square & \square & \curlywedge & \square & \curlywedge & \square \\ & 2 & & & & 3 & & 4 & & & 5 & & 6 & \end{array}$$

Thus, we have to select an $(n-1)$-element subset from an $(r-1)$-element set. The number of possibilities to do so is $\binom{r-1}{n-1}$. If we are unfair, we have more possibilities:

Proposition 1.5. *The number of integer solutions to the equation*

$$x_1 + \cdots + x_n = r$$

under the condition that $x_i \geq 0$ for all $i = 1, \ldots, n$, is $\binom{n+r-1}{r}$.

Proof. In this situation we are unfair and allow that some of the children may be left without a sweet. With the following trick we can reduce the problem of counting the number of such distributions to the problem we just solved: we borrow one sweet from each child, and then distribute the whole amount of $n+r$ sweets to the children so that each child gets at least one sweet. This way every child gets back the sweet we borrowed from him/her, and the lucky ones get some more. This "more" is exactly r sweets distributed to n children. We already know that the number of ways to distribute $n+r$ sweets to n children in a fair way is $\binom{n+r-1}{n-1}$, which by (1.1) equals $\binom{n+r-1}{r}$. □

1.3 Partitions

A *partition* of n objects is a collection of its mutually disjoint subsets, called *blocks*, whose union gives the whole set. Let $S(n; k_1, k_2, \ldots, k_n)$ denote the number of all partitions of n objects with k_i i-element blocks ($i = 1, \ldots, n$; $k_1 + 2k_2 + \ldots + nk_n = n$). That is,

$$k_i = \text{the number of } i\text{-element blocks in a partition.}$$

Proposition 1.6.

$$S(n; k_1, k_2, \ldots, k_n) = \frac{n!}{k_1! \cdots k_n! (1!)^{k_1} \cdots (n!)^{k_n}}\,.$$

Proof. If we consider any arrangement (i.e., a permutation) of the n objects we can get such a partition by taking the first k_1 elements as 1-element blocks, the next $2k_2$ elements as 2-element blocks, etc. Since we have $n!$ possible arrangements, it remains to show that we get any given partition exactly

$$k_1! \cdots k_n! (1!)^{k_1} \cdots (n!)^{k_n}$$

times. Indeed, we can construct an arrangement of the objects by putting the 1-element blocks first, then the 2-element blocks, etc. However, there are $k_i!$ possible ways to order the i-element blocks and $(i!)^{k_i}$ possible ways to order the elements in the i-element blocks. □

1.4 Double counting

The *double counting* principle states the following "obvious" fact: if the elements of a set are counted in two different ways, the answers are the same.

In terms of matrices the principle is as follows. Let M be an $n \times m$ matrix with entries 0 and 1. Let r_i be the number of 1s in the i-th row, and c_j be the number of 1s in the j-th column. Then

$$\sum_{i=1}^{n} r_i = \sum_{j=1}^{m} c_j = \text{the total number of 1s in } M\,.$$

The next example is a standard demonstration of double counting. Suppose a finite number of people meet at a party and some shake hands. Assume that no person shakes his or her own hand and furthermore no two people shake hands more than once.

Handshaking Lemma. *At a party, the number of guests who shake hands an odd number of times is even.*

Proof. Let $P_1, \ldots, P_n$ be the persons. We apply double counting to the set of ordered pairs (P_i, P_j) for which P_i and P_j shake hands with each other at the party. Let x_i be the number of times that P_i shakes hands, and y the total number of handshakes that occur. On one hand, the number of pairs is $\sum_{i=1}^{n} x_i$, since for each P_i the number of choices of P_j is equal to x_i. On the other hand, each handshake gives rise to two pairs (P_i, P_j) and (P_j, P_i); so the total is $2y$. Thus $\sum_{i=1}^{n} x_i = 2y$. But, if the sum of n numbers is even, then evenly many of the numbers are odd. (Because if we add an odd number of odd numbers and any number of even numbers, the sum will be always odd). □

This lemma is also a direct consequence of the following general identity, whose special version for graphs was already proved by Euler. For a point x, its *degree* or *replication number* $d(x)$ in a family $\mathcal{F}$ is the number of members of $\mathcal{F}$ containing x.

Proposition 1.7. *Let $\mathcal{F}$ be a family of subsets of some set X. Then*

$$\sum_{x\in X} d(x) = \sum_{A\in\mathcal{F}} |A|\,. \tag{1.10}$$

Proof. Consider the *incidence matrix* $M = (m_{x,A})$ of $\mathcal{F}$. That is, M is a 0-1 matrix with $|X|$ rows labeled by points $x \in X$ and with $|\mathcal{F}|$ columns labeled by sets $A \in \mathcal{F}$ such that $m_{x,A} = 1$ if and only if $x \in A$. Observe that $d(x)$ is exactly the number of 1s in the x-th row, and $|A|$ is the number of 1s in the A-th column. □

Graphs are families of 2-element sets, and the degree of a vertex x is the number of edges incident to x, i.e., the number of vertices in its neighborhood. Proposition 1.7 immediately implies

Theorem 1.8 (Euler 1736)**.** *In every graph the sum of degrees of its vertices is two times the number of its edges, and hence, is even.*

The following identities can be proved in a similar manner (we leave their proofs as exercises):

$$\sum_{x\in Y} d(x) = \sum_{A\in\mathcal{F}} |Y\cap A| \quad \text{for any } Y\subseteq X. \tag{1.11}$$

$$\sum_{x\in X} d(x)^2 = \sum_{A\in\mathcal{F}}\sum_{x\in A} d(x) = \sum_{A\in\mathcal{F}}\sum_{B\in\mathcal{F}} |A\cap B|\,. \tag{1.12}$$

Turán's number $T(n, k, l)$ $(l \le k \le n)$ is the smallest number of l-element subsets of an n-element set X such that every k-element subset of X contains at least one of these sets.

Proposition 1.9. *For all positive integers $l \le k \le n$,*

$$T(n,k,l) \ge \binom{n}{l}\Big/\binom{k}{l}\,.$$

Proof. Let $\mathcal{F}$ be a smallest l-uniform family over X such that every k-subset of X contains at least one member of $\mathcal{F}$. Take a 0-1 matrix $M = (m_{A,B})$ whose rows are labeled by sets A in $\mathcal{F}$, columns by k-element subsets B of X, and $m_{A,B} = 1$ if and only if $A \subseteq B$.

Let r_A be the number of 1s in the A-th row and c_B be the number of 1s in the B-th column. Then, $c_B \ge 1$ for every B, since B must contain at least one member of $\mathcal{F}$. On the other hand, r_A is precisely the number of k-element subsets B containing a fixed l-element set A; so $r_A = \binom{n-l}{k-l}$ for every $A \in \mathcal{F}$. By the double counting principle,

$$|\mathcal{F}|\cdot\binom{n-l}{k-l} = \sum_{A\in\mathcal{F}} r_A = \sum_{B} c_B \ge \binom{n}{k},$$

which yields

$$T(n,k,l) = |\mathcal{F}| \geq \binom{n}{k} \Big/ \binom{n-l}{k-l} = \binom{n}{l} \Big/ \binom{k}{l},$$

where the last equality is another property of binomial coefficients (see Exercise 1.12). □

Our next application of double counting is from number theory: How many numbers divide at least one of the first n numbers $1, 2, \ldots, n$? If $t(n)$ is the number of divisors of n, then the behavior of this function is rather non-uniform: $t(p) = 2$ for every prime number, whereas $t(2^m) = m + 1$. It is therefore interesting that the *average* number

$$\tau(n) = \frac{t(1) + t(2) + \cdots + t(n)}{n}$$

of divisors is quite stable: It is about $\ln n$.

Proposition 1.10. *$|\tau(n) - \ln n| \leq 1$.*

Proof. To apply the double counting principle, consider the 0-1 $n \times n$ matrix $M = (m_{ij})$ with $m_{ij} = 1$ iff j is divisible by i:

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1
2		1		1		1		1		1		1
3			1			1			1			1
4				1				1				1
5					1					1		
6						1						1
7							1					
8								1				

The number of 1s in the j-th column is exactly the number $t(j)$ of divisors of j. So, summing over columns we see that the total number of 1s in the matrix is $T_n = t(1) + \cdots + t(n)$.

On the other hand, the number of 1s in the i-th row is the number of multipliers $i, 2i, 3i, \ldots, ri$ of i such that $ri \leq n$. Hence, we have exactly $\lfloor n/i \rfloor$ ones in the i-th row. Summing over rows, we obtain that $T_n = \sum_{i=1}^{n} \lfloor n/i \rfloor$. Since $x - 1 < \lfloor x \rfloor \leq x$ for every real number x, we obtain that

$$H_n - 1 \leq \tau(n) = \frac{1}{n} T_n \leq H_n\,,$$

where

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \ln n + \gamma_n, \qquad 0 \leq \gamma_n \leq 1 \tag{1.13}$$

is the n-th harmonic number. □

1.5 The averaging principle

Suppose we have a set of m objects, the i-th of which has "size" l_i, and we would like to know if at least one of the objects is large, i.e., has size $l_i \geq t$ for some given t. In this situation we can try to consider the average size $\bar{l} = \sum l_i/m$ and try to prove that $\bar{l} \geq t$. This would immediately yield the result, because we have the following

Averaging Principle. Every set of numbers must contain a number at least as large ($\geq$) as the average and a number at least as small ($\leq$) as the average.

This principle is a prototype of a very powerful technique – the probabilistic method – which we will study in Part 4. The concept is very simple, but the applications can be surprisingly subtle. We will use this principle quite often.

To demonstrate the principle, let us prove the following sufficient condition that a graph is disconnected.

A (connected) *component* in a graph is a set of its vertices such that there is a path between any two of them. A graph is *connected* if it consists of one component; otherwise it is *disconnected.*

Proposition 1.11. *Every graph on n vertices with fewer than $n-1$ edges is disconnected.*

Proof. Induction by n. When $n = 1$, the claim is vacuously satisfied, since no graph has a negative number of edges.

When $n = 2$, a graph with less than 1 edge is evidently disconnected.

Suppose now that the result has been established for graphs on n vertices, and take a graph $G = (V, E)$ on $|V| = n+1$ vertices such that $|E| \leq n-1$. By Euler's theorem (Theorem 1.8), the average degree of its vertices is

$$\frac{1}{|V|}\sum_{x\in V} d(x) = \frac{2|E|}{|V|} \leq \frac{2(n-1)}{n+1} < 2\,.$$

By the averaging principle, some vertex x has degree 0 or 1. If $d(x) = 0$, x is a component disjoint from the rest of G, so G is disconnected. If $d(x) = 1$, suppose the unique neighbor of x is y. Then, the graph H obtained from G by deleting x and its incident edge has $|V| - 1 = n$ vertices and $|E| - 1 \leq (n-1) - 1 = n-2$ edges; by the induction hypothesis, H is disconnected. The restoration of an edge joining a vertex y in one component to a vertex x which is outside of a second component cannot reconnect the graph. Hence, G is also disconnected. □

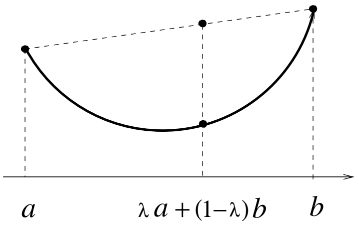


Fig. 1.1 A convex function.

We mention one important inequality, which is especially useful when dealing with averages.

A real-valued function $f(x)$ is *convex* if

$$f(\lambda a+(1-\lambda)b)\leq \lambda f(a)+(1-\lambda)f(b)\,,$$

for any $0\leq\lambda\leq 1$. From a geometrical point of view, the convexity of f means that if we draw a line l through points $(a,f(a))$ and $(b,f(b))$, then the graph of the curve $f(z)$ must lie below that of $l(z)$ for $z\in[a,b]$. Thus, for a function f to be convex it is sufficient that its second derivative is nonnegative.

Proposition 1.12 (Jensen's Inequality). *If* $0\leq\lambda_i\leq 1$, $\sum_{i=1}^n\lambda_i=1$ *and* f *is convex, then*

$$f\left(\sum_{i=1}^n\lambda_i x_i\right)\leq\sum_{i=1}^n\lambda_i f(x_i)\,. \tag{1.14}$$

Proof. Easy induction on the number of summands n. For $n=2$ this is true, so assume the inequality holds for the number of summands up to n, and prove it for $n+1$. For this it is enough to replace the sum of the first two terms in $\lambda_1x_1+\lambda_2x_2+\ldots+\lambda_{n+1}x_{n+1}$ by the term

$$(\lambda_1+\lambda_2)\left(\frac{\lambda_1}{\lambda_1+\lambda_2}x_1+\frac{\lambda_2}{\lambda_1+\lambda_2}x_2\right)\,,$$

and apply the induction hypothesis. □

If $a_1,\ldots,a_n$ are non-negative then, taking $f(x)=x^2$ and $\lambda_i=1/n$, we obtain a useful inequality (which is also an easy consequence of the Cauchy–Schwarz inequality):

$$\sum_{i=1}^n a_i^2\geq\frac{1}{n}\left(\sum_{i=1}^n a_i\right)^2\,. \tag{1.15}$$

Jensen's inequality (1.14) yields the following useful inequality between the arithmetic and geometric means: for any be non-negative numbers $a_1,\ldots,a_n$,

$$\frac{1}{n}\sum_{i=1}^{n} a_i \geq \left(\prod_{i=1}^{n} a_i\right)^{1/n} . \tag{1.16}$$

To show this, apply Jensen's inequality with $f(x) = 2^x$, $\lambda_1 = \ldots = \lambda_n = 1/n$ and $x_i = \log_2 a_i$, for all $i = 1, \ldots, n$. Then

$$\frac{1}{n}\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \lambda_i f(x_i) \geq f\left(\sum_{i=1}^{n} \lambda_i x_i\right) = 2^{(\sum_{i=1}^{n} x_i)/n} = \left(\prod_{i=1}^{n} a_i\right)^{1/n} .$$

1.6 The inclusion-exclusion principle

The *principle of inclusion and exclusion* (sieve of Eratosthenes) is a powerful tool in the theory of enumeration as well as in number theory. This principle relates the cardinality of the union of certain sets to the cardinalities of intersections of some of them, these latter cardinalities often being easier to handle.

For any two sets A and B we have

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

In general, given n subsets $A_1, \ldots, A_n$ of a set X, we want to calculate the number $|A_1 \cup \cdots \cup A_n|$ of points in their union. As the first approximation of this number we can take the sum

$$|A_1| + \cdots + |A_n|. \tag{1.17}$$

However, in general, this number is too large since if, say, $A_i \cap A_j \neq \emptyset$ then each point of $A_i \cap A_j$ is counted two times in (1.17): once in $|A_i|$ and once in $|A_j|$. We can try to correct the situation by subtracting from (1.17) the sum

$$\sum_{1 \leq i < j \leq n} |A_i \cap A_j|. \tag{1.18}$$

But then we get a number which is too small since each of the points in $A_i \cap A_j \cap A_k \neq \emptyset$ is counted three times in (1.18): once in $|A_i \cap A_j|$, once in $|A_j \cap A_k|$, and once in $|A_i \cap A_k|$. We can therefore try to correct the situation by adding the sum

$$\sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k|,$$

but again we will get a too large number, etc. Nevertheless, it turns out that after n steps we will get the correct result. This result is known as the *inclusion-exclusion principle.* The following notation will be handy: if I is a subset of the index set $\{1, \dots, n\}$, we set

$$A_I := \bigcap_{i \in I} A_i,$$

with the convention that $A_\emptyset = X$.

Proposition 1.13 (Inclusion-Exclusion Principle)**.** *Let $A_1, \dots, A_n$ be subsets of X. Then the number of elements of X which lie in none of the subsets A_i is*

$$\sum_{I \subseteq \{1,\dots,n\}} (-1)^{|I|} |A_I|. \tag{1.19}$$

Proof. The sum is a linear combination of cardinalities of sets A_I with coefficients $+1$ and -1. We can re-write this sum as

$$\sum_I (-1)^{|I|} |A_I| = \sum_I \sum_{x \in A_I} (-1)^{|I|} = \sum_x \sum_{I : x \in A_I} (-1)^{|I|}.$$

We calculate, for each point of X, its contribution to the sum, that is, the sum of the coefficients of the sets A_I which contain it.

First suppose that $x \in X$ lies in none of the sets A_i. Then the only term in the sum to which x contributes is that with $I = \emptyset$; and this contribution is 1.

Otherwise, the set $J := \{i : x \in A_i\}$ is non-empty; and $x \in A_I$ precisely when $I \subseteq J$. Thus, the contribution of x is

$$\sum_{I \subseteq J} (-1)^{|I|} = \sum_{i=0}^{|J|} \binom{|J|}{i} (-1)^i = (1-1)^{|J|} = 0$$

by the binomial theorem.

Thus, points lying in no set A_i contribute 1 to the sum, while points in some A_i contribute 0; so the overall sum is the number of points lying in none of the sets, as claimed. □

For some applications the following form of the inclusion-exclusion principle is more convenient.

Proposition 1.14. *Let $A_1, \dots, A_n$ be a sequence of (not necessarily distinct) sets. Then*

$$|A_1 \cup \dots \cup A_n| = \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} |A_I|\,. \tag{1.20}$$

Proof. The left-hand of (1.20) is $|A_\emptyset|$ minus the number of elements of $X = A_\emptyset$ which lie in none of the subsets A_i. By Proposition 1.13 this number is

$$|A_\emptyset| - \sum_{I \subseteq \{1,\dots,n\}} (-1)^{|I|} |A_I| = \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} |A_I|\,,$$

as desired. □

Suppose we would like to know, given a set of indices I, how many elements belong to all the sets A_i with $i \in I$ and do not belong to any of the remaining sets. Proposition 1.13 (which corresponds to the case when $I = \emptyset$) can be generalized for this situation.

Proposition 1.15. *Let $A_1, \ldots, A_n$ be sets, and I a subset of the index set $\{1, \ldots, n\}$. Then the number of elements which belong to A_i for all $i \in I$ and for no other values is*

$$\sum_{J \supseteq I} (-1)^{|J \setminus I|} |A_J| \,. \tag{1.21}$$

Proof. Consider the set $X := \bigcap_{i \in I} A_i$ and its subsets $B_k := X \cap A_k$, for all $k \in N \setminus I$, where $N := \{1, \ldots, n\}$. The proposition asks us to calculate the number of elements of X lying in none of B_k. By Proposition 1.13, this number is

$$\begin{aligned} \sum_{K \subseteq N \setminus I} (-1)^{|K|} \left| \bigcap_{k \in K} B_k \right| &= \sum_{K \subseteq N \setminus I} (-1)^{|K|} \left| \bigcap_{i \in K \cup I} A_i \right| \\ &= \sum_{J \supseteq I} (-1)^{|J \setminus I|} |A_J| \,. \quad \square \end{aligned}$$

What is the probability that if n people randomly search a dark closet to retrieve their hats, no person will pick his own hat? Using the principle of inclusion and exclusion it can be shown that this probability is very close to $\mathrm{e}^{-1} = 0.3678....$

This question can be formalized as follows. A *permutation* is a bijective mapping f of the set $\{1, \ldots, n\}$ into itself. We say that f *fixes* a point i if $f(i) = i$. A *derangement* is a permutation which fixes none of the points. We have exactly $n!$ permutations. How many of them are derangements?

Proposition 1.16. *The number of derangements of $\{1, \ldots, n\}$ is equal to*

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} (n-i)! = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!} \,. \tag{1.22}$$

The sum $\sum_{i=0}^{n} \frac{(-1)^i}{i!}$ is the initial part of the Taylor expansion of e^{-1}; so about an e^{-1} fraction of all permutations are derangements.

Proof. We are going to apply the inclusion-exclusion formula (1.19). Let X be the set of all permutations, and A_i the set of permutations fixing the point i; so $|A_i| = (n-1)!$, and more generally, $|A_I| = (n - |I|)!$, since permutations in A_I fix every point in I and permute the remaining points arbitrarily. A permutation is a derangement if and only if it lies in none of the sets A_i; so by (1.19), the number of derangements is

$$\sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|}(n-|I|)! = \sum_{i=0}^{n}(-1)^i\binom{n}{i}(n-i)!$$

putting $i = |I|$. □

Exercises

1.1. In how many ways can we distribute k balls to n boxes so that each box has at most one ball?

1.2. Show that for every k the product of any k consecutive natural numbers is divisible by $k!$. *Hint*: Consider $\binom{n+k}{k}$.

1.3. Show that the number of pairs (A, B) of distinct subsets of $\{1, \ldots, n\}$ with $A \subset B$ is $3^n - 2^n$. *Hint*: Use the binomial theorem to evaluate $\sum_{k=0}^{n}\binom{n}{k}(2^k - 1)$.

1.4. Show that

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}.$$

Hint: Count in two ways the number of pairs (x, M), where M is a k-element subset of $\{1, \ldots, n\}$ and $x \in M$.

1.5. Prove that

$$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}.$$

Hint: Count in two ways the number of pairs (x, M) with $x \in M \subseteq \{1, \ldots, n\}$.

1.6. There is a set of $2n$ people: n male and n female. A good party is a set with the same number of male and female. How many possibilities are there to build such a good party?

1.7. Use Proposition 1.3 to show that

$$\sum_{i=0}^{r}\binom{n+i-1}{i} = \binom{n+r}{r}.$$

1.8. Let $0 \le a \le m \le n$ be integers. Use Proposition 1.3 to show that

$$\sum_{i=m}^{n}\binom{i}{a} = \binom{n+1}{a+1} - \binom{m}{a+1}.$$

1.9. Prove the Cauchy–Vandermonde identity:

$$\binom{p+q}{k} = \sum_{i=0}^{k}\binom{p}{i}\binom{q}{k-i}.$$

Hint: Take a set of $p+q$ people (p male and q female) and make a set of k people (with i male and $k-i$ female).

1.10. Show that

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

Hint: Exercise 1.9 and Eq. (1.1).

1.11. Prove the following analogy of the binomial theorem for factorials:

$$(x+y)_n = \sum_{k=0}^{n} \binom{n}{k} (x)_k (y)_{n-k}.$$

Hint: Divide both sides by $n!$, and use the Cauchy–Vandermonde identity.

1.12. Let $0 \leq l \leq k \leq n$. Show that

$$\binom{n}{k}\binom{k}{l} = \binom{n}{l}\binom{n-l}{k-l}.$$

Hint: Count in two ways the number of all pairs (L, K) of subsets of $\{1, \ldots, n\}$ such that $L \subseteq K$, $|L| = l$ and $|K| = k$.

1.13. Use combinatorics (not algebra) to prove that, for $0 \leq k \leq n$,

$$\binom{n}{2} = \binom{k}{2} + k(n-k) + \binom{n-k}{2}.$$

Hint: $\binom{n}{2}$ is the number of edges in a complete graph on n vertices.

1.14. One of Euclid's theorems says that, if a prime number divides a product $a \cdot b$ of two integers, then p must divide at least one of these integers. Use this to show that:

(i) If $1 \leq k < p$, then $\binom{p}{k} \equiv 0 \bmod p$.
(ii) If $1 \leq k \leq n < p$, then $\binom{n}{k} \not\equiv 0 \bmod p$.

Hint: (i) Let $x = n(n-1)\cdots(n-k+1)$. Note that $x = a \cdot b$ with $a = \binom{n}{k}$ and $b = k!$.

1.15. Prove Fermat's Little theorem: if p is a prime and if a is a natural number, then $a^p \equiv a \bmod p$. In particular, if p does not divide a, then $a^{p-1} \equiv 1 \bmod p$. *Hint*: Apply the induction on a. For the induction step, use the binomial theorem to show that $(a+1)^p \equiv a^p + 1 \bmod p$.

1.16. Let $0 < \alpha < 1$ be a real number, and αn be an integer. Using Stirling's formula show that

$$\binom{n}{\alpha n} = \frac{1 + o(1)}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot 2^{n \cdot H(\alpha)},$$

where $H(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$ is the binary entropy function. *Hint*: $H(\alpha) = \log_2 h(\alpha)$, where $h(\alpha) = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}$.

1.17. Prove that, for $s \leq n/2$,

(1) $$\sum_{k=0}^{s} \binom{n}{k} \leq \binom{n}{s}\left(1 + \frac{s}{n-2s+1}\right);$$

(2) $$\sum_{k=0}^{s} \binom{n}{k} \leq 2^{n\cdot H(s/n)}.$$

Hint: To (1): observe that $\binom{n}{k-1}/\binom{n}{k} = k/(n-k+1)$ does not exceed $\alpha := s/(n-s+1)$, and use the identity $\sum_{i=0}^{\infty} \alpha^i = 1/(1-\alpha)$.
To (2): set $p = s/n$ and apply the binomial theorem to show that

$$p^s(1-p)^{n-s}\sum_{k=0}^{s}\binom{n}{k} \leq 1\,.$$

See also Corollary 22.9 for another proof.

1.18. Prove the following estimates: If $k \leq k + x < n$ and $y < k \leq n$, then

$$\left(\frac{n-k-x}{n-x}\right)^x \leq \binom{n-x}{k}\binom{n}{k}^{-1} \leq \left(\frac{n-k}{n}\right)^x \leq \mathrm{e}^{-(k/n)x} \tag{1.23}$$

and

$$\left(\frac{k-y}{n-y}\right)^y \leq \binom{n-y}{k-y}\binom{n}{k}^{-1} \leq \left(\frac{k}{n}\right)^y .$$

1.19. Prove that if $1 \leq k \leq n/2$, then

$$\binom{n}{k} \geq \gamma \cdot \left(\frac{n\mathrm{e}}{k}\right)^k \text{, where } \gamma = \frac{1}{\sqrt{2\pi k}}\,\mathrm{e}^{-k^2/n - 1/(6k)}. \tag{1.24}$$

Hint: Use Stirling's formula to show that

$$\binom{n}{k} \geq \frac{1}{\sqrt{2\pi}\,\mathrm{e}^{1/(6k)}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}\left(\frac{n}{k(n-k)}\right)^{1/2},$$

and apply the estimate $\ln(1+t) \geq t - t^2/2$ valid for all $t \geq 0$.

1.20. In how many ways can we choose a subset $S \subseteq \{1, 2, \ldots, n\}$ such that $|S| = k$ and no two elements of S precede each other, i.e., $x \neq y + 1$ for all $x, y \in S$? *Hint*: If $S = \{a_1, \ldots, a_k\}$ is such a subset with $a_1 < a_2 < \ldots < a_k$, then $a_1 < a_2 - 1 < \ldots < a_k - (k-1)$.

1.21. Let $k \geq 2n$. In how many ways can we distribute k sweets to n children, if each child is supposed to get at least 2 of them?

1.22. Let $\mathcal{F} = \{A_1, \dots, A_m\}$ be a family of subsets of a finite set X. For $x \in X$, let $d(x)$ be the number of members of $\mathcal{F}$ containing x. Show that $\sum_{i,j=1}^{m} |A_i \cap A_j| = \sum_{x \in X} d(x)^2$.

1.23. *Bell's number* B_n is the number of all possible partitions of an n-element set X (we assume that $B_0 = 1$). Prove that $B_{n+1} = \sum_{i=0}^{n} \binom{n}{i} B_i$. *Hint*: For every subset $A \subseteq X$ there are precisely $B_{|X \setminus A|}$ partitions of X containing A as one of its blocks.

1.24. Let $|N| = n$ and $|X| = x$. Show that there are x^n mappings from N to X, and that $S(n,k)x(x-1)\cdots(x-k+1)$ of these mappings have a range of cardinality k; here $S(n,k)$ is the Stirling number (the number of partitions of an n-element set into exactly k blocks). *Hint*: We have $x(x-1)\cdots(x-k+1)$ possibilities to choose a sequence of k elements in X, and we can specify $S(n,k)$ ways in which elements of N are mapped onto these chosen elements.

1.25. Let $\mathcal{F}$ be a family of subsets of an n-element set X with the property that any two members of $\mathcal{F}$ meet, i.e., $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{F}$. Suppose also that no other subset of X meets all of the members of $\mathcal{F}$. Prove that $|\mathcal{F}| = 2^{n-1}$. *Hint*: Consider sets and their complements.

1.26. Let $\mathcal{F}$ be a family of k-element subsets of an n-element set X such that every l-element subset of X is contained in at least one member of $\mathcal{F}$. Show that $|\mathcal{F}| \geq \binom{n}{l} / \binom{k}{l}$. *Hint*: Argue as in the proof of Proposition 1.9.

1.27. (Sperner 1928). Let $\mathcal{F}$ be a family of k-element subsets of $\{1, \dots, n\}$. Its *shadow* is the family of all those $(k-1)$-element subsets which lie entirely in at least one member of $\mathcal{F}$. Show that the shadow contains at least $k|\mathcal{F}|/(n-k+1)$ sets. *Hint*: Argue as in the proof of Proposition 1.9.

1.28. (Counting in bipartite graphs). Let $G = (A \cup B, E)$ be a bipartite graph, d be a minimum degree of a vertex in A and D the maximum degree of a vertex in B. Assume that $|A|d \geq |B|D$. Show that then, for every subset $A_0 \subseteq A$ of density $\alpha := |A_0|/|A|$, there is a subset $B_0 \subseteq B$ such that: (i) $|B_0| \geq \alpha|B|/2$, (ii) every vertex of B_0 has at least $\alpha D/2$ neighbors in A_0, and (iii) at least half of the edges leaving A_0 go to B_0. *Hint*: Let B_0 consist of all vertices in B having $> \alpha D/2$ neighbors in A_0.

1.29. Let $a_1, \dots, a_n$ be nonnegative numbers. Define

$$f(t) = \left(\frac{a_1^t + \cdots + a_n^t}{n} \right)^{1/t}.$$

Use Jensen's inequality to show that $s \leq t$ implies $f(s) \leq f(t)$.

1.30. (Quine 1988). The famous Fermat's Last Theorem states that if $n > 2$, then $x^n + y^n = z^n$ has no solutions in nonzero integers x, y and z. This theorem can be stated in terms of sorting objects into a row of bins, some of which are red, some blue, and the rest unpainted. The theorem amounts to saying that when there are more than two objects, the following statement is never true: *The number of ways of sorting them that shun both colors is equal to the number of ways that shun neither.* Show that this statement is equivalent to Fermat's equation $x^n + y^n = z^n$. *Hint*: Let n be the number of objects, z the number of bins, x the number of bins that are not red and y the number of bins that are not blue. There are z^n ways of sorting the objects into bins; x^n of these ways shun red and y^n of them shun blue.

1.31. Use the principle of inclusion and exclusion to determine the number of ways in which three women and their three spouses may be seated around a round table under each of the following two restrictions:

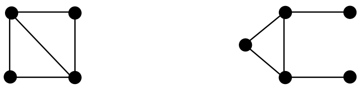
(i) no woman sits beside her spouse (on either side);
(ii) no two women may sit opposite one another at the table (i.e., with two people between them on either side).

Hint: To (i): two seatings are equivalent if one can be rotated into the other; so the underlying set consists of all circular permutations, 5! in number. Let A_i ($i = 1, 2, 3$) be the subset of permutations in which the members of the i-th couple sit side by side. Show that $|A_i| = 2 \cdot 4!$, $|A_i \cap A_j| = 2^2 \cdot 3!$, $|A_1 \cap A_2 \cap A_3| = 2^3 \cdot 2!$ and apply the inclusion-exclusion formula. To (ii): distinguish two cases, according to whether there exist two women sitting side by side or not.

1.32. Let $m \geq n$. A function $f \colon [m] \to [n]$ is a *surjection* (or a mapping of $[m]$ *onto* $[n]$) if f maps at least one element of $[m]$ to each element of $[n]$. Prove that the number of such functions is $\sum_{k=0}^{n-1} (-1)^k \binom{n}{k} (n-k)^m$. *Hint*: Let $A_i = \{f \ : \ f(j) \neq i \text{ for all } j\}$ and apply the inclusion-exclusion formula.

1.33. Let n and $k \geq l$ be positive integers. How many different integer solutions are there to the equation $x_1 + x_2 + \cdots + x_n = k$, with all $0 \leq x_i < l$? *Hint*: Consider the universum $X = X_{n,k}$ of all solutions with all $x_i \geq 0$, let A_i be the set of all solutions with $x_i \geq l$, and apply the inclusion-exclusion formula (1.19). Observe that $|A_i| = |X_{n,k-l}|$, where the size of $X_{n,k}$ is given by Proposition 1.5.

1.34. Let $r \geq 5$. How many ways are there to color the vertices with r colors in the following graphs such that adjacent vertices get different colors?



Hint: For the first graph, the universe X is the set of all r^4 ways to color the vertices. Associate with each edge e the set A_e of all colorings, which assign the same color to its ends, and apply the inclusion-exclusion formula (1.19).

1.35. Say that a permutation π on $[2n]$ has property P if for some $i \in [2n]$, $|\pi(i) - \pi(i+1)| = n$, where $i + 1$ is taken modulo 2. Show that, for each n, there are more permutations with property P than without it. *Hint*: Consider the sets $A_i = \{\pi \ : \ |\pi(i) - \pi(i+1)| = n\}$. Show that $|A_i| = 2n(2n-2)!$ and $A_i \cap A_{i+1} = \emptyset$.

1.36. Prove that for any two sets $I \subseteq J$,

$$\sum_{I \subseteq K \subseteq J} (-1)^{|K \setminus I|} = \begin{cases} 1, & \text{if } I = J \\ 0, & \text{if } I \neq J. \end{cases}$$

1.37. Prove the following *Bonferroni inequalities* for each even $k \geq 2$:

$$\sum_{\nu=1}^{k} (-1)^{\nu+1} \sum_{|I|=\nu} |A_I| \leq |\bigcup_{i=1}^{n} A_i| \leq \sum_{\nu=1}^{k+1} (-1)^{\nu+1} \sum_{|I|=\nu} |A_I|$$

where $A_I := \bigcap_{i \in I} A_i$. What about an odd k?

1.38. Let M be an $n \times n$ boolean matrix (with entries 0 and 1). A *covering* of M is a set $R_1, \ldots, R_t$ of rank-1 boolean matrices such that every 1-entry of M is a 1-entry in at least one of these matrices, and every 0-entry of M is a 0-entry in all these matrices. That is, M must be an entry-wise Or $M = \bigvee_{i=1}^{t} R_i$ of the R_i's. Let $t(A)$ be the smallest number of the R_i's in such a covering of M. For a boolean matrix $B \leq M$ (again, inequality is entry-wise), let $w_M(B)$ denote the largest possible number of 1-entries in B that can be covered by some all-1 submatrix R of M. (Note that R need not be a *submatrix* of B.) Set

$$\mu(M) = \max_{B \leq M} \frac{|B|}{w_M(B)},$$

where $|B|$ is the number of 1s in B. Prove that

$$\mu(M) \leq t(M) \leq \mu(M) \cdot \ln |M| + 1.$$

Hint: For the upper bound, consider a greedy covering $R_1, ..., R_t$ of M by all-1 submatrices: in the i-th step choose an all-1 submatrix $R_i \leq M$ covering the largest number of all yet uncovered 1s in M. Let $B_i \leq M$ be the submatrix containing all 1-entries of M that are left uncovered after the i-th step. Observe that $|B_i|/w_A(B_i) \leq \mu(A)$ for all i.

1.39. For a boolean matrix M and an integer $k \geq 1$, let $t_k(M)$ denote the smallest number t of rank-1 boolean matrices $R_1, \ldots, R_t$ in a covering of M with a restriction that $\sum_{i=1}^{t} R_i \leq kJ$, where J is an all-1 matrix. That is, we now require that no 1-entry of M is covered more than k times. Prove that then

$$\sum_{i=1}^{k} \binom{t}{i} \geq \mathrm{rk}(M).$$

Hint: For a subset $I \subseteq \{1, \ldots, t\}$, let R_I be a $(0,1)$ matrix with $R_I[x,y] = 1$ iff $R_i[x,y] = 1$ for all $i \in I$. Use the inclusion-exclusion principle to write M as

$$M = \sum_{I \neq \emptyset} (-1)^{|I|+1} R_I.$$

1.40. The *determinant* $\det(A)$ of an $n \times n$ matrix $A = (a_{ij})$ is a sum of $n!$ signed products $\pm a_{1i_1} a_{2i_2} \cdots a_{ni_n}$, where $(i_1, i_2, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$, the sign being $+1$ or -1, according to whether the number of *inversions* of $(i_1, i_2, \ldots, i_n)$ is even or odd. An inversion occurs when $i_r > i_s$ but $r < s$. Prove the following: let A be a matrix of even order n with 0s on the diagonal and arbitrary entries from $\{+1, -1\}$ elsewhere. Then $\det(A) \neq 0$.
Hint: Observe that for such matrices, $\det(A)$ is congruent modulo 2 to the number of derangements on n points, and show that for even n, the sum (1.22) is odd.

2. Advanced Counting

When properly applied, the (double) counting argument can lead to more subtle results than those discussed in the previous chapter.

2.1 Bounds on intersection size

How many r-element subsets of an n-element set can we choose under the restriction that no two of them share more than k elements? Intuitively, the smaller k is, the fewer sets we can choose. This intuition can be made precise as follows. (We address the optimality of this bound in Exercise 2.5.)

Lemma 2.1 (Corrádi 1969)**.** *Let $A_1, \ldots, A_N$ be r-element sets and X be their union. If $|A_i \cap A_j| \leq k$ for all $i \neq j$, then*

$$|X| \geq \frac{r^2 N}{r + (N-1)k}. \tag{2.1}$$

Proof. Just count. By (1.11), we have for each $i = 1, \ldots, N$,

$$\sum_{x \in A_i} d(x) = \sum_{j=1}^{N} |A_i \cap A_j| = |A_i| + \sum_{j \neq i} |A_i \cap A_j| \leq r + (N-1)k\,. \tag{2.2}$$

Summing over all sets A_i and using Jensen's inequality (1.15) we get

$$\sum_{i=1}^{N} \sum_{x \in A_i} d(x) = \sum_{x \in X} d(x)^2 \geq \frac{1}{n}\Big(\sum_{x \in X} d(x)\Big)^2 = \frac{1}{n}\Big(\sum_{i=1}^{n} |A_i|\Big)^2 = \frac{(Nr)^2}{n}\,.$$

Using (2.2) we obtain $(Nr)^2 \leq N \cdot |X|\,(r + (N-1)k)$, which gives the desired lower bound on $|X|$. □

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_2,

Given a family of sets $A_1, \dots, A_N$, their *average size* is

$$\frac{1}{N}\sum_{i=1}^{N}|A_i|.$$

The following lemma says that, if the average size of sets is large, then some two of them must share many elements.

Lemma 2.2. *Let X be a set of n elements, and let $A_1, \dots, A_N$ be subsets of X of average size at least n/w. If $N \geq 2w^2$, then there exist $i \neq j$ such that*

$$|A_i \cap A_j| \geq \frac{n}{2w^2}. \tag{2.3}$$

Proof. Again, let us just count. On the one hand, using Jensen's inequality (1.15) and equality (1.10), we obtain that

$$\sum_{x\in X} d(x)^2 \geq \frac{1}{n}\Big(\sum_{x\in X} d(x)\Big)^2 = \frac{1}{n}\Big(\sum_{i=1}^{N}|A_i|\Big)^2 \geq \frac{nN^2}{w^2}\,.$$

On the other hand, assuming that (2.3) is false and using (1.11) and (1.12) we would obtain

$$\begin{aligned}\sum_{x\in X} d(x)^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N}|A_i \cap A_j| = \sum_{i}|A_i| + \sum_{i\neq j}|A_i\cap A_j| \\ &< nN + \frac{nN(N-1)}{2w^2} = \frac{nN^2}{2w^2}\Big(1+\frac{2w^2}{N}-\frac{1}{N}\Big) \leq \frac{nN^2}{w^2},\end{aligned}$$

a contradiction. □

Lemma 2.2 is a very special (but still illustrative) case of the following more general result.

Lemma 2.3 (Erdős 1964b)**.** *Let X be a set of n elements $x_1, \dots, x_n$, and let $A_1, \dots, A_N$ be N subsets of X of average size at least n/w. If $N \geq 2kw^k$, then there exist $A_{i_1}, \dots, A_{i_k}$ such that $|A_{i_1} \cap \cdots \cap A_{i_k}| \geq n/(2w^k)$.*

The proof is a generalization of the one above and we leave it as an exercise (see Exercises 2.8 and 2.9).

2.2 Graphs with no 4-cycles

Let H be a fixed graph. A graph is H-free if it does not contain H as a subgraph. (Recall that a *subgraph* is obtained by deleting edges and vertices.) A typical question in graph theory is the following one:

How many edges can a H-free graph with n vertices have?

That is, one is interested in the maximum number $\mathrm{ex}(n, H)$ of edges in a H-free graph on n vertices. The graph H itself is then called a "forbidden subgraph."

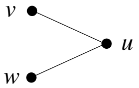
Let us consider the case when forbidden subgraphs are *cycles*. Recall that a cycle C_k of length k (or a k-cycle) is a sequence $v_0, v_1, \ldots, v_k$ such that $v_k = v_0$ and each subsequent pair v_i and v_{i+1} is joined by an edge.

If $H = C_3$, a triangle, then $\mathrm{ex}(n, C_3) \geq n^2/4$ for every even $n \geq 2$: a complete bipartite $r \times r$ graph $K_{r,r}$ with $r = n/2$ has no triangles but has $r^2 = n^2/4$ edges. We will show later that this is already optimal: any n-vertex graph with more than $n^2/4$ edges must contain a triangle (see Theorem 4.7). Interestingly, $\mathrm{ex}(n, C_4)$ is much smaller, smaller than $n^{3/2}$.

Theorem 2.4 (Reiman 1958). *If $G = (V, E)$ on n vertices has no 4-cycles, then*

$$|E| \leq \frac{n}{4}(1 + \sqrt{4n - 3}).$$

Proof. Let $G = (V, E)$ be a C_4-free graph with vertex-set $V = \{1, \ldots, n\}$, and $d_1, d_2, \ldots, d_n$ be the degrees of its vertices. We now count in two ways the number of elements in the following set S. The set S consists of all (ordered) pairs $(u, \{v, w\})$ such that $v \neq w$ and u is adjacent to both v and w in G. That is, we count all occurrences of "cherries"



in G. For each vertex u, we have $\binom{d_u}{2}$ possibilities to choose a 2-element subset of its d_u neighbors. Thus, summing over u, we find $|S| = \sum_{u=1}^{n} \binom{d_u}{2}$. On the other hand, the C_4-freeness of G implies that no pair of vertices $v \neq w$ can have more than one common neighbor. Thus, summing over all pairs we obtain that $|S| \leq \binom{n}{2}$. Altogether this gives

$$\sum_{i=1}^{n} \binom{d_i}{2} \leq \binom{n}{2}$$

or

$$\sum_{i=1}^{n} d_i^2 \leq n(n-1) + \sum_{i=1}^{n} d_i\,. \tag{2.4}$$

Now, we use the Cauchy–Schwarz inequality

$$\Big(\sum_{i=1}^{n} x_i y_i\Big)^2 \leq \Big(\sum_{i=1}^{n} x_i^2\Big)\Big(\sum_{i=1}^{n} y_i^2\Big)$$

with $x_i = d_i$ and $y_i = 1$, and obtain

$$\left(\sum_{i=1}^{n} d_i\right)^2 \leq n \sum_{i=1}^{n} d_i^2$$

and hence by (2.4)

$$\left(\sum_{i=1}^{n} d_i\right)^2 \leq n^2(n-1) + n \sum_{i=1}^{n} d_i \,.$$

Euler's theorem gives $\sum_{i=1}^{n} d_i = 2|E|$. Invoking this fact, we obtain

$$4|E|^2 \leq n^2(n-1) + 2n|E|$$

or

$$|E|^2 - \frac{n}{2}|E| - \frac{n^2(n-1)}{4} \leq 0\,.$$

Solving the corresponding quadratic equation yields the desired upper bound on $|E|$. □

Example 2.5 (Construction of dense C_4-free graphs)*.* The following construction shows that the bound of Theorem 2.4 is optimal up to a constant factor.

Let p be a prime number and take $V = (\mathbb{Z}_p \setminus \{0\}) \times \mathbb{Z}_p$, that is, vertices are pairs (a, b) of elements of a finite field with $a \neq 0$. We define a graph G on these vertices, where (a, b) and (c, d) are joined by an edge iff $ac = b + d$ (all operations modulo p). For each vertex (a, b), there are $p - 1$ solutions of the equation $ax = b + y$: pick any $x \in \mathbb{Z}_p \setminus \{0\}$, and y is uniquely determined. Thus, G is a $(p - 1)$-regular graph on $n = p(p - 1)$ vertices (some edges are loops). The number of edges in it is $n(p - 1)/2 = \Omega(n^{3/2})$.

To verify that the graph is C_4-free, take any two its vertices (a, b) and (c, d). The unique solution (x, y) of the system

$$\begin{cases} ax = b + y \\ cx = d + y \end{cases} \quad \text{is given by} \quad \begin{aligned} x &= (b - d)(a - c)^{-1} \\ 2y &= x(a + c) - b - d \end{aligned}$$

which is only defined when $a \neq c$, and has $x \neq 0$ only when $b \neq d$. Hence, if $a \neq c$ and $b \neq d$, then the vertices (a, b) and (c, d) have precisely one common neighbor, and have no common neighbors at all, if $a = c$ or $b = d$.

2.3 Graphs with no induced 4-cycles

Recall that an *induced subgraph* is obtained by deleting vertices together with all the edges incident to them (see Fig. 2.1).

Theorem 2.4 says that a graph cannot have many edges, unless it contains C_4 as a (not necessarily induced) subgraph. But what about graphs that

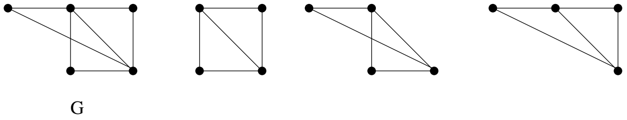


Fig. 2.1 Graph G contains several copies of C_4 as a subgraph, but none of them as an *induced* subgraph.

do not contain C_4 as an *induced* subgraph? Let us call such graphs *weakly C_4-free.*

Note that such graphs can already have many more edges. In particular, the complete graph K_n is weakly C_4-free: in any 4-cycle there are edges in K_n between non-neighboring vertices of C_4. Interestingly, any(!) dense enough weakly C_4-free graph must contain large complete subgraphs.

Let $\omega(G)$ denote the maximum number of vertices in a complete subgraph of G. In particular, $\omega(G) \leq 3$ for every C_4-free graph. In contrast, for *weakly* C_4-free graphs we have the following result, due to Gyárfás, Hubenko and Solymosi (2002).

Theorem 2.6. *If an n-vertex graph $G = (V, E)$ is weakly C_4-free, then*

$$\omega(G) \geq 0.4 \frac{|E|^2}{n^3} .$$

The proof of Theorem 2.6 is based on a simple fact, relating the average degree with the minimum degree, as well as on two facts concerning independent sets in weakly C_4-free graphs.

For a graph $G = (V, E)$, let $e(G) = |E|$ denote the number of its edges, $d_{\min}(G)$ the smallest degree of its vertices, and $d_{\text{ave}}(G) = 2e(G)/|V|$ the average degree. Note that, by Euler's theorem, $d_{\text{ave}}(G)$ is indeed the sum of all degrees divided by the total number of vertices.

Proposition 2.7. *Every graph G has an induced subgraph H with*

$$d_{\text{ave}}(H) \geq d_{\text{ave}}(G) \qquad \textit{and} \qquad d_{\min}(H) \geq \frac{1}{2}\, d_{\text{ave}}(G) .$$

Proof. We remove vertices one-by-one. To avoid the danger of ending up with the empty graph, let us remove a vertex $v \in V$ if this does not decrease the average degree $d_{\text{ave}}(G)$. Thus, we should have

$$d_{\text{ave}}(G - v) = \frac{2(e(G) - d(v))}{|V| - 1} \geq d_{\text{ave}}(G) = \frac{2e(G)}{|V|}$$

which is equivalent to $d(v) \leq d_{\text{ave}}(G)/2$. So, when we stick, each vertex in the resulting graph H has minimum degree at least $d_{\text{ave}}(G)/2$. □

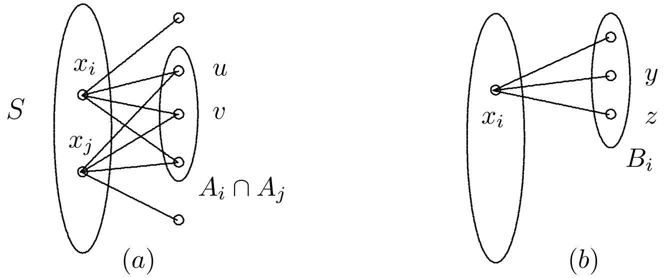


Fig. 2.2 (a) If u and v were non-adjacent, we would have an induced 4-cycle $\{x_i, x_j, u, v\}$. (b) If y and z were non-adjacent, then $(S \setminus \{x_i\}) \cup \{y, z\}$ would be a larger independent set.

Recall that a set of vertices in a graph is *independent* if no two of its vertices are adjacent. Let $\alpha(G)$ denote the largest number of vertices in such a set.

Proposition 2.8. *For every weakly C_4-free graph G on n vertices, we have*

$$\omega(G) \geq \frac{n}{\binom{\alpha(G)+1}{2}}.$$

Proof. Fix an independent set $S = \{x_1, \ldots, x_\alpha\}$ with $\alpha = \alpha(G)$. Let A_i be the set of neighbors of x_i in G, and B_i the set of vertices whose only neighbor in S is x_i. Consider the family $\mathcal{F}$ consisting of all α sets $\{x_i\} \cup B_i$ and $\binom{\alpha}{2}$ sets $A_i \cap A_j$. We claim that:

(i) each member of $\mathcal{F}$ forms a clique in G, and
(ii) the members of $\mathcal{F}$ cover all vertices of G.

The sets $A_i \cap A_j$ are cliques because G is weakly C_4-free: Any two vertices $u \neq v \in A_i \cap A_j$ must be joined by an edge, for otherwise $\{x_i, x_j, u, v\}$ would form a copy of C_4 as an induced subgraph. The sets $\{x_i\} \cup B_i$ are cliques because S is a maximal independent set: Otherwise we could replace x_i in S by any two vertices from B_i. By the same reason (S being a *maximal* independent set), the members of $\mathcal{F}$ must cover all vertices of G: If some vertex v were not covered, then $S \cup \{v\}$ would be a larger independent set.

Claims (i) and (ii), together with the averaging principle, imply that

$$\omega(G) \geq \frac{n}{|\mathcal{F}|} = \frac{n}{\alpha + \binom{\alpha}{2}} = \frac{n}{\binom{\alpha+1}{2}}. \qquad \square$$

Proposition 2.9. *Let G be a weakly C_4-free graph on n vertices, and $d = d_{\min}(G)$. Then, for every $t \leq \alpha(G)$,*

$$\omega(G) \geq \frac{d \cdot t - n}{\binom{t}{2}}.$$

Proof. Take an independent set $S = \{x_1, \ldots, x_t\}$ of size t and let A_i be the set of neighbors of x_i in G. Let m be the maximum of $|A_i \cap A_j|$ over all $1 \le i < j \le t$. We already know that each $A_i \cap A_j$ must form a clique; hence, $\omega(G) \ge m$. On the other hand, by the Bonferroni inequality (Exercise 1.37) we have that

$$n \ge \left| \bigcup_{i=1}^{t} A_i \right| \ge td - \sum_{i<j} |A_i \cap A_j| \ge td - \binom{t}{2} m ,$$

from which the desired lower bound on $\omega(G)$ follows. □

Now we are able to prove Theorem 2.6.

Proof of Theorem 2.6. Let a be the average degree of G; hence, $a = 2|E|/n$. By Proposition 2.7, we know that G has an induced subgraph of average degree $\ge a$ and minimum degree $\ge a/2$. So, we may assume w.l.o.g. that the graph G itself has these two properties. We now consider the two possible cases.

If $\alpha(G) \ge 4n/a$, then we apply Proposition 2.9 with* $t = 4n/a$ and obtain

$$\omega(G) \ge \frac{(a/2) \cdot t - n}{\binom{t}{2}} = \frac{n}{\binom{4n/a}{2}} .$$

If $\alpha(G) \le 4n/a$, then we apply Proposition 2.8 and obtain

$$\omega(G) \ge \frac{n}{\binom{\alpha(G)+1}{2}} \ge \frac{n}{\binom{4n/a+1}{2}} .$$

In both cases we obtain

$$\omega(G) \ge \frac{n}{\binom{4n/a+1}{2}} = \frac{a^2}{8n+2a} \ge 0.1 \frac{a^2}{n} .$$ □

2.4 Zarankiewicz's problem

At most how many 1s can an $n \times n$ 0-1 matrix contain if it has no $a \times b$ submatrix whose entries are all 1s? Zarankiewicz (1951) raised the problem of the estimation of this number for $a = b = 3$ and $n = 4, 5, 6$ and the general problem became known as *Zarankiewicz's problem.*

It is worth reformulating this problem in terms of bipartite graphs. A bipartite graph with parts of size n is a triple $G = (V_1, V_2, E)$, where V_1 and V_2 are disjoint n-element sets of *vertices* (or *nodes*), and $E \subseteq V_1 \times V_2$ is the set of *edges.* We say that the graph contains an $a \times b$ *clique* if there exist an

* For simplicity, we ignore ceilings and floors.

a-element subset $A \subseteq V_1$ and a b-element subset $B \subseteq V_2$ such that $A \times B \subseteq E$. (Note that an $a \times b$ clique is not the same as a $b \times a$ clique, unless $a = b$.)

Let $k_a(n)$ be the minimal integer k such that *any* bipartite graph with parts of size n and more than k edges contains at least one $a \times a$ clique. Using the probabilistic argument, it can be shown (see Exercise 20.6) that

$$k_a(n) \geq c \cdot n^{2-2/a},$$

where $c > 0$ is a constant, depending only on a. It turns out that this bound is not very far from the best possible, and this can be proved using the double counting argument. The result is essentially due to Kővári, Sós and Turán (1954). For $a = 2$, a lower bound $k_2(n) \leq 3n^{3/2}$ was proved by Erdős (1938). He used this to prove that, if a set $A \subseteq [n]$ is such that the products of any two of its different members are different, then $|A| \leq \pi(n) + O(n^{3/4})$, where $\pi(n)$ is the number of primes not exceeding n.

Theorem 2.10. *For all natural numbers $n \geq a \geq 2$ we have*

$$k_a(n) \leq (a-1)^{1/a} n^{2-1/a} + (a-1)n.$$

Proof. The proof is a direct generalization of a double counting argument we used in the proof of Theorem 2.4. Our goal is to prove the following: let $G = (V_1, V_2, E)$ be a bipartite graph with parts of size n, and suppose that G *does not* contain an $a \times a$ clique; then $|E| \leq (a-1)^{1/a} n^{2-1/a} + (a-1)n$.

By a *star* in the graph G we will mean a set of any a of its edges incident with one vertex $x \in V_1$, i.e., a set of the form

$$S(x, B) := \{(x, y) \in E \ : \ y \in B\},$$

where $B \subseteq V_2$, $|B| = a$. Let $\varDelta$ be the total number of such stars in G. We may count the stars $S(x, B)$ in two ways, by fixing either the vertex x or the subset B.

For a fixed subset $B \subseteq V_2$, with $|B| = a$, we can have at most $a - 1$ stars of the form $S(x, B)$, because otherwise we would have an $a \times a$ clique in G. Thus,

$$\varDelta \leq (a-1) \cdot \binom{n}{a}. \tag{2.5}$$

On the other hand, for a fixed vertex $x \in V_1$, we can form $\binom{d(x)}{a}$ stars $S(x, B)$, where $d(x)$ is the degree of vertex x in G (i.e., the number of vertices adjacent to x). Therefore,

$$\sum_{x \in V_1} \binom{d(x)}{a} \leq (a-1) \cdot \binom{n}{a}. \tag{2.6}$$

We are going to estimate the left-hand side from below using Jensen's inequality. Unfortunately, the function $\binom{x}{a} = x(x-1)\cdots(x-a+1)/a!$ is convex only for $x \geq a - 1$. But we can set $f(z) := \binom{x}{a}$ if $x \geq a - 1$, and $f(x) := 0$

otherwise. Then Jensen's inequality (1.14) (with $\lambda_x = 1/n$ for all $x \in V_1$) yields

$$\sum_{x\in V_1}\binom{d(x)}{a} \geq \sum_{x\in V_1} f(d(x)) \geq n\cdot f\Big(\sum_{x\in V_1} d(x)/n\Big) = n\cdot f(|E|/n)\,.$$

If $|E|/n < a-1$, there is nothing to prove. So, we can suppose that $|E|/n \geq a-1$. Then we have that

$$n\cdot\binom{|E|/n}{a} = n\cdot f(|E|/n) \leq \sum_{x\in V_1}\binom{d(x)}{a} \leq (a-1)\binom{n}{a}\,.$$

Expressing the binomial coefficients as quotients of factorials, this inequality implies

$$n\left(|E|/n-(a-1)\right)^a \leq (a-1)n^a,$$

and therefore $|E|/n \leq (a-1)^{1/a}n^{1-1/a}+a-1$, from which the desired upper bound on $|E|$ follows. □

The theorem above says that any bipartite graph with many edges has large cliques. In order to destroy such cliques we can try to remove some of their vertices. We would like to remove as few vertices as possible. Just how few says the following result.

Theorem 2.11 (Ossowski 1993)**.** *Let $G=(V_1,V_2,E)$ be a bipartite graph with no isolated vertices, $|E|<(k+1)r$ edges and $d(y)\leq r$ for all $y\in V_2$. Then we can delete at most k vertices from V_1 so that the resulting graph has no $(r-a+1)\times a$ clique for $a=1,2,\ldots,r$.*

For a vertex x, let $N(x)$ denote the set of its neighbors in G, that is, the set of all vertices adjacent to x; hence, $|N(x)|$ is the degree $d(x)$ of x. We will use the following lemma relating the degree to the total number of vertices.

Lemma 2.12. *Let (X,Y,E) be a bipartite graph with no isolated vertices, and $f : Y \to [0,\infty)$ be a function. If the inequality $d(y)\leq d(x)\cdot f(y)$ holds for each edge $(x,y)\in E$, then $|X|\leq\sum_{y\in Y} f(y)$.*

Proof. By double counting,

$$\begin{aligned}|X| &= \sum_{x\in X}\sum_{y\in N(x)}\frac{1}{d(x)} \leq \sum_{x\in X}\sum_{y\in N(x)}\frac{f(y)}{d(y)}\\ &= \sum_{y\in Y}\sum_{x\in N(y)}\frac{f(y)}{d(y)} = \sum_{y\in Y}\frac{f(y)}{d(y)}\cdot|N(y)| = \sum_{y\in Y} f(y). \quad \square\end{aligned}$$

Proof of Theorem 2.11. (Due to F. Galvin 1997). For a set of vertices $Y\subseteq V_2$, let $N(Y) := \bigcap_{y\in Y} N(y)$ denote the set of all its *common neighbors* in G, that is, the set of all those vertices in V_1 which are joined to each vertex of Y;

hence $|N(Y)| \leq r$ for all $Y \subseteq V_2$. Let $X \subseteq V_1$ be a minimal set with the property that $|N(Y) \setminus X| \leq r - |Y|$ whenever $Y \subseteq V_2$ and $1 \leq |Y| \leq r$. Put otherwise, X is a minimal set of vertices in V_1, the removal of which leads to a graph without $(r-a+1) \times a$ cliques, for all $a = 1, \ldots, r$.

Our goal is to show that $|X| \leq k$.

Note that, for each $x \in X$ we can choose $Y_x \subseteq V_2$ so that $1 \leq |Y_x| \leq r$, $x \in N(Y_x)$ and

$$|N(Y_x) \setminus X| = r - |Y_x|;$$

otherwise X could be replaced by $X \setminus \{x\}$, contradicting the minimality of X. We will apply Lemma 2.12 to the bipartite graph $G' = (X, V_2, F)$, where

$$F = \{(x,y) \;:\; y \in Y_x\}.$$

All we have to do is to show that the hypothesis of the lemma is satisfied by the function (here $N(y)$ is the set of neighbors of y in the original graph G):

$$f(y) := \frac{|N(y)|}{r},$$

because then

$$|X| \leq \sum_{y \in V_2} f(y) = \frac{1}{r} \sum_{y \in V_2} |N(y)| = \frac{|E|}{r} < k+1.$$

Consider an edge $(x,y) \in F$; we have to show that $d(y) \leq d(x) \cdot f(y)$, where

$$d(x) = |Y_x| \;\text{ and } d(y) = |\{x \in X \;:\; y \in Y_x\}|$$

are the degrees of x and y in the graph $G' = (X, V_2, F)$. Now, $y \in Y_x$ implies $N(Y_x) \subseteq N(y)$, which in its turn implies

$$|N(y) \setminus X| \geq |N(Y_x) \setminus X| = r - |Y_x|;$$

hence

$$\begin{aligned} d(y) &\leq |N(y) \cap X| = |N(y)| - |N(y) \setminus X| \\ &\leq |N(y)| - r + |Y_x| = r \cdot f(y) - r + d(x), \end{aligned}$$

and so

$$\begin{aligned} d(x) \cdot f(y) - d(y) &\geq d(x) \cdot f(y) - r \cdot f(y) + r - d(x) \\ &= (r - d(x)) \cdot (1 - f(y)) \geq 0\,. \quad \square \end{aligned}$$

2.5 Density of 0-1 matrices

Let H be an $m \times n$ 0-1 matrix. We say that H is *α-dense* if at least an α-fraction of all its mn entries are 1s. Similarly, a row (or column) is *α-dense* if at least an α-fraction of all its entries are 1s.

The next result says that any dense 0-1 matrix must either have one "very dense" row or there must be many rows which are still "dense enough."

Lemma 2.13 (Grigni and Sipser 1995)**.** *If H is 2α-dense then either*
(a) *there exists a row which is $\sqrt{\alpha}$-dense, or*
(b) *at least $\sqrt{\alpha} \cdot m$ of the rows are α-dense.*

Note that $\sqrt{\alpha}$ is larger than α when $\alpha < 1$.

Proof. Suppose that the two cases do not hold. We calculate the density of the entire matrix. Since (b) does not hold, less than $\sqrt{\alpha} \cdot m$ of the rows are α-dense. Since (a) does not hold, each of these rows has less than $\sqrt{\alpha} \cdot n$ 1s; hence, the fraction of 1s in α-dense rows is strictly less than $(\sqrt{\alpha})(\sqrt{\alpha}) = \alpha$. We have at most m rows which are not α-dense, and each of them has less than αn ones. Hence, the fraction of 1s in these rows is also less than α. Thus, the total fraction of 1s in the matrix is less than 2α, contradicting the 2α-density of H. □

Now consider a slightly different question: if H is α-dense, how many of its rows *or* columns are "dense enough"? The answer is given by the following general estimate due to Johan Håstad. This result appeared in the paper of Karchmer and Wigderson (1990) and was used to prove that the graph connectivity problem cannot be solved by monotone circuits of logarithmic depth.

Suppose that our universe is a Cartesian product $A = A_1 \times \cdots \times A_k$ of some finite sets $A_1, \ldots, A_k$. Hence, elements of A are strings $\boldsymbol{a} = (a_1, \ldots, a_k)$ with $a_i \in A_i$. Fix now a subset of strings $H \subseteq A$ and a point $b \in A_i$. The *degree* of b in H is the number $d_H(b) = |\{\boldsymbol{a} \in H : a_i = b\}|$ of strings in H whose i-th coordinate is b.

Say that a point $b \in A_i$ from the i-th set is *popular* in H if its degree $d_H(b)$ is at least a $1/2k$ fraction of the average degree of an element in A_i, that is, if

$$d_H(b) \geq \frac{1}{2k}\frac{|H|}{|A_i|}\,.$$

Let $P_i \subseteq A_i$ be the set of all popular points in the i-th set A_i, and consider the Cartesian product of these sets:

$$P := P_1 \times P_2 \times \cdots \times P_k\,.$$

Lemma 2.14 (Håstad)**.** $|P| > \frac{1}{2}|H|$.

Proof. It is enough to show that $|H \setminus P| < \frac{1}{2}|H|$. For every non-popular point $b \in A_i$, we have that

$$|\{\boldsymbol{a} \in H \ : \ a_i = b\}| < \frac{1}{2k}\frac{|H|}{|A_i|}\,.$$

Since the number of non-popular points in each set A_i does not exceed the total number of points $|A_i|$, we obtain

$$\begin{aligned} |H \setminus P| &\le \sum_{i=1}^{k} \sum_{b \notin P_i} |\{\boldsymbol{a} \in H \ : \ a_i = b\}| < \sum_{i=1}^{k} \sum_{b \notin P_i} \frac{1}{2k}\frac{|H|}{|A_i|} \\ &\le \sum_{i=1}^{k} \frac{1}{2k}|H| = \frac{1}{2}|H|\,. \quad \square \end{aligned}$$

Corollary 2.15. *In any* 2α*-dense* 0-1 *matrix* H *either a* $\sqrt{\alpha}$*-fraction of its rows or a* $\sqrt{\alpha}$*-fraction of its columns (or both) are* $(\alpha/2)$*-dense.*

Proof. Let H be an $m \times n$ matrix. We can view H as a subset of the Cartesian product $[m] \times [n]$, where $(i, j) \in H$ iff the entry in the i-th row and j-th column is 1. We are going to apply Lemma 2.14 with $k = 2$. We know that $|H| \ge 2\alpha mn$. So, if P_1 is the set of all rows with at least $\frac{1}{4}|H|/|A_1| = \alpha n/2$ ones, and P_2 is the set of all columns with at least $\frac{1}{4}|H|/|A_2| = \alpha m/2$ ones, then Lemma 2.14 implies that

$$\frac{|P_1|}{m} \cdot \frac{|P_2|}{n} \ge \frac{1}{2}\frac{|H|}{mn} \ge \frac{1}{2} \cdot \frac{2\alpha mn}{mn} = \alpha\,.$$

Hence, either $|P_1|/m$ or $|P_2|/n$ must be at least $\sqrt{\alpha}$, as claimed. □

2.6 The Lovász–Stein theorem

This theorem was used by Stein (1974) and Lovász (1975) in studying some combinatorial covering problems. The advantage of this result is that it can be used to get existence results for some combinatorial problems using constructive methods rather than probabilistic methods.

Given a family $\mathcal{F}$ of subsets of some finite set X, its *cover number* of $\mathcal{F}$, $\mathrm{Cov}\,(\mathcal{F})$, is the minimum number of members of $\mathcal{F}$ whose union covers all points (elements) of X.

Theorem 2.16. *If each member of* $\mathcal{F}$ *has at most* a *elements, and each point* $x \in X$ *belongs to at least* v *of the sets in* $\mathcal{F}$*, then*

$$\mathrm{Cov}\,(\mathcal{F}) \le \frac{|\mathcal{F}|}{v}(1 + \ln a)\,.$$

Proof. Let $N = |X|$, $M = |\mathcal{F}|$ and consider the $N \times M$ 0-1 matrix $A = (a_{x,i})$, where $a_{x,i} = 1$ iff $x \in X$ belongs to the i-th member of $\mathcal{F}$. By our assumption, each row of A has at least v ones and each column at most a ones. By double counting, we have that $Nv \geq Ma$, or equivalently,

$$\frac{M}{v} \leq \frac{N}{a}\,. \tag{2.7}$$

Our goal is to show that then A must contain an $N \times K$ submatrix C with no all-0 rows and such that

$$K \leq N/a + (M/v)\ln a \leq (M/v)(1 + \ln a)\,.$$

We describe a constructive procedure for producing the desired submatrix C. Let $A_a = A$ and define A'_a to be any maximal set of columns from A_a whose supports[†] are pairwise disjoint and whose columns each have a ones. Let $K_a = |A'_a|$. Discard from A_a the columns of A'_a and any row with a one in A'_a. We are left with a $k_a \times (M - K_a)$ matrix A_{a-1}, where $k_a = N - aK_a$. Clearly, the columns of A_{a-1} have at most $a-1$ ones (indeed, otherwise such a column could be added to the previously discarded set, contradicting its maximality). We continue by doing to A_{a-1} what we did to A_a. That is we define A'_{a-1} to be any maximal set of columns from A_{a-1} whose supports are pairwise disjoint and whose columns each have $a - 1$ ones. Let $K_{a-1} = |A'_{a-1}|$. Then discard from A_{a-1} the columns of A'_{a-1} and any row with a one in A'_{a-1} getting a $k_{a-1} \times (M - K_a - K_{a-1})$ matrix A_{a-2}, where $k_{a-1} = N - aK_a - (a-1)K_{a-1}$.

The process will terminate after at most a steps (when we have a matrix containing only zeros). The union of the columns of the discarded sets form the desired submatrix C with $K = \sum_{i=1}^{a} K_i$. The first step of the algorithm gives $k_a = N - aK_a$, which we rewrite, setting $k_{a+1} = N$, as

$$K_a = \frac{k_{a+1} - k_a}{a}\,.$$

Analogously,

$$K_i = \frac{k_{i+1} - k_i}{i} \qquad \text{for } i = 1, \ldots, a.$$

Now we derive an upper bound for k_i by counting the number of ones in A_{i-1} in two ways: every row of A_{i-1} contains at least v ones, and every column at most $i - 1$ ones, thus

$$vk_i \leq (i-1)(M - K_a - \cdots - K_{i+1}) \leq (i-1)M\,,$$

or equivalently,

$$k_i \leq \frac{(i-1)M}{v}\,.$$

[†] The *support* of a vector is the set of its nonzero coordinates.

So,

$$\begin{aligned} K &= \sum_{i=1}^{a} K_i = \sum_{i=1}^{a} \frac{k_{i+1}-k_i}{i} \\ &= \frac{k_{a+1}}{a} + \frac{k_a}{a(a-1)} + \frac{k_{a-1}}{(a-1)(a-2)} + \cdots + \frac{k_2}{2\cdot 1} - k_1 \\ &\le \frac{N}{a} + \frac{M}{v}\left(\frac{1}{a} + \frac{1}{a-1} + \cdots + \frac{1}{2}\right) \le \frac{N}{a} + \frac{M}{v}\ln a\,. \end{aligned}$$

The last inequality here follows because $1 + 1/2 + 1/3 + \cdots + 1/n$ is the n-th harmonic number which is known to lie between $\ln n$ and $\ln n + 1$. Together with (2.7), this yields $K \le (M/v)(1 + \ln a)$, as desired. □

The advantage of this proof is that it can be turned into a simple greedy algorithm which constructs the desired $N \times K$ submatrix A' with column-set C, $|C| = K$:

```
1. Set $C := \emptyset$ and $A' := A$.
2. While $A'$ has at least one row do:
     - find a column $c$ in $A'$ having a maximum number of ones;
     - delete all rows of $A'$ that contain a 1 in column $c$;
     - delete column $c$ from $A'$;
     - set $C := C \cup \{c\}$.
```

2.6.1 Covering designs

An (n, k, l) *covering design* is a family $\mathcal{F}$ of k-subsets of an n-element set (called *blocks*) such that every l-subset is contained in at least one of these blocks. Let $M(n, k, l)$ denote the minimal cardinality of such a design. A simple counting argument (Exercise 1.26) shows that $M(n, k, l) \ge \binom{n}{l}/\binom{k}{l}$.

In 1985, Rödl proved a long-standing conjecture of Erdős and Hanani that for fixed k and l, coverings of size $\binom{n}{l}/\binom{k}{l}(1 + o(1))$ exist. Rödl used non-constructive probabilistic arguments. We will now use the Lovász–Stein theorem to show how to *construct* an (n, k, l) covering design with only $\ln\binom{k}{l}$ times more blocks. This is not as sharp as Rödl's celebrated result, but it is constructive. A polynomial-time covering algorithm, achieving Rödl's bound, was found by Kuzjurin (2000).

Theorem 2.17. $M(n, k, l) \le \binom{n}{l}/\binom{k}{l}\left[1 + \ln\binom{k}{l}\right]$.

Proof. Let $X = (x_{S,T})$ be an $N \times M$ 0-1 matrix with $N = \binom{n}{l}$ and $M = \binom{n}{k}$. Rows of X are labeled by l-element subsets $S \subseteq [n]$, columns by k-element subsets $T \subseteq [n]$, and $x_{S,T} = 1$ iff $S \subseteq T$. Note that each row contains exactly $v = \binom{n-l}{k-l}$ ones, and each column contains exactly $a = \binom{k}{l}$ ones.

By the Lovász–Stein theorem, there is an $N \times K$ submatrix X' such that X' does not contain an all-0 row and

$$K \leq (M/v)(1+\ln a) = \binom{n}{k} \Big/ \binom{n-l}{k-l}\left[1+\ln\binom{k}{l}\right]$$
$$= \binom{n}{l} \Big/ \binom{k}{l}\left[1+\ln\binom{k}{l}\right],$$

as $\binom{n}{l}\binom{n-l}{k-l} = \binom{n}{k}\binom{k}{l}$ (see Exercise 1.12). By the definition of X and the property of X' (no all-0 row), the k-subsets that correspond to the columns of X' form an (n,k,l) covering design. □

Exercises

2.1. Let $A_1, \ldots, A_m$ be subsets of an n-element set such that $|A_i \cap A_j| \leq t$ for all $i \neq j$. Prove that $\sum_{i=1}^{m} |A_i| \leq n + t \cdot \binom{m}{2}$.

2.2. Let $A = (a_{ij})$ be an $n \times n$ matrix ($n \geq 4$ even). The matrix is filled with integers and each integer appears exactly twice. Show that there exists a permutation π of $\{1, \ldots, n\}$ such that all the numbers $a_{i,\pi(i)}$, $i = 1, \ldots, n$ are distinct. (Such a permutation π is also called a *Latin transversal* of A.) *Hint*: Look at how many pairs of entries are "bad," i.e., contain the same number, and show that strictly less than $n!$ of all permutations can go through such pairs.

2.3. Let $\mathcal{F}$ be a family of m subsets of a finite set X. For $x \in X$, let $p(x)$ be the number of pairs (A, B) of sets $A, B \in \mathcal{F}$ such that either $x \in A \cap B$ or $x \notin A \cup B$. Prove that $p(x) \geq m^2/2$ for every $x \in X$. *Hint*: Let $d(x)$ be the degree of x in $\mathcal{F}$, and observe that $p(x) = d(x)^2 + (m - d(x))^2$.

2.4. Let $\mathcal{F}$ be a family of nonempty subsets of a finite set X that is closed under union (i.e., $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$). Prove or give a counterexample: there exists $x \in X$ such that $d(x) \geq |\mathcal{F}|/2$. (Open conjecture, due to Peter Frankl.)

2.5. A *projective plane* of order $r-1$ is a family of $n = r^2 - r + 1$ r-element subsets (called *lines*) of an n-element set of points such that each two lines intersect at precisely one point and each point belongs to precisely r lines (cf. Sect. 12.4). Use this family to show that the bound given by Corrádi's lemma (Lemma 2.1) is optimal.

2.6. Theorem 2.10 gives a sufficient condition for a bipartite graph with parts of the same size n to contain an $a \times a$ clique. Extend this result to not necessarily balanced graphs. Let $k_{a,b}(m,n)$ be the minimal integer k such that any bipartite graph with parts of size m and n and more than k edges contains at least one $a \times b$ clique. Prove that for any $0 \leq a \leq m$ and $0 \leq b \leq n$,

$$k_{a,b}(m,n) \leq (a-1)^{1/b} n m^{1-1/b} + (b-1)m.$$

2.7. (Paturi–Zane 1998). Extend Theorem 2.10 to r-partite graphs as follows. An r-partite m-clique is a Cartesian product $V_1 \times V_2 \times \cdots \times V_r$ of m-element sets $V_1, \ldots, V_r$. An r-partite graph with parts of size m is a subset E of an r-partite m-clique. Let $\mathrm{ex}(m, r, 2)$ denote the maximum size $|E|$ of such a graph E which does not contain an r-partite 2-clique. Erdős (1959, 1964b) proved that

$$cm^{r-r/2^{r-1}} \leq \mathrm{ex}(m, r, 2) \leq m^{r-1/2^{r-1}},$$

where $c = c(r) > 0$ is a constant depending only on r. A slightly weaker upper bound $\mathrm{ex}(m, r, 2) < 2m^{r-1/2^{r-1}}$ can be derived from Lemma 2.2. Show how to do this. *Hint*: Argue by induction on r. For the induction step take $X = V_1 \times \cdots \times V_{r-1}$ and consider m subsets $A_v = \{x \in X : (x, v) \in E\}$ with $v \in V_r$. Apply Lemma 2.2 with $n = m^{r-1}$, $N = m$ and $w = \frac{1}{2}m^{1/2^{r-1}}$, to obtain a pair of points $u \neq v \in V_k$ for which the graph $E' = A_u \cap A_v$ is large enough, and use the induction hypothesis.

2.8. Let $\mathcal{F} = \{A_1, \ldots, A_N\}$ be a family of subsets of some set X. Use (1.11) to prove that for every $1 \leq s \leq N$,

$$\sum_{x \in X} d(x)^s = \sum_{(i_1, i_2, \ldots, i_s)} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}|,$$

where the last sum is over all s-tuples $(i_1, i_2, \ldots, i_s)$ of (not necessarily distinct) indices.

2.9. Use the previous exercise and the argument of Lemma 2.2 to prove Lemma 2.3.

2.10. Let $A_1, \ldots, A_N$ be subsets of some n-element set X, and suppose that these sets have average size at least αn. Show that for every $s \leq (1 - \epsilon)\alpha N$ with $0 < \epsilon < 1$, there are indices $i_1, i_2, \ldots, i_s$ such that

$$|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}| \geq (\epsilon\alpha)^s n.$$

Hint: Consider the bipartite graph $G = (X, V, E)$ where $V = \{1, \ldots, N\}$, and $(x, i) \in E$ if and only if $x \in A_i$. Observe that $|E| \geq \alpha n N$ and argue as in the proof of Theorem 2.10.

2.11. Prove the following very useful averaging principle for partitions. Let $X = A_1 \cup A_2 \cup \cdots \cup A_m$ be a partition of a finite set X into m mutually disjoint sets (blocks), and $a = \sum_{i=1}^{m} |A_i|/m$ be the average size of a block in this partition. Show that for every $1 \leq b \leq a$, at least $(1 - 1/b)|X|$ elements of X belong to blocks of size at least a/b. How many elements of X belong to blocks of size at most ab? *Hint*: $m \cdot (a/b) \leq |X|/b$.

2.12. Let $A_1, \ldots, A_r$ be a sequence of (not necessarily distinct) subsets of an n-element set X such that each set has size n/s and each element $x \in X$ belongs to least one and to at most k of them; hence $r \leq ks$. Let $K := \sum_{i=0}^{k} \binom{r}{i}$ and assume that $s > 2k$. Prove that there exist two disjoint subsets X_1 and X_2 of X such that $|X_i| \geq n/(2K)$ for both $i = 1, 2$, and none of the sets $A_1, \dots, A_r$ contains points from both sets X_1 and X_2. *Hint*: Associate with each $x \in X$ its *trace* $T(x) = \{i : x \in A_i\}$ and partition the elements of X according to their traces. Use the previous exercise to show that at least $n/2$ elements belong to blocks of size at least $n/(2K)$. Show that some two of these elements x and y must have *disjoint* traces, $T(x) \cap T(y) = \emptyset$.

2.13. Let $X = A_1 \cup A_2 \cup \cdots \cup A_m$ be a partition of a finite set X into mutually disjoint blocks. Given a subset $Y \subseteq X$, we obtain its partition $Y = B_1 \cup B_2 \cup \cdots \cup B_m$ into blocks $B_i = A_i \cap Y$. Say that a block B_i is λ-*large* if $|B_i|/|A_i| \geq \lambda \cdot |Y|/|X|$. Show that, for every $\lambda > 0$, at least $(1-\lambda) \cdot |Y|$ elements of Y belong to λ-large blocks.

2.14. Given a family $S_1, \dots, S_n$ of subsets of $V = \{1, \dots, n\}$, its *intersection graph* $G = (V, E)$ is defined by: $\{i, j\} \in E$ if and only if $S_i \cap S_j \neq \emptyset$. Suppose that: (i) the sets have average size at least r, and (ii) the average size of their pairwise intersections does not exceed k. Show that $|E| \geq \frac{n}{k} \cdot \binom{r}{2}$. *Hint*: Consider the sum $\sum_{i<j} |S_i \cap S_j|$.

2.15. Let H be a 2α-dense 0-1 matrix. Prove that at least an $\alpha/(1-\alpha)$ fraction of its rows must be α-dense.

2.16. (Alon 1986). Let S be a set of strings of length n over some alphabet. Suppose that every two strings of S differ in at least d coordinates. Let k be such that $d > n(1 - 1/\binom{k}{2})$. Show that any k distinct strings $v_1, \dots, v_k$ of S attain k distinct values in at least one coordinate. *Hint*: Assume the opposite and count the sum of distances between the $\binom{k}{2}$ pairs of v_i's.

3. Probabilistic Counting

Roughly speaking, the *probabilistic method* works as follows: trying to prove that an object with certain properties exists, one defines an appropriate probability space of objects and shows that a randomly chosen element of this space has the desired properties with a positive probability. A prototype of this method is the following averaging (counting) argument:

If $x_1, \ldots, x_n \in \mathbb{R}$ *and*

$$\frac{x_1 + \cdots + x_n}{n} \geq a \tag{3.1}$$

then for some j

$$x_j \geq a. \tag{3.2}$$

The usefulness of the method lies in the fact that the average (3.1) is often easier to compute than to exhibit a specific x_j for which (3.2) can be proved to hold.

The goal of this chapter is to demonstrate the probabilistic method on simple examples (more impressive applications will be given in Part IV devoted to this method). In its simplest applications, probabilistic argument can be replaced by a straightforward counting, "counting with weights." However, as soon as one gets away from the simplest examples, the heart and soul of the method is the probabilistic point of view rather than the act of counting.

3.1 Probabilistic preliminaries

We briefly recall some basic definitions of (discrete) probability.

A *finite probability space* consists of a finite set Ω and a function (called also *probability distribution*) $\Pr : \Omega \to [0, 1]$, such that $\sum_{x \in \Omega} \Pr[x] = 1$. A probability space is a representation of a random experiment, where we choose a member of Ω at random and $\Pr[x]$ is the probability that x is chosen. Subsets $A \subseteq \Omega$ are called *events*. The probability of an event is

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science. An EATCS Series, DOI 10.1007/978-3-642-17364-6_3,

defined by $\Pr[A] := \sum_{x \in A} \Pr[x]$, i.e., the probability that a member of A is chosen.

We call Ω the *domain* (or a *sample space*) and we call Pr a *probability distribution*. The most common probability distribution is the *uniform distribution*, which is defined as $\Pr[x] = 1/|\Omega|$ for each $x \in \Omega$; the corresponding sample space is then called *symmetric*.

Some elementary properties follow directly from the definitions. In particular, for any two events* A and B we have that

1. $\Pr[\Omega] = 1$, $\Pr[\emptyset] = 0$ and $\Pr[A] \geq 0$ for all $A \subseteq \Omega$;
2. $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] \leq \Pr[A] + \Pr[B]$;
3. $\Pr[A \cup B] = \Pr[A] + \Pr[B]$ if A and B are disjoint;
4. $\Pr\left[\overline{A}\right] = 1 - \Pr[A]$;
5. $\Pr[A \setminus B] = \Pr[A] - \Pr[A \cap B]$;
6. $\Pr[A \cap B] \geq \Pr[A] - \Pr\left[\overline{B}\right]$;
7. If $B_1, \ldots, B_m$ is a partition of Ω then $\Pr[A] = \sum_{i=1}^{m} \Pr[A \cap B_i]$.

For two events A and B, the *conditional probability of A given B*, denoted $\Pr[A|B]$, is the probability that one would assign to A if one knew that B occurs. Formally,

$$\Pr[A|B] := \frac{\Pr[A \cap B]}{\Pr[B]},$$

when $\Pr[B] \neq 0$. For example, if we are choosing a uniform integer from $\{1, \ldots, 6\}$, A is the event that the number is 2 and B is the event that the number is even, then $\Pr[A|B] = 1/3$, whereas $\Pr[B|A] = 1$.

Two events A and B are *independent* if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$. If $B \neq \emptyset$, this is equivalent to $\Pr[A \mid B] = \Pr[A]$. It is very important to note that the "independence" has nothing to do with the "disjointness" of the events: if, say, $0 < \Pr[A] < 1$, then the events A and $\overline{A}$ *are* dependent!

Let Γ be finite set, and $0 \leq p \leq 1$. A *random subset* $\boldsymbol{S}$ of Γ is obtained by flipping a coin, with probability p of success, for each element of Γ to determine whether the element is to be included in $\boldsymbol{S}$; the distribution of $\boldsymbol{S}$ is the probability distribution on $\Omega = 2^{\Gamma}$ given by $\Pr[S] = p^{|S|}(1-p)^{|\Gamma|-|S|}$ for $S \subseteq \Gamma$. We will mainly consider the case when $\boldsymbol{S}$ is uniformly distributed, that is, when $p = 1/2$. In this case each subset $S \subseteq \Gamma$ receives the same probability $\Pr[S] = 2^{-|\Gamma|}$. If $\mathcal{F}$ is a family of subsets, then its *random member* $\boldsymbol{S}$ is a uniformly distributed member; in this case, $\Omega = \mathcal{F}$ and $\boldsymbol{S}$ has the probability distribution $\Pr[S] = 1/|\mathcal{F}|$. Note that, for $p = 1/2$, a random subset of Γ is just a random member of 2^{Γ}.

A *random variable* is a variable defined as a function $X : \Omega \to \mathbb{R}$ of the domain of a probability space. For example, if X is a uniform integer chosen from $\{1, \ldots, n\}$, then $Y := 2X$ and $Z :=$ "the number of prime divisors of X" are both random variables, and so is X itself. In what follows, $\Pr[X = s]$ denotes the probability of the event $X^{-1}(s) = \{x \in \Omega : X(x) = s\}$. One

* Here and throughout $\overline{A} = \Omega \setminus A$ stands for the complement of A.

says in this case that X takes value $s \in \mathbb{R}$ with probability $\Pr[X = s]$. It is clear that events are a special type of random variables taking only two values 0 and 1. Namely, one can identify an event $A \subseteq \Omega$ with its *indicator random variable* X_A such that $X_A(x) = 1$ if and only if $x \in A$.

One of the most basic probabilistic notions is the expected value of a random variable. This is defined for any real-valued random variable X, and intuitively, it is the value that we would expect to obtain if we repeated a random experiment several times and took the average of the outcomes of X. Namely, if X takes values $s_1, \ldots, s_m$, then the *mean* or *expectation* of X is defined as the weighted average of these values:

$$\mathrm{E}[X] := \sum_{i=1}^{m} s_i \cdot \Pr[X = s_i] = \sum_{x \in \Omega} X(x) \cdot \Pr[x] .$$

For example, if X is the number of spots on the top face when we roll a fair die, then the expected number of spots is $\mathrm{E}[X] = \sum_{i=1}^{6} i(1/6) = 3.5$. In this book we will only consider random variables with *finite* ranges S, so that we will not be faced with the convergence issue of the corresponding series.

Note that the probability distribution $\Pr : \Omega \to [0, 1]$ itself is a random variable, and its expectation is

$$\mathrm{E}[\Pr] = \sum_{x \in \Omega} \Pr[x]^2 .$$

In particular, if Pr is a uniform distribution of a set with n elements, then its expectation is $1/n$, as it should be. The expectation of the indicator random variable X_A of an event A is just its probability:

$$\mathrm{E}[X_A] = 0 \cdot \Pr[X_A = 0] + 1 \cdot \Pr[X_A = 1] = \Pr[X_A = 1] = \Pr[A] .$$

The probabilistic method is most striking when it is applied to prove theorems whose statement does not seem to suggest the need for probability at all. It is therefore surprising what results may be obtained from such simple principles like the *union bound*: The probability of a union of events is at most the sum of the probabilities of the events,

$$\Pr[A_1 \cup A_2 \cup \cdots \cup A_n] \le \Pr[A_1] + \Pr[A_2] + \cdots + \Pr[A_n] . \tag{3.3}$$

Thus, if A_i's are some "bad" events and $\sum \Pr[A_i] < 1$ then we know that $\Pr\left[\cap_i \overline{A_i}\right] = \Pr\left[\overline{\cup_i A_i}\right] = 1 - \Pr[\cup_i A_i] > 0$, that is, with positive probability, *none* of these bad events happens. Already this simple fact often allows to show that some object with desired "good" properties exists.

A next useful property is the *linearity of expectation*: If $X_1, \ldots, X_n$ are random variables and $a_1, \ldots, a_n$ real numbers, then

$$\mathrm{E}[a_1 X_1 + \cdots + a_n X_n] = a_1 \mathrm{E}[X_1] + \cdots + a_n \mathrm{E}[X_n] .$$

The equality follows directly from the definition $\mathrm{E}[X]$. The power of this principle comes from there being no restrictions on the X_i's.

A general framework for the probabilistic method is the following. Many extremal problems can be defined by a pair (M, f), where M is some finite set of objects and $f : M \to \mathbb{R}$ some function assigning each object $x \in M$ its “value”. For example, M could be a set of graphs, satisfying some conditions, and $f(x)$ could be the maximum size of a clique in x. Given a threshold value t, the goal is to show that an object $x \in M$ with $f(x) \geq t$ *exists.* That is, we want to show that $\max_{x \in M} f(x) \geq t$.

To solve this task, one defines an appropriate probability distribution $\mathrm{Pr} : M \to [0, 1]$ and considers the resulting probability space. In this space the target function f becomes a random variable. One tries then to show that either $\mathrm{E}[f] \geq t$ or $\mathrm{Pr}[f(x) \geq t] > 0$ holds. If at least one of these inequalities holds, then the existence of $x \in M$ with $f(x) \geq t$ is already shown. Indeed, would $f(x) < t$ hold for *all* $x \in M$, then we would have

$$\mathrm{Pr}[f(x) \geq t] = \mathrm{Pr}[\emptyset] = 0$$

and

$$\mathrm{E}[f] = \sum_i i \cdot \mathrm{Pr}[f = i] < \sum_i t \cdot \mathrm{Pr}[f = i] = t\,.$$

The property

$$\mathrm{E}[f] \geq t \text{ implies } f(x) \geq t \text{ for at least one } x \in M$$

is sometimes called the *pigeonhole principle* of expectation: a random variable cannot always be smaller (or always greater) than its expectation.

In the next sections we give some simplest applications of the probabilistic method (more applications are given in Part IV).

3.2 Tournaments

A *tournament* is an oriented graph $T = (V, E)$ such that $(x, x) \not\in E$ for all $x \in V$, and for any two vertices $x \neq y$ exactly one of (x, y) and (y, x) belongs to E. That is, each tournament is obtained from a complete graph by orienting its edges. The name tournament is natural, since one can think of the set V as a set of players in which each pair participates in a single match, where $(x, y) \in E$ iff x beats y.

Say that a tournament has the property P_k if for every set of k players there is one who beats them all, i.e., if for any subset $S \subseteq V$ of k players there exists a player $y \not\in S$ such that $(y, x) \in E$ for all $x \in S$.

Theorem 3.1 (Erdős 1963a)**.** *If $n \geq k^2 2^{k+1}$, then there is a tournament of n players that has the property P_k.*

Proof. Consider a random tournament of n players, i.e., the outcome of every game is determined by the flip of fair coin. For a set S of k players, let A_S be the event that *no* $y \notin S$ *beats all of* S. Each $y \notin S$ has probability 2^{-k} of beating all of S and there are $n-k$ such possible y, all of whose chances are mutually independent. Hence $\Pr[A_S] = (1-2^{-k})^{n-k}$ and

$$\Pr\left[\bigcup A_S\right] \le \binom{n}{k}(1-2^{-k})^{n-k} < \frac{n^k}{k!}\mathrm{e}^{-(n-k)/2^k} \le n^k \mathrm{e}^{-n/2^k} .$$

If $n \ge k^2 2^{k+1}$, this probability is strictly smaller than 1. Thus, for such an n, with positive probability no event A_S occurs. This means that there is a point in the probability space for which none of the events A_S happens. This point is a tournament T and this tournament has the property P_k. □

3.3 Universal sets

A set of 0-1 strings of length n is (n,k)-*universal* if, for any subset of k coordinates $S = \{i_1, \ldots, i_k\}$, the projection

$$A\restriction_S := \{(a_{i_1}, \ldots, a_{i_k}) \ : \ (a_1, \ldots, a_n) \in A\}$$

of A onto the coordinates in S contains all possible 2^k configurations.

In Sects. 10.5 and 17.4 we will present two *explicit* constructions of such sets of size about n, when $k \le (\log n)/3$, and of size $n^{O(k)}$, for arbitrary k. On the other hand, a simple probabilistic argument shows that (n,k)-universal sets of size $k2^k \log_2 n$ exist (note that 2^k is a trivial lower bound).

Theorem 3.2 (Kleitman–Spencer 1973). *If* $\binom{n}{k}2^k(1-2^{-k})^r < 1$*, then there is an* (n,k)*-universal set of size* r*.*

Proof. Let $\boldsymbol{A}$ be a set of r random 0-1 strings of length n, each entry of which takes values 0 or 1 independently and with equal probability 1/2. For every fixed set S of k coordinates and for every fixed vector $v \in \{0,1\}^k$,

$$\Pr[v \notin \boldsymbol{A}\restriction_S] = \prod_{a \in \boldsymbol{A}} \Pr[v \ne a\restriction_S] = \prod_{a \in \boldsymbol{A}} \left(1 - 2^{-|S|}\right) = \left(1-2^{-k}\right)^r .$$

Since there are only $\binom{n}{k}2^k$ possibilities to choose a pair (S,v), the set $\boldsymbol{A}$ is *not* (n,k)-universal with probability at most $\binom{n}{k}2^k(1-2^{-k})^r$, which is strictly smaller than 1. Thus, at least one set A of r vectors must be (n,k)-universal, as claimed. □

3.4 Covering by bipartite cliques

A *biclique covering* of a graph G is a set $H_1, \ldots, H_t$ of its complete bipartite subgraphs such that each edge of G belongs to at least one of these subgraphs. The *weight* of such a covering is the sum $\sum_{i=1}^{t} |V(H_i)|$ of the number of vertices in these subgraphs. Let $\mathrm{bc}(G)$ be the smallest weight of a biclique covering of G. Let K_n be a complete graph on n vertices.

Theorem 3.3. *If n is a power of two, then* $\mathrm{bc}(K_n) = n \log_2 n$.

Proof. Let $n = 2^m$. We can construct a covering of K_n as follows. Assign to each vertex v its *own* vector $x_v \in \{0,1\}^m$, and consider $m = \log_2 n$ bipartite cliques $H_1, \ldots, H_m$, where two vertices u and v are adjacent in H_i iff $x_u(i) = 0$ and $x_v(i) = 1$. Since every two distinct vectors must differ in at least one coordinate, each edge of K_n belongs to at least one of these bipartite cliques. Moreover, each of the cliques has weight $(n/2) + (n/2) = n$, since exactly $2^{m-1} = n/2$ of the vectors in $\{0,1\}^m$ have the same value in the i-th coordinate. So, the total weight of this covering is $mn = n \log_2 n$.

To prove the lower bound we use a probabilistic argument. Let $A_1 \times B_1, \ldots, A_t \times B_t$ be a covering of K_n by bipartite cliques. For a vertex v, let m_v be the number of these cliques containing v. By the double-counting principle,

$$\sum_{i=1}^{t} (|A_i| + |B_i|) = \sum_{v=1}^{n} m_v$$

is the weight of the covering. So, it is enough to show that the right-hand sum is at least $n \log_2 n$.

To do this, we throw a fair 0-1 coin for each of the cliques $A_i \times B_i$ and remove all vertices in A_i from the graph if the outcome is 0; if the outcome is 1, then we remove B_i. Let $X = X_1 + \cdots + X_n$, where X_v is the indicator variable for the event “the vertex v survives.”

Since any two vertices of K_n are joined by an edge, and since this edge is covered by at least one of the cliques, at most one vertex can survive at the end. This implies that $\mathrm{E}[X] \leq 1$. On the other hand, each vertex v will survive with probability 2^{-m_v}: there are m_v steps that are “dangerous” for v, and in each of these steps the vertex v will survive with probability $1/2$. By the linearity of expectation,

$$\sum_{v=1}^{n} 2^{-m_v} = \sum_{v=1}^{n} \Pr[v \text{ survives}] = \sum_{v=1}^{n} \mathrm{E}[X_v] = \mathrm{E}[X] \leq 1\,.$$

We already know that the arithmetic mean of numbers $a_1, \ldots, a_n$ is at least their geometric mean (1.16):

$$\frac{1}{n} \sum_{v=1}^{n} a_v \geq \Bigl(\prod_{v=1}^{n} a_v \Bigr)^{1/n}.$$

When applied with $a_v = 2^{-m_v}$, this yields

$$\frac{1}{n} \geq \frac{1}{n} \sum_{v=1}^{n} 2^{-m_v} \geq \Big(\prod_{v=1}^{n} 2^{-m_v} \Big)^{1/n} = 2^{-\frac{1}{n}\sum_{v=1}^{n} m_v} ,$$

from which $2^{\frac{1}{n}\sum_{v=1}^{n} m_v} \geq n$, and hence, also $\sum_{v=1}^{n} m_v \geq n \log_2 n$ follows. □

3.5 2-colorable families

Let $\mathcal{F}$ be a family of subsets of some finite set. Can we color the elements of the underlying set in red and blue so that no member of $\mathcal{F}$ will be monochromatic? Such families are called 2-*colorable.*

Recall that a family is k-*uniform* if each member has exactly k elements.

Theorem 3.4 (Erdős 1963b)**.** *Every k-uniform family with fewer than 2^{k-1} members is 2-colorable.*

Proof. Let $\mathcal{F}$ be an arbitrary k-uniform family of subsets of some finite set X. Consider a random 2-coloring obtained by coloring each point independently either red or blue, where each color is equally likely. Informally, we have an experiment in which a fair coin is flipped to determine the color of each point. For a member $A \in \mathcal{F}$, let X_A be the indicator random variable for the event that A is monochromatic. So, $X = \sum_{A\in\mathcal{F}} X_A$ is the total number of monochromatic members.

For a member A to be monochromatic, all its $|A| = k$ points must receive the same color. Since the colors are assigned at random and independently, this implies that each member of $\mathcal{F}$ will be monochromatic with probability at most $2 \cdot 2^{-k} = 2^{1-k}$ (factor 2 comes since we have two colors). Hence,

$$\mathrm{E}[X] = \sum_{A\in\mathcal{F}} \mathrm{E}[X_A] = \sum_{A\in\mathcal{F}} 2^{1-k} = |\mathcal{F}| \cdot 2^{1-k} .$$

Since points in our probability space are 2-colorings, the pigeonhole property of expectation implies that a coloring, leaving at most $|\mathcal{F}| \cdot 2^{1-k}$ members of $\mathcal{F}$ monochromatic, must exist.

In particular, if $|\mathcal{F}| < 2^{k-1}$ then no member of $\mathcal{F}$ will be left monochromatic. □

The proof was quite easy. So one could ask whether we can replace 2^{k-1} by, say, 4^k? By turning the probabilistic argument "on its head" it can be shown that this is *not* possible. The sets now become random and each coloring defines an event.

Theorem 3.5 (Erdős 1964a)**.** *If k is sufficiently large, then there exists a k-uniform family $\mathcal{F}$ such that $|\mathcal{F}| \leq k^2 2^k$ and $\mathcal{F}$ is not 2-colorable.*

Proof. Set $r = \lfloor k^2/2 \rfloor$. Let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$ be independent random members of $\binom{[r]}{k}$, that is, $\boldsymbol{A}_i$ ranges over the set of all $A \subseteq \{1, \ldots, r\}$ with $|A| = k$, and $\Pr[\boldsymbol{A}_i = A] = \binom{r}{k}^{-1}$. Consider the family $\boldsymbol{\mathcal{F}} = \{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_b\}$, where b is a parameter to be specified later. Let χ be a coloring of $\{1, \ldots, r\}$ in red and blue, with a red points and $r - a$ blue points. Using Jensen's inequality (see Proposition 1.12), for any such coloring and any i, we have

$$\begin{aligned} \Pr[\boldsymbol{A}_i \text{ is monochromatic}] &= \Pr[\boldsymbol{A}_i \text{ is red}] + \Pr[\boldsymbol{A}_i \text{ is blue}] \\ &= \frac{\binom{a}{k} + \binom{r-a}{k}}{\binom{r}{k}} \geq 2\binom{r/2}{k} \Big/ \binom{r}{k} := p, \end{aligned}$$

where, by the asymptotic formula (1.9) for the binomial coefficients, p is about $\mathrm{e}^{-1}2^{1-k}$. Since the members $\boldsymbol{A}_i$ of $\boldsymbol{\mathcal{F}}$ are independent, the probability that a given coloring χ is legal for $\boldsymbol{\mathcal{F}}$ equals

$$\prod_{i=1}^{b} (1 - \Pr[\boldsymbol{A}_i \text{ is monochromatic}]) \leq (1-p)^b .$$

Hence, the probability that at least one of all 2^r possible colorings will be legal for $\boldsymbol{\mathcal{F}}$ does not exceed $2^r(1-p)^b < \mathrm{e}^{r \ln 2 - pb}$, which is less than 1 for $b = (r \ln 2)/p = (1 + o(1))k^2 2^{k-2}\mathrm{e} \ln 2$. But this means that there must be at least one realization of the (random) family $\boldsymbol{\mathcal{F}}$, which has only b sets and which cannot be colored legally. □

Let $B(k)$ be the minimum possible number of sets in a k-uniform family which is not 2-colorable. We have already shown that

$$2^{k-1} \leq B(k) \leq k^2 2^k .$$

As for exact values of $B(k)$, only the first two $B(2) = 3$ and $B(3) = 7$ are known. The value $B(2) = 3$ is realized by the graph K_3. We address the inequality $B(3) \leq 7$ in Exercise 3.9.

There is yet another class of 2-colorable families, without any uniformity restriction.

Theorem 3.6. *Let $\mathcal{F}$ be an arbitrary family of subsets of a finite set, each of which has at least two elements. If every two non-disjoint members of $\mathcal{F}$ share at least two common elements, then $\mathcal{F}$ is 2-colorable.*

Proof. Let $X = \{x_1, \ldots, x_n\}$ be the underlying set. We will color the points $x_1, \ldots, x_n$ one-by-one so that we do not color all points of any set in $\mathcal{F}$ with the same color. Color the first point x_1 arbitrarily. Suppose that $x_1, \ldots, x_i$ are already colored. If we cannot color the next element x_{i+1} in red then this means that there is a set $A \in \mathcal{F}$ such that $A \subseteq \{x_1, \ldots, x_{i+1}\}$, $x_{i+1} \in A$ and all the points in $A \setminus \{x_{i+1}\}$ are red. Similarly, if we cannot color the next element x_{i+1} in blue, then there is a set $B \in \mathcal{F}$ such that $B \subseteq \{x_1, \ldots, x_{i+1}\}$,

$x_{i+1} \in B$ and all the points in $B \setminus \{x_{i+1}\}$ are blue. But then $A \cap B = \{x_{i+1}\}$, a contradiction. Thus, we *can* color the point x_{i+1} either red or blue. Proceeding in this way we will finally color all the points and no set of $\mathcal{F}$ becomes monochromatic. □

3.6 The choice number of graphs

The *choice number* (or *list-coloring number*) of a graph G, denoted by $ch(G)$, is the minimum integer k such that for every assignment of a set $S(v)$ of k colors to every vertex v of G, there is a legal coloring of G that assigns to each vertex v a color from $S(v)$. Recall that a coloring is *legal* if adjacent vertices receive different colors. Obviously, this number is at least the chromatic number $\chi(G)$ of G.

Theorem 3.7 (Alon 1992)**.** *For every bipartite $n \times n$ graph G with $n \geq 3$, we have that $ch(G) \leq 2\log_2 n$.*

Proof. Let $G = (V_0 \cup V_1, E)$ be a bipartite graph with $|V_0| = |V_1| = n$. Suppose that each vertex v is assigned a set $S(v)$ of at least $2\log_2 n$ colors, and let $S = \cup_v S(v)$ be the set of all colors. Since the graph is bipartite, it is enough to show that there is a partition $S = S_0 \cup S_1$ of S such that

$$S_i \cap S(v) \neq \emptyset \text{ for both } i = 0, 1 \text{ and all } v \in V_i. \tag{3.4}$$

Then, for every two (even not necessarily adjacent) vertices $u \in V_0$ and $v \in V_1$, we can choose arbitrary colors $c_v \in S_0 \cap S(v)$ and $c_u \in S_1 \cap S(v)$; since $S_0 \cap S_1 = \emptyset$, these colors are clearly distinct.

To define such a partition $S = S_0 \cup S_1$ just flip, for each color $c \in S$, a fair 0-1 coin to decide whether to include this color in the set S_0; let also $S_1 = S \setminus S_0$. For a fixed $i \in \{0, 1\}$ and $v \in V_i$ we have that

$$\Pr[S_i \cap S(v) = \emptyset] = 2^{-|S(v)|} \leq 2^{-2\log_2 n} = \frac{1}{n^2} < \frac{1}{2n}\,.$$

The number of pairs (i, v) with $i \in \{0, 1\}$ and $v \in V_i$ is $2n$. Hence, by the union bound, the probability that our random partition $S = S_0 \cup S_1$ does not satisfy (3.4) is strictly smaller than 1, implying that a desired partition exists. □

Note that the proof says a bit more. If $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ are any two sequences of not necessarily distinct $2\log_2 n$-element subsets of some set of vertices, then it is possible to color the vertices in red and blue so that each of the A_i receives at least one red color and each of the B_i receives at least one blue color.

Exercises

3.1. Let $\Omega = B_1 \cup B_2$ be a partition of a sample space, and $A \subseteq \Omega$ be an event. Prove that then $\Pr[A]$ does not exceed the maximum of $\Pr[A \mid B_1]$ and $\Pr[A \mid B_2]$. *Hint*: Show that $\Pr[A] = \Pr[B_1] \cdot \Pr[A \mid B_1] + \Pr[B_2] \cdot \Pr[A \mid B_2]$.

3.2. Prove the following *Bonferroni inequality*:

$$\Pr[A_1 \cap \cdots \cap A_n] \geq \Pr[A_1] + \cdots + \Pr[A_n] - n + 1 .$$

3.3. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. The *variance* of a random variable X is defined as $\mathrm{Var}[X] := \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$. Prove that

1. $\mathrm{E}[a \cdot X + b \cdot Y] = a \cdot \mathrm{E}[X] + b \cdot \mathrm{E}[Y]$ for any constants a and b.
2. If X and Y are independent then $\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$ and $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.
3. $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$. *Hint*: $\mathrm{E}[X \cdot \mathrm{E}[X]] = \mathrm{E}[X]^2$.

3.4. Let X be a non-negative integer-valued random variable. Show that $\mathrm{E}[X^2] \geq \mathrm{E}[X]$, $\Pr[X = 0] \geq 1 - \mathrm{E}[X]$ and $\mathrm{E}[X] = \sum_{x=1}^{\infty} \Pr[X \geq x]$.

3.5. Use the Cauchy–Schwarz inequality $(\sum_{i=1}^n a_i b_i)^2 \leq (\sum_{i=1}^n a_i^2)(\sum_{i=1}^n b_i^2)$ to show that, for any random variable X, $\mathrm{E}[X]^2 \leq \mathrm{E}[X^2]$.

3.6. Let $X_1, \ldots, X_n$ be n independent 0-1 random variables such that $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Let $X = \sum_{i=1}^n X_i \bmod 2$. Prove that $\Pr[X = 1] = \frac{1}{2}\left[1 - \prod_i (1 - 2p_i)\right]$. *Hint*: Consider the random variable $Y = Y_1 \cdots Y_n$, where $Y_i = 1 - 2X_i$, and observe that $\mathrm{E}[Y] = 1 - 2 \cdot \Pr[Y = -1]$.

3.7. For a graph G let, as before, $\mathrm{bc}(G)$ denote the smallest weight of a biclique covering of G. Show that if an n-vertex graph G has no independent set of size larger than α then $\mathrm{bc}(G) \geq n \log_2(n/\alpha)$. *Hint*: Argue as in the proof of the lower bound in Theorem 3.3, and show that $\mathrm{E}[X] \leq \alpha$.

3.8. For a graph $G = (V, E)$, let μ_G be the minimum of $(a + b)/ab$ over all pairs of integers $a, b \geq 1$ such that G contains a copy of a complete bipartite $a \times b$ graph $K_{a,b}$. Show that $\mathrm{bc}(G) \geq \mu_G \cdot |E|$.

3.9. Prove that $B(3) \leq 7$. That is, exhibit a family of seven 3-element sets which is not 2-colorable. *Hint*: Consider the Fano configuration (Fig. 12.1).

3.10. (Razborov 1990). Consider the family of all pairs (A, B) of disjoint k-element subsets of $\{1, \ldots, n\}$. A set Y *separates* the pair (A, B) if $A \subseteq Y$ and $B \cap Y = \emptyset$. Prove that there exist $\ell = 2k4^k \ln n$ sets such that every pair (A, B) is separated by at least one of them. *Hint*: Pick subsets $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\ell$ of $\{1, \ldots, n\}$ randomly and independently, each with probability 2^{-n}. Show that the probability that none of them separates a given pair (A, B) is at most $\left(1 - 2^{-2k}\right)^\ell$ and use the counting sieve.

3.11. Let X be a set of $n = kr$ points and consider their colorings $c \colon X \to \{1, \ldots, k\}$ by k colors. Say that such a coloring c is *balanced* if each color is used for the same number of points, i.e., if $|c^{-1}(i)| = r$ for every color $i = 1, \ldots, k$. Given a k-element set of points, say that it is *differently colored* if no two of its points get the same color. Prove that there exist $\ell = O(k\mathrm{e}^k \log n)$ balanced colorings $c_1, \ldots, c_\ell$ such that every k-element subset of X is differently colored by at least one of them. *Hint*: Consider independent copies $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_\ell$ of a balanced coloring $\boldsymbol{c}$ selected at random from the set of all $n!/(r!)^k$ such colorings. Show that for every k-element subset S of X, $\boldsymbol{c}$ colors S differently with probability $p = r^k \cdot \binom{n}{k}^{-1}$. Use the counting sieve to show that, with probability at least $1 - \binom{n}{k}(1-p)^\ell$, every k-element subset S will be colored differently by at least one of $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_\ell$. Recall that $r = n/k$ and use Proposition 1.4 to show that this probability is nonzero for some $\ell = O(k\mathrm{e}^k \log n)$

3.12. (Khasin 1969). Consider the *k-threshold function* $T_k^n(x_1, \ldots, x_n)$ which outputs 1 if and only if $x_1 + \cdots + x_n \geq k$. In Sect. 6.3.2 we will show that any depth-3 *Or-And-Or formula* for T_k^n must have size exponential in k. What about the upper bounds? Use the previous exercise to show that T_k^n can be computed by a monotone *Or-And-Or formula* of size $O(k\mathrm{e}^k n \log n)$. *Hint*: Each balanced k-coloring c of $\{1, \ldots, n\}$ gives us an And-Or formula $F_c = \bigwedge_{i=1}^k \bigvee_{c(j)=i} x_j$. Use the previous exercise to combine them into an Or-And-Or formula for T_k^n.

3.13. Let $\mathcal{F}$ be a family, each member of which has ≥ 3 points and any two members share exactly one point in common. Suppose also that $\mathcal{F}$ is not 2-colorable. Prove that: (i) every point x belongs to at least two members of $\mathcal{F}$, and (ii) any two points x, y belong to at least one member of $\mathcal{F}$. *Hint*: (i) Take $x \in A \in \mathcal{F}$, color $A \setminus \{x\}$ red and the rest blue. (ii) Select sets A, B such that $x \in A \setminus B$ and $y \in B \setminus A$; color $(A \cup B) \setminus \{x, y\}$ red and everything else blue.

3.14. (Lovász 1973). Let $\mathcal{F}$ be 3-uniform family on $n \geq 5$ points, in which each pair of points occurs in the same number of sets. Prove that $\mathcal{F}$ is not 2-colorable. *Hint*: Suppose there is a 2-coloring, count the members of $\mathcal{F}$ in two ways: by the monochromatic pairs contained in them and also by the bichromatic pairs contained in them. Let n_1 and n_2 denote the number of red and blue points, respectively, and let a be the number of members of $\mathcal{F}$ containing a given pair of points. We have $a\binom{n_1}{2}$ sets in $\mathcal{F}$ containing a pair $\{x, y\}$ of red points, and $a\binom{n_2}{2}$ sets containing a blue pair of points. Hence, $|\mathcal{F}|$ is the sum of these two numbers. On the other hand, each set of $\mathcal{F}$ contains exactly two pairs $\{x, y\}$ where x is blue and y is red; so $2|\mathcal{F}| = an_1n_2$. Compare these numbers, and use the arithmetic-geometric mean inequality (1.16) to show that the equality can hold only if $n \leq 4$.

4. The Pigeonhole Principle

The *pigeonhole principle* (also known as *Dirichlet's principle*) states the "obvious" fact that $n + 1$ pigeons cannot sit in n holes so that every pigeon is alone in its hole. More generally, the pigeonhole principle states the following:

> *If a set consisting of at least $rs + 1$ objects is partitioned into r classes, then some class receives at least $s + 1$ objects.*

Its truth is easy to verify: if every class receives at most s objects, then a total of at most rs objects have been distributed. To see that the result is best possible, observe that a set with at most rs points can be divided into r groups with at most s points in each group; hence none of the groups contains $s + 1$ points.

This is one of the oldest "non-constructive" principles: it states only the *existence* of a pigeonhole with more than k items and says nothing about how to *find* such a pigeonhole. Today we have powerful and far reaching generalizations of this principle (Ramsey-like theorems, the probabilistic method, etc.). We will talk about them later.

As trivial as the pigeonhole principle itself may sound, it has numerous nontrivial applications. The hard part in applying this principle is to decide what to take as pigeons and what as pigeonholes. Let us illustrate this by several examples.

4.1 Some quickies

To "warm-up," let us start with the simplest applications. The *degree* of a vertex x in a graph G is the number $d(x)$ of edges of G adjacent to x.

Proposition 4.1. *In any graph there exist two vertices of the same degree.*

Proof. Given a graph G on n vertices, make n pigeonholes labeled from 0 up to $n - 1$ and put a vertex x into the k-th pigeonhole iff $d(x) = k$. If some

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_4,

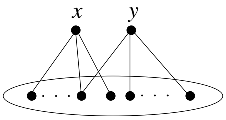


Fig. 4.1 There are only $n-2$ vertices and at least $n-1$ edges going to them.

pigeonhole contains more than one vertex, we are done. So, assume that no pigeonhole has more than one vertex. There are n vertices going into the n pigeonholes; hence each pigeonhole has *exactly one* vertex. Let x and y be the vertices lying in the pigeonholes labeled 0 and $n-1$, respectively. The vertex x has degree 0 and so has no connection with other vertices, including y. But y has degree $n-1$ and hence, is connected with all the remaining vertices, including x, a contradiction. □

If G is a finite graph, the *independence number* $\alpha(G)$ is the maximum number of pairwise nonadjacent vertices of G. The *chromatic number* $\chi(G)$ of G is the minimum number of colors in a coloring of the vertices of G with the property that no two adjacent vertices have the same color.

Proposition 4.2. *In any graph G with n vertices, $n \leq \alpha(G) \cdot \chi(G)$.*

Proof. Consider the vertices of G partitioned into $\chi(G)$ color classes (sets of vertices with the same color). By the pigeonhole principle, one of the classes must contain at least $n/\chi(G)$ vertices, and these vertices are pairwise nonadjacent. Thus $\alpha(G) \geq n/\chi(G)$, as desired. □

A graph is *connected* if there is a path between any two of its vertices.

Proposition 4.3. *Let G be an n-vertex graph. If every vertex has a degree of at least $(n-1)/2$ then G is connected.*

Proof. Take any two vertices x and y. If these vertices are not adjacent, then at least $n-1$ edges join them to the remaining vertices, because both x and y have a degree of at least $(n-1)/2$.

Since there are only $n-2$ other vertices, the pigeonhole principle implies that one of them must be adjacent to both x and y (see Fig. 4.1). We have proved that every pair of vertices is adjacent or has a common neighbor, so G is connected. □

Remark 4.4. A result is *best possible* if the conclusion no longer holds when we weaken one of the conditions. Such is, for example, the result above: let n be even and G be a union of two vertex disjoint complete graphs on $n/2$ vertices; then every vertex has degree $(n-2)/2$, but the graph is disconnected.

Note that, in fact, we have proved more: if every vertex of an n-vertex graph has degree at least $(n-1)/2$ then the graph has diameter at most two. The *diameter* of a graph is the smallest number k such that every two vertices are connected by a path with at most k edges.

4.2 The Erdős–Szekeres theorem

Let $A = (a_1, a_2, \dots, a_n)$ be a sequence of n different numbers. A *subsequence of k terms* of A is a sequence B of k distinct terms of A appearing in the same order in which they appear in A. In symbols, we have $B = (a_{i_1}, a_{i_2}, \dots, a_{i_k})$, where $i_1 < i_2 < \cdots < i_k$. A subsequence B is said to be *increasing* if $a_{i_1} < a_{i_2} < \cdots < a_{i_k}$, and *decreasing* if $a_{i_1} > a_{i_2} > \cdots > a_{i_k}$.

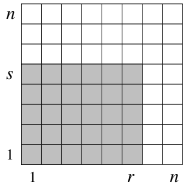
We will be interested in the length of the *longest* increasing and decreasing subsequences of A. It is intuitively plausible that there should be some kind of tradeoff between these lengths. If the longest increasing subsequence is short, say has length s, then *any* subsequence of A of length $s+1$ must contain a pair of decreasing elements, so there are lots of pairs of decreasing elements. Hence, we would expect the longest decreasing sequence to be large. An extreme case occurs when $s = 1$. Then the whole sequence A is decreasing.

How can we quantify the feeling that the length of *both*, longest increasing and longest decreasing subsequences, cannot be small? A famous result of Erdős and Szekeres (1935) gives an answer to this question and was one of the first results in extremal combinatorics.

Theorem 4.5 (Erdős–Szekeres 1935). *Let $A = (a_1, \dots, a_n)$ be a sequence of n different real numbers. If $n \geq sr + 1$ then either A has an increasing subsequence of $s+1$ terms or a decreasing subsequence of $r+1$ terms (or both).*

Proof (due to Seidenberg 1959). Associate to each term a_i of A a pair of "scores" (x_i, y_i) where x_i is the number of terms in the longest *increasing* subsequence *ending* at a_i, and y_i is the number of terms in the longest *decreasing* subsequence *starting* at a_i. Observe that no two terms have the same score, i.e., that $(x_i, y_i) \neq (x_j, y_j)$ whenever $i \neq j$. Indeed, if we have $\cdots a_i \cdots a_j \cdots$, then either $a_i < a_j$ and the longest increasing subsequence ending at a_i can be extended by adding on a_j (so that $x_i < x_j$), or $a_i > a_j$ and the longest decreasing subsequence starting at a_j can be preceded by a_i (so that $y_i > y_j$).

Now make a grid of n^2 pigeonholes:



Place each term a_i in the pigeonhole with coordinates (x_i, y_i). Each term of A can be placed in some pigeonhole, since $1 \leq x_i, y_i \leq n$ for all $i = 1, \dots, n$. Moreover, no pigeonhole can have more than one term because $(x_i, y_i) \neq (x_j, y_j)$ whenever $i \neq j$. Since $|A| = n \geq sr + 1$, we have more items than the pigeonholes shaded in the above picture. So some term a_i will lie outside this shaded region. But this means that either $x_i \geq s+1$ or $y_i \geq r+1$ (or both), exactly what we need. □

The set of real numbers is *totally ordered.* That is, for any two distinct numbers x and y, either $x < y$ or $y < x$. The following lemma, due to Dilworth, generalizes the Erdős–Szekeres theorem to sets in which two elements may or may not be comparable.

A *partial order* on a set P is a binary relation $<$ between its elements which is transitive and irreflexive: if $x < y$ and $y < z$ then $x < z$, but $x < y$ and $y < x$ cannot both hold. We write $x \leq y$ if $x < y$ or $x = y$. Elements x and y are *comparable* if either $x \leq y$ or $y \leq x$ (or both) hold. A *chain* in a poset P is a subset $C \subseteq P$ such that any two of its points are comparable. Dually, an *antichain* is a subset $A \subseteq P$ such that no two of its points are comparable.

Lemma 4.6 (Dilworth 1950)**.** *In any partial order on a set P of $n \geq sr+1$ elements, there exists a chain of length $s+1$ or an antichain of size $r+1$.*

Proof. A chain is *maximal* if it cannot be prolonged by adding a new element. Let $C_1, \ldots, C_m$ be all maximal chains in P, and suppose there is no chain of length $s+1$. Since the chains C_i must cover all n points of P, the pigeonhole principle implies that we must have $m \geq r+1$ such chains. Let $x_i \in C_i$ be the greatest element of C_i. Then no two elements x_i and x_j with $i \neq j$ can be comparable: if $x_i \leq x_j$ then $C_i \cup \{x_j\}$ would also be a chain, a contradiction with the maximality of C_i. Thus, the elements $x_1, \ldots, x_m$ form an antichain of size $m \geq r+1$. □

This lemma implies the Erdős–Szekeres theorem (we address this question in Exercise 4.10).

4.3 Mantel's theorem

Here we discuss one typical *extremal* property of graphs. How many edges are possible in a triangle-free graph G on n vertices? A triangle is a set of three vertices, each two of which are connected by an edge. Certainly, G can have $n^2/4$ edges without containing a triangle: just let G be the bipartite complete graph consisting of two sets of $n/2$ vertices each and all the edges between the two sets. Indeed, $n^2/4$ turns out to be the maximum possible number of edges: if we take one more edge then the graph will have a triangle.

We give four proofs of this beautiful result: the first (original) proof is based on double counting, the second uses the inequality $\sqrt{ab} \leq (a+b)/2$ of the arithmetic and geometric mean, the third uses the pigeonhole principle, and the fourth employs the so-called "shifting argument" (we will give this last proof in the Sect. 4.7 devoted to this argument).

Theorem 4.7 (Mantel 1907)**.** *If a graph G on n vertices contains more than $n^2/4$ edges, then G contains a triangle.*

First proof. Let G be a graph on a set V of n vertices containing $m > n^2/4$ edges. Assume that G has no triangles. Then adjacent vertices have no common neighbors, so $d(x) + d(y) \leq n$ for each edge $\{x, y\} \in E$. Summing over all edges of G, we have (cf. Equation (1.12))

$$\sum_{x \in V} d(x)^2 = \sum_{\{x,y\} \in E} (d(x) + d(y)) \leq mn\,.$$

On the other hand, using Cauchy–Schwarz inequality (see Notation or Proposition 13.4) and Euler's equality $\sum_{x \in V} d(x) = 2m$ (see Theorem 1.8), we obtain

$$\sum_{x \in V} d(x)^2 \geq \frac{\left(\sum_{x \in V} d(x)\right)^2}{|V|} = \frac{4m^2}{n}\,.$$

These two inequalities imply that $m \leq n^2/4$, contradicting the hypothesis. □

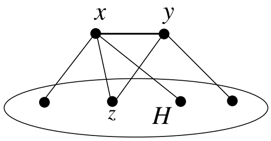
Second proof. Let $G = (V, E)$ be a graph on a set V of n vertices and assume that G has no triangles. Let $A \subseteq V$ be the largest independent set, i.e., a maximal set of vertices, no two of which are adjacent in G. Since G is triangle-free, the neighbors of a vertex $x \in V$ form an independent set, and we infer $d(x) \leq |A|$ for all x.

The set $B = V \setminus A$ meets every edge of G. Counting the edges of G according to their end-vertices in B, we obtain $|E| \leq \sum_{x \in B} d(x)$. The inequality of the arithmetic and geometric mean (1.16) yields

$$|E| \leq \sum_{x \in B} d(x) \leq |A| \cdot |B| \leq \left(\frac{|A| + |B|}{2}\right)^2 = \frac{n^2}{4}\,.$$

□

Third proof. To avoid ceilings and floorings, we will prove the theorem for graphs on an *even* number $2n$ of vertices. We want to prove that every such graph with at least $n^2 + 1$ edges must contain a triangle. We argue by induction on n. If $n = 1$, then G cannot have $n^2 + 1$ edges; hence the statement is true. Assuming the result for n, we now consider a graph G on $2(n + 1)$ vertices with $(n + 1)^2 + 1$ edges. Let x and y be adjacent vertices in G, and let H be the induced subgraph on the remaining $2n$ vertices. If H contains at least $n^2 + 1$ edges then we are done by the induction hypothesis. Suppose that H has at most n^2 edges, and therefore at least $2n + 1$ edges of G emanate from x and y to vertices in H:



By the pigeonhole principle, among these $2n+1$ edges there must be an edge from x and an edge from y to the same vertex z in H. Hence G contains the triangle $\{x, y, z\}$. □

4.4 Turán's theorem

A *k-clique* is a graph on k vertices, every two of which are connected by an edge. For example, triangles are 3-cliques. Mantel's theorem says that, if a graph on n vertices has no 3-clique then it has at most $n^2/4$ edges. What about $k > 3$?

The answer is given by a fundamental result of Paul Turán, which initiated extremal graph theory.

Theorem 4.8 (Turán 1941)**.** *If a graph $G = (V, E)$ on n vertices has no $(k+1)$-clique, $k \geq 2$, then*

$$|E| \leq \left(1 - \frac{1}{k}\right) \frac{n^2}{2}. \tag{4.1}$$

Like Mantel's theorem, this result was rediscovered many times with various different proofs. Here we present the original one due to Turán. The proof based on so-called "weight shifting" argument is addressed in Exercise 4.9. In Sect. 18.4 we will give a proof which employs ideas of a totally different nature – the probabilistic argument.

Proof. We use induction on n. Inequality (4.1) is trivially true for $n = 1$. The case $k = 2$ is Mantel's theorem. Suppose now that the inequality is true for all graphs on at most $n-1$ vertices, and let $G = (V, E)$ be a graph on n vertices without $(k+1)$-cliques and with a maximal number of edges. This graph certainly contains k-cliques, since otherwise we could add edges. Let A be a k-clique, and set $B = V \setminus A$.

Since each two vertices of A are joined by an edge, A contains $e_A = \binom{k}{2}$ edges. Let e_B be the number of edges joining the vertices of B and $e_{A,B}$ the number of edges between A and B. By induction, we have

$$e_B \leq \left(1 - \frac{1}{k}\right) \frac{(n-k)^2}{2}.$$

Since G has no $(k+1)$-clique, every $x \in B$ is adjacent to at most $k-1$ vertices in A, and we obtain

$$e_{A,B} \leq (k-1)(n-k).$$

Summing up and using the identity

$$\left(1-\frac{1}{k}\right)\frac{n^2}{2} = \binom{k}{2}\left(\frac{n}{k}\right)^2$$

we conclude that

$$\begin{aligned} |E| \leq e_A + e_B + e_{A,B} &\leq \binom{k}{2} + \binom{k}{2}\left(\frac{n-k}{k}\right)^2 + (k-1)(n-k) \\ &= \binom{k}{2}\left(1+\frac{n-k}{k}\right)^2 = \left(1-\frac{1}{k}\right)\frac{n^2}{2}. \qquad \square \end{aligned}$$

An n-vertex graph $T(n,k)$ that does not contain any $(k+1)$-clique may be formed by partitioning the set of vertices into k parts of equal or nearly-equal size, and connecting two vertices by an edge whenever they belong to two different parts. Thus, Turán's theorem states that the graph $T(n,k)$ has the largest number of edges among all n-vertex graphs without $(k+1)$-cliques.

4.5 Dirichlet's theorem

Here is the application of the pigeonhole principle which Dirichlet made, resulting in his name being attached to the principle. It concerns the existence of good rational approximations to irrational numbers. The result belongs to number theory, but the argument is combinatorial.

Theorem 4.9 (Dirichlet 1879)**.** *Let x be a real number. For any natural number n, there is a rational number p/q such that $1 \leq q \leq n$ and*

$$\left| x - \frac{p}{q} \right| < \frac{1}{nq} \leq \frac{1}{q^2}.$$

Note that it is easy to get an approximation whose error is at most $1/n$, by fxing the denominator to be $q = n$. The improved approximation uses the pigeonhole principle.

Proof. For this proof, we let $\{x\}$ denote the *fractional part* of the real number x, that is, $\{x\} := x - \lfloor x \rfloor$. Consider the $n+1$ numbers $\{ax\}$, $a = 1, 2, \ldots, n+1$. We put these numbers into the n pigeonholes

$$[0, 1/n), [1/n, 2/n), \ldots, [1-1/n, 1)\ .$$

By the pigeonhole principle, some interval contains more than one of the numbers, say $\{ax\}$ and $\{bx\}$ with $a > b$, which therefore differ by less than

$1/n$. Letting $q = a - b$, we see that there exists an integer $p = \lfloor ax \rfloor - \lfloor bx \rfloor$ such that $|qx - p| < 1/n$, from which the result follows on division by q. Moreover, q is the difference between two integers in the range $1, \ldots, n+1$, so $q \leq n$. □

4.6 Swell-colored graphs

Let us color the edges of the complete graph K_n on n vertices. We say that the graph is *swell-colored* if each triangle contains exactly 1 or 3 colors, but never 2 colors and if the graph contains more than one color. That is, we must use at least two colors, and for every triangle, either all its three edges have the same color or each of them has a different color.

It can be shown (do this!) that K_n can never be swell-colored with exactly two colors. A simple investigation shows that K_3 and K_4 are the only K_n swell-colorable with 3 colors; the other K_n require more colors since they are more highly connected.

Using the pigeonhole principle we can prove the following lower bound.

Theorem 4.10 (Ward–Szabó 1994). *The complete graph on n vertices cannot be swell-colored with fewer than $\sqrt{n} + 1$ colors.*

Proof. Let K_n be swell-colored with r distinct colors. Let $N(x, c)$ denote the number of edges incident to vertex x which have color c. Fix x_0 and c_0 for which $N(x_0, c_0)$ is maximal, and denote this maximum by N.

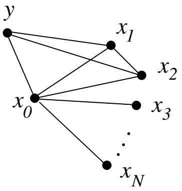
The $n - 1$ edges incident to x_0 can be partitioned into $\leq r$ color classes, each of which with N or fewer members. By the pigeonhole principle,

$$N \cdot r \geq n - 1.$$

Let $x_1, x_2, \ldots, x_N$ be the vertices connected to x_0 by the N edges of color c_0. Let G denote the (complete) subgraph of K_n induced by the vertex set $\{x_0, x_1, \ldots, x_N\}$. The swell-coloredness of K_n is inherited by G and so all edges of G have color c_0. Since K_n is assumed to have at least two colors, there must be some vertex y of K_n not in subgraph G and such that at least one edge joining y to G has a color different from c_0.

Claim 4.11. The $N + 1$ edges connecting y to G all are distinctly colored with colors other than c_0.

The claim implies that $r \geq N + 2$, which together with $N \cdot r \geq n - 1$ yields $r(r - 2) \geq n - 1$, and hence, $r \geq \sqrt{n} + 1$, as desired. So, it remains to prove the claim.



If an edge connecting y to G, say $\{y, x_1\}$ (see the figure above), has color c_0 then by the swell-coloredness of G, edge $\{y, x_0\}$ would have color c_0, contrary to the definition of x_0 (recall that $x_1, x_2, \ldots, x_N$ are *all* the edges incident to x_0 and colored by c_0). Furthermore, if any two edges connecting y to G, say $\{y, x_1\}$ and $\{y, x_2\}$, have the same color, then the swell-coloredness of K_n implies that the edge $\{x_1, x_2\}$ shares the same color. But $\{x_1, x_2\}$ belongs to G, and hence has color c_0 and so $\{y, x_1\}$ would have color c_0 which we have seen is impossible. This completes the proof of the claim, and thus, of the theorem. □

The optimality of the lower bound given by Theorem 4.10, can be shown using a configuration known as "affine plane." We will investigate these configurations in Chap. 12. For our current purposes it is enough to know that an *affine plane* $\mathrm{AG}(2, q)$ of order q contains exactly q^2 points and exactly $q+1$ classes (also called "pencils") of parallel lines, each containing q lines (two lines are parallel if they share no point). Moreover, each two points lie on a unique line.

Having such a plane, we can construct a swell-coloring of K_{q^2} with $q+1$ colors as follows. Identify the vertices of K_{q^2} with the points in $\mathrm{AG}(2, q)$ and associate some unique color with each of the $q+1$ pencils of parallel lines. In order to define a swell-coloring, consider two distinct vertices x and y of K_{q^2}. These points lie on a unique line which, in its turn, belongs to exactly one of the pencils. Color the edge $\{x, y\}$ with the color of this pencil. Since any two points lie on a unique line and parallel lines do not meet in a point, all three edges of a triangle will receive different colors, and hence, the coloring is swell, as desired.

In fact, Ward and Szabó (1994) have proved that the converse also holds: if the graph K_{q^2} $(q \geq 2)$ can be swell-colored using $q+1$ colors then this coloring can be used to construct an affine plane of order q.

4.7 The weight shifting argument

A version of the pigeonhole principle is the *averaging principle* which we formulated in Sect. 1.5: *every set of numbers contains a number at least as large as the average (and one at least as small).*

Trying to show that some "good" object exists, we can try to assign objects their "weights" so that objects with a large enough (or small enough) weight are good, and try to show that the average weight is large (or small). The averaging principle then guarantees that at least one of the objects is good. The main difficulty is to *define* the weights relevant for the desired application. After this we face the problem of how to *compute* the weights and accumulate their sum. At this step the so-called "shifting argument" can help. Let us illustrate this by three examples (the first is trivial, whereas the next two are not).

Proposition 4.12. *Let $n \leq m < 2n$. Then for any distribution of m pigeons among n pigeonholes so that no hole is left empty, at most $2(m-n)$ of the pigeons will be happy, i.e., will sit not alone in their holes.*

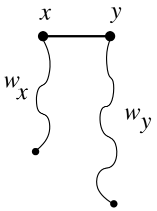
Proof. If some hole contains more than two pigeons then, by removing a pigeon from this hole and placing it in a hole which had contained exactly one pigeon, we arrive to a new distribution with one more happy pigeon. Thus, the maximum number of happy pigeons is achieved when each hole has at most two pigeons, and in this case this number is $\leq 2(m-n)$, as desired. □

A *trail* in a graph is a walk without repeated edges.

Theorem 4.13 (Graham–Kleitman 1973)**.** *If the edges of a complete graph on n vertices are labeled arbitrarily with the integers $1, 2, \ldots, \binom{n}{2}$, each edge receiving its own integer, then there is a trail of length at least $n-1$ with an increasing sequence of edge-labels.*

Proof. To each vertex x, assign its weight w_x equal to the length of the longest increasing trail ending at x. If we can show that $\sum_x w_x \geq n(n-1)$, then the averaging principle guarantees a vertex with a large enough weight.

We accumulate the weights and their sum iteratively, growing the graph from the trivial graph; at each step we add a new edge whose label is *minimal* among the remaining ones. Initially, the graph has no edges, and the weights are all 0. At the i-th step we take a new edge $e = \{x, y\}$ labeled by i. Let w_x and w_y be the weights of x and y accumulated so far.



If $w_x = w_y$ then increase both weights by 1. If $w_x < w_y$ then the edge e prolongs the longest increasing trail ending at y by 1; so the new weights are $w'_x = w_y + 1$ and $w'_y = w_y$. In either case, when an edge is added, the sum of the weights of the vertices increases by at least 2. Therefore, when all the $\binom{n}{2}$ steps are finished, the sum of the vertex weights is at least $n(n-1)$, as desired. □

Finally, we illustrate the shifting argument by the fourth proof of Mantel's theorem: If a graph G on $2n$ vertices contains n^2+1 edges, then G contains a triangle.

Fourth proof of Mantel's theorem (Motzkin–Straus 1965). Let G be a graph on $2n$ vertices, and let m be the number of edges in G. Assume that G has no triangles. Our goal is to prove that then $m \leq n^2$. We assign a nonnegative w_x to each vertex x such that $\sum_x w_x = 1$. We seek to maximize

$$S := \sum w_x w_y,$$

where the sum is taken over all edges $\{x, y\}$ of G. One way of assigning the weights is to let $w_x = 1/(2n)$ for each x. This gives

$$S \geq \frac{m}{(2n)^2}. \tag{4.2}$$

We are going to show that, on the other hand, S never exceeds $1/4$, which together with the previous lower bound will imply that $m \leq n^2$, as desired.

And now comes the "shifting argument." Suppose that x and y are two nonadjacent vertices and W_x and W_y are the total weights of vertices connected to x and y, respectively. Suppose also that $W_x \geq W_y$. Then for any $\epsilon \geq 0$,

$$(w_x + \epsilon)W_x + (w_y - \epsilon)W_y \geq w_x W_x + w_y W_y.$$

This, in particular, means that we do not decrease the value of S if we shift all of the weight of vertex y to the vertex x. It follows that S is maximized when all of the weight is concentrated on a complete subgraph of G. But we have assumed that G has no triangles; so G cannot have complete subgraphs other than single edges. Hence, S is maximized when all of the weight is concentrated on two adjacent vertices, say x and y. Therefore

$$S \leq \max\left\{w_x \cdot w_y \;:\; w_x + w_y = 1\right\} = 1/4$$

which, together with (4.2), yield the desired upper bound $m \leq n^2$. □

4.8 Schur's theorem

The famous Fermat's Last Theorem states that if $n > 2$, then $x^n + y^n = z^n$ has no solutions in nonzero integers x, y and z. This theorem was first conjectured by Pierre de Fermat in 1637, but was not proven until 1995 despite the efforts of many mathematicians. The last step in its proof was done by Andrew Wiles.

As early as 1916, Issai Schur used the pigeonhole principle to show that Fermat's Last Theorem is false in the finite field $\mathbb{Z}_p$ for any sufficiently large prime p. He derived this from the following combinatorial result about colorings of numbers. The result may perhaps be considered as the earliest result in Ramsey theory.

An r-*coloring* of a set assigns one of the colors $1, 2, \ldots, r$ to each element of the set.

Theorem 4.14 (Schur 1916). *For any $r \geq 2$ and for any r-coloring of $\{1, 2, \ldots, n\}$, where $n = \lceil \mathrm{e}r! \rceil$, there are three integers x, y, z of the same color and such that $x + y = z$.*

Proof. Let $\chi : \{1, \ldots, n\} \to \{1, \ldots, r\}$ be an r-coloring of the first n positive integers. Assume that there *do not* exist positive integers x, y with $x + y \leq n$ such that $\chi(x) = \chi(y) = \chi(x + y)$. Our goal is to show that then $n < \mathrm{e}r!$.

Let c_0 be a color which appears most frequently among the n elements, and let $x_0 < x_1 < \ldots < x_{n_1 - 1}$ be the elements of color c_0. By the pigeonhole principle, we know that $n \leq r n_1$,

Consider the set $A_0 = \{x_i - x_0 : 1 \leq i < n_1\}$. By our assumption, no number in A_0 can receive color c_0. So, the set A_0 is colored by $r - 1$ colors. Let c_1 be a color which appears most frequently among the elements of A_0, and let $y_0 < y_1 < \ldots < y_{n_2 - 1}$ be its elements of color c_1. Observe that $n_1 - 1 \leq (r - 1) n_2$.

Consider the set $A_1 = \{y_i - y_0 : 1 \leq i < n_2\}$. By the assumption, no number in A_1 can receive any of colors c_0 and c_1. So, the set A_1 is colored by $r - 2$ colors. Let c_2 be a color which appears most frequently among the elements of A_1, and let $z_0 < z_1 < \ldots < z_{n_3 - 1}$ be its elements of color c_2. Observe that $n_2 - 1 \leq (r - 2) n_3$.

Continue this procedure until some n_k becomes 1. Since we have only r colors, this happens at the latest for $k = r$. Thus, we obtained the inequalities $n \leq r n_1$ and $n_i \leq (r - i) n_{i+1} + 1$ for $i = 1, \ldots, k - 1$, with $n_k = 1$. Putting them together we obtain that

$$n \leq \sum_{i=0}^{r-1} r(r-1)(r-2)\cdots(r-i) = \sum_{i=0}^{r-1} \frac{r!}{i!} < r! \sum_{i=0}^{\infty} \frac{1}{i!} = \mathrm{e}r! \qquad \square$$

Schur (1916) used Theorem 4.14 to show that Fermat's Last Theorem is false in the finite field $\mathbb{Z}_p$ for any sufficiently large prime p.

Theorem 4.15. *For every integer $n \geq 1$, there exists p_0 such that for any prime $p \geq p_0$, the congruence*

$$x^n + y^n = z^n \mod p$$

has a solution.

Proof. The multiplicative group $\mathbb{Z}_p^* = \{1, 2, \ldots, p - 1\}$ is known to be cyclic and hence it has a generator g. Each element of $\mathbb{Z}_p^*$ can be written as $x = g^{nj+i}$ where $0 \leq i < n$. We color the elements of $\mathbb{Z}_p^*$ by n colors, where $\chi(x) = i$

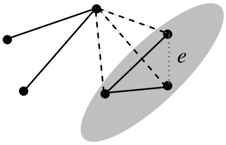


Fig. 4.2 What is the color of e?

if $x = g^{nj+i}$. By Schur's theorem, for p sufficiently large, there are elements $x', y', z' \in \mathbb{Z}_p^*$ such that $x' + y' = z'$ and $\chi(x') = \chi(y') = \chi(z')$. Therefore, $x' = g^{nj_x+i}$, $y' = g^{nj_y+i}$, $z' = g^{nj_z+i}$ and

$$g^{nj_x+i} + g^{nj_y+i} = g^{nj_z+i}\,.$$

Setting $x = g^{j_x}$, $y = g^{j_y}$ and $z = g^{j_z}$, we get a solution of $x^n + y^n = z^n$ in $\mathbb{Z}_p^*$. □

4.9 Ramseyan theorems for graphs

How many people can we invite to a party where among each three people there are two who know each other and two who don't know each other? It turns out that at most five persons can attend such a party.

To show this, let us consider the following simple game. Mark six points on the paper, no three in line. There are two players; one has a Red pencil the other Blue. Each player's turn consists in drawing a line with his/her pencil between two of the points which haven't already been joined. (The crossing of lines is allowed). The player's goal is to create a triangle in his/her color. If you try to play it with a friend, you will notice that it always end in a win for one player: a draw is not possible. Prove this! (Hint: see Fig. 4.2.)

We can generalize this argument to *arbitrary* graphs, not only those with up to six vertices.

Let $G = (V, E)$ be an undirected graph. A subset $S \subseteq V$ is a *clique* of G if any two vertices of S are adjacent. Similarly, a subset $T \subseteq V$ is an *independent set* of G if no two vertices of T are adjacent in G.

For integers $s, t \geq 1$, let $R(s, t)$ denote the smallest number n such that in any(!) graph on n or more vertices, there exists either a clique of s vertices or an independent set of t vertices.

Theorem 4.16.

$$R(s,t) \leq \binom{s+t-2}{s-1} = \binom{s+t-2}{t-1}.$$

Proof. By induction on $s + t$. It is clear form the definition that $R(1, t) = R(s, 1) = 1$. For $s > 1$ and $t > 1$, let us prove that

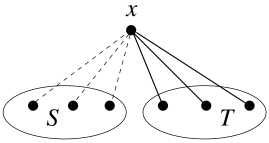


Fig. 4.3 Splitting the graph into neighbors and non-neighbors of x

$$R(s,t) \leq R(s,t-1) + R(s-1,t). \tag{4.3}$$

Let $G = (V, E)$ be a graph on $n = R(s,t-1) + R(s-1,t)$ vertices. Take an arbitrary vertex $x \in V$, and split $V \setminus \{x\}$ into two subsets S and T, where each vertex of S is nonadjacent to x and each vertex of T is adjacent to x (see Fig. 4.3). Since

$$R(s,t-1) + R(s-1,t) = |S| + |T| + 1,$$

we have either $|S| \geq R(s,t-1)$ or $|T| \geq R(s-1,t)$.

Let $|S| \geq R(s,t-1)$, and consider the induced subgraph $G[S]$ of G: this is a graph on vertices S, in which two vertices are adjacent if and only if they are such in G. Since the graph $G[S]$ has at least $R(s,t-1)$ vertices, by the induction hypothesis, it contains either a clique on s vertices or an independent set of $t-1$ vertices. Moreover, we know that x is not adjacent to any vertex of S in G. By adding this vertex to S, we conclude that the subgraph $G[S \cup \{x\}]$ (and hence, the graph G itself) contains either a clique of s vertices or an independent set of t vertices. The case when $|T| \geq R(s-1,t)$ is analogous.

Since $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$ (see Proposition 1.3), the recurrence (4.3) implies

$$R(s,t) \leq R(s,t-1)+R(s-1,t) \leq \binom{s+t-3}{s-1}+\binom{s+t-3}{s-2} = \binom{s+t-2}{s-1}.$$

□

We have proved Theorem 4.16 by induction on $s+t$. The same result can also be proved using so-called *induced coloring argument.* This argument is encountered frequently in Ramsey theory. To explain the idea, let us prove the following weaker bound for $s = t$:

$$R(t,t) \leq 2^{2t}\,.$$

That is, any graph on 4^t or more vertices must contain either a clique or an independent set on t vertices.

Proof via induced coloring argument. Take a complete graph on 2^{2t} vertices, and fix an arbitrary coloring of its edges in red and blue. Let us suppose for convenience that the vertices are totally ordered. Let x_1 be the first vertex.

Then by the pigeonhole principle there is a set of vertices S_1 of size at least 2^{2t-1} such that every edge from x_1 to S_1 has the same color. Now let x_2 be the least vertex of S_1. By the pigeonhole principle again there is a set $S_2 \subseteq S_1$ of size at least 2^{2t-2} such that every edge from x_2 to S_2 has the same color. Continuing this process, we obtain a sequence $x_1, \ldots, x_{2t}$ of vertices and a sequence $S_0 \supset S_1 \supset S_2 \supset \ldots \supset S_{2t}$ of sets such that $x_i \in S_{i-1}$ for every i, and every edge from x_i to S_i has the same color. (Here S_0 is the set of all vertices.) It follows that the color of the edge joining x_i to x_j depends only on $\min\{i, j\}$. That is, for each $i = 0, 1, \ldots, 2t-1$, all edges joining x_i with the subsequent vertices $x_{i+1}, \ldots, x_{2t}$ have the same color $c_i \in \{\text{red, blue}\}$. Since we have $2t$ distinct values for i and only two colors, the pigeonhole principle implies that there must be a subset $T \subseteq \{1, \ldots, 2t\}$ of size $|T| \geq (2t)/2 = t$ such that $c_i = c_j$ for all $i, j \in T$. Thus, all edges joining vertices in $\{x_i : i \in T\}$ have the same color. □

A simple probabilistic argument yields the following *lower* bound.

Theorem 4.17 (Erdős 1947). *$R(t,t) > 2^{t/2}$ for all $t \geq 3$.*

That is, the edges of K_n can be colored in two colors so that we get no monochromatic $K_{2\log n}$.

Proof. Consider a random 2-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equally likely. For any fixed set T of t vertices, the probability that all $\binom{t}{2}$ edges between these vertices receive the same color (i.e., that either all edges are red or they are blue) is $2 \cdot 2^{-\binom{t}{2}}$. The number of t-element subsets of vertices $\binom{n}{t}$ and therefore the probability that there is at least one monochromatic t-clique is at most

$$\binom{n}{t} \cdot 2^{1-\binom{t}{2}} < \frac{n^t}{t!} \cdot \frac{2^{1+t/2}}{2^{t^2/2}},$$

which is < 1 if $n \leq 2^{t/2}$ and $t \geq 3$. □

Using Stirling's Formula, the lower bound on $R(t,t)$ can be improved to about $t2^{t/2}$. On the other hand, Theorem 4.16 gives an upper bound $\binom{2t}{t}$ on $R(t+1, t+1)$. This bound was recently improved by Conlon (2009) to about $t^{-\ell}\binom{2t}{t}$ with $\ell \geq c \log n / \log\log n$. The gap is still large, and tight bounds are known only for $s = 3$:

$$c_1 \frac{t^2}{\log t} \leq R(3,t) \leq c_2 \frac{t^2}{\log t}.$$

The upper bound is due to Ajtai, Komlós, and Szemerédi (1980) and the lower bound was proved by Kim (1995) using a probabilistic argument.

In the case of *bipartite* graphs the following bounds are known. Let $b(t)$ be the smallest number n such that, in any two-coloring of the complete bipartite $n \times n$ graph $K_{n,n}$ there is a monochromatic $K_{t,t}$. The best known lower bound $b(t) = \Omega(k2^{k/2})$ is the same as for ordinary graphs. The best known upper bound $b(t) = O(2^k \log k)$ was proved by Conlon (2008).

4.10 Ramsey's theorem for sets

We now consider colorings of k-element subsets of $[n]$ for $k > 2$. The Ramsey theorem for graphs (we just proved) speaks about the case $k = 2$: for every $s \geq 1$, there is an n such that it is not possible to color 2-element subsets of $[n]$ (edges) in red and blue so that every s-element subset of $[n]$ will contain two 2-element subsets of different colors. (In this case $n \geq 4^s$ is enough.) In its unabridged form, the celebrated result of Ramsey (1930) speaks about colorings of larger subsets using any number of colors.

Theorem 4.18 (Ramsey's theorem)**.** *For every natural numbers $1 \leq k \leq s$ and $r \geq 2$ there exists a natural number $n = R_r(k; s)$ such that whenever k-subsets of $[n]$ are colored in r colors, there is an s-subset of $[n]$ whose all k-subsets receive the same color.*

Proof. We first observe that it is enough to consider the case of $r = 2$ colors.

Claim 4.19. $R_{r+1}(k; s) \leq R_r\,(k;\ R_2(k; s))$.

Proof. Let $N = R_r\,(k;\ R_2(k; s))$ and let an arbitrary coloring of k-subsets of an N-element set X with $r + 1$ colors $0, 1, \ldots, r$ be given. Then consider this as an r-coloring simply by identifying the colors 0 and 1. (This is known as the "mixing colors" trick.) By the choice of N, either there exists an $R_2(k; s)$-element subset, all whose k-subsets receive one of the colors $2, \ldots, r$ (and we are done), or there exists an $R_2(k; s)$-element subset Y with each its k-subsets in color 0 or 1. According to the size of Y, all k-subsets of some its s-element subset must be monochromatic. □

By Claim 4.19, it is enough to show that $R_2(k; s)$ exists.

In order to argue by induction, we define a more "granulated" version of the Ramsey number $R_2(k; s)$. Namely let $R(k; s, t)$ be the smallest number n with the following property: If k-subsets of an n-set are colored with two colors 0 and 1, then all k-subsets of some s-subset receive color 0 or all k-subsets of some t-subset receive color 1. Thus, the theorem claims that $R(k; s, s)$ exists for all $s \geq k$. We will prove a stronger statement that $R(k; s, t) \leq n$, where

$$n := R\,(k - 1; R(k; s - 1, t), R(k; s, t - 1)) + 1.$$

We prove this recurrence by induction on k and on s, t. Observe that, by the pigeonhole principle, $R(1; s, t) = s + t - 1$ for all s and t and, moreover, $R(k; x, k) = R(k; k, x) = x$ for all k and $x \geq k$. By induction, we may assume that the numbers $R(k; s-1, t)$ and $R(k; s, t-1)$ exist, and take an arbitrary n-element set X, where n is defined above.

Let χ be a coloring of k-subsets of X with two colors 0 and 1. Fix a point $x \in X$, and let $X' := X \setminus \{x\}$. We define a new coloring χ' of the $(k-1)$-subsets A of X' by

$$\chi'(A) := \chi\left(A \cup \{x\}\right).$$

By the choice of n and by symmetry, we can assume to have found a subset $Y \subseteq X'$ such that $|Y| = R(k; s-1, t)$ and

$$\chi'(A) = 0 \text{ for all } (k-1)\text{-subsets } A \text{ of } Y.$$

Now consider how the original coloring χ acts on the k-subsets of Y. According to its size, the set Y must either contain a t-element subset, all whose k-subsets receive color 1 (and we are done), or it must contain an $(s-1)$-element subset Z, all whose k-subsets receive color 0. In this last case consider the s-element subset $Z \cup \{x\}$ and take an arbitrary its subset B of size k. If $x \not\in B$ then B is a k-element subset of Z, and hence, $\chi(B) = 0$. If $x \in B$ then the set $A = B \setminus \{x\}$ is a $(k-1)$-subset of Y, and hence again, $\chi(B) = \chi\left(A \cup \{x\}\right) = \chi'(A) = 0$. □

One of the earliest and most popular applications of Ramsey's theorem is due to Erdős and Szekeres (1935). In fact, this application was a first step in popularizing Ramsey's theorem.

Theorem 4.20 (Erdős–Szekeres 1935)**.** *Let $m \geq 3$ be a positive integer. Then there exists a positive integer n such that any set of n points in the Euclidean plane, no three of which are collinear, contains m points which are the vertices of a convex m-gon.*

Proof (due to Johnson 1986). Choose $n = R_2(3; m)$, the number from the Ramsey's Theorem 4.18, and let A be any set of n points in the plane, no three of which are collinear (i.e., lie on a line). For $a, b, c \in A$, let $|abc|$ denote the number of points of A which lie in the interior of the triangle spanned by a, b and c. Define the 2-coloring χ of triples of points in A by $\chi(a, b, c) = 0$ if $|abc|$ is even and $\chi(a, b, c) = 1$ otherwise. By the choice of n, there exists an m-element subset $B \subseteq A$ such that all its 3-element subsets receive the same color. Then the points of B form a convex m-gon. Otherwise, there would be four points $a, b, c, d \in B$ such that d lies in the interior of the triangle abc (see Fig. 4.4). Since no three points of B are collinear, we have $|abc| = |abd| + |acd| + |bcd| + 1$, contradicting that the coloring χ is constant on all triples from B. □

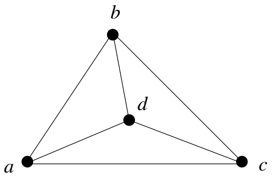


Fig. 4.4 Point d lies in none of the lines ab, bc and ac.

Exercises

4.1. Suppose five points are chosen inside an equilateral triangle with side-length 1. Show that there is at least one pair of points whose distance apart is at most 1/2. *Hint*: Divide the triangle into four suitable boxes.

4.2. (D.R. Karger). Jellybeans of 8 different colors are in 6 jars. There are 20 jellybeans of each color. Use the pigeonhole principle to prove that there must be a jar containing two pairs of jellybeans from two different colors of jellybeans. *Hint*: For each color there is a jar containing a pair of jellybeans of that color, and we have more colors than jars.

4.3. Show that for any positive integer n, there is a multiple of n that contains only the digits 7 or 0. *Hint*: Consider the values modulo n of all the numbers a_i of the form $77\ldots7$, with i sevens, $i = 1, \ldots, n+1$.

4.4. Prove that every set of $n+1$ distinct integers chosen from $\{1, 2, \ldots, 2n\}$ contains a pair of consecutive numbers and a pair whose sum is $2n+1$. For each n, exhibit two sets of size n to show that these results are the best possible. *Hint*: Use pigeonholes $(2i, 2i-1)$ and $(i, 2n-i+1)$, $i = 1, \ldots, n$.

4.5. Prove that every set of $n+1$ distinct integers chosen from $\{1, 2, \ldots, 2n\}$ contains two numbers such that one divides the other. *Sketch*: (due to Lajos Pósa): Write every number x in the form $x = k_x 2^a$, where k_x is an odd number between 1 and $2n-1$. Take odd pigeonholes $1, 3, 5, \ldots, 2n-1$ and put x into the pigeonhole number k_x. Some hole must have two numbers $x < y$.

4.6. Coin-weighing problem (Erdős–Spencer 1974). Let n coins of weights 0 and 1 be given. We are also given a scale with which we may weigh any subset of the coins. The information from previous weighings may be used. The object is to determine the weights of the coins with the minimal number of weighings. Formally, the problem may be stated as follows. A collection $S_1, \ldots, S_m$ of subsets of $[n]$ is called *determining* if an arbitrary subset T of $[n]$ can be uniquely determined by the cardinalities $|S_i \cap T|$, $1 \leq i \leq m$. Let $D(n)$ be the minimum m for which such a determining collection exists. By weighting each coin separately ($S_i = \{i\}$) we see that $D(n) \leq n$. Show that $D(n) \geq n/(\log_2(n+1))$. *Hint*: Take a determining collection $S_1, \ldots, S_m$, observe that for each i there are only $n+1$ possible $|S_i \cap T|$, and apply the pigeonhole principle.

4.7. Suppose that n is a multiple of k. Construct a graph without $(k+1)$-cliques, in which the number of edges achieves the upper bound (4.1) given by Turán's theorem. *Hint*: Split the n vertices into k equal size parts and join all pairs of vertices from different parts (this is a complete k-partite graph).

4.8. Recall that the *independence number* $\alpha(G)$ of a graph G is the maximum number of pairwise nonadjacent vertices of G. Prove the following dual version of Turán's theorem: if G is a graph with n vertices and $nk/2$ edges, $k \geq 1$, then $\alpha(G) \geq n/(k+1)$.

4.9. (Motzkin–Straus 1965). Prove Turán's theorem using the shifting argument described in the fourth proof of Mantel's theorem. *Hint*: Let G be a graph with n vertices and m edges, and suppose that G has no $(k+1)$-clique. Assign weights w_x to the vertices as before. Setting $w_x = 1/n$ for all vertices, we obtain $S \geq m/n^2$. On the other hand, the same shifting argument yields that the weight is concentrated on some clique U with $|U| = t \leq k$ vertices. Setting $w_x = 1/t$ for $x \in U$, and $w_x = 0$ otherwise, the total weight becomes $\binom{t}{2}/t^2 = (1-1/t)/2$. Since this expression is increasing in t, the best we can do is to set $t = k$.

4.10. Derive the Erdős–Szekeres theorem from Lemma 4.6. *Hint*: Given a sequence $A = (a_1, \ldots, a_n)$ of $n \geq rs+1$ real numbers, define a partial order $\preccurlyeq$ on A by $a_i \preccurlyeq a_j$ if $a_i \leq a_j$ and $i \leq j$, and apply Dilworth's lemma.

4.11. Let n^2+1 points be given in $\mathbb{R}^2$. Prove that there is a sequence of $n+1$ points $(x_1, y_1), \ldots, (x_{n+1}, y_{n+1})$ for which $x_1 \leq x_2 \leq \cdots \leq x_{n+1}$ and $y_1 \geq y_2 \geq \cdots \geq y_{n+1}$, or a sequence of $n+1$ points for which $x_1 \leq x_2 \leq \cdots \leq x_{n+1}$ and $y_1 \leq y_2 \leq \cdots \leq y_{n+1}$.

4.12. Show that, if $n > srp$, then any sequence of n real numbers must contain either a strictly increasing subsequence of length greater than s, a strictly decreasing subsequence of length greater than r, or a constant subsequence of length greater than p. *Hint*: By the pigeonhole principle, if only sr or fewer distinct values occur, then some value must be taken by more than p numbers in the sequence. Otherwise, we can argue as in the Erdős–Szekeres theorem.

4.13. Let $0 < a_1 < a_2 < \cdots < a_{sr+1}$ be $sr+1$ integers. Prove that we can select either $s+1$ of them, no one of which divides any other, or $r+1$ of them, each dividing the following one. *Hint*: Apply Dilworth's lemma.

4.14. Show that the bound in the Erdős–Szekeres' theorem is best possible. *Hint*: Consider the sequence $A = (B_{s-1}, B_{s-2}, \ldots, B_0)$, where

$$B_i = (ir+1, ir+2, \ldots, ir+r).$$

4.15. Use the pigeonhole principle to prove the following fact, known as *Chinese remainder theorem*. Let $a_1, \ldots, a_k, b$ be integers, and $m = m_1 \cdots m_k$ where m_i and m_j are relatively prime, for all $i \neq j$. Then there exists exactly one integer a, $b \leq a < b+m$, such that $a \equiv a_i \mod m_i$ for all $i = 1, \ldots, k$. *Hint*: The integers $x \in \{b, b+1, \ldots b+m-1\}$ are different modulo m; hence their residues $(x \bmod m_1, \ldots, x \bmod m_k)$ run through all m possible values.

4.16. (Moon–Moser 1962). Let $G = (V, E)$ be a graph on n vertices and $t(G)$ the number of triangles in it. Show that

$$t(G) \geq \frac{|E|}{3n} \left(4 \cdot |E| - n^2\right).$$

Hint: For an edge $e = \{x, y\}$, let $t(e)$ be the number of triangles containing e. Let $B = V \setminus \{x, y\}$. Among the vertices in B there are precisely $t(e)$ vertices which are adjacent to both x and y. Every other vertex in B is adjacent to at most one of these two vertices. We thus obtain $d(x) + d(y) - t(e) \leq n$. Summing over all edges $e = \{x, y\}$ we obtain

$$\sum_{e \in E} (d(x) + d(y)) - \sum_{e \in E} t(e) \leq n \cdot |E|.$$

Apply the Cauchy–Schwarz inequality to estimate the first sum.

Comment: This implies that a graph G on an even number n of vertices with $|E| = n^2/4 + 1$ edges not only contains one triangle (as it must be by Mantel's theorem), but more than $n/3$.

4.17. (Goodman 1959). Let G be a graph with n vertices and m edges. Let $t(G)$ denote the number of triangles contained in the graph G or in its complement. Prove that

$$t(G) \geq \binom{n}{3} + \frac{2m^2}{n} - m(n-1).$$

Hint: Let t_i be the number of triples of vertices $\{i, j, k\}$ such that the vertex i is adjacent to precisely one of j or k. Observe that $t(G) \geq \binom{n}{3} - \frac{1}{2}\sum_i t_i$ and that $t_i = d_i(n-1-d_i)$, where d_i is the degree of the vertex i in G. Use the Cauchy–Schwarz inequality (13.3) and Euler's theorem (Theorem 1.8) to show that $\sum d_i^2 \geq \frac{1}{n}\left(\sum d_i\right)^2 = \frac{4m^2}{n}$.

4.18. A set $S \subseteq V$ of vertices in a graph $G = (V, E)$ *spans* an edge $e \in E$ if both endpoints of e belong to S. Say that a graph is (k, r)-*sparse* if every subset of k vertices spans at most r of its edges. Turán's theorem (Theorem 4.8) gives an upper bound on the maximal possible number of edges in a (k, r)-sparse graph for $r = \binom{k}{2} - 1$. Show that every (k, r)-sparse graph on n vertices has at most $\alpha \cdot \binom{n}{2}$ edges, where $\alpha = r \cdot \binom{k}{2}^{-1}$. *Hint*: Observe that every edge is spanned by precisely $\binom{n-2}{k-2}$ of k-element subsets and use Exercise 1.12.

4.19. Color all non-empty subsets (not the points!) of $[n] = \{1, \ldots, n\}$ with r colors. Prove that, if n is large enough, then there are two disjoint non-empty subsets A, B such that A, B and $A \cup B$ have the same color. *Hint*: Take $n = R_r(2; 3)$. Assume the non-empty subsets of $[n]$ are colored with r colors. Now color each pair $\{i, j\}$ $(1 \leq i < j \leq n)$ by the color of the interval $\{i, i+1, \ldots, j-1\}$. By Theorem 4.18, there exists a monochromatic triangle $x < y < z$. Take $A = \{x, x+1, \ldots, y-1\}$ and $B = \{y, y+1, \ldots, z-1\}$.

4.20. Show that for every $r \geq 2$ there exists a constant $c = c(r)$ such that, if n is large enough, then for every r-coloring of the points $1, \ldots, n$, at least $c \cdot n^2$

of the pairs $\{i, j\}$ of points will receive the same color. *Hint*: By the pigeonhole principle, every $(r+1)$-subset of points contributes at least one monochromatic pair, and every pair is contained only in $\binom{n-2}{r-1}$ of such subsets.

4.21. Prove that $R(3,4) \leq 9$. *Hint*: Color the edges of K_9 in red and blue, and assume that there are no red triangles and no blue 4-cliques. Then each vertex is incident to precisely three red edges and five blue edges. Thus, there are exactly $(9 \cdot 3)/2$ many red edges. But this should be an integer!

4.22. Derive the following weaker version of Schur's theorem (Theorem 4.14) from Ramsey's theorem (Theorem 4.18): For any $r \geq 2$ there is $n > 3$ such that for any r-coloring of $\{1, 2, \ldots, n\}$, there are three integers of the same color and such that $x + y = z$. *Hint*: Choose $n = R_r(2; 3)$. Given a coloring $\chi : [n] \to [r]$ of the points in $[n]$, consider the coloring χ' of the pairs defined by: $\chi'(\{x, y\}) = \chi(|x - y|)$ What does it means to have a χ'-monochromatic triangle with vertices $x < y < z$?

4.23. Use the previous exercise to show that $R(4,4) \leq 18$. *Hint*: (4.3).

The next exercises are about the *chromatic number* $\chi(G)$ of graphs. Recall that this is the smallest number of colors we need in order to color the vertices of G in such a way that no two adjacent vertices receive the same color.

4.24. Show that any graph G must have at least $\binom{\chi(G)}{2}$ edges.

4.25. Let G_1, G_2 be two graphs. Prove that $\chi(G_1 \cup G_2) \leq \chi(G_1) \cdot \chi(G_2)$. *Hint*: Use pairs of colors to color $G_1 \cup G_2$.

4.26. Let G be a graph on n vertices. A complement $\overline{G}$ of a graph G is a graph on the same set of vertices in which two vertices are adjacent if and only if they are non-adjacent in G. Prove that $\chi(G) \cdot \chi(\overline{G}) \geq n$ and $\chi(G) + \chi(\overline{G}) \geq 2\sqrt{n}$. *Hint*: $(\chi(G) - \chi(\overline{G}))^2 \geq 0$.

4.27. Prove that $\chi(G) \leq \Delta(G) + 1$, where $\Delta(G)$ is the maximum degree of a vertex in G. *Hint*: Order the vertices $v_1, \ldots, v_n$ and use greedy coloring: assign to v_i the smallest-indexed color not already used on its lower-indexed neighbors.

4.28. (Welsh–Powell 1967). Let G be a graph on n vertices, whose degrees are $d_1 \geq d_2 \geq \ldots \geq d_n$. Prove that $\chi(G) \leq 1 + \max_i \min\{d_i, i-1\}$. *Hint*: Apply the greedy algorithm from the previous exercise. When we color the i-th vertex, at most $\min\{d_i, i-1\}$ of its neighbors have already been colored, so its color is at most $1 + \min\{d_i, i-1\}$.

4.29. Let $G = (V, E)$ be a graph and $S \subseteq V$ a subset of its vertices. The induced subgraph of G is the graph $G[S]$ on vertices S, in which two vertices are adjacent if and only if they are such in the original graph G. Prove that for any graph G we can find a partition $V = S \cup T$ of its vertices into two disjoint non-empty subsets S and T such that $\chi(G[S]) + \chi(G[T]) = \chi(G)$.

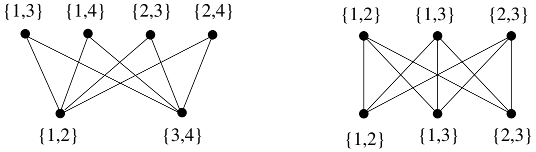


Fig. 4.5 The graphs $K_{2,4}$ and $K_{3,3}$ with a particular lists of color sets

4.30. A graph G is *k-critical* if $\chi(G) = k$ but $\chi(H) < k$ for every proper subgraph H of G. Let $\delta(G)$ denote the minimum degree of a vertex in G. Prove the following: if G is a k-critical graph, then $\delta(G) \geq k-1$. *Hint*: Assume there is a vertex $x \in V$ of degree at most $k-2$, and consider the induced subgraph $H = G[V \setminus \{x\}]$. Graph H must have a legal $(k-1)$-coloring, and at least one of these $k-1$ colors is not used to color the neighbors of x; we can use it for x.

4.31. (Szekeres–Wilf 1968). Prove that $\chi(G) \leq 1 + \max_{H \subseteq G} \delta(H)$ holds for any graph G. *Hint*: Let $k = \chi(G)$, take a k-critical subgraph H of G and use the previous estimate.

4.32. Let G be a directed graph without cycles and suppose that G has no path of length k. Prove that then $\chi(G) \leq k$. *Hint*: Let $c(x)$ denote the maximum length of a path starting from x. Then c is a coloration with colors $0, 1, \ldots, k-1$. Show that it is legal.

4.33. Let G be a graph on n vertices, and $\alpha(G)$ be its *independence number*, i.e., the maximal number of vertices, no two of which are joined by an edge. Show that $n/\alpha(G) \leq \chi(G) \leq n - \alpha(G) + 1$.

4.34. It is clear that $\chi(G) \geq \omega(G)$, where $\omega(G)$ is the *clique number* of G, i.e., the maximum size of a clique in G. Erdős (1947) has proved that, for every large enough n, there exists an n-vertex graph G such that $\omega(G) \leq 2\log_2 n$ and $\omega(\overline{G}) \leq 2\log_2 n$ (see Theorem 4.17 for a proof). Use this result to show that the gap between $\chi(G)$ and $\omega(G)$ can be quite large: the maximum of $\chi(G)/\omega(G)$ over all n-vertex graphs G is $\Omega\left(n/(\log_2 n)^2\right)$. *Hint*: $\chi(G) \geq n/\omega(\overline{G})$.

4.35. Let $G = (V, E)$ be a graph, and $(C_v)_{v \in V}$ be a sequence of (not necessarily disjoint) sets. We can look at each set C_v as a color set (or a "palette") for the vertex v. Given such a list of color sets, we consider only colorings c such that $c(v) \in C_v$ for all $v \in V$, and call them *list colorings* of G. As before, a coloring is *legal* if no two adjacent vertices receive the same color. The *list chromatic number* $\chi_\ell(G)$ is the smallest number k such that for *any* list of color sets C_v with $|C_v| = k$ for all $v \in V$, there always exists a legal list coloring of G. Of course, $\chi_\ell(G) \leq |V|$. Show that $\chi(G) \leq \chi_\ell(G) \leq \Delta(G) + 1$.

4.36. Let $K_{2,4}$ be a complete bipartite graph with parts of size 2 and 4 (see Fig. 4.5). Show that $\chi(K_{2,4}) = 2$ but $\chi_\ell(K_{2,4}) = 3$. What is $\chi_\ell(K_{3,3})$? *Hint*: Use the list of color sets given in Fig 4.5.

4.37. Generalize the above construction for $K_{3,3}$ to find graphs G where $\chi(G) = 2$, but $\chi_\ell(G)$ is arbitrarily large. For this, consider the complete bipartite graph $G = V_1 \times V_2$ whose parts V_1 and V_2 consist of all k-subsets v of $\{1, \ldots, 2k-1\}$. Define the pallete C_v of a vertex (k-subset) v to be the subset v itself. Show that $\chi_\ell(G) > k$. *Hint*: Observe that we need at least k colors to color V_1 and at least k colors to color V_2.

4.38. Let S_n be a graph which has vertex set the n^2 entries of an $n \times n$ matrix with two entries adjacent if and only if they are in the same row or in the same column. Show that $\chi_\ell(S_n) \geq n$. *Hint*: Any legal coloring of S_n corresponds to Latin square.

> *Comment*: The problem, whether $\chi_\ell(S_n) = n$, was raised by Jeff Dinitz in 1978. Janssen (1992) has proved that $\chi_\ell(S_n) \leq n + 1$, and the final solution $\chi_\ell(S_n) = n$ was found by Galvin (1995).

5. Systems of Distinct Representatives

A *system of distinct representatives* for a sequence of (not necessarily distinct) sets $S_1, S_2, \ldots, S_m$ is a sequence of distinct elements $x_1, x_2, \ldots, x_m$ such that $x_i \in S_i$ for all $i = 1, 2, \ldots, m$.

When does such a system exist? This problem is called the "marriage problem" because an easy reformulation of it asks whether we can marry each of m girls to a boy she knows; boys are the elements and S_i is the set of boys known to the i-th girl.

Clearly, if the sets $S_1, S_2, \ldots, S_m$ have a system of distinct representatives then the following *Hall's Condition* is fulfilled:

$(*)$ for every $k = 1, 2, \ldots, m$ the union of any k sets has at least k elements:

$$\left| \bigcup_{i \in I} S_i \right| \geq |I| \quad \text{for all } I \subseteq \{1, \ldots, m\}.$$

Surprisingly, this obvious necessary condition is also sufficient.

5.1 The marriage theorem

The following fundamental result is known as *Hall's marriage theorem* (Hall 1935), though an equivalent form of it was discovered earlier by König (1931) and Egerváry (1931), and the result is also a special case of Menger's theorem (1927). The case when we have the same number of girls as boys was proved by Frobenius (1917).

Theorem 5.1 (Hall's Theorem). *The sets $S_1, S_2, \ldots, S_m$ have a system of distinct representatives if and only if $(*)$ holds.*

Proof. We prove the sufficiency of Hall's condition $(*)$ by induction on m. The case $m = 1$ is clear. Assume that the claim holds for any collection with less than m sets.

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_5,

Case 1: For each k, $1 \leq k < m$, the union of any k sets contains more than k elements.

Take any of the sets, and choose any of its elements x as its representative, and remove x from all the other sets. The union of any $s \leq m-1$ of the remaining $m-1$ sets has at least s elements, and therefore the remaining sets have a system of distinct representatives, which together with x give a system of distinct representatives for the original family.

Case 2: The union of some k, $1 \leq k < m$, sets contains exactly k elements.

By the induction hypothesis, these k sets have a system of distinct representatives. Remove these k elements from the remaining $m-k$ sets. Take any s of these sets. Their union contains at least s elements, since otherwise the union of these s sets and the k sets would have less than $s+k$ elements. Consequently, the remaining $m-k$ sets also have a system of distinct representatives by the induction hypothesis. Together these two systems of distinct representatives give a system of distinct representatives for the original family. □

In general, Hall's condition $(*)$ is hard to verify: we must check if the union of *any* k, $1 \leq k \leq m$, of the sets $S_1, \ldots, S_m$ contains at least k elements. But if we know more about these sets, then (sometimes) the situation is much better. Here is an example.

Corollary 5.2. *Let $S_1, \ldots, S_m$ be r-element subsets of an n-element set such that each element belongs to the same number d of these sets. If $m \leq n$, then the sets $S_1, \ldots, S_m$ have a system of distinct representatives.*

Proof. By the double counting argument (1.10), $mr = nd$, and hence, $m \leq n$ implies that $d \leq r$. Now suppose that $S_1, \ldots, S_m$ does not have a system of distinct representatives. By Hall's theorem, the union $Y = S_{i_1} \cup \cdots \cup S_{i_k}$ of some k $(1 \leq k \leq m)$ sets contains strictly less than k elements. For $x \in Y$, let d_x be the number of these sets containing x. Then, again, using (1.10), we obtain

$$rk = \sum_{j=1}^{k} |S_{i_j}| = \sum_{x \in Y} d_x \leq d|Y| < dk,$$

a contradiction with $d \leq r$. □

Hall's theorem was generalized in different ways. Suppose, for example, that each of the elements of the underlying set is colored either in red or in blue. Interpret red points as "bad" points. Given a system of subsets of this (colored) set, we would like to come up with a system of distinct representatives which has as few bad elements as possible.

Theorem 5.3 (Chvátal–Szemerédi 1988)**.** *The sets $S_1, \ldots, S_m$ have a system of distinct representatives with at most t red elements if and only if they have a system of distinct representatives and for every $k = 1, 2, \ldots, m$ the union of any k sets has at least $k - t$ blue elements.*

Proof. The "only if" part is obvious. To prove the "if" part, let R be the set of red elements. We may assume that $|R| > t$ (otherwise the conclusion is trivial). Now enlarge $S_1, \ldots, S_m$ to $S_1, \ldots, S_m, S_{m+1}, \ldots, S_{m+r}$ by adding $r = |R| - t$ copies of the set R. Observe that the sequence $S_1, \ldots, S_m$ has a system of distinct representatives with at most t red elements if and only if the extended sequence has a system of distinct representatives (without any restriction). Hence, Hall's theorem reduces our task to proving that the extended sequence fulfills Hall's condition $(*)$, i.e., that for any set of indices $I \subseteq \{1, \ldots, m+r\}$, the union $Y = \bigcup_{i \in I} S_i$ contains at least $|I|$ elements. Let $J = I \cap \{1, \ldots, m\}$. If $J = I$ then, by the first assumption, the sets S_i $(i \in I)$ have a system of distinct representatives, and hence, $|Y| \geq |I|$. Otherwise, by the second assumption,

$$\begin{aligned} |Y| &= \left| \bigcup_{i \in J} (S_i \setminus R) \right| + |R| \geq (|J| - t) + |R| \\ &= |J| + (|R| - t) \geq |J| + |I \setminus J| = |I|; \end{aligned}$$

hence $(*)$ holds again. □

5.2 Two applications

In this section we present two applications of Hall's theorem to prove results whose statement does not seem to be related at all to set systems and their representatives.

5.2.1 Latin rectangles

An $r \times n$ *Latin rectangle* is an $r \times n$ matrix with entries in $\{1, \ldots, n\}$ such that each of the numbers $1, 2, \ldots, n$ occurs once in each row and at most once in each column. A *Latin square* is a Latin $r \times n$-rectangle with $r = n$. This is one of the oldest combinatorial objects, whose study goes back to ancient times.

Suppose somebody gives us an $n \times n$ matrix, *some* of whose entries are filled with the numbers from $\{1, \ldots, n\}$ so that no number occurs more than once in a row or column. Our goal is to fill the remaining entries so that to get a Latin square. When is this possible? Of course, the fewer entries are filled, the more chances we have to complete the matrix. Fig. 5.1 shows that, in general, it is possible to fill n entries so that the resulting partial matrix cannot be completed.

In 1960, Trevor Evans raised the following question: if fewer than n entries in an $n \times n$ matrix are filled, can one then always complete it to obtain a Latin square? The assertion that a completion is always possible became known as

1	5	2	4	?
				3

Fig. 5.1 A partial 2×5 Latin square that cannot be completed

the Evans conjecture, and was proved by Smetaniuk (1981) using a quite subtle induction argument.

On the other hand, it was long known that if a partial Latin square has no partially filled rows (that is, each row is either completely filled or completely free) then it can always be completed. That is, we can build Latin squares by adding rows one-by-one. And this can be easily derived from Hall's theorem.

Theorem 5.4 (Ryser 1951)**.** *If $r < n$, then any given $r \times n$ Latin rectangle can be extended to an $(r+1) \times n$ Latin rectangle.*

Proof. Let R be an $r \times n$ Latin rectangle. For $j = 1, \ldots, n$, define S_j to be the set of those integers $1, 2, \ldots, n$ which do *not* occur in the j-th column of R. It is sufficient to prove that the sets $S_1, \ldots, S_n$ have a system of distinct representatives. But this follows immediately from Corollary 5.2, because: every set S_j has precisely $n-r$ elements, and each element belongs to precisely $n-r$ sets S_j (since it appears in precisely r columns of the rectangle R). □

5.2.2 Decomposition of doubly stochastic matrices

Using Hall's theorem we can obtain a basic result of polyhedral combinatorics, due to Birkhoff (1949) and von Neumann (1953).

An $n \times n$ matrix $A = \{a_{ij}\}$ with real non-negative entries $a_{ij} \geq 0$ is *doubly stochastic* if the sum of entries along any row and any column equals 1. A *permutation matrix* is a doubly stochastic matrix with entries 0 and 1; such a matrix has exactly one 1 in each row and in each column. Doubly stochastic matrices arise in the theory of Markov chains: a_{ij} is the transition probability from the state i to the state j. A matrix A is a convex combination of matrices $A_1, \ldots, A_s$ if there exist non-negative reals $\lambda_1, \ldots, \lambda_s$ such that $A = \sum_{i=1}^{s} \lambda_i A_i$ and $\sum_{i=1}^{s} \lambda_i = 1$.

Birkhoff–Von Neumann Theorem. *Every doubly stochastic matrix is a convex combination of permutation matrices.*

Proof. We will prove a more general result that every $n \times n$ non-negative matrix $A = (a_{ij})$ having all row and column sums equal to some positive value $\gamma > 0$ can be expressed as a linear combination $A = \sum_{i=1}^{s} \lambda_i P_i$ of permutation matrices $P_1, \ldots, P_s$, where $\lambda_1, \ldots, \lambda_s$ are non-negative reals such that $\sum_{i=1}^{s} \lambda_i = \gamma$.

To prove this, we apply induction on the number of non-zero entries in A. Since $\gamma > 0$, we have at least n such entries. If there are exactly n non-zero entries then $A = \gamma P$ for some permutation matrix P, and we are done. Now suppose that A has more than n non-zero entries and that the result holds for matrices with a smaller number of such entries. Define

$$S_i = \{j \, : \, a_{ij} > 0\}, \; i = 1, 2, \ldots, n,$$

and observe that the sets $S_1, \ldots, S_n$ fulfill Hall's condition. Indeed, if the union of some k $(1 \leq k \leq n)$ of these sets contained less than k elements, then all the non-zero entries of the corresponding k rows of A would occupy no more than $k-1$ columns; hence, the sum of these entries by columns would be at most $(k-1)\gamma$, whereas the sum by rows is $k\gamma$, a contradiction.

By Hall's theorem, there is a system of distinct representatives

$$j_1 \in S_1, \ldots, j_n \in S_n.$$

Take the permutation matrix $P_1 = \{p_{ij}\}$ with entries $p_{ij} = 1$ if and only if $j = j_i$. Let $\lambda_1 = \min\{a_{1j_1}, \ldots, a_{nj_n}\}$, and consider the matrix $A_1 = A - \lambda_1 P_1$. By the definition of the sets S_i, $\lambda_1 > 0$. So, this new matrix A_1 has less non-zero entries than A. Moreover, the matrix A_1 satisfies the condition of the theorem with $\gamma_1 = \gamma - \lambda_1$. We can therefore apply the induction hypothesis to A_1, which yields a decomposition $A_1 = \lambda_2 P_2 + \cdots + \lambda_s P_s$, and hence, $A = \lambda_1 P_1 + A_1 = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_s P_s$, as desired. □

5.3 Min–max theorems

The early results of Frobenius and König have given rise to a large number of *min-max theorems* in combinatorics, in which the minimum of one quantity equals the maximum of another. Celebrated among these are:

- *Menger's theorem* (Menger 1927): the minimum number of vertices separating two given vertices in a graph is equal to the maximum number of vertex-disjoint paths between them;
- *König–Egerváry's min-max theorem* (König 1931, Egerváry 1931): the size of a largest matching in a bipartite graph is equal to the smallest set of vertices which together touch every edge;
- *Dilworth's theorem* for partially ordered sets (Dilworth 1950): the minimum number of *chains* (totally ordered sets) which cover a partially ordered set is equal to the maximum size of an antichain (set of incomparable elements).

Here we present the proof of König–Egerváry's theorem (stated not for bipartite graphs but for their adjacency matrices); the proof of Dilworth's theorem is given in Sect. 8.1.

By Hall's theorem, we know whether each of the girls can be married to a boy she knows. If so, all are happy (except for the boys not chosen ...). But what if not? In this sad situation it would be nice to make as many happy marriages as possible. So, given a sequence of sets $S_1, S_2, \ldots, S_m$, we try to find a system of distinct representatives for as many of these sets as possible. In terms of 0-1 matrices this problem is solved by the following result.

Let A be an $m \times n$ matrix, all whose entries have value 0 or 1. Two 1s are *dependent* if they are on the same row or on the same column; otherwise, they are *independent*. The size of the largest set of independent 1s is also known as the *term rank* of A.

Theorem 5.5 (König 1931, Egerváry 1931)**.** *Let A be an $m \times n$ 0-1 matrix. The maximum number of independent 1s is equal to the minimum number of rows and columns required to cover all the 1s in A.*

Proof. Let r denote the maximum number of independent 1s and R the minimum number of rows and columns required to cover all the 1s. Clearly, $R \geq r$, because we can find r independent 1s in A, and any row or column covers at most one of them.

We need to prove that $r \geq R$. Assume that some a rows and b columns cover all the 1s and $a + b = R$. Because permuting the rows and columns changes neither r nor R, we may assume that the first a rows and the first b columns cover the 1s. Write A in the form

$$A = \begin{pmatrix} B_{a\times b} & C_{a\times(n-b)} \\ D_{(m-a)\times b} & E_{(m-a)\times(n-b)} \end{pmatrix}.$$

We know that there are no 1s in E. We will show that there are a independent 1s in C. The same argument shows – by symmetry – that there are b independent 1s in D. Since altogether these $a + b$ 1s are independent, this shows that $r \geq a + b = R$, as desired.

We use Hall's theorem. Define

$$S_i = \{j \ : \ c_{ij} = 1\} \subseteq \{1, 2, \ldots, n - b\},$$

as the set of locations of the 1s in the i-th row of $C = (c_{ij})$. We claim that the sequence $S_1, S_2, \ldots, S_a$ has a system of distinct representatives, i.e., we can choose a 1 from each row, no two in the same column. Otherwise, Hall's theorem tells us that the 1s in some k $(1 \leq k \leq a)$ of these rows can all be covered by less than k columns. But then we obtain a covering of all the 1s in A with fewer than $a + b$ rows and columns, a contradiction. □

5.4 Matchings in bipartite graphs

Let G be a bipartite graph with bipartition A, B. Two edges are *disjoint* if they have no vertex in common. A *matching* in G is a set of pairwise disjoint edges. The vertices belonging to the edges of a matching are *matched*, others are *free*. We may ask whether G has a matching which matches all the vertices from A; we call this a matching of A *into* B. A *perfect matching* is a matching of A into B in the case when $|A| = |B|$.

The answer is given by Hall's theorem. A vertex $x \in A$ is a *neighbor* of a vertex $y \in B$ in the graph G if $(x, y) \in E$. Let S_x be the set of all neighbors of x in G. Observing that there is a matching of A into B if and only if the sets S_x with $x \in A$ have a system of distinct representatives, Hall's theorem immediately yields the following:

Theorem 5.6. *If G is a bipartite graph with bipartition A, B, then G has a matching of A into B if and only if, for every $k = 1, 2, \ldots, |A|$, every subset of k vertices from A has at least k neighbors.*

To illustrate this form of Hall's theorem, we prove the following (simple but non-trivial!) fact.

Proposition 5.7. *Let X be an n-element set. For any $k \leq (n-1)/2$ it is possible to extend every k-element subset of X to a $(k+1)$-element subset (by adding some element to that set) so that the extensions of no two sets coincide.*

Proof. Consider the bipartite graph $G = (A, B, E)$, where A consists of all k-element subsets, B consists of all $(k+1)$-element subsets of X and $(x, y) \in E$ if and only if $x \subset y$. What we need is to prove that this graph has a matching of A into B. Is the condition of Theorem 5.6 satisfied? Certainly, since for $I \subseteq A$, every vertex of I is joined to $n-k$ vertices in B and every vertex of B is joined to at most $k+1$ vertices in I. So, if $S(I)$ is the union of all neighbors of the vertices from I, and $E' = E \cap (I \times B)$ is the set of edges in the corresponding subgraph, then

$$|I|(n-k) = |E'| \leq |S(I)|(k+1).$$

Thus,

$$|S(I)| \geq |I|(n-k)/(k+1) \geq |I|$$

for every $I \subseteq A$, and Theorem 5.6 gives the desired matching of A into B. □

In terms of (bipartite) graphs, the König–Egerváry theorem is as follows. A *vertex cover* in a bipartite graph G with bipartition A, B is a set of vertices $S \subseteq A \cup B$ such that every edge is incident to at least one vertex from S. A *maximum matching* is a matching of maximum size.

Theorem 5.8. *The maximum size of a matching in a bipartite graph equals the minimum size of a vertex cover.*

How can we find such a matching of maximal size? To obtain a large matching, we could iteratively select an edge disjoint from those previously selected. This yields a matching which is "maximal" in a sense that no more

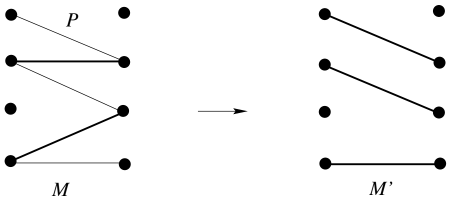


Fig. 5.2 Enlarging the matching M by the M-augmenting path P

edges can be added to it. But this matching does not need to be a maximum matching: some other matching may have more edges. A better idea is to jump between different matchings so that the new matching will always have one edge more, until we exhaust the "quota" of possible edges, i.e., until we reach the maximal possible number of edges in a matching. This idea employs the notion of "augmenting paths."

Assume that M is a (not necessarily maximum) matching in a given graph G. The edges of M are called *matched* and other edges are called *free.* Similarly, vertices which are endpoints of edges in M are called *matched* (in M); all other vertices are called *free* (in M). An *augmenting path* with respect to M (or M*-augmenting path*) is a path in G such that its edges are alternatively matched and free, and the endpoints of the path are free.

If P is an M-augmenting path, then M is certainly not a maximum size matching: the set M' of all *free* edges along this path form a matching with one more edge (see Fig. 5.2). Thus, the presence of an augmenting path implies that a matching is not a maximum matching. Interestingly (and it is a key for the matching algorithm), the converse is also valid: the *absence* of an augmenting path implies that the matching is, in fact, a maximum matching. This result was proved by Berge (1957), and holds for arbitrary graphs.

Theorem 5.9 (Berge 1957)**.** *A matching M in a graph G is a maximum matching if and only if G has no M-augmenting path.*

Proof. We have noted that an M-augmenting path produces a larger matching. For the converse, suppose that G has a matching M' larger than M; we want to construct an M-augmenting path. Consider the graph $H = M \oplus M'$, where $\oplus$ is the symmetric difference of sets. That is, H consists of precisely those edges which appear in *exactly one* of the matchings M and M'.

Since M and M' are matchings, every vertex has at most one incident edge in each of them. This means that in H, every vertex has at most degree 2, and hence, the graph H consists of disjoint paths and cycles. Furthermore, every path or cycle in H alternates between edges of M and edges of M'. This implies that each cycle in H has even length. As $|M'| > |M|$, the graph H must have a component with more edges of M' than of M. Such a component can only be a path that starts and ends with an edge of M'; it remains to observe that every such path in H is an M-augmenting path in G. □

This theorem suggests the following algorithm to find a maximum matching in a graph G: start with the empty matching $M = \emptyset$, and at each step search for an M-augmenting path in G. In one step the matching is enlarged by one, and we can have at most ℓ such steps, where ℓ is the size of a maximum matching. In general, the computation of augmenting paths is not a trivial task, but for bipartite graphs it is quite easy.

Given a bipartite graph $G = (A, B, E)$ and a matching M in it, construct a *directed* graph G_M by directing all matched edges from A to B and other edges from B to A. Let A_0, B_0 denote the sets of free vertices in A and B, respectively.

Proposition 5.10. *A bipartite graph G has an M-augmenting path if and only if there is a directed path in G_M from a vertex in B_0 to a vertex in A_0.*

We leave the proof of this fact as an exercise.

Using this fact, one may easily design an augmenting path algorithm running in time $O(n^2)$, where n is the total number of vertices. (One can apply, for example, the "depth-first search" algorithm to find a path from B_0 to A_0.) We need to find an augmenting path at most $n/2$ times, hence, the complexity of this matching algorithm is $O(n^3)$. Using a trickier augmenting path algorithm, Hopcroft and Karp (1973) have found a faster algorithm using time $O(n^{5/2})$.

Exercises

5.1. Let $S_1, \ldots, S_m$ be a sequence of sets such that: (i) each set contains at least r elements (where $r > 0$) and (ii) no element is in more than r of the sets. Show that these sets have a system of distinct representatives. *Hint*: See the proof of Corollary 5.2.

5.2. Show that in a group of m girls and n boys there exist some t girls for whom husbands can be found if and only if any subset of the girls (k of them, say) between them know at least $k + t - m$ of the boys. *Hint*: Invite additional $m - t$ "very popular" boys who are known to all the girls. Show that at least t girls can find husbands in the original situation if and only if *all* the girls can find husbands in the new situation. Then apply Hall's theorem to the new situation.

5.3. Show that any bipartite graph with maximum degree d is a union of d matchings. *Hint*: Argue by induction on d and use Theorem 5.6 in the induction step.

5.4. Let $S_1, \ldots, S_m$ be a sequence of sets satisfying Hall's condition $(*)$. Suppose that for some $1 \leq k < m$, the union $S_1 \cup \cdots \cup S_k$ of the first k sets has precisely k elements. Show that none of the remaining sets $S_{k+1}, \ldots, S_m$ can lie entirely in this union.

5.5. In Theorem 5.4 we have shown that (as long as $r < n$) we can add a new row to every $r \times n$ Latin rectangle such that the resulting $(r+1) \times n$ matrix is still Latin. Prove that this can be done in at least $(n-r)!$ ways.

5.6. Let G be a bipartite graph with bipartition A, B. Let a be the minimum degree of a vertex in A, and b the maximum degree of a vertex in B. Prove the following: if $a \geq b$ then there exists a matching of A into B.

5.7. Let $S_1, \ldots, S_m$ be a sequence of sets each of cardinality at least r, and assume that it has a system of distinctive representatives. Prove that then it has at least

$$f(r,m) = \prod_{i=1}^{\min\{r,m\}} (r+1-i)$$

systems of distinctive representatives. *Hint*: Follow the proof of Hall's theorem. Case 1 gives at least $r \cdot f(r-1, m-1) \geq f(r,m)$ and Case 2 at least $f(r,k) \cdot f(\max\{r-k,1\}, m-k) = f(r,m)$ systems of distinctive representatives.

5.8. Prove that every bipartite graph G with ℓ edges has a matching of size at least $\ell/\Delta(G)$, where $\Delta(G)$ is the maximum degree of a vertex in G. *Hint*: Use Theorem 5.8.

5.9. Suppose that M, M' are matchings in a bipartite graph G with bipartition A, B. Suppose that all the vertices of $S \subseteq A$ are matched by M and that all the vertices of $T \subseteq B$ are matched by M'. Prove that G contains a matching that matches all the vertices of $S \cup T$.

5.10. (Lovász et al. 1995). Let $\mathcal{F}$ be a family of sets, each of size at least 2. Let A, B be two sets such that $|A| = |B|$, both A and B intersect all the members of $\mathcal{F}$, and no set of fewer than $|A|$ elements does this. Consider a bipartite graph G with parts A and B, where $a \in A$ is connected to $b \in B$ if there is an $F \in \mathcal{F}$ containing both a and b. Show that this graph has a perfect matching. *Hint*: For $I \subseteq A$, let $S(I) \subseteq B$ be the set of neighbors of I in G; show that the set $A' = (A \setminus I) \cup S(I)$ intersects all the members of $\mathcal{F}$.

5.11. (Sperner 1928). Let $t < n/2$ and let $\mathcal{F}$ be a family of subsets of an n-element set X. Suppose that: (i) each member of $\mathcal{F}$ has size at most t, and (ii) $\mathcal{F}$ is an *antichain*, i.e., no member of $\mathcal{F}$ is a subset of another one. Let $\mathcal{F}_t$ be the family of all those t-element subsets of X, which contain at least one member of $\mathcal{F}$. Prove that then $|\mathcal{F}| \leq |\mathcal{F}_t|$. *Hint*: Use Proposition 5.7 to extend each member of $\mathcal{F}$ to a unique member in the family $\mathcal{F}_t$.

5.12. Let A be a 0-1 matrix with m 1s. Let s be the maximal number of 1s in a row or column of A, and suppose that A has no square $r \times r$ all-1 sub-matrix. Use the König–Egerváry theorem to show that we then need at least $m/(sr)$ all-1 (not necessarily square) sub-matrices to cover all 1s in A. *Hint*: There are at least m/s independent 1s, and at most r of them can be covered by one all-1 sub-matrix.

Part II
Extremal Set Theory

6. Sunflowers

One of most beautiful results in extremal set theory is the so-called *Sunflower Lemma* discovered by Erdős and Rado (1960) asserting that in a sufficiently large uniform family, some highly regular configurations, called "sunflowers," must occur, regardless of the size of the universe. In this chapter we will consider this result as well as some of its modifications and applications.

6.1 The sunflower lemma

A *sunflower* (or *Δ-system*) with k *petals* and a *core* Y is a collection of sets $S_1, \ldots, S_k$ such that $S_i \cap S_j = Y$ for all $i \neq j$; the sets $S_i \setminus Y$ are petals, and we require that none of them is empty. Note that a family of pairwise disjoint sets is a sunflower (with an empty core).

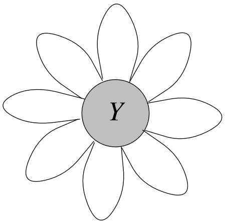


Fig. 6.1 A sunflower with 8 petals

Sunflower Lemma. *Let $\mathcal{F}$ be family of sets each of cardinality s. If $|\mathcal{F}| > s!(k-1)^s$ then $\mathcal{F}$ contains a sunflower with k petals.*

Proof. We proceed by induction on s. For $s = 1$, we have more than $k - 1$ points (disjoint 1-element sets), so any k of them form a sunflower with k

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_6,

petals (and an empty core). Now let $s \geq 2$, and take a maximal family $\mathcal{A} = \{A_1, \ldots, A_t\}$ of pairwise disjoint members of $\mathcal{F}$.

If $t \geq k$, these sets form a sunflower with $t \geq k$ petals (and empty core), and we are done.

Assume that $t \leq k-1$, and let $B = A_1 \cup \cdots \cup A_t$. Then $|B| \leq s(k-1)$. By the maximality of $\mathcal{A}$, the set B intersects every member of $\mathcal{F}$. By the pigeonhole principle, some point $x \in B$ must be contained in at least

$$\frac{|\mathcal{F}|}{|B|} > \frac{s!(k-1)^s}{s(k-1)} = (s-1)!(k-1)^{s-1}$$

members of $\mathcal{F}$. Let us delete x from these sets and consider the family

$$\mathcal{F}_x := \{S \setminus \{x\} \ : \ S \in \mathcal{F},\, x \in S\}.$$

By the induction hypothesis, this family contains a sunflower with k petals. Adding x to the members of this sunflower, we get the desired sunflower in the original family $\mathcal{F}$. □

It is not known if the bound $s!(k-1)^s$ is the best possible. Let $f(s,k)$ denote the least integer so that any s-uniform family of $f(s,k)$ sets contains a sunflower with k petals. Then

$$(k-1)^s < f(s,k) \leq s!(k-1)^s + 1. \tag{6.1}$$

The upper bound is the sunflower lemma, the lower bound is Exercise 6.2. The gap between the upper and lower bound for $f(s,k)$ is still huge (by a factor of $s!$).

Conjecture 6.1 (Erdős and Rado). For every fixed k there is a constant $C = C(k)$ such that $f(s,k) < C^s$.

The conjecture remains open even for $k = 3$ (note that in this case the sunflower lemma requires at least $s!2^s \approx s^s$ sets). Several authors have slightly improved the bounds in (6.1). In particular, J. Spencer has proved

$$f(s,3) \leq \mathrm{e}^{c\sqrt{s}} s!.$$

For s fixed and k sufficiently large, Kostochka et al. (1999) have proved

$$f(s,k) \leq k^s \left(1 + ck^{-2^{-s}}\right),$$

where c is a constant depending only on s.

But the proof or disproof of the conjecture is nowhere in sight.

A family $\mathcal{F} = \{S_1, \ldots, S_m\}$ is called a *weak Δ-system* if there is some λ such that $|S_i \cap S_j| = \lambda$ whenever $i \neq j$. Of course, not every such system is a sunflower: in a weak Δ-system it is enough that all the cardinalities of mutual intersections coincide whereas in a sunflower we require that these intersections all have the same elements. However, the following interesting result due to M. Deza states that if a weak $\varDelta$-system has many members then it is, in fact, "strong," i.e., forms a sunflower. We state this result without proof.

Theorem 6.2 (Deza 1973)**.** *Let $\mathcal{F}$ be an s-uniform weak $\varDelta$-system. If $|\mathcal{F}| \geq s^2 - s + 2$ then $\mathcal{F}$ is a sunflower.*

The family of lines in a projective plane of order $s - 1$ shows that this bound is optimal (see Exercise 6.1).

A related problem is to estimate the maximal possible number $F(n, k)$ of members in a family $\mathcal{F}$ of subsets of an n-element set such that $\mathcal{F}$ does not contain a weak $\varDelta$-system with k members. It is known that

$$2^{0.01(n \ln n)^{1/3}} \leq F(n, 3) \leq 1.99^n.$$

The upper bound was proved by Frankl and Rödl (1987), and the lower bound by Kostochka and R"odl (1998).

6.2 Modifications

Due to its importance, the sunflower lemma was modified in various directions. If $S_1, \ldots, S_k$ form a sunflower with a core Y, then we have two nice properties:

(a) the core Y lies *entirely* in all the sets $S_1, \ldots, S_k$;
(b) the sets $S_1 \setminus Y, \ldots, S_k \setminus Y$ are mutually disjoint.

It is therefore natural to look at what happens if we relax any of these two conditions.

6.2.1 Relaxed core

We can relax property (a) and require that only the differences $S_i \setminus Y$ be non-empty and mutually disjoint for *some* set Y.

Given distinct finite sets $S_1, \ldots, S_k$, their *common part* is the set

$$Y := \bigcup_{i \neq j} (S_i \cap S_j)\,.$$

Note that, if $|Y| < \min_i |S_i|$ then all the sets $S_1 \setminus Y, \ldots, S_k \setminus Y$ are nonempty and mutually disjoint.

Lemma 6.3 (Füredi 1980)**.** *Let $\mathcal{F}$ be a finite family of sets, and $s = \max_{S \in \mathcal{F}} |S|$. If $|\mathcal{F}| > (k-1)^s$ then the common part of some k of its members has fewer than s elements.*

Proof. We prove a contraposition of this claim: If the common part of every k members of $\mathcal{F}$ has at least $s = \max_{S\in\mathcal{F}} |S|$ elements, then $|\mathcal{F}| \le (k-1)^s$. The cases $k = 2$ and $s = 1$ are trivial. Apply induction on k. Once k is fixed, apply induction on s. Let S_0 be an arbitrary member of $\mathcal{F}$ of size s. We have

$$|\mathcal{F}| = 1 + \sum_{X\subset S_0} |\{S \in \mathcal{F} \,:\, S \cap S_0 = X\}| \,. \tag{6.2}$$

Fix now an arbitrary $X \subset S_0$, and consider the family

$$\mathcal{F}_X := \{S \setminus S_0 \colon\; S \in \mathcal{F}, S \cap S_0 = X\} \,.$$

The maximum size of its member is $s' \le s - |X|$. Moreover, the common part of any its $k' = k-1$ members $S_1 \setminus S_0, \ldots, S_{k-1} \setminus S_0$ is the common part of k members $S_0, S_1, \ldots, S_{k-1}$ of $\mathcal{F}$ minus X, and hence is at least $s - |X| \ge s'$. We can therefore apply the induction hypothesis with $s' \le s - |X|$, $k' = k-1$ and deduce

$$|\mathcal{F}_X| \le (k-2)^{s-|X|} \,. \tag{6.3}$$

Combining (6.2) and (6.3) we obtain

$$|\mathcal{F}| \le 1 + \sum_{X\subset S_0} (k-2)^{s-|X|} \le \sum_{i=0}^{s} \binom{s}{i} (k-2)^{s-i} = (k-1)^s \,,$$

as desired. □

6.2.2 Relaxed disjointness

What if we relax the disjointness property (b) of sunflowers, and only require that the differences $S_1 \setminus Y, \ldots, S_k \setminus Y$ cannot be intersected (blocked) by a set of size smaller than some number t? In this case we say that sets $S_1, \ldots, S_k$ form a "flower" with t petals.

A *blocking set* of a family $\mathcal{F}$ is a set which intersects all the members of $\mathcal{F}$; the minimum number of elements in a blocking set is the *blocking number* of $\mathcal{F}$ and is denoted by $\tau(\mathcal{F})$; if $\emptyset \in \mathcal{F}$ then we set $\tau(\mathcal{F}) = 0$. A *restriction* of a family $\mathcal{F}$ onto a set Y is the family

$$\mathcal{F}_Y := \{S \setminus Y \;:\; S \in \mathcal{F},\, S \supseteq Y\} \,.$$

A *flower* with k *petals* and a *core* Y is a family $\mathcal{F}$ such that $\tau(\mathcal{F}_Y) \ge k$. Note that every sunflwover is a flower with the same number of petals, but not every flower is a sunflower (give an example).

Håstad et al. (1995) observed that the proof of the sunflower lemma can be easily modified to yield a similar result for flowers.

Lemma 6.4. *Let $\mathcal{F}$ be a family of sets each of cardinality s, and $k \ge 1$ and integer. If $|\mathcal{F}| > (k-1)^s$ then $\mathcal{F}$ contains a flower with k petals.*

Proof. Induction on s. The basis $s = 1$ is trivial since then $\mathcal{F}$ consists of at least k distinct single-element sets. Now suppose that the lemma is true for $s-1$ and prove it for s. Take a family $\mathcal{F}$ of sets each of cardinality s, and assume that $|\mathcal{F}| > (k-1)^s$. If $\tau(\mathcal{F}) \geq k$ then the family $\mathcal{F}$ itself is a flower with at least $(k-1)^s+1 \geq k$ petals (and an empty core). Otherwise, some set of size $k-1$ intersects all the members of $\mathcal{F}$, and hence, at least $|\mathcal{F}|/(k-1)$ of the members must contain some point x. The family

$$\mathcal{F}_x := \{S \setminus \{x\} \ : \ S \in \mathcal{F},\, x \in S\}$$

has

$$|\mathcal{F}_x| \geq \frac{|\mathcal{F}|}{k-1} > (k-1)^{s-1}$$

members, each of cardinality $s-1$. By the induction hypothesis, the family $\mathcal{F}_x$ contains a flower with k petals and some core Y, $x \not\in Y$. Adding the element x back to the sets in this flower, we obtain a flower in $\mathcal{F}$ with the same number of petals and the core $Y \cup \{x\}$. □

6.3 Applications

The sunflower lemma and its modifications have many applications in complexity theory. In particular, the combinatorial part of the celebrated lower bounds argument for monotone circuits, found by Razborov (1985), is based on this lemma and on its modification due to Füredi (Lemma 6.3). Andreev (1987) has also used his modification (Exercise 6.5) to prove exponential lower bounds for such circuits. In this section we will show how the last modification (Lemma 6.4) can be used to obtain some information about the number of minterms and to prove lower bounds for small depth non-monotone circuits.

6.3.1 The number of minterms

Let $x_1, \ldots, x_n$ be boolean variables taking their values in $\{0,1\}$. A *monomial* is an And of literals, and a *clause* is an Or of literals, where a *literal* is either a variable x_i or its negation $\overline{x}_i = x_i \oplus 1$. Thus, we have $2^s\binom{n}{s}$ monomials and that many clauses of size s.

A 1-*term* of a boolean function $f : \{0,1\}^n \to \{0,1\}$ is a monomial M such that $M(a) \leq f(a)$ for all inputs $a \in \{0,1\}^n$. That is, if we set all literals of M to 1, then the function f is forced to take value 1 independent on what values we assign to the remaining variables. Dually, a 0-*term* of f is a clause C such that $C(a) \geq f(a)$ for all inputs $a \in \{0,1\}^n$. A *minterm* of f is a 1-term M of f which is minimal in the sense that deleting every single literal from M already violates this property.

A boolean function f is a *t-And-Or* (or a *t-CNF*) if it can be written as an And of an arbitrary number of clauses, each of size at most t.

Lemma 6.5. *Let f be a t-And-Or function on n variables. Then for every $s = 1, \dots, n$ the function f has at most t^s minterms of size s.*

Proof. Let $f = C_1 \wedge \cdots \wedge C_m$, where each clause C_i has size at most t. We interpret the clauses as sets of their literals, and let $\mathcal{C} = \{C_1, \dots, C_m\}$ be the corresponding family of these sets. Let $\mathcal{F}$ be the family of all minterms of f that have size s (we look at minterms as sets of their literals). Then every set in $\mathcal{C}$ intersects each set in $\mathcal{F}$ (see Exercise 6.9).

Suppose that $|\mathcal{F}| > t^s$. Then, by Lemma 6.4, $\mathcal{F}$ has a flower with $t + 1$ petals. That is, there exists a set of literals Y such that no set of at most t literals can intersect all the members of the family

$$\mathcal{F}_Y = \{M \setminus Y \ : \ M \in \mathcal{F}, M \supseteq Y\}.$$

The set Y is a proper part of at least one minterm of f, meaning that Y cannot intersect all the clauses in $\mathcal{C}$. Take a clause $C \in \mathcal{C}$ such that $C \cap Y = \emptyset$. Since this clause intersects all the sets in $\mathcal{F}$, this means that it must intersect all the sets in $\mathcal{F}_Y$. But this is impossible because C has size at most t. □

6.3.2 Small depth formulas

An *s-threshold function* is a monotone boolean function T_s^n which accepts a 0-1 vector if and only if it has at least s ones. That is,

$$T_s^n(x_1, \dots, x_n) = 1 \text{ if and only if } x_1 + \cdots + x_n \geq s.$$

This function can be computed by the following formula:

$$T_s^n(x_1, \dots, x_n) = \bigvee_{I \,:\, |I|=s} \bigwedge_{i \in I} x_i.$$

This formula is monotone (has no negated literals) and has depth 2 (there are only two alternations between And and Or operations). But the size of this formula (the number of literals in it) is $s\binom{n}{s}$. Can T_s^n be computed by a substantially smaller formula if we allow negated literals and/or a larger depth?

Håstad (1986) proved that, for $s = \lfloor n/2 \rfloor$, each such formula computing T_s^n must have size exponential in n, even if we allow any *constant* depth, i.e., any constant number of alternations of And's and Or's. Razborov (1987) has proved that the same holds even if we allow sum modulo 2 as an additional operation. Both these proofs employ non-trivial machinery: the switching lemma and approximations of boolean functions by low-degree polynomials.

On the other hand, Håstad et al. (1995) have shown that, at least for depth-3, one can deduce the same lower bound in an elementary way using the flower lemma (Lemma 6.4). In fact, their proof holds for depth-3 *circuits* but, to demonstrate the idea, it is enough to show how it works for special depth-3 *formulas.*

An *Or-And-Or formula* is a formula of the form

$$F = F_1 \vee F_2 \vee \cdots \vee F_t, \tag{6.4}$$

where each F_i is an And-Or formula, that is, each F_i is an And of an arbitrary number of clauses, each clause being an Or of literals (variables or their negations). We say that such a formula has *bottom fan-in* k if each of its clauses has at most k positive literals (the number of negated variables may be arbitrary). The *size* of a formula is the total number of literals in it.

At this point, let us note that the condition on bottom fan-in is not crucial: if the size of F is not too large then it is possible to set some small number of variables to constant 1 so that the resulting formula will already satisfy this condition (see Exercise 6.10).

The idea of Håstad et al. (1995) is accumulated in the following lemma.

Lemma 6.6. *Let $F = F_1 \vee F_2 \vee \cdots \vee F_t$ be an Or-And-Or formula of bottom fan-in k. Suppose that F rejects all vectors with fewer than s ones. Then F cannot accept more than tk^s vectors with precisely s ones.*

Note that this lemma immediately implies that every Or-And-Or formula of bottom fan-in k computing the threshold function T^n_s has size at least

$$\binom{n}{s}k^{-s} > \left(\frac{n}{ks}\right)^s .$$

Proof. Suppose that F accepts more than tk^s vectors with precisely s ones. Then some of its And-Or subformulas F_i accepts more than k^s of such vectors. Let A be this set of vectors with s ones accepted by F_i; hence

$$|A| > k^s.$$

The formula F_i has the form

$$F_i = C_1 \wedge C_2 \wedge \cdots \wedge C_r,$$

where $C_1, \ldots, C_r$ are clauses with at most k positive literals in each of them. Let B be the set of all vectors with at most $s-1$ ones. All these vectors must be rejected by F_i, since they are rejected by the whole formula F. Our goal is to show that the set B contains a vector v on which each of the clauses $C_1, \ldots, C_r$ outputs the same value as on some vector from A; this will mean that the formula F_i makes an error on this input – it is forced to accept v.

Say that a vector v is a *k-limit* for A if, for every subset S of k coordinates, there exists a vector $u \in A$ such that $v \leq u$ (vector comparision) and v coincides with u in all the coordinates from S; that is, $v_i \leq u_i$ for all i and $v_i = u_i$ for all $i \in S$.

Claim 6.7. There exists a vector $v \in B$ which is a k-limit for A.

Proof of Claim 6.7. For a vector $u \in \{0,1\}^n$, let E_u be the corresponding subset of $\{1,\ldots,n\}$, whose incidence vector is u, that is, $E_u = \{i : u_i = 1\}$. Consider the family $\mathcal{F} = \{E_u : u \in A\}$. This family is s-uniform and has more than k^s members. By Lemma 6.4, $\mathcal{F}$ has a flower with $k+1$ petals. That is, there exists a set Y such that no set of size at most k can intersect all the members of the family $\mathcal{F}_Y = \{E \setminus Y : E \in \mathcal{F},\ E \supseteq Y\}$. Let v be the incidence vector of Y. We claim that v is a k-limit for A.

To show this, take an arbitrary subset S of $\{1,\ldots,n\}$ of size at most k. Then

$$S \cap (E_u \setminus Y) = \emptyset \tag{6.5}$$

for at least one set $E_u \in \mathcal{F}$ such that $Y \subseteq E_u$. The last condition implies that $v \leq u$, and hence, v coincides with u on all coordinates from $S \setminus E_u$ and from $S \cap Y$. But, by (6.5), there are no other coordinates in S, and hence, v coincides with u on all coordinates from S, as desired. □

Fix a vector v guaranteed by the claim. To get the desired contradiction we will show that the formula F_i will be forced to (wrongly) accept this vector. Suppose the opposite that v is rejected by F_i. Then $C(v) = 0$ for some clause C of F_i. This clause has a form

$$C = \left(\bigvee_{i \in S} x_i\right) \vee \left(\bigvee_{j \in T} \overline{x}_j\right)$$

for some two disjoint sets of S, T such that $|S| \leq k$. By Claim 6.7, there is a vector u in A such that $v \leq u$ and v coincides with u on all the coordinates from S. The vector u must be accepted by the formula F_i, and hence, by the clause C. This can happen only if this vector has a 1 in some coordinate $i \in S$ or has a 0 in some coordinate $j \in T$ (or both). In the first case $C(v) = 1$ because v coincides with u on S, and in the second case $C(v) = 1$ because, due to the condition $v \leq u$, vector v has 0s in all coordinates where vector u has them. Thus, in both cases, $C(v) = 1$, a contradiction. □

Exercises

6.1. A projective plane of order $s-1$ is a family of $n = s^2 - s + 1$ s-element subsets (called *lines*) of an n-element set of points such that each two lines intersect in precisely one point and each point belongs to precisely s lines (cf. Sect. 12.4). Show that the equality in Deza's theorem (Theorem 6.2) is attained when a projective plane of order $s-1$ exists.

6.2. Take s pairwise disjoint $(k-1)$-element sets $V_1, \ldots, V_s$ and consider the family

$$\mathcal{F} = \{S : |S| = s \text{ and } |S \cap V_i| = 1 \text{ for all } i = 1, \ldots, s\}.$$

This family has $(k-1)^s$ sets. Show that it has no sunflower with k petals.

6.3. Show that the bounds in Lemmas 6.3, and 6.4 are optimal. *Hint*: Consider the family defined in the previous exercise.

6.4. A *matching* of size k in a graph is a set of its k pairwise disjoint edges (two edges are disjoint if they have no vertex in common). A *star* of size k is a set of k edges incident to one vertex. Argue as in the proof of the sunflower lemma to show that any set of more than $2(k-1)^2$ edges either contains a matching of size k or a star of size k.

6.5. (Andreev 1987). Let $\mathcal{F}$ be a family of sets each of cardinality at most s, and suppose that $|\mathcal{F}| > (k-1)^s$. Use the argument of Lemma 6.3 to prove that then there exist k sets $S_1, \ldots, S_k$ in $\mathcal{F}$ such that all the sets $S_i \setminus (S_1 \cap S_2)$, $i = 1, \ldots, k$ are pairwise disjoint.

6.6. Let $n-k+1 < s \leq n$ and consider the family $\mathcal{F}$ of all s-element subsets of a n-element set. Prove that $\mathcal{F}$ has no sunflower with k petals. *Hint*: Suppose the opposite and count the number of elements used in such a sunflower.

6.7. Given a graph $G = (V, E)$ and a number $2 \leq s \leq |V|$, let $\mathcal{G}_s$ denote the graph whose vertices are all s-element subsets of V, and two such subsets A and B are connected by an edge if and only if there is an edge $(u, v) \in E$ such that $u \in A \setminus B$ and $v \in B \setminus A$. Suppose that the graph G is "sparse" in the following sense: every subset of at most ks vertices spans fewer that $\binom{k}{2}$ edges. Use Lemma 6.4 to show that then $\mathcal{G}_s$ has no clique of size larger than $(k-1)^s$. *Hint*: Let $\mathcal{F}$ be a clique in $\mathcal{G}_s$, and suppose that $\mathcal{F}$ forms a flower with k petals. Then each member $A \in \mathcal{F}$ contains an element v_A which is not contained in any other member of $\mathcal{F}$. Use the fact that $\mathcal{F}$ was a clique in $\mathcal{G}_s$ to argue that $\{v_A : A \in \mathcal{F}\}$ is a clique in G.

6.8. For a graph G, let $\mathcal{G}$ be a graph whose vertices are all maximum cliques of G, and where two such cliques A and B are connected by an edge if and only if there is an edge $(u, v) \in E$ such that $u \in A \setminus B$ and $v \in B \setminus A$. Recall that a clique is a *maximum clique* if each of remaining vertices is not connected to at least one of its edges (i.e. we cannot add any new vertices). Let $\alpha(G)$ denote the independence number of G, that is, the maximum number of vertices no two of which are adjacent in G. Show that $\alpha(\mathcal{G}) \leq \alpha(G)$. *Hint*: Let $\mathcal{F}$ be an independent set in $\mathcal{G}$. Show that then for any three distinct members A, B and K of $\mathcal{F}$, the intersections $A \cap K$ and $B \cap K$ must be comparable by set-inclusion. Argue that then each member $K \in \mathcal{F}$ must contain an element v_K which belongs to none of the remaining members. Why is then the set $\{v_K : K \in \mathcal{F}\}$ an independent set in G?

6.9. Show that every 0-term C and every 1-term K of a boolean function f must share at least one literal in common. *Hint*: Take a restriction (a partial assignment to variables) which evaluates all the literals of K to 1. If C has no literal of K, then this restriction can be extended to an input a such that $f(a) = 0$.

6.10. Let F be a set of clauses on n variables. Say that a clause is *long* if it has at least $k+1$ positive literals. Let ℓ be the number of long clauses in F, and suppose that

$$\ell < \left(\frac{n+1}{m+1}\right)^k .$$

Prove that then it is possible to assign some $n-m$ variables to constant 1 so that the resulting set F' will have no long clauses. *Hint*: Construct the desired set assignment via the following "greedy" procedure: Take the variable x_{i_1} which occurs in the largest number of long clauses and set it to 1; then take the variable x_{i_2} which occurs in the largest number of remaining long clauses and set it to 1, and so on, until all long clauses dissapear (get value 1). In computations use the estimate $\sum_{i=1}^n i^{-1} \sim \ln n$.

6.11. Consider the following function on $n = sr$ variables:

$$f = \bigwedge_{i=1}^{s} \bigvee_{j=1}^{r} x_{ij} .$$

Let F be an Or-And-Or formula of bottom fan-in k $(k \leq r)$ computing this function. Show that then F has size at least $(r/k)^s$. *Hint*: Observe that f rejects all vectors with fewer than s ones and accepts r^s vectors with precisely s ones; apply Lemma 6.6.

6.12. (Håstad et al. 1995). Consider the function on $n = m^2$ variables defined by the formula

$$f = \bigwedge_{i=1}^{m} \bigvee_{j=1}^{m} x_{ij} \wedge y_{ij} .$$

This formula is a depth-3 And-Or-And formula of size only $2n$. Prove that any depth-3 Or-And-Or formula for this function has size at least $2^{\Omega(\sqrt{n})}$. *Hint*: Assume that f has such a formula F of size at most $2^{m/3}$. Reduce the bottom fan-in of F to $k = \lceil m/2 \rceil$ by setting one half of the variables to constants at random as follows: for each pair of variables x_{ij}, y_{ij}, pick one of them at random (with probability 1/2) and set it to 1. If some clause has more than k positive literals, then none of these literals is set to 1 with probability at most 2^{-k-1}. The probability, that some of the clauses with more than k positive literals is not evaluated to 1, does not exceed $2^{m/3} \cdot 2^{-(k+1)} \leq 2^{-m/6} < 1$, and in particular such a setting exists. The resulting function has the same form as that considered in the previous exercise.

7. Intersecting Families

A basic interrelation between sets is their intersection. The size (or other characteristics) of mutual intersections between the members of a given family reflects some kind of "dependence" between them. In this chapter we will study the weakest kind of this dependence – the members are required to be non-disjoint. A family is *intersecting* if any two of its sets have a non-empty intersection.

7.1 Ultrafilters and Helly property

We start with two simple structural properties of intersecting families.

An *ultrafilter* over a set X is a collection $\mathcal{F}$ of its subsets such that: (i) $\mathcal{F}$ is upwards-closed, that is, $A \in \mathcal{F}$ and $A \subseteq B$ implies $B \in \mathcal{F}$, and (ii) for every subset A of X, exactly one of A or its complement $\overline{A} = X \setminus A$ belongs to $\mathcal{F}$.

Theorem 7.1. *Every ultrafilter is an intersecting family, and every intersecting family is contained in some ultrafilter.*

Proof. To prove the first claim, let $\mathcal{F}$ be an ultrafilter. If some two members A, B of $\mathcal{F}$ were disjoint, then the complement of B would contain the set A, and hence, would belong to $\mathcal{F}$ (by (i)). But this is impossible since, by (ii), $\mathcal{F}$ cannot contain the set B together with its complement.

To prove the second claim, take an arbitrary intersecting family and extend it to an ultrafilter as follows. If there are some sets not in the family and such that their addition does not destruct the intersection property, add all them. After that, add all supersets of the sets we already have. We claim that the resulting family $\mathcal{F}$ is an ultrafilter. Indeed, if it is not, there must be a set A such that neither A nor its complement $\overline{A}$ belongs to $\mathcal{F}$. By the construction, A must be disjoint from at least one member B of our initial family (for otherwise A would be added during the first phase), and hence,

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_7,

B is contained in the complement $\overline{A}$. But $B \in \mathcal{F}$ and $\mathcal{F}$ is upwards-closed, a contradiction. □

In 1923, E. Helly proved the following result: if $n \geq k+1$ convex sets in $\mathbb{R}^k$ have the property that any $k+1$ of them have a nonempty intersection, then there is a point common to all of them.

It is natural to ask if objects other than convex sets obey Helly-type laws. For arbitrary families of sets we have the following Helly-type result.

Theorem 7.2. *Let $\mathcal{F}$ be a family and k be the minimum size of its member. If any $k+1$ members of $\mathcal{F}$ intersect (i.e., share a common point) then all of them do.*

Proof. Suppose the opposite that the intersection of all sets in $\mathcal{F}$ is empty, and take a set $A = \{x_1, \ldots, x_k\} \in \mathcal{F}$. For every $i = 1, \ldots, k$ there must be a set $B_i \in \mathcal{F}$ such that $x_i \notin B_i$. Hence, $A \cap B_1 \cap \cdots \cap B_k = \emptyset$, a contradiction. □

7.2 The Erdős–Ko–Rado theorem

Let $\mathcal{F}$ be an intersecting family of k-element subsets of $\{1, \ldots, n\}$. The basic question is: how large can such a family be? To avoid trivialities, we assume $n \geq 2k$ since otherwise any two k-element sets intersect, and there is nothing to prove.

We can obtain an intersecting family by taking all $\binom{n-1}{k-1}$ k-element subsets containing a fixed element. Can we find larger intersecting families? The whole number of k-element subsets is $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$, so the question is not trivial.

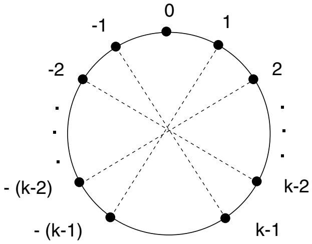
The following result, found by Erdős, Ko, and Rado in 1938 (but published only 23 years later), answers the question.

Theorem 7.3 (Erdős–Ko–Rado 1961)**.** *If $2k \leq n$ then every intersecting family of k-element subsets of an n-element set has at most $\binom{n-1}{k-1}$ members.*

Proof. (Due to G.O.H. Katona 1972.) Let $[n] = \{0, 1, \ldots, n-1\}$ be the underlying set. The idea is to study all permutations of the elements of $[n]$, estimating how often the consecutive elements of these permutations can constitute one of the sets in our family. For $s \in [n]$, let B_s denote the set of k consecutive numbers $s, s+1, \ldots, s+k-1$, where the addition is modulo n.

Claim 7.4. At most k of the sets B_s can belong to $\mathcal{F}$.

We can suppose that $B_0 \in \mathcal{F}$. The only sets B_s that intersect B_0 other than B_0 itself are the $2k-2$ sets B_s with $-(k-1) \leq s \leq k-1$, $s \neq 0$ (where the indices are taken modulo n). These sets can be partitioned into $k-1$ pairs of disjoint sets, B_i, B_{i+k}, where $-(k-1) \leq i \leq -1$.



Since $\mathcal{F}$ can contain at most one set of each such pair the assertion of the claim follows.

We now count in two ways the number L of pairs (f, s), where f is a permutation of $[n]$ and s is a point in $[n]$, such that the set

$$f(B_s) := \{f(s), f(s+1), \ldots, f(s+k-1)\}$$

belongs to $\mathcal{F}$. By the claim, for each *fixed* permutation f, the family $\mathcal{F}$ can contain at most k of the sets $f(B_s)$. Hence, $L \leq kn!$. On the other hand, exactly $nk!(n-k)!$ of the pairs (f, s) yield the same set $f(B_s)$: there are n possibilities for s, and for each fixed s, there are $k!(n-k)!$ possibilities to choose the permutation f. Hence, $L = |\mathcal{F}| \cdot nk!(n-k)!$. Combining this with the previous estimate, we obtain

$$|\mathcal{F}| \leq \frac{kn!}{nk!(n-k)!} = \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}. \qquad \square$$

7.3 Fisher's inequality

A fundamental result of design theory–known as *Fisher's inequality*—states that, if each two clubs in a town share the same number of members in common, then the number of clubs cannot exceed the total number of inhabitants in the town. In the proof of this result we will use (for the first time) a powerful tool: linear algebra.

The general frame for the *linear algebra method* in combinatorics is the following: if we want to come up with an upper bound on the size of a set of objects, associate them with elements in a vector space V of relatively low dimension, and show that these elements are linearly independent; hence, we cannot have more objects in our set than the dimension of V. This fact—there cannot be more linearly independent vectors in V than the dimension of V—is usually called the "linear algebra bound." We will consider this tool in great details in Part III. Here we restrict ourselves with just one important application.

Theorem 7.5 (Fisher's inequality). *Let $A_1, \dots, A_m$ be distinct subsets of $\{1, \dots, n\}$ such that $|A_i \cap A_j| = k$ for some fixed $1 \le k \le n$ and every $i \neq j$. Then $m \le n$.*

Proof. For two vectors $x, y \in \mathbb{R}^n$, let $\langle x, y\rangle = x_1y_1 + \cdots + x_ny_n$ denote their scalar product. Let $v_1, \dots, v_m \in \{0,1\}^n$ be incidence vectors of $A_1, \dots, A_m$. By the linear algebra bound it is enough to show that these vectors are linearly independent over the reals. Assume the contrary, i.e., that the linear relation $\sum_{i=1}^m \lambda_i v_i = 0$ exists, with not all coefficients being zero. Obviously, $\langle v_i, v_j\rangle = |A_i|$ if $j = i$, and $\langle v_i, v_j\rangle = k$ if $j \neq i$. Consequently,

$$\begin{aligned} 0 &= \Big(\sum_{i=1}^m \lambda_i v_i\Big)\Big(\sum_{j=1}^m \lambda_j v_j\Big) = \sum_{i=1}^m \lambda_i^2 \langle v_i, v_i\rangle + \sum_{1\le i\neq j\le m} \lambda_i\lambda_j \langle v_i, v_j\rangle \\ &= \sum_{i=1}^m \lambda_i^2 |A_i| + \sum_{1\le i\neq j\le m} \lambda_i\lambda_j k = \sum_{i=1}^m \lambda_i^2 (|A_i| - k) + k \cdot \Big(\sum_{i=1}^m \lambda_i\Big)^2 . \end{aligned}$$

Clearly, $|A_i| \ge k$ for all i and $|A_i| = k$ for at most one i, since otherwise the intersection condition would not be satisfied. But then the right-hand is greater than 0 (because the last sum can vanish only if at least two of the coefficients λ_i are nonzero), a contradiction. □

This theorem was first proved by the statistician R. A. Fisher in 1940 for the case when $k = 1$ and all sets A_i have the same size (such configurations are known as balanced incomplete block designs). In 1948, de Bruijn and Erdős relaxed the uniformity condition for the sets A_i (see Theorem 12.4). This was generalized by R. C. Bose in 1949, and later by several other authors. But it was the two-page paper of Bose where the linear argument was first applied to solve a combinatorial problem. The general version, stated above, was first proved by Majumdar (1953); the proof we presented is a variation of a simplified argument found by Babai and Frankl (1992).

7.4 Maximal intersecting families

Let $\mathcal{F}$ be a k-uniform family of sets of some n-element set. Say that $\mathcal{F}$ is *maximal intersecting* if

(i) $\mathcal{F}$ is intersecting;
(ii) the addition of any new k-element set to $\mathcal{F}$ destroys this property, that is, for every k-element subset $E \notin \mathcal{F}$, the family $\mathcal{F} \cup \{E\}$ is no longer intersecting.

The case when $n \le 2k - 1$ is not interesting, because then the only maximal intersecting family is the family of all k-element subsets. But what if $n \ge 2k$? Intuitively, any maximal intersecting family must be large enough, because *every* k-element set not in the family must be avoided by at least one of its members. It is therefore interesting to investigate the minimal possible number $f(k)$ of members which such a family can have.

To give an upper bound on $f(k)$, consider the family $\mathcal{F}$ of lines in a projective plane of order $k-1$ (see Sect. 12.4). For our current purposes it is enough to know that $\mathcal{F}$ is a family of $|\mathcal{F}| = n = k^2 - k + 1$ k-element subsets (called *lines*) of an n-element set of points such that each two lines intersect in precisely one point and each point belongs to precisely k lines. It is easy to show that this family is maximal intersecting (see Exercise 7.7). Hence, $f(k) \le k^2 - k + 1$ for all those values of k for which a projective plane of order $k-1$ exists.

In the case of projective planes, we have a k-uniform family with about k^2 sets. But the number of points in this case is also the same. What if we take a lot fewer than k^2 points? Can we then still find a k-uniform and maximal intersecting family of size at most k^2? Using double-counting we can answer this question negatively.

Theorem 7.6 (Füredi 1980)**.** *Let $\mathcal{F}$ be a maximal intersecting family of k-element sets of an n-element set. If $n \le k^2/2\log k$, then $\mathcal{F}$ must have more than k^2 members.*

Proof. To simplify computations, we only prove the theorem under a slightly stronger assumption that $n \le k^2/(1+2\log k)$. The idea is to count in two ways the number N of pairs (F, E) where $F \in \mathcal{F}$ and E is a k-element subset disjoint from F (and hence, $E \notin \mathcal{F}$). Since every such set E must be avoided by at least one member of $\mathcal{F}$, $N \ge \binom{n}{k} - |\mathcal{F}|$. On the other hand, each member of $\mathcal{F}$ can avoid at most $\binom{n-k}{k}$ of the sets E; hence $N \le |\mathcal{F}| \cdot \binom{n-k}{k}$. These two inequalities, together with the estimate (1.23), imply

$$|\mathcal{F}| \ge \frac{\binom{n}{k}}{1+\binom{n-k}{k}} \ge \frac{1}{2}\cdot\left(\frac{n}{n-k}\right)^k > \mathrm{e}^{k^2/n-1} \ge \mathrm{e}^{2\log k} \ge k^2 . \qquad \square$$

Now suppose that $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are intersecting (not necessarily uniform) families of an n-element set $\{1, \ldots, n\}$. How many sets can we have in their union?

Taking each $\mathcal{F}_i$ to be the family of all 2^{n-1} subsets containing the element i, we see that the union will have $2^n - 2^{n-m}$ sets.

A beautiful result, due to Kleitman, says that this bound is best possible.

Theorem 7.7 (Kleitman 1966)**.** *The union of m intersecting families contains at most $2^n - 2^{n-m}$ sets.*

Proof. We apply induction on m. The case $m = 1$ being trivial, we turn to the induction step. We say that a family $\mathcal{A}$ is *monotone increasing* (*monotone decreasing*) if $A \in \mathcal{A}$ and $B \supseteq A$ (respectively, $B \subseteq A$) implies $B \in \mathcal{A}$. A famous result, also due to Kleitman (we will prove it in Sect. 10.2; see

Theorem 10.6 and Exercise 10.8) says that, if $\mathcal{A}$ is a monotone decreasing and $\mathcal{B}$ is a monotone increasing family of subsets of an n-element set, then

$$|\mathcal{A} \cap \mathcal{B}| \leq 2^{-n}|\mathcal{A}| \cdot |\mathcal{B}|\,. \tag{7.1}$$

Now let $\mathcal{F} = \bigcup_{i=1}^{m} \mathcal{F}_i$, with each $\mathcal{F}_i$ being an intersecting family. Since our aim is to bound $|\mathcal{F}|$ from the above, we may assume that each $\mathcal{F}_i$ is maximal intersecting family; in particular, $|\mathcal{F}_m| = 2^{n-1}$. Let $\mathcal{A}$ be the complement of $\mathcal{F}_m$, i.e., the family of all $|\mathcal{A}| = 2^{n-1}$ subsets not in $\mathcal{F}_m$, and $\mathcal{B} = \bigcup_{i=1}^{m-1} \mathcal{F}_i$. Since $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are maximal intersecting families, $\mathcal{A}$ is monotone decreasing and $\mathcal{B}$ is monotone increasing. By the induction hypothesis, $|\mathcal{B}| \leq 2^n - 2^{n-m+1}$, and, by (7.1),

$$|\mathcal{A} \cap \mathcal{B}| \leq 2^{-n}2^{n-1}(2^n - 2^{n-m+1}) = 2^{n-1} - 2^{n-m}\,.$$

Therefore,

$$|\mathcal{B} \cap \mathcal{F}_m| = |\mathcal{B}| - |\mathcal{A} \cap \mathcal{B}| \geq |\mathcal{B}| - 2^{n-1} + 2^{n-m}$$

and

$$|\mathcal{F}| = \left|\bigcup_{i=1}^{m} \mathcal{F}_i\right| = |\mathcal{B}| + |\mathcal{F}_m| - |\mathcal{B} \cap \mathcal{F}_m| \leq 2^n - 2^{n-m}\,. \qquad \square$$

7.5 Cross-intersecting families

A pair of families $\mathcal{A}, \mathcal{B}$ is *cross-intersecting* if every set in $\mathcal{A}$ intersects every set in $\mathcal{B}$. The *rank* of $\mathcal{A}$ is the maximum cardinality of a set in $\mathcal{A}$. The *degree* $d_{\mathcal{A}}(x)$ of a point x in $\mathcal{A}$ is the number of sets in $\mathcal{A}$ containing x.

If $\mathcal{A}$ has rank a, then, by the pigeonhole principle, each set in $\mathcal{A}$ contains a point x which is "popular" for the members of $\mathcal{B}$ in that $d_{\mathcal{B}}(x) \geq |\mathcal{B}|/a$. Similarly, if $\mathcal{B}$ has rank b, then each member of $\mathcal{B}$ contains a point y for which $d_{\mathcal{A}}(y) \geq |\mathcal{A}|/b$. However, this alone does not imply that we can find a point which is popular in *both* families $\mathcal{A}$ and $\mathcal{B}$. It turns out that if we relax the "degree of popularity" by one-half, then such a point exists.

Theorem 7.8 (Razborov–Vereshchagin 1999)**.** *Let $\mathcal{A}$ be a family of rank a and $\mathcal{B}$ be a family of rank b. Suppose that the pair $\mathcal{A}, \mathcal{B}$ is cross-intersecting. Then there exists a point x such that*

$$d_{\mathcal{A}}(x) \geq \frac{|\mathcal{A}|}{2b} \quad \textit{and} \quad d_{\mathcal{B}}(x) \geq \frac{|\mathcal{B}|}{2a}.$$

Proof. Assume the contrary and let $\boldsymbol{A}, \boldsymbol{B}$ be independent random sets that are uniformly distributed in $\mathcal{A}, \mathcal{B}$ respectively. That is, for each $A \in \mathcal{A}$ and $B \in \mathcal{B}$, $\Pr[\boldsymbol{A} = A] = 1/|\mathcal{A}|$ and $\Pr[\boldsymbol{B} = B] = 1/|\mathcal{B}|$. Since the pair $\mathcal{A}, \mathcal{B}$ is cross-intersecting, the probability of the event "$\exists x (x \in \boldsymbol{A} \cap \boldsymbol{B})$" is equal to

1. Since the probability of a disjunction of events is at most the sum of the probabilities of the events, we have

$$\sum_x \Pr[x \in \boldsymbol{A} \cap \boldsymbol{B}] \geq 1.$$

Let X_0 consist of those points x for which

$$\frac{d_{\mathcal{A}}(x)}{|\mathcal{A}|} = \Pr[x \in \boldsymbol{A}] < \frac{1}{2b},$$

and X_1 consist of the remaining points. Note that by our assumption, for any $x \in X_1$,

$$\Pr[x \in \boldsymbol{B}] = \frac{d_{\mathcal{B}}(x)}{|\mathcal{B}|} < \frac{1}{2a}$$

holds. By double counting (see Proposition 1.7), $\sum_x d_{\mathcal{A}}(x) = \sum_{A \in \mathcal{A}} |A|$. Hence,

$$\begin{aligned}
&\sum_{x \in X_1} \Pr[x \in \boldsymbol{A} \cap \boldsymbol{B}] = \sum_{x \in X_1} \Pr[x \in \boldsymbol{A}] \cdot \Pr[x \in \boldsymbol{B}] \\
&< \frac{1}{2a} \cdot \sum_{x \in X_1} \Pr[x \in \boldsymbol{A}] \leq \frac{1}{2a} \cdot \sum_x \Pr[x \in \boldsymbol{A}] \\
&= \frac{1}{2a} \cdot \sum_x \frac{d_{\mathcal{A}}(x)}{|\mathcal{A}|} = \frac{1}{2a|\mathcal{A}|} \cdot \sum_x d_{\mathcal{A}}(x) = \frac{1}{2a|\mathcal{A}|} \cdot \sum_{A \in \mathcal{A}} |A| \leq \frac{a|\mathcal{A}|}{2a|\mathcal{A}|} = \frac{1}{2}.
\end{aligned}$$

In a similar way we obtain

$$\sum_{x \in X_0} \Pr[x \in \boldsymbol{A} \cap \boldsymbol{B}] < \frac{1}{2},$$

a contradiction. □

We mention (without proof) the following related result.

Theorem 7.9 (Füredi 1995). *Let $a + b \leq n$, $\mathcal{A}$ be a family of a-element sets and $\mathcal{B}$ a family of b-element sets on the common underlying set $[n]$ such that $|\mathcal{A}| \geq \binom{n-1}{a-1} - \binom{n-b-1}{a-1} + 1$ and $|\mathcal{B}| \geq \binom{n-1}{b-1} - \binom{n-a-1}{b-1} + 1$. If the pair $\mathcal{A}, \mathcal{B}$ is cross-intersecting, then some element $x \in [n]$ belongs to all members of $\mathcal{A}$ and $\mathcal{B}$.*

Exercises

7.1. Let $\mathcal{F}$ be a family of subsets of an n-element set. Prove that if $\mathcal{F}$ is intersecting then $|\mathcal{F}| \leq 2^{n-1}$. Is this the best bound? If so, then exhibit an intersecting family with $|\mathcal{F}| = 2^{n-1}$. *Hint*: A set and its complement cannot both be the members of $\mathcal{F}$.

7.2. Let $\mathcal{F}$ be an intersecting family of subsets of an n-element set X. Show that there is an intersecting family $\mathcal{F}' \supseteq \mathcal{F}$ such that $|\mathcal{F}| = 2^{n-1}$. *Hint*: Show that for any set A such that neither A nor $\overline{A}$ belongs to $\mathcal{F}$, exactly one of A and $\overline{A}$ can be added to $\mathcal{F}$.

7.3. Let $n \leq 2k$ and let $A_1, \ldots, A_m$ be a family of k-element subsets of $[n]$ such that $A_i \cup A_j \neq [n]$ for all i, j. Show that $m \leq \left(1 - \frac{k}{n}\right)\binom{n}{k}$. *Hint*: Apply the Erdős–Ko–Rado theorem to the complements $\overline{A_i} = [n] - A_i$.

7.4. The upper bound $\binom{n-1}{k-1}$ given by Erdős–Ko–Rado theorem is achieved by the families of sets containing a fixed element. Show that for $n = 2k$ there are other families achieving this bound. *Hint*: Include one set out of every pair of sets formed by a k-element set and its complement.

7.5. One can generalize the intersection property and require that $|A \cap B| \geq t$ for all $A \neq B \in \mathcal{F}$. Such families are called *t-intersecting*. The first example of a t-intersecting family which comes to mind, is the family of all subsets of $[n]$ containing some fixed set of t elements. This family has 2^{n-t} sets. Are there larger t-intersecting families? *Hint*: Let $n + t$ be even and take $\mathcal{F} = \left\{A \subseteq [n] \,:\, |A| = \frac{n+t}{2}\right\}$.

7.6. Let $\underset{.}{=}\{B_1, \ldots, B_b\}$ be a (v, k, λ) design, i.e., a family of k-element subsets of a v-element set of points $X = \{x_1, \ldots, x_v\}$ such that every two points belong to exactly λ sets. Use Fisher's inequality to show that $b \geq v$. *Hint*: Take $A_i := \{j \,:\, x_i \in B_j\}$.

7.7. Consider the k-uniform family of all $n = k^2 - k + 1$ lines in the set of points of a projective plane of order $k - 1$. Clearly, this family is intersecting. Show that it is also maximal intersecting, i.e., that every k-element set E, which intersects all the lines, must be a line. *Hint*: Assume that E is not a line, draw a line L through some two points $x \neq y$ of E, and take a point $z \in L \setminus \{x, y\}$. This point belongs to k lines, and each of them intersect E.

7.8. (Razborov–Vereshchagin 1999). Show that the bound in Theorem 7.8 is tight up to a multiplicative factor of 2. *Hint*: Consider the following pair of families

$$\begin{aligned} \mathcal{A} &= \{A_1, \ldots, A_b\}, \quad \text{where} \quad A_i = \{(i,1), (i,2), \ldots, (i,a)\}, \\ \mathcal{B} &= \{B_1, \ldots, B_a\}, \quad \text{where} \quad B_j = \{(1,j), (2,j), \ldots, (b,j)\}. \end{aligned}$$

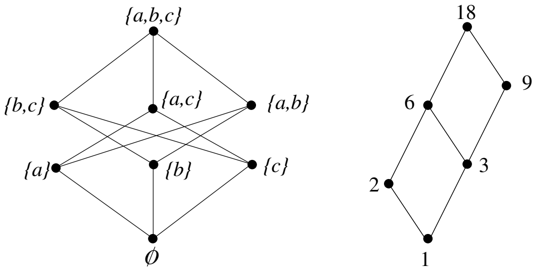
8. Chains and Antichains

Partial ordered sets provide a common frame for many combinatorial configurations. Formally, a *partially ordered set* (or *poset*, for short) is a set P together with a binary relation $<$ between its elements which is transitive and antysymmetric: if $x < y$ and $y < z$ then $x < z$, but $x < y$ and $y < x$ cannot both hold. We write $x \leq y$ if $x < y$ or $x = y$. Elements x and y are *comparable* if either $x \leq y$ or $y \leq x$ (or both) hold.

A *chain* in a poset P is a subset $C \subseteq P$ such that any two of its points are comparable. Dually, an *antichain* is a subset $A \subseteq P$ such that no two of its points are comparable. Observe that $|C \cap A| \leq 1$, i.e., every chain C and every antichain A can have at most one element in common (for two points in their intersection would be both comparable and incomparable).

Here are some frequently encountered examples of posets: a family of sets is partially ordered by set inclusion; a set of positive integers is partially ordered by division; a set of vectors in $\mathbb{R}^n$ is partially ordered by $(a_1, \ldots, a_n) < (b_1, \ldots, b_n)$ iff $a_i \leq b_i$ for all i, and $a_i < b_i$ for at least one i.

Small posets may be visualized by drawings, known as *Hasse diagrams*: x is lower in the plane than y whenever $x < y$ and there is no other point $z \in P$ for which both $x < z$ and $z < y$. For example:



S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_8,
© Springer-Verlag Berlin Heidelberg 2011

8.1 Decomposition in chains and antichains

A decomposition of a poset is its partition into mutually disjoint chains or antichains. Given a poset P, our goal is to decompose it into as few chains (or antichains) as possible. One direction is easy: if a poset P has a chain (antichain) of size r then it cannot be partitioned into fewer than r antichains (chains). The reason here is simple: any two points of the same chain must lie in different members of a partition into antichains.

Is this optimal? If P has no chain (or antichain) of size greater than r, is it then possible to partition P into r antichains (or chains, respectively)? One direction is straightforward (see Exercise 8.8 for an alternative proof):

Theorem 8.1. *Suppose that the largest chain in the poset P has size r. Then P can be partitioned into r antichains.*

Proof. Let A_i be the set of points $x \in P$ such that the longest chain, whose greatest element is x, has i points (including x). Then, by the hypothesis, $A_i = \emptyset$ for $i \geq r+1$, and hence, $P = A_1 \cup A_2 \cup \cdots \cup A_r$ is a partition of P into r mutually disjoint subsets (some of them may be also empty). Moreover, each A_i is an antichain, since if $x, y \in A_i$ and $x < y$, then the longest chain $x_1 < \ldots < x_i = x$ ending in x could be prolonged to a longer chain $x_1 < \ldots < x_i < y$, meaning that $y \notin A_i$. □

The dual result looks similar, but its proof is more involved. This result, uniformly known as Dilworth's Decomposition Theorem (Dilworth 1950) has played an important role in motivating research into posets. There are several elegant proofs; the one we present is due to F. Galvin.

Theorem 8.2 (Dilworth's theorem)**.** *Suppose that the largest antichain in the poset P has size r. Then P can be partitioned into r chains.*

Proof (due to Galvin 1994). We use induction on the cardinality of P. Let a be a maximal element of P, and let r be the size of a largest antichain in $P' = P \setminus \{a\}$. Then P' is the union of r disjoint chains $C_1, \ldots, C_r$. We have to show that P either contains an $(r+1)$-element antichain or else is the union of r disjoint chains. Now, every r-element antichain in P' consists of one element from each C_i. Let a_i be the maximal element in C_i which belongs to some r-element antichain in P'. It is easy to see that $A = \{a_1, \ldots, a_r\}$ is an antichain in P'. If $A \cup \{a\}$ is an antichain in P, we are done: we have found an antichain of size $r+1$). Otherwise, we have $a > a_i$ for some i. Then $K = \{a\} \cup \{x \in C_i : x \leq a_i\}$ is a chain in P, and there are no r-element antichains in $P \setminus K$ (since a_i was the maximal element of C_i participating in such an antichain), whence $P \setminus K$ is the union of $r-1$ chains. □

To recognize the power of this theorem, let us show that it contains Hall's Marriage Theorem 5.1 as a special case!

Suppose that $S_1, \ldots, S_m$ are sets satisfying Hall's condition, i.e., $|S(I)| \geq |I|$ for all $I \subseteq \{1, \ldots, m\}$, where $S(I) := \bigcup_{i \in I} S_i$. We construct a poset P as follows. The points of P are the elements of $X := S_1 \cup \dots \cup S_m$ and symbols $y_1, \dots, y_m$, with $x < y_i$ if $x \in S_i$, and no other comparabilities. It is clear that X is an antichain in P. We claim that there is no larger antichain. To show this, let A be an antichain, and set $I := \{i \,:\, y_i \in A\}$. Then A contains no point of $S(I)$, for if $x \in S_i$ then x is comparable with y_i, and hence, A cannot contain both of these points. Hence, Hall's condition implies that $|A| \leq |I| + |X| - |S(I)| \leq |X|$, as claimed.

Now, Dilworth's theorem implies that P can be partitioned into $|X|$ chains. Since the antichain X is maximal, each of the chains in the partition must contain a point of X. Let the chain through y_i be $\{x_i, y_i\}$. Then $(x_1, \dots, x_m)$ is a desired system of distinct representatives: for $x_i \in S_i$ (since $x_i < y_i$) and $x_i \neq x_j$ (since the chains are disjoint).

In general, Dilworth's theorem says nothing more about the chains, forming the partition, except that they are mutually disjoint. However, if we consider special posets then we can extract more information about the partition. To illustrate this, let us consider now the poset 2^X whose points are all subsets of an n-element set X partially ordered by set inclusion. De Bruijn, Tengbergen, and Kruyswijk (1952) have shown that 2^X can be partitioned into disjoint chains that are also "symmetric."

Let $\mathcal{C} = \{A_1, \dots, A_k\}$ be a chain in 2^X, i.e., $A_1 \subset A_2 \subset \ldots \subset A_k$. This chain is *symmetric* if $|A_1| + |A_k| = n$ and $|A_{i+1}| = |A_i| + 1$ for all $i = 1, \dots, k-1$. "Symmetric" here means symmetric positioned about the middle level $\frac{n}{2}$. Symmetric chains with $k = n$ are *maximal.* Maximal chains are in one-to-one correspondence with the permutations of the underlying set: every permutation $(x_1, \dots, x_n)$ gives the maximal chain

$$\{x_1\} \subset \{x_1, x_2\} \subset \ldots \subset \{x_1, \dots, x_n\}.$$

Theorem 8.3. *The family of all subsets of an n-element set can be partitioned into $\binom{n}{\lfloor n/2 \rfloor}$ mutually disjoint symmetric chains.*

Proof. Take an n-element set X, and assume for a moment that we already have some partition of 2^X into symmetric chains. Every such chain contains exactly one set from the middle level; hence there are $\binom{n}{\lfloor n/2 \rfloor}$ chains in that partition.

Let us now prove that such a partition is possible at all. We argue by the induction on $n = |X|$. Clearly the result holds for the one point set X. So, suppose that it is true for all sets with fewer points then n. Pick a point $x \in X$, and let $Y := X \setminus \{x\}$. By induction, we can partition 2^Y into symmetric chains $\mathcal{C}_1, \dots, \mathcal{C}_r$. Each of these chains over Y

$$\mathcal{C}_i \;:=\; A_1 \subset A_2 \subset \ldots \subset A_k$$

produce the following two chains over the whole set X:

$$\mathcal{C}_i' := A_1 \subset A_2 \subset \ldots \subset A_{k-1} \subset A_k \subset A_k \cup \{x\}$$

$$\mathcal{C}_i'' := A_1 \cup \{x\} \subset A_2 \cup \{x\} \subset \ldots \subset A_{k-1} \cup \{x\}.$$

These chains are symmetric since

$$|A_1| + |A_k \cup \{x\}| = (|A_1| + |A_k|) + 1 = (n-1) + 1 = n$$

and

$$|A_1 \cup \{x\}| + |A_{k-1} \cup \{x\}| = (|A_1| + |A_{k-1}|) + 2 = (n-2) + 2 = n.$$

Is this a partition? It is indeed. If $A \subseteq Y$ then only $\mathcal{C}_i'$ contains A where $\mathcal{C}_i$ is the chain in 2^Y containing A. If $A = B \cup \{x\}$ where $B \subseteq Y$ then $B \in \mathcal{C}_i$ for some i. If B is the maximal element of $\mathcal{C}_i$ then $\mathcal{C}_i'$ is the only chain containing A, otherwise A is contained only in $\mathcal{C}_i''$. □

8.2 Application: the memory allocation problem

The following problem arises in information storage and retrieval. Suppose we have some list (a sequence) $L = (a_1, a_2, \ldots, a_m)$ of not necessarily distinct elements of some set X. We say that this list *contains* a subset A if it contains A as a subsequence of consecutive terms, that is, if

$$A = \{a_i, a_{i+1}, \ldots, a_{i+|A|-1}\}$$

for some i. A sequence is *universal* for X if it contains all the subsets of X. For example, if $X = \{1, 2, 3, 4, 5\}$ then the list

$$L = (1\,2\,3\,4\,5\,1\,2\,4\,1\,3\,5\,2\,4)$$

of length $m = 13$ is universal for X.

What is the length of a shortest universal sequence for an n-element set? Since any two sets of equal cardinality must start from different places of this string, the trivial lower bound for the length of universal sequence is $\binom{n}{\lfloor n/2 \rfloor}$, which is about $\sqrt{\frac{2}{\pi n}}2^n$, according to Stirling's formula (1.7). A trivial upper bound for the length of the shortest universal sequence is obtained by considering the sequence obtained simply by writing down each subset one after the other. Since there are 2^n subsets of average size $n/2$, the length of the resulting universal sequence is at most $n2^{n-1}$. Using Dilworth's theorem, we can obtain a universal sequence, which is n times (!) shorter than this trivial one.

Theorem 8.4 (Lipski 1978)**.** *There is a universal sequence for* $\{1, \ldots, n\}$ *of length at most* $\frac{2}{\pi}2^n$.

Proof. We consider the case when n is even, say $n = 2k$ (the case of odd n is similar). Let $S = \{1, \ldots, k\}$ be the set of the first k elements and $T =$

$\{k+1,\ldots,2k\}$ the set of the last k elements. By Theorem 8.3, both S and T have symmetric chain decompositions of their posets of subsets into $m = \binom{k}{k/2}$ symmetric chains: $2^S = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_m$ and $2^T = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_m$. Corresponding to the chain

$$\mathcal{C}_i = \{x_1,\ldots,x_j\} \subset \{x_1,\ldots,x_j,x_{j+1}\} \subset \ldots \subset \{x_1,\ldots,x_h\} \quad (j+h=k)$$

we associate the sequence (not the set!) $C_i = (x_1, x_2, \ldots x_h)$. Then every subset of S occurs as an *initial* part of one of the sequences $C_1,\ldots,C_m$. Similarly let $D_1,\ldots,D_m$ be sequences corresponding to the chains $\mathcal{D}_1,\ldots,\mathcal{D}_m$. If we let $\overline{D}_i$ denote the sequence obtained by writing D_i in reverse order, then every subset of T occurs as a *final* part of one of the $\overline{D}_i$. Next, consider the sequence

$$L = \overline{D}_1 C_1 \overline{D}_1 C_2 \ldots \overline{D}_1 C_m \ldots \overline{D}_m C_1 \overline{D}_m C_2 \ldots \overline{D}_m C_m.$$

We claim that L is a universal sequence for the set $\{1,\ldots,n\}$. Indeed, each of its subsets A can be written as $A = E \cup F$ where $E \subseteq S$ and $F \subseteq T$. Now F occurs as the final part of some $\overline{D}_f$ and E occurs as the initial part of some C_e; hence, the whole set A occurs in the sequence L as the part of $\overline{D}_f C_e$. Thus, the sequence L contains every subset of $\{1,\ldots,n\}$. The length of the sequence L is at most $km^2 = k\binom{k}{k/2}^2$. Since, by Stirling's formula, $\binom{k}{k/2} \sim 2^k\sqrt{\frac{2}{k\pi}}$, the length of the sequence is $km^2 \sim k\frac{2}{k\pi} \cdot 2^{2k} = \frac{2}{\pi}2^n$. □

8.3 Sperner's theorem

A set system $\mathcal{F}$ is an *antichain* (or *Sperner system*) if no set in it contains another: if $A, B \in \mathcal{F}$ and $A \neq B$ then $A \not\subseteq B$. It is an antichain in the sense that this property is the other extreme from that of the chain in which every pair of sets is comparable.

Simplest examples of antichains over $\{1,\ldots,n\}$ are the families of all sets of fixed cardinality k, $k = 0,1,\ldots,n$. Each of these antichains has $\binom{n}{k}$ members. Recognizing that the maximum of $\binom{n}{k}$ is achieved for $k = \lfloor n/2 \rfloor$, we conclude that there are antichains of size $\binom{n}{\lfloor n/2 \rfloor}$. Are these antichains the largest ones?

The positive answer to this question was found by Emanuel Sperner in 1928, and this result is known as Sperner's Theorem.

Theorem 8.5 (Sperner 1928). *Let $\mathcal{F}$ be a family of subsets of an n element set. If $\mathcal{F}$ is an antichain then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

A considerably sharper result, Theorem 8.6 below, is due to Lubell (1966). The same result was discovered by Meshalkin (1963) and (not so explicitly) by Yamamoto (1954). Although Lubell's result is also a rather special case of an earlier result of Bollobás (see Theorem 8.8 below), inequality (8.1) has become known as the *LYM inequality.*

Theorem 8.6 (LYM Inequality)**.** *Let $\mathcal{F}$ be an antichain over a set X of n elements. Then*

$$\sum_{A\in\mathcal{F}} \binom{n}{|A|}^{-1} \leq 1. \tag{8.1}$$

Note that Sperner's theorem follows from this bound: recognizing that $\binom{n}{k}$ is maximized when $k = \lfloor n/2 \rfloor$, we obtain

$$|\mathcal{F}| \cdot \binom{n}{\lfloor n/2 \rfloor}^{-1} \leq \sum_{A\in\mathcal{F}} \binom{n}{|A|}^{-1} \leq 1.$$

We will give an elegant proof of Theorem 8.6 due to Lubell (1966) together with one of its reformulations which is pregnant with further extensions.

First proof. For each subset A, exactly $|A|!(n - |A|)!$ maximal chains over X contain A. Since none of the $n!$ maximal chains meet $\mathcal{F}$ more than once, we have $\sum_{A\in\mathcal{F}} |A|!(n - |A|)! \leq n!$. Dividing this inequality by $n!$ we get the desired result. □

Second proof. The idea is to associate with each subset $A \subseteq X$, a permutation on X, and count their number. For an a-element set A let us say that a permutation $(x_1, x_2, \ldots, x_n)$ of X *contains* A if $\{x_1, \ldots, x_a\} = A$. Note that A is contained in precisely $a!(n-a)!$ permutations. Now if $\mathcal{F}$ is an antichain, then each of $n!$ permutations contains at most one $A \in \mathcal{F}$. Consequently, $\sum_{A\in\mathcal{F}} a!(n - a)! \leq n!$, and the result follows. To recover the first proof, simply identify a permutation $(x_1, x_2, \ldots, x_n)$ with the maximal chain $\{x_1\} \subset \{x_1, x_2\} \subset \ldots \subset \{x_1, x_2 \ldots, x_n\} = X$. □

8.4 The Bollobás theorem

The following theorem due to B. Bollobás is one of the cornerstones in extremal set theory. Its importance is reflected, among other things, by the list of different proofs published as well as the list of different generalizations. In particular, this theorem implies both Sperner's theorem and the LYM inequality.

Theorem 8.7 (Bollobás' theorem)**.** *Let $A_1, \ldots, A_m$ be a-element sets and $B_1, \ldots, B_m$ be b-element sets such that $A_i \cap B_j = \emptyset$ if and only if $i = j$. Then $m \leq \binom{a+b}{a}$.*

This is a special case of the following result.

Theorem 8.8 (Bollobás 1965). *Let $A_1, \dots, A_m$ and $B_1, \dots, B_m$ be two sequences of sets such that $A_i \cap B_j = \emptyset$ if and only if $i = j$. Then*

$$\sum_{i=1}^{m} \binom{a_i + b_i}{a_i}^{-1} \le 1, \tag{8.2}$$

where $a_i = |A_i|$ and $b_i = |B_i|$.

As we already mentioned, due to its importance, there are several different proofs of this theorem. We present two of them.

First proof. Our goal is to prove that (8.2) holds for every family $\mathcal{F} = \{(A_i, B_i) : i = 1, \dots, m\}$ of pairs of sets such that $A_i \cap B_j = \emptyset$ precisely when $i = j$. Let X be the union of all sets $A_i \cup B_i$. We argue by induction on $n = |X|$. For $n = 1$ the claim is obvious, so assume it holds for $n - 1$ and prove it for n. For every point $x \in X$, consider the family of pairs

$$\mathcal{F}_x := \{(A_i, B_i \setminus \{x\}) : x \notin A_i\}.$$

Since each of these families $\mathcal{F}_x$ has less than n points, we can apply the induction hypothesis for each of them, and sum the corresponding inequalities (8.2). The resulting sum counts $n - a_i - b_i$ times the term $\binom{a_i+b_i}{a_i}^{-1}$, corresponding to points $x \notin A_i \cup B_i$, and b_i times the term $\binom{a_i+b_i-1}{a_i}^{-1}$, corresponding to points $x \in B_i$; the total is $\le n$. Hence we obtain that

$$\sum_{i=1}^{m} (n - a_i - b_i) \binom{a_i + b_i}{a_i}^{-1} + b_i \binom{a_i + b_i - 1}{a_i}^{-1} \le n.$$

Since $\binom{k-1}{l} = \frac{k-l}{k}\binom{k}{l}$, the i-th term of this sum is equal to $n \cdot \binom{a_i+b_i}{a_i}^{-1}$. Dividing both sides by n we get the result. □

Second proof. Lubell's method of counting permutations. Let, as before, X be the union of all sets $A_i \cup B_i$. If A and B are disjoint subsets of X then we say that a permutation $(x_1, x_2, \dots, x_n)$ of X *separates* the pair (A, B) if no element of B precedes an element of A, i.e., if $x_k \in A$ and $x_l \in B$ imply $k < l$.

Each of the $n!$ permutations can separate at most one of the pairs (A_i, B_i), $i = 1, \dots, m$. Indeed, suppose that $(x_1, x_2, \dots, x_n)$ separates two pairs (A_i, B_i) and (A_j, B_j) with $i \ne j$, and assume that $\max\{k : x_k \in A_i\} \le \max\{k : x_k \in A_j\}$. Since the permutation separates the pair (A_j, B_j),

$$\min\{l : x_l \in B_j\} > \max\{k : x_k \in A_j\} \ge \max\{k : x_k \in A_i\}$$

which implies that $A_i \cap B_j = \emptyset$, contradicting the assumption.

We now estimate the number of permutations separating one fixed pair. If $|A| = a$ and $|B| = b$ and A and B are disjoint then the pair (A, B) is separated by exactly

$$\binom{n}{a+b} a!b!(n-a-b)! = n! \binom{a+b}{a}^{-1}$$

permutations. Here $\binom{n}{a+b}$ counts the number of choices for the positions of $A \cup B$ in the permutation; having chosen these positions, A has to occupy the first a places, giving $a!$ choices for the order of A, and $b!$ choices for the order of B; the remaining elements can be chosen in $(n-a-b)!$ ways.

Since no permutation can separate two different pairs (A_i, B_i), summing up over all m pairs we get all permutations at most once

$$\sum_{i=1}^{m} n! \binom{a_i+b_i}{a_i}^{-1} \leq n!$$

and the desired bound (8.2) follows. □

Tuza (1984) observed that Bollobás's theorem implies both Sperner's theorem and the LYM inequality. Let $A_1, \ldots, A_m$ be an antichain over a set X. Take the complements $B_i = X \setminus A_i$ and let $a_i = |A_i|$ for $i = 1, \ldots, m$. Then $b_i = n - a_i$ and by (8.2)

$$\sum_{i=1}^{m} \binom{n}{|A_i|}^{-1} = \sum_{i=1}^{m} \binom{a_i+b_i}{a_i}^{-1} \leq 1.$$

Due to its importance, the theorem of Bollobás was extended in several ways.

Theorem 8.9 (Tuza 1985)**.** *Let $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ be collections of sets such that $A_i \cap B_i = \emptyset$ and for all $i \neq j$ either $A_i \cap B_j \neq \emptyset$ or $A_j \cap B_i \neq \emptyset$ (or both) holds. Then for any real number $0 < p < 1$, we have*

$$\sum_{i=1}^{m} p^{|A_i|}(1-p)^{|B_i|} \leq 1.$$

Proof. Let X be the union of all sets $A_i \cup B_i$. Choose a subset $\boldsymbol{Y} \subseteq X$ at random in such a way that each element $x \in X$ is included in $\boldsymbol{Y}$ independently and with the same probability p. Let E_i be the event that $A_i \subseteq \boldsymbol{Y} \subseteq X \setminus B_i$. Then for their probabilities we have $\Pr[E_i] = p^{|A_i|}(1-p)^{|B_i|}$ for every $i = 1, \ldots, m$ (see Exercise 8.4). We claim that, for $i \neq j$, the events E_i and E_j cannot occur at the same time. Indeed, otherwise we would have $A_i \cup A_j \subseteq \boldsymbol{Y} \subseteq X \setminus (B_i \cup B_j)$, implying $A_i \cap B_j = A_j \cap B_i = \emptyset$, which contradicts our assumption.

Since the events $E_1, \ldots, E_m$ are mutually disjoint, we conclude that $\Pr[E_1] + \cdots + \Pr[E_m] = \Pr[E_1 \cup \cdots \cup E_m] \leq 1$, as desired. □

The theorem of Bollobás also has other important extensions. We do not intend to give a complete account here; we only mention some of these results without proof. More information about Bollobás-type results can be found, for example, in a survey by Tuza (1994).

A typical generalization of Bollobás's theorem is its following "skew version." This result was proved by Frankl (1982) by modifying an argument of Lovász (1977) and was also proved in an equivalent form by Kalai (1984).

Theorem 8.10. *Let $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ be finite sets such that $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ if $i < j$. Also suppose that $|A_i| \leq a$ and $|B_i| \leq b$. Then $m \leq \binom{a+b}{a}$.*

We also have the following "threshold version" of Bollobás's theorem.

Theorem 8.11 (Füredi 1984)**.** *Let $A_1, \ldots, A_m$ be a collection of a-sets and $B_1, \ldots, B_m$ be a collection of b-sets such that $|A_i \cap B_i| \leq s$ and $|A_i \cap B_j| > s$ for every $i \neq j$. Then $m \leq \binom{a+b-2s}{a-s}$.*

8.5 Strong systems of distinct representatives

Recall that a system of distinct representatives for the sets $S_1, S_2, \ldots, S_k$ is a k-tuple $(x_1, x_2, \ldots, x_k)$ where the elements x_i are distinct and $x_i \in S_i$ for all $i = 1, 2, \ldots, k$. Such a system is *strong* if we additionally have $x_i \notin S_j$ for all $i \neq j$.

Theorem 8.12 (Füredi–Tuza 1985)**.** *In any family of more than $\binom{r+k}{k}$ sets of cardinality at most r, at least $k+2$ of its members have a strong system of distinct representatives.*

Proof. Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a family of sets, each of cardinality at most r. Suppose that *no* $k+2$ of these sets have a strong system of distinct representatives. We will apply the theorem of Bollobás to prove that then $m \leq \binom{r+k}{k}$. Let us make an additional assumption that our sets form an antichain, i.e., that no of them is a subset of another one. By Theorem 8.8 it is enough to prove that, for every $i = 1, \ldots, m$ there exists a set B_i, such that $|B_i| \leq k$, $B_i \cap A_i = \emptyset$ and $B_i \cap A_j \neq \emptyset$ for all $j \neq i$.

Fix an i and let $B_i = \{x_1, \ldots, x_t\}$ be a minimal set which intersects all the sets $A_j \setminus A_i$, $j = 1, \ldots, m$, $j \neq i$. (Such a set exists because none of these differences is empty.) By the minimality of B_i, for every $\nu = 1, \ldots, t$ there exists a set $S_\nu \in \mathcal{F}$ such that $B_i \cap S_\nu = \{x_\nu\}$. Fix an arbitrary element $y_i \in A_i$. Then $(y_i, x_1, \ldots, x_t)$ is a strong system of distinct representatives for $t+1$ sets $A_i, S_1, \ldots, S_t$. By the indirect assumption, we can have at most $k+1$ such sets. Therefore, $|B_i| = t \leq k$, as desired.

In the case when our family $\mathcal{F}$ is not an antichain, it is enough to order the sets so that $A_i \not\subseteq A_j$ for $i < j$, and apply the skew version of Bollobás's theorem. □

8.6 Union-free families

A family of sets $\mathcal{F}$ is called *r-union-free* if $A_0 \not\subseteq A_1 \cup A_2 \cup \cdots \cup A_r$ holds for all distinct $A_0, A_1, \ldots, A_r \in \mathcal{F}$. Thus, antichains are r-union-free for $r = 1$.

Let $T(n,r)$ denote the maximum cardinality of an r-union-free family $\mathcal{F}$ over an n-element underlying set. This notion was introduced by Kautz and Singleton (1964). They proved that

$$\Omega(1/r^2) \le \frac{\log_2 T(n,r)}{n} \le O(1/r).$$

This result was rediscovered several times in information theory, in combinatorics by Erdős, Frankl, and Füredi (1985), and in group testing by Hwang and Sós (1987). Dyachkov and Rykov (1982) obtained, with a rather involved proof, that

$$\frac{\log_2 T(n,r)}{n} \le O(\log_2 r/r^2).$$

Recently, Ruszinkó (1994) gave a purely combinatorial proof of this upper bound. Shortly after, Füredi (1996) found a very elegant argument, and we present it below.

Theorem 8.13 (Füredi 1996). *Let $\mathcal{F}$ be a family of subsets of an n-element underlying set X, and $r \ge 2$. If $\mathcal{F}$ is r-union-free then $|\mathcal{F}| \le r + \binom{n}{t}$ where*

$$t := \left\lceil (n-r)/\binom{r+1}{2} \right\rceil .$$

That is,

$$\log_2 |\mathcal{F}|/n \le O\left(\log_2 r/r^2\right).$$

Proof. Let $\mathcal{F}_t$ be the family of all members of $\mathcal{F}$ having their *own* t-subset. That is, $\mathcal{F}_t$ contains all those members $A \in \mathcal{F}$ for which there exists a t-element subset $T \subseteq A$ such that $T \not\subseteq A'$ for every other $A' \in \mathcal{F}$. Let $\mathcal{T}_t$ be the family of these t-subsets; hence $|\mathcal{T}_t| = |\mathcal{F}_t|$. Let $\mathcal{F}_0 := \{A \in \mathcal{F} : |A| < t\}$, and let $\mathcal{T}_0$ be the family of *all* t-subsets of X containing a member of $\mathcal{F}_0$, i.e.,

$$\mathcal{T}_0 := \{T : T \subseteq X, |T| = t \text{ and } T \supset A \text{ for some } A \in \mathcal{F}_0\}.$$

The family $\mathcal{F}$ is an antichain. This implies that $\mathcal{T}_t$ and $\mathcal{T}_0$ are disjoint. The family $\mathcal{F}_0$ is also an antichain, and since $t < n/2$, we know from Exercise 5.11 that $|\mathcal{F}_0| \le |\mathcal{T}_0|$. Therefore,

$$|\mathcal{F}_0 \cup \mathcal{F}_t| \le |\mathcal{T}_t| + |\mathcal{T}_0| \le \binom{n}{t}. \tag{8.3}$$

It remains to show that the family

$$\mathcal{F}' := \mathcal{F} \setminus (\mathcal{F}_0 \cup \mathcal{F}_t)$$

has at most r members. Note that $A \in \mathcal{F}'$ if and only if $A \in \mathcal{F}$, $|A| \geq t$ and for every t-subset $T \subseteq A$ there is an $A' \in \mathcal{F}$ such that $A' \neq A$ and $A' \supseteq T$. We will use this property to prove that $A \in \mathcal{F}'$, $A_1, A_2, \ldots, A_i \in \mathcal{F}$ $(i \leq r)$ imply

$$|A \setminus (A_1 \cup \cdots \cup A_i)| \geq t(r-i)+1. \tag{8.4}$$

To show this, assume the opposite. Then the set $A \setminus (A_1 \cup \cdots \cup A_i)$ can be written as the union of some $(r-i)$ t-element sets $T_{i+1}, \ldots T_r$. Therefore, A lies entirely in the union of $A_1, \ldots, A_i$ and these sets $T_{i+1}, \ldots, T_r$. But, by the choice of A, each of the sets T_j lies in some other set $A_j \in \mathcal{F}$ different from A. Therefore, $A \subseteq A_1 \cup \cdots \cup A_r$, a contradiction.

Now suppose that $\mathcal{F}'$ has more than r members, and take any $r+1$ of them $A_0, A_1, \ldots, A_r \in \mathcal{F}'$. Applying (8.4) we obtain

$$\begin{aligned} |\bigcup_{i=0}^{r} A_i| &= |A_0| + |A_1 \setminus A_0| + |A_2 \setminus (A_0 \cup A_1)| + \cdots \\ &\quad + |A_r \setminus (A_0 \cup A_1 \cup \cdots \cup A_{r-1})| \\ &\geq \big(tr+1\big) + \big(t(r-1)+1\big) + \big(t(r-2)+1\big) + \cdots + \big(t \cdot 0 + 1\big) \\ &= t \cdot \frac{r(r+1)}{2} + r + 1 = t\binom{r+1}{2} + r + 1. \end{aligned}$$

By the choice of t, the right-hand side exceeds the total number of points n, which is impossible. Therefore, $\mathcal{F}'$ cannot have more than r distinct members. Together with (8.3), this yields the desired upper bound on $|\mathcal{F}|$. □

Exercises

8.1. Let $\mathcal{F}$ be an antichain consisting of sets of size at most $k \leq n/2$. Show that $|\mathcal{F}| \leq \binom{n}{k}$.

8.2. Derive from Bollobás's theorem the following weaker version of Theorem 8.11. Let $A_1, \ldots, A_m$ be a collection of a-element sets and $B_1, \ldots, B_m$ be a collection of b-element sets such that $|A_i \cap B_i| = t$ for all i, and $|A_i \cap B_j| > t$ for $i \neq j$. Then $m \leq \binom{a+b-t}{a-t}$.

8.3. Show that the upper bounds in Bollobás's and Füredi's theorems (Theorems 8.7 and 8.11) are tight. *Hint*: Take two disjoint sets X and S of respective sizes $a+b-2s$ and s. Arrange the s-element subsets of X in any order: $Y_1, Y_2, \ldots$. Let $A_i = S \cup Y_i$ and $B_i = S \cup (X \setminus Y_i)$.

8.4. Use the binomial theorem to prove the following. Let $0 < p < 1$ be a real number, and $C \subset D$ be any two fixed subsets of $\{1, \ldots, n\}$. Then the sum of $p^{|A|}(1-p)^{n-|A|}$ over all sets A such that $C \subseteq A \subseteq D$, equals $p^{|C|}(1-p)^{n-|D|}$.

8.5. (Frankl 1986). Let $\mathcal{F}$ be a k-uniform family, and suppose that it is *intersection free*, i.e., that $A \cap B \not\subseteq C$ for any three sets A, B and C of $\mathcal{F}$. Prove that $|\mathcal{F}| \leq 1 + \binom{k}{\lfloor k/2 \rfloor}$. *Hint*: Fix a set $B_0 \in \mathcal{F}$, and observe that $\{A \cap B_0 : A \in \mathcal{F}, A \neq B_0\}$ is an antichain over B_0.

8.6. Let $A_1, \ldots, A_m$ be a family of subsets of an n-element set, and suppose that it is *convex* in the following sense: if $A_i \subseteq B \subseteq A_j$ for some i, j, then B belongs to the family. Prove that the absolute value of the sum $\sum_{i=1}^{m} (-1)^{|A_i|}$ does not exceed $\binom{n}{\lfloor n/2 \rfloor}$. *Hint*: Use the chain decomposition theorem. Observe that the contribution to the sum from each of the chains is of the form $\pm(1 - 1 + 1 - 1 \ldots)$, and so this contribution is $1, -1$ or 0.

8.7. Let $x_1, \ldots, x_n$ be real numbers, $x_i \geq 1$ for each i, and let S be the set of all numbers, which can be obtained as a linear combinations $\alpha_1 x_1 + \ldots + \alpha_n x_n$ with $\alpha_i \in \{-1, +1\}$. Let $I = [a, b)$ be any interval (in the real line) of length $b - a = 2$. Show that $|I \cap S| \leq \binom{n}{\lfloor n/2 \rfloor}$. *Hint*: Associate with each such sum $\xi = \alpha_1 x_1 + \ldots + \alpha_n x_n$ the corresponding set $A_\xi = \{i : \alpha_i = +1\}$ of indices i for which $\alpha_i = +1$. Show that the family of sets A_ξ for which $\xi \in I$, forms an antichain and apply Sperner's theorem. Note: Erdős (1945) proved a more general result that if $b - a = 2t$ then $|I \cap S|$ is less than or equal to the sum of the t largest binomial coefficients $\binom{n}{i}$.

8.8. Let P be a finite poset and suppose that the largest chain in it has size r. We know (see Theorem 8.1) that P can be partitioned into r antichains. Show that the following argument also gives the desired decomposition: let A_1 be the set of all maximal elements in P; remove this set from P, and let A_2 be the set of all maximal elements in the reduced set $P \setminus A_1$, etc.

8.9. Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ and suppose that

$$|A_i \cap A_j| < \frac{1}{r} \min\{|A_i|, |A_j|\} \quad \text{for all } i \neq j.$$

Show that $\mathcal{F}$ is r-union-free.

8.10. Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be an r-union-free family. Show that then $\bigcup_{i \in I} A_i \neq \bigcup_{j \in J} A_j$ for any two distinct non-empty subsets I, J of size at most r.

9. Blocking Sets and the Duality

In this chapter we will consider one of the most basic properties of set systems — their duality. The dual of a family $\mathcal{F}$ consists of all (minimal under set-inclusion) sets that intersect all members of $\mathcal{F}$. Dual families play an important role in many applications, boolean function complexity being just one example.

9.1 Duality

A *blocking set* of a family $\mathcal{F}$ is a set T that intersects (blocks) every member of $\mathcal{F}$. A blocking set of $\mathcal{F}$ is *minimal* if none of its proper subsets is such. (Minimal blocking sets are also called *transversals* of $\mathcal{F}$.) The family of all minimal blocking sets of $\mathcal{F}$ is called its *dual* and is denoted by $b\,(\mathcal{F})$.

Proposition 9.1. *For every family $\mathcal{F}$ we have $b\,(b\,(\mathcal{F})) \subseteq \mathcal{F}$. Moreover, if $\mathcal{F}$ is an antichain then $b\,(b\,(\mathcal{F})) = \mathcal{F}$.*

Proof. To prove the first claim, take a set $B \in b\,(b\,(\mathcal{F}))$. Observe that none of the sets $A \setminus B$ with $A \in \mathcal{F}$ can be empty: Since B is a *minimal* blocking set of $b\,(\mathcal{F})$, it cannot contain any member A of $\mathcal{F}$ as a proper subset, just because each member of $\mathcal{F}$ is a blocking set of $b\,(F)$. Assume now that $B \not\in \mathcal{F}$. Then, for each set $A \in \mathcal{F}$, there is a point $x_A \in A \setminus B$. The set $\{x_A\ :\ A \in \mathcal{F}\}$ of all such points is a blocking set of $\mathcal{F}$, and hence, contains at least one minimal blocking set $T \in b\,(\mathcal{F})$. But this is impossible, because then B must intersect the set T which, by it definition, can contain no element of B.

To prove the second claim, let $\mathcal{F}$ be an antichain, and take any $A \in \mathcal{F}$. We want to show A is in $b\,(b\,(\mathcal{F}))$. Each element of $b\,(\mathcal{F})$ intersects A, so A is a blocking set for $b\,(\mathcal{F})$. Therefore A contains (as a subset) some minimal blocking set $B \in b\,(b\,(\mathcal{F}))$. Since $b\,(b\,(\mathcal{F}))$ is a subset of $\mathcal{F}$ (by the first part of the proof), the set B must belong to $\mathcal{F}$. Hence, A and its subset B are both in $\mathcal{F}$. But $\mathcal{F}$ is an antichain, therefore $A = B$, so $A \in b\,(b\,(\mathcal{F}))$. □

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_9,

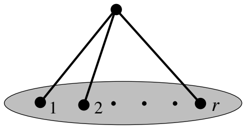


Fig. 9.1 Example of a self-dual family.

Let us consider the following problem of "keys of the safe" (Berge 1989). An administrative council is composed of a set X of individuals. Each of them carries a certain weight in decisions, and it is required that only subsets $A \subseteq X$ carrying a total weight greater than some threshold fixed in advance, should have access to documents kept in a safe with multiply locks. The minimal "coalitions" which can open the safe constitute an antichain $\mathcal{F}$. The problem consists in determining the minimal number of locks necessary so that by giving one or more keys to every individual, the safe can be opened if and only if at least one of the coalitions of $\mathcal{F}$ is present.

Proposition 9.2. *For every family $\mathcal{F}$ of minimal coalitions, $|b(\mathcal{F})|$ locks are enough.*

Proof. Let $b(\mathcal{F}) = \{T_1, \ldots, T_\ell\}$. Then give the key of the i-th lock to all the members of T_i. It is clear that then every coalition $A \in \mathcal{F}$ will have the keys to all ℓ locks, and hence, will be able to open the safe. On the other hand, if some set B of individuals does not include a coalition then, by Proposition 9.1, the set B is not a blocking set of $b(\mathcal{F})$, that is, $B \cap T_i = \emptyset$ for some i. But this means that people in B lack the i-th key, as desired. □

A family $\mathcal{F}$ is called *self-dual* if $b(\mathcal{F}) = \mathcal{F}$.

For example, the family of all k-element subsets of a $(2k-1)$-element set is self-dual. Another example is the family of $r+1$ sets, one of which has r elements and the remaining r sets have 2 elements (see Fig. 9.1).

What other families are self-dual? Our nearest goal is to show that a family is self-dual if and only if it is intersecting and not 2-colorable. Let us first recall the definition of these two concepts.

A family is *intersecting* if any two of its sets have a non-empty intersection. The *chromatic number* $\chi(\mathcal{F})$ of $\mathcal{F} \subseteq 2^X$ is the smallest number of colors necessary to color the points in X so that no set of $\mathcal{F}$ of cardinality > 1 is monochromatic. It is clear that $\chi(\mathcal{F}) \geq 2$ (as long as $\mathcal{F}$ is non-trivial, i.e., contains at least one set with more than one element).

The families with $\chi(\mathcal{F}) = 2$ are of special interest and are called 2-*colorable*. In other words, $\mathcal{F}$ is 2-colorable iff there is a subset S such that neither S nor its complement $X \setminus S$ contain a member of $\mathcal{F}$. It turns out that $\chi(\mathcal{F}) > 2$ is a necessary condition for a family $\mathcal{F}$ to be self-dual.

For families of sets $\mathcal{F}$ and $\mathcal{G}$, we write $\mathcal{F} \succ \mathcal{G}$ if every member of $\mathcal{F}$ contains at least one member of $\mathcal{G}$.

Proposition 9.3. (i) *A family $\mathcal{F}$ is intersecting if and only if $\mathcal{F} \succ b\,(\mathcal{F})$.*
(ii) $\chi(\mathcal{F}) > 2$ *if and only if* $b\,(\mathcal{F}) \succ \mathcal{F}$.

Proof. (i) If $\mathcal{F}$ is intersecting then every $A \in \mathcal{F}$ is also a blocking set of $\mathcal{F}$, and hence, contains at least one minimal blocking set. Conversely, if $\mathcal{F} \succ b\,(\mathcal{F})$ then every set of $\mathcal{F}$ contains a blocking set of $\mathcal{F}$, and hence, intersects all other sets of $\mathcal{F}$.

(ii) Let us prove that $\chi(\mathcal{F}) > 2$ implies $b\,(\mathcal{F}) \succ \mathcal{F}$. If not, then there must be a blocking set T of $\mathcal{F}$ which contains no set of $\mathcal{F}$. But its complement $X \setminus T$ also contains no set of $\mathcal{F}$, since otherwise T would not block all the members of $\mathcal{F}$. Thus $(T, X \setminus T)$ is a 2-coloring of $\mathcal{F}$ with no monochromatic set, a contradiction with $\chi(\mathcal{F}) > 2$.

For the other direction, assume that $b\,(\mathcal{F}) \succ \mathcal{F}$ but $\chi(\mathcal{F}) = 2$. By the definition of $\chi(\mathcal{F})$ there exists a set S such that neither S nor $X \setminus S$ contain a set of $\mathcal{F}$. This, in particular, means that S is a blocking set of $\mathcal{F}$ which together with $b\,(\mathcal{F}) \succ \mathcal{F}$ implies that $S \supseteq A$ for some $A \in \mathcal{F}$, a contradiction. □

Corollary 9.4. *Let $\mathcal{F}$ be an antichain. Then the following three conditions are equivalent:*

(1) $b\,(\mathcal{F}) = \mathcal{F}$;
(2) *$\mathcal{F}$ is intersecting and $\chi(\mathcal{F}) > 2$;*
(3) *both $\mathcal{F}$ and $b\,(\mathcal{F})$ are intersecting.*

Proof. Equivalence of (1) and (2) follows directly from Proposition 9.3. Equivalence of (1) and (3) follows from the fact that both $\mathcal{F}$ and $b\,(\mathcal{F})$ are antichains. □

9.2 The blocking number

Recall that the *blocking number* $\tau(\mathcal{F})$ of a family $\mathcal{F}$ is the minimum number of elements in a blocking set of $\mathcal{F}$, that is,

$$\tau(\mathcal{F}) := \min\{|T| \,:\, T \cap A \neq \emptyset \text{ for every } A \in \mathcal{F}\}.$$

We make two observations concerning this characteristic:

If $\mathcal{F}$ contains a *k-matching*, i.e., k mutually disjoint sets, then $\tau(\mathcal{F}) \geq k$.
If $\mathcal{F}$ is intersecting, then $\tau(\mathcal{F}) \leq \min_{A \in \mathcal{F}} |A|$.

A family $\mathcal{F}$ can have many smallest blocking sets, i.e., blocking sets of size $\tau(\mathcal{F})$. The following result says how many. The *rank* of a family $\mathcal{F}$ is the maximum cardinality of a set in $\mathcal{F}$.

Theorem 9.5 (Gyárfás 1987)**.** *Let $\mathcal{F}$ be a family of rank r, and let $\tau = \tau(\mathcal{F})$. Then the number of blocking sets of $\mathcal{F}$ with τ elements is at most r^{τ}.*

Proof. We will prove by backward induction on i that every i-element set I is contained in at most $r^{\tau-i}$ τ-element blocking sets. It is obvious for $i = \tau$ and the case $i = 0$ gives the theorem. If $i < \tau$ then there exists a set $A \in \mathcal{F}$ such that $A \cap I = \emptyset$ (because $|I| < \tau(\mathcal{F})$). Now apply the induction hypothesis for the sets $I \cup \{x\}$, $x \in A$. Observe that every blocking set T of $\mathcal{F}$, containing the set I, must contain at least one of the extended sets $I \cup \{x\}$, with $x \in A$ (because $I \cap A = \emptyset$ whereas $T \cap A \neq \emptyset$). By the induction hypothesis, each of the sets $I \cup \{x\}$ with $x \in A$, is contained in at most $r^{\tau-(i+1)}$ τ-element blocking sets of $\mathcal{F}$. Thus, the set I itself is contained in at most $|A| \cdot r^{\tau-i-1} \leq r^{\tau-i}$ τ-element blocking sets, as desired. □

Considering τ pairwise disjoint sets of size r shows that Theorem 9.5 is best possible.

Corollary 9.6 (Erdős–Lovász 1975). *Let $\mathcal{F}$ be an intersecting r-uniform family with $\tau(\mathcal{F}) = r$. Then $|\mathcal{F}| \leq r^r$.*

Proof. Each $A \in \mathcal{F}$ is a blocking set of size r. □

9.3 Helly-type theorems

In terms of the blocking number τ, the simplest Helly-type result for families of sets (Theorem 7.2) says that if $\mathcal{F}$ is r-uniform and each set of $\leq r+1$ of its members intersect then $\tau(\mathcal{F}) = 1$. This result can be generalized as follows.

Theorem 9.7 (Lovász 1979). *Let $\mathcal{F}$ be r-uniform. If each collection of k members $(k \geq 2)$ of $\mathcal{F}$ intersect then $\tau(\mathcal{F}) \leq (r-1)/(k-1)+1$.*

Proof. By construction. For $j = 1, \ldots, k$ we will select j sets $A_1, \ldots, A_j$ in $\mathcal{F}$ such that

$$1 \leq |A_1 \cap \cdots \cap A_j| \leq r - (j-1)(\tau(\mathcal{F}) - 1), \tag{9.1}$$

which for $j = k$ gives the desired upper bound on $\tau = \tau(\mathcal{F})$.

For $j = 1$ take $A_1 \in \mathcal{F}$ arbitrarily.

Assume $A_1, \ldots, A_j$ have been selected $(j \leq k-1)$. The set $A_1 \cap \cdots \cap A_j$ intersects every set of $\mathcal{F}$ (why?), thus $|A_1 \cap \cdots \cap A_j| \geq \tau(\mathcal{F})$. Take a subset $S \subseteq A_1 \cap \cdots \cap A_j$ with $|S| = \tau - 1$. Since $|S| < \tau(\mathcal{F})$, there must be a set $A_{j+1} \in \mathcal{F}$ such that $S \cap A_{j+1} = \emptyset$ (this set is different from $A_1, \ldots, A_j$ since S intersects all of them). Thus

$$\begin{aligned}
&|A_1 \cap \cdots \cap A_j \cap A_{j+1}| \leq |A_1 \cap \cdots \cap A_j \cap \overline{S}| = |A_1 \cap \cdots \cap A_j - S| \\
&= |A_1 \cap \cdots \cap A_j| - (\tau - 1) \leq r - (j-1)(\tau - 1) - \tau + 1 = r - j(\tau - 1).
\end{aligned}$$

□

For graphs (i.e., for 2-uniform families) Helly's theorem (Theorem 7.2) says that, if in a finite graph any three edges share a common vertex, then this graph is a star. Erdős, Hajnal, and Moon (1964) generalized this easy observation about graphs in a different direction. A set of vertices S *covers* a set of edges $F \subseteq E$ of a graph $G = (V, E)$ if every edge in F has at least one of its endpoints in S.

Theorem 9.8. *If each family of at most $\binom{s+2}{2}$ edges of a graph can be covered by s vertices, then all edges can.*

The complete graph on $s+2$ vertices shows that this bound is best possible. Graphs are 2-uniform families. The question was how to generalize the result to r-uniform families for arbitrary r. The conjecture was easy to formulate: the formula $\binom{s+r}{r}$. This turns out to be the correct answer.

Theorem 9.9 (Bollobás 1965)**.** *If each family of at most $\binom{s+r}{r}$ members of an r-uniform family can be blocked by s points then all members can.*

Proof. Let $\mathcal{F}$ be an r-uniform family, satisfying the assumption of the theorem, and suppose that $\tau(\mathcal{F}) \geq s+1$. Then there is a subfamily $\mathcal{F}' = \{A_1, \ldots, A_m\} \subseteq \mathcal{F}$ such that $\tau(\mathcal{F}') = s+1$ and $\mathcal{F}'$ is τ-*critical*, that is,

$$\tau(\mathcal{F}' \setminus \{A_i\}) \leq s$$

for all $i = 1, \ldots, m$. Our goal is to show that $m \leq \binom{r+s}{s}$, contradicting the assumption (that every subfamily with so few members *can* be blocked by s points).

Since $\tau(\mathcal{F}') = s+1$ and $\mathcal{F}'$ is τ-critical, for each $i = 1, \ldots, m$, the family $\mathcal{F}' \setminus \{A_i\}$ has a blocking set B_i of size s. Hence, $A_j \cap B_i \neq \emptyset$ for all $j \neq i$. Moreover, $A_i \cap B_i = \emptyset$ since B_i has too few elements to intersect all the members of $\mathcal{F}'$. Thus, we can apply the Bollobás theorem (Theorem 8.8) with $a_1 = \ldots = a_m = r$ and $b_1 = \ldots = b_m = s$, which yields

$$m \cdot \binom{s+r}{s}^{-1} = \sum_{i=1}^{m} \binom{a_i + b_i}{a_i}^{-1} \leq 1,$$

and the desired upper bound on m follows. □

9.4 Blocking sets and decision trees

Blocking sets play an important role in the theory of boolean functions. In the next sections we will present some results in that direction.

Fix an arbitrary boolean function $f(x_1, \ldots, x_n)$. Given a vector $a = (a_1, \ldots, a_n)$ in $\{0,1\}^n$, a *certificate for a* (with respect to the function f) is a subset $S \subseteq [n] = \{1, \ldots, n\}$ of positions such that $f(b) = f(a)$ for all vectors $b \in \{0,1\}^n$ with $b_i = a_i$ for all $i \in S$. That is, if we set the variables x_i with $i \in S$ to the corresponding bits of a, then the function will take the value $f(a)$ *independent of the values of other variables*. The certificate complexity of f on a vector a, $C(f,a)$, is the minimum size $|S|$ of a certificate S for a. Define

$$C_1(f) = \max\{C(f,a)\colon f(a)=1\} \quad \text{and} \quad C_0(f) = \max\{C(f,a)\colon f(a)=0\}\,.$$

That is, $C_1(f)$ is the smallest number k such that, for every input a with $f(a)=1$, there is a subset S of $|S| \le k$ positions such that, if we set $x_i := a_i$ for all $i \in S$, then the function f takes value 1 independent of the values of other variables.

Let $\mathcal{F}_i$ be the family of all certificates for inputs $a \in f^{-1}(i)$, $i = 0,1$. Then we have the following cross-intersection property:

$$S \cap T \neq \emptyset \quad \text{for all } S \in \mathcal{F}_0 \text{ and } T \in \mathcal{F}_1. \tag{9.2}$$

Proof. Assume that there is a certificate S for a vector $a \in f^{-1}(0)$ and a certificate T for a vector $b \in f^{-1}(1)$ such that $S \cap T = \emptyset$. Take a vector $c \in \{0,1\}^n$ such that $c_i = a_i$ for all $i \in S$, $c_i = b_i$ for all $i \in T$, and $c_i = 0$ for all $i \notin S \cup T$. Since S is a certificate for a, and since vector c coincides with a in all position $i \in S$, we have that $f(c) = f(a) = 0$. But by the same reason we also have that $f(c) = f(b) = 1$, a clear contradiction. □

One can describe the certificates of a given boolean function f by so-called "decision trees."

A *decision tree* for a boolean function $f(x_1, \ldots, x_n)$ is a binary tree whose internal nodes have labels from $x_1, \ldots, x_n$ and whose leaves have labels from $\{0,1\}$. If a node has label x_i then the test performed at that node is to examine the i-th bit of the input. If the result is 0, one descends into the left subtree, whereas if the result is 1, one descends into the right subtree. The label of the leaf so reached is the value of the function (on that particular input). The *depth* of a decision tree is the number of edges in a longest path from the root to a leaf, or equivalently, the maximum number of bits tested on such a path. Let $DT(f)$ denote the minimum depth of a decision tree computing f.

It is not difficult to show (do this!) that, for every boolean function f, we have that

$$\max\{C_0(f), C_1(f)\} \le DT(f)\,.$$

This upper bound is, however, not optimal: there are boolean functions f for which

$$\max\{C_0(f), C_1(f)\} \le \sqrt{DT(f)}\,.$$

Such is, for example, the monotone boolean function $f(X)$ on $n = m^2$ boolean variables defined by: $f = \bigwedge_{i=1}^{m} \bigvee_{j=1}^{m} x_{ij}$. For this function we have $C_0(f) = C_1(f) = m$ but $DT(f) = m^2$ (see Exercise 9.9), implying that $DT(f) =$

$C_0(f) \cdot C_1(f)$. It turns out that the example given above is, in fact, the worst possible case.

Theorem 9.10. *For every boolean function f,*

$$DT(f) \le C_0(f) \cdot C_1(f)\,.$$

Proof. Induction on the number of variables n. If $n = 1$ then the inequality is trivial.

Let (say) $f(0,\dots,0) = 0$; then some set Y of $k \le C_0(f)$ variables can be chosen such that by fixing their value to 0, the function f is 0 independently of the other variables. We can assume w.l.o.g. that the set

$$Y = \{x_1, \dots, x_k\}$$

of the first k variables has this property.

Take a complete deterministic decision tree T_0 of depth k on these k variables. Each of its leaves corresponds to a unique input $a = (a_1, \dots, a_k) \in \{0,1\}^k$ reaching this leaf. Replace such a leaf by a minimal depth deterministic decision tree T_a for the sub-function

$$f_a := f(a_1, \dots, a_k, x_{k+1}, \dots, x_n)\,.$$

Obviously, $D_0(f_a) \le C_0(f)$ and $D_1(f_a) \le C_1(f)$. We claim that the latter inequality can be strengthened:

$$C_1(f_a) \le C_1(f) - 1 \tag{9.3}$$

The argument is essentially the same as that in the proof of (9.2). Take an arbitrary input $(a_{k+1}, \dots, a_n)$ of f_a which is accepted by f_a. Together with the bits $(a_1, \dots, a_k)$, this gives an input of the whole function f with $f(a_1, \dots, a_n) = 1$. According to the definition of the quantity $C_1(f)$, there must be a set $Z = \{x_{i_1}, \dots, x_{i_m}\}$ of $m \le C_1(f)$ variables such that fixing them to the corresponding values $x_{i_1} = a_{i_1}, \dots, x_{i_m} = a_{i_m}$, the value of f becomes 1 independently of the other variables. A simple (but crucial) observation is that

$$Y \cap Z \ne \emptyset. \tag{9.4}$$

Indeed, if $Y \cap Z = \emptyset$ then the value of $f(0,\dots,0,a_{k+1},\dots,a_n)$ should be 0 because fixing the variables in Y to 0 forces f to be 0, but should be 1, because fixing the variables in Z to the corresponding values of a_i forces f to be 1, a contradiction.

By (9.4), only $|Z \setminus Y| \le m-1$ of the bits of $(a_{k+1}, \dots, a_n)$ must be fixed to force the sub-function f_a to obtain the constant function 1. This completes the proof of (9.3).

Applying the induction hypothesis to each of the sub-functions f_a with $a \in \{0,1\}^k$, we obtain

$$DT(f_a) \leq C_0(f_a) \cdot C_1(f_a) \leq C_0(f)(C_1(f) - 1)\,.$$

Altogether,

$$DT(f) \leq k + \max_a DT(f_a) \leq C_0(f) + C_0(f)(C_1(f) - 1) = C_0(f)C_1(f)\,.$$

□

9.5 Blocking sets and monotone circuits

A boolean function $f(x_1, \ldots, x_n)$ is *monotone* if $f(x_1, \ldots, x_n) = 1$ and $x_i \leq y_i$ for all i, imply $f(y_1, \ldots, y_n) = 1$. A *monotone circuit* is a sequence $f_1, \ldots, f_t$ of monotone boolean functions, called *gates*, each of which is either one of the variables $x_1, \ldots, x_n$ or is obtained from some previous gates via an And or Or operation. That is, each gate f_i has either the form $f_i = x_l$ for some $1 \leq l \leq n$, or one of the forms $f = g \vee h$ or $f = g \wedge h$ for some $g, h \in \{0, 1, f_1, \ldots, f_{i-1}\}$. The *size* of a circuit is the number t of gates in it. The function *computed* by such a circuit is the last function f_t.

The problem (known as the *lower bounds problem*) is, given an explicit boolean function, to prove that it cannot be computed by a circuit of small size. It is clear that every function can be computed by a circuit of size exponential in the number of variables. However, even in the case of monotone circuits, it is difficult to show that some function is indeed hard, i.e., requires many gates.

In this section we will show that, using some combinatorial properties of blocking sets, one may obtain exponential lower bounds in a relatively easy and direct way.

A monotone k-CNF (conjunctive normal form) is an And of an arbitrary number of monotone clauses, each being an Or of at most k variables. Dually, a monotone k-DNF is an Or of an arbitrary number of monomials, each being an And of at most k variables. Note that in k-CNFs we allow clauses shorter than k.

In an *exact k-CNF* we require that all clauses have *exactly k* distinct variables; *exact k-DNF* is defined similarly. For two boolean functions f and g in n variables, we write $f \leq g$ if $f(x) \leq g(x)$ for all input vectors x. For a CNF/DNF C we will denote by $|C|$ the number of clauses/monomials in it.

Our goal is to show that complex monotone functions, that is, monotone functions requiring large monotone circuits cannot be "simple" in the sense that they cannot be approximated by small CNFs and DNFs. The proof of this is based on the following "switching lemma" allowing us to switch between CNFs and DNFs, and vice versa.

Lemma 9.11 (Monotone Switching Lemma)**.** *For every s-CNF f_0 there is an r-DNF f_1 and an exact $(r+1)$-DNF D such that*

$$f_1 \le f_0 \le f_1 \vee D \quad \textit{and} \quad |D| \le s^{r+1}\,. \tag{9.5}$$

Dually, for every r-DNF f_1 there is an s-CNF f_0 and an exact $(s+1)$-CNF C such that

$$f_0 \wedge C \le f_1 \le f_0 \quad \textit{and} \quad |C| \le r^{s+1}\,. \tag{9.6}$$

Proof. We prove the first claim (the second is dual). Let $f_0 = C_1 \wedge \cdots \wedge C_\ell$ be an s-CNF; hence, each clause C_i has $|C_i| \le s$ variables. It will be convenient to identify clauses and monomials with the *sets* of indices of their variables.

We associate with the CNF f_0 the following tree T of fan-out at most s. The first node of T corresponds to the first clause C_1, and the outgoing $|C_1|$ edges are labeled by the variables from C_1. Suppose we have reached a node v, and let M be the monomial consisting of the labels of edges on the path from the root to v. If M intersects all the clauses of f_0, then v is a leaf. Otherwise, let C_i be the *first* clause such that $M \cap C_i = \emptyset$. Then the node v has $|C_i|$ outgoing edges labeled by the variables in C_i.

Note that each path from the root to a leaf of T corresponds to a monomial of f_0 (since each such path intersects all clauses). More important is that also the converse holds: each monomial of f_0 must contain all labels of at least one path from the root to a leaf. Thus, we have just represented the DNF of f_0 as a tree, implying that $T(x) = f_0(x)$ for all input vectors $x \in \{0,1\}^n$. But some paths (monomials) may be longer than $r+1$. So, we now cut off these long paths.

Namely, let f_1 be the Or of all paths of length at most r ending in leafs, and D be the set of all paths of length exactly $r+1$. Observe that:

(i) every monomial of f_1 is also a monomial of f_0, and
(ii) every monomial of f_0, which is not a monomial of f_1, must contain (be an extension of) at least one monomial of D.

For every input $x \in \{0,1\}^n$, we have $f_1(x) \le f_0(x)$ by (i), and $f_0(x) \le f_1(x) \vee D(x)$ by (ii). Finally, we also have that $|D| \le s^{r+1}$, because every node of T has fan-out at most s. □

Most important in the Switching Lemma is that the $(r+1)$-DNF D, correcting possible errors, contains only s^{r+1} monomials instead of all $\binom{n}{r+1}$ possible monomials.

We now give a general lower bounds criterion for monotone circuits.

Definition 9.12. Let $f(x_1, \ldots, x_n)$ be a monotone boolean function. We say that f is *t-simple* if for every pair of integers $1 \le r, s \le n-1$ there exists an exact $(s+1)$-CNF C, an exact $(r+1)$-DNF D, and a subset $I \subseteq \{1, \ldots, n\}$ of size $|I| \le s$ such that

(a) $|C| \le t \cdot r^{s+1}$ and $|D| \le t \cdot s^{r+1}$, and
(b) either $C \le f$ or $f \le D \vee \bigvee_{i \in I} x_i$ (or both) hold.

Theorem 9.13 (Lower bounds criterion)**.** *If a monotone boolean function can be computed by a monotone circuit of size t, then it is t-simple.*

Proof. Given a monotone circuit, the idea is to approximate every intermediate gate (more exactly – the function computed at the gate) by an s-CNF and an r-DNF, and to show that when doing so we do not introduce too many errors. If the function computed by the whole circuit is not t-simple, then it cannot be approximated well by such a CNF/DNF pair meaning that every such pair must make many errors. Since the number of errors introduced at each separate gate is small, the total number of gates must be large. To make as few errors at each gate as possible we will use the Switching Lemma (Lemma 9.11) which allows us to approximate an s-CNF by small r-DNFs and vice versa.

Let $F(x_1, \ldots, x_n)$ be a monotone boolean function, and suppose that F can be computed by a monotone circuit of size t. Our goal is to show that the function F is t-simple. To do this, fix an arbitrary pair of integers $1 \leq s, r \leq n-1$.

Let $f = g * h$ be a gate in our circuit. By an *approximator* of this gate we will mean a pair f_0, f_1, where f_0 is an s-CNF (a *left* approximator of f) and f_1 is an r-DNF (a *right* approximator of f) such that $f_1 \leq f_0$.

We say that such an approximator f_0, f_1 of f introduces a new error on input $x \in \{0,1\}^n$ if the approximators of g and of h did not make an error on x, but the approximator of f does. That is, $g_0(x) = g_1(x) = g(x)$ and $h_0(x) = h_1(x) = h(x)$, but either $f_0(x) \neq f(x)$ or $f_1(x) \neq f(x)$.

We define approximators inductively as follows.

Case 1: f is an input variable, say, $f = x_i$. In this case we take $f_0 = f_1 := x_i$. It is clear that this approximator introduces no errors.

Case 2: f is an And gate, $f = g \wedge h$. In this case we take $f_0 := g_0 \wedge h_0$ as the left approximator of f; hence, f_0 introduces no new errors. To define the right approximator of f we use Lemma 9.11 to convert f_0 into an r-DNF f_1; hence, $f_1 \leq f_0$. Let E_f be the set of inputs on which f_1 introduces a new error, i.e.,

$$E_f := \{x \colon f(x) = f_0(x) = 1 \ \text{ but } \ f_1(x) = 0\}\,.$$

By Lemma 9.11, all these errors can be "corrected" by adding a relatively small exact $(r+1)$-DNF: there is an exact $(r+1)$-DNF D such that $|D| \leq s^{r+1}$ and $D(x) = 1$ for all $x \in E_f$.

Case 3: f is an Or gate, $f = g \vee h$. This case is dual to Case 2. We take $f_1 := g_1 \vee h_1$ as the right approximator of f; hence, f_1 introduces no new errors. To define the left approximator of f we use Lemma 9.11 to convert f_1 into an s-CNF f_0; hence, $f_1 \leq f_0$. Let E_f be the set of inputs on which f_0 introduces a new error, i.e.,

$$E_f := \{x \colon f(x) = f_1(x) = 0 \ \text{ but } \ f_0(x) = 1\}\,.$$

By Lemma 9.11, all these errors can be "corrected" by adding a relatively small exact $(s+1)$-CNF: there is an exact $(s+1)$-CNF C such that $|C| \leq r^{s+1}$ and $C(x) = 0$ for all $x \in E_f$.

Proceeding in this way we will reach the last gate of our circuit computing the given function F. Let F_0, F_1 be its approximator, and let E be the set of all inputs $x \in \{0,1\}^n$ on which F differs from at least one of the functions F_0 or F_1. Since at input gates (= variables) no error was made, for every such input $x \in E$, the corresponding error must be introduced at some intermediate gate. That is, for every $x \in E$ there is a gate f such that $x \in E_f$ (approximator of f introduces an error on x for the first time). But we have shown that, for each gate, all these errors can be corrected by adding an exact $(s+1)$-CNF of size at most r^{s+1} or an exact $(r+1)$-DNF of size at most s^{r+1}. Since we have only t gates, all such errors $x \in E$ can be corrected by adding an exact $(s+1)$-CNF C of size at most $t \cdot r^{s+1}$ and an exact $(r+1)$-DNF D of size at most $t \cdot s^{r+1}$, that is, for all inputs $x \in \{0,1\}^n$, we have

$$C(x) \wedge F_0(x) \leq F(x) \leq F_1(x) \vee D(x)\,.$$

This already implies that the function F is t-simple. Indeed, if the CNF F_0 is empty (i.e., if $F_0 \equiv 1$) then $C \leq F$, and we are done. Otherwise, F_0 must contain some clause S of length at most s, say, $S = \bigvee_{i \in I} x_i$ for some I of size $|I| \leq s$. Since $F_0 \leq S$, the condition $F_1 \leq F_0$ implies $F \leq F_1 \vee D \leq F_0 \vee D \leq S \vee D$, as desired. This completes the proof of Theorem 9.13. □

In applications, boolean functions f are usually defined as set-theoretic predicates. In this case we say that f accepts a set $S \subseteq \{1, \ldots, n\}$ if and only if f accepts its incidence vector.

A set S is a *positive input* for f if $f(S) = 1$, and a *negative input* if $f(\overline{S}) = 0$, where $\overline{S}$ is the complement of S. Put otherwise, a positive (negative) input is a set of variables which, if assigned the value 1 (0), forces the function to take the value 1 (0) regardless of the values assigned to the remaining variables. Note that one set S can be both positive and negative input! For example, if $f(x_1, x_2, x_3)$ outputs 1 iff $x_1 + x_2 + x_3 \geq 2$, then $S = \{1, 2\}$ is both positive and negative input for f, because $f(1,1,x_3) = 1$ and $f(0,0,x_3) = 0$.

To translate the definition of t-simplicity of f (Definition 9.12) in terms of positive/negative inputs, note that if C is a CNF, then $C \leq f$ means that every negative input of f must contain at least one clause of C (looked at as set of indices of its variables). Similarly, $f \leq D \vee \bigvee_{i \in I} x_i$ means that every positive input must either intersect the set I or contain at least one monomial of D.

We begin with the simplest example. We will also present a more respectable applications—a $2^{\Omega(n^{1/4})}$ lower bound—but this special case already demonstrates the common way of reasoning pretty well.

Let us consider a monotone boolean function Δ_m, whose input is an undirected graph on m vertices, represented by $n = \binom{m}{2}$ variables, one for each possible edge. The value of the function is 1 if and only if the graph contains a triangle (three incident vertices). Clearly, there is a monotone circuit of size $O(m^3)$ computing this function: just test whether any of $\binom{m}{3}$ triangles is present in the graph. Thus, the following theorem is tight, up to a poly-logarithmic factor.

Theorem 9.14. *Any monotone circuit, detecting whether a given m-vertex graph is triangle-free, must have $\Omega\left(m^3/\log^4 m\right)$ gates.*

Proof. Let t be the minimal number for which Δ_m is t-simple. By Theorem 9.13, it is enough to show that $t \geq \Omega\left(m^3/\log^4 m\right)$. For this proof we take

$$s := \lfloor 5\log^2 m \rfloor \quad \text{and} \quad r := 1\,.$$

According to the definition of t-simplicity, we have only two possibilities.

Case 1: Every positive input for Δ_m either intersects a fixed set I of s edges, or contains at least one of $L \leq ts^{r+1} = ts^2$ 2-element sets of edges $R_1, \ldots, R_L$.

As positive inputs for Δ_m we take all triangles, i.e., graphs on m vertices with exactly one triangle; we have $\binom{m}{3}$ such graphs. At most $s(m-2)$ of them will have an edge in I. Each of the remaining triangles must contain one of ts^2 given pairs of edges R_i. Since two edges can lie in at most one triangle, we conclude that, in this case,

$$t \geq \frac{\binom{m}{3} - s(m-2)}{s^2} = \Omega\left(m^3/\log^4 m\right)\,.$$

Case 2: Every negative input for Δ_m contains at least one of $tr^{s+1} = t$ sets of edges $S_1, \ldots, S_t$, each of size $|S_i| = s+1$.

In this case we consider the graphs $E = E_1 \cup E_2$ consisting of two disjoint non-empty cliques E_1 and E_2 (we look at graphs as sets of their edges). Each such graph E is a negative input for Δ_m, because its complement is a bipartite graph, and hence, has no triangles. The number of such graphs is a half of the number 2^m of all binary strings of length m excluding $\mathbf{0}$ and $\mathbf{1}$. Hence, We have $2^{m-1}-1$ such graphs, and each of them must contain at least one of the sets $S_1, \ldots, S_t$. Every of these sets of edges S_i is incident to at least $\sqrt{2s}$ vertices, and if $E \supseteq S_i$ then all these vertices must belong to one of the cliques E_1 or E_2. Thus, at most $2^{m-\sqrt{2s}}-1$ of our negative inputs E can contain one fixed set S_i, implying that, in this case,

$$t \geq \frac{2^{m-1}-1}{2^{m-\sqrt{2s}}-1} \geq 2^{\sqrt{2s}-1} \geq 2^{3\log m} \geq m^3\,.$$

Thus, in both cases, $t \geq \Omega\left(m^3/\log^4 m\right)$, and we are done. □

Our next example is the following monotone boolean function introduced by Andreev (1985). Let $q \geq 2$ be a prime power, and set $d := \lfloor (q/\ln q)^{1/2}/2 \rfloor$. Consider $q \times q$ $(0,1)$ matrices $A = (a_{i,j})$. Given such a matrix A, we are interested in whether it contains a graph of a polynomial $h : GF(q) \to GF(q)$, that is, whether $a_{i,h(i)} = 1$ for all rows $i \in GF(q)$.

Let f_n be a monotone boolean function in $n = q^2$ variables such that $f_n(A) = 1$ iff A contains a graph of at least one polynomial over $GF(q)$ of degree at most $d-1$. That is,

$$f_n(X) = \bigvee_{h} \bigwedge_{i \in GF(q)} x_{i,h(i)} \,,$$

where h ranges over all polynomials over $GF(q)$ of degree at most $d-1$. Since we have at most q^d such polynomials, the function f_n can be computed by a monotone boolean circuit of size at most q^{d+1}, which is at most $n^{O(d)} = 2^{O(n^{1/4}\sqrt{\ln n})}$. We will now show that this trivial upper bound is almost optimal.

Theorem 9.15. *Any monotone circuit computing the function f_n has size at least $2^{\Omega(n^{1/4}\sqrt{\ln n})}$.*

Proof. Take a minimal t for which the function f_n is t-simple. Since $n = q^2$ and (by our choice) $d = \Theta(n^{1/4}\sqrt{\ln n})$, it is enough by Theorem 9.13 to show that $t \geq q^{\Omega(d)}$. For this proof we take

$$s := \lceil d \ln q \rceil \quad \text{and} \quad r := d \,,$$

and look at input matrices as bipartite $q \times q$ graphs. In the proof we will essentially use the well-known fact that no two distinct polynomials of degree at most $d-1$ can coincide on d points. According to the definition of t-simplicity, we have only two possibilities.

Case 1: Every positive input for f_n either intersects a fixed set I of at most s edges, or contains at least one of $L \leq ts^{r+1}$ $(r+1)$-element sets of edges $R_1, \ldots, R_L$.

Graphs of polynomials of degree at most $d-1$ are positive inputs for f_n. Each set of l $(1 \leq l \leq d)$ edges is contained in either 0 or precisely q^{d-l} of such graphs. Hence, at most sq^{d-1} of these graphs can contain an edge in I, and at most $q^{d-(r+1)}$ of them can contain any of the given graphs R_i. Therefore, in this case we again have

$$t \geq \left(1 - \frac{s}{q}\right) \frac{q^d}{s^{r+1} \cdot q^{d-(r+1)}} \geq \left(\frac{q}{s}\right)^{\Omega(r)} \geq q^{\Omega(d)} \,.$$

Case 2: Every negative input for f_n contains at least one of $K \leq tr^{s+1}$ $(s+1)$-element sets of edges $S_1, \ldots, S_K$.

Let $\boldsymbol{E}$ be a random bipartite graph, with each edge appearing in $\boldsymbol{E}$ independently with probability $\gamma := (2d \ln q)/q$. Since there are only q^d polynomials of degree at most $d-1$, the probability that the complement of $\boldsymbol{E}$ will contain the graph of at least one of them does not exceed $q^d(1-\gamma)^q \leq q^{-d}$, by our choice of γ. Hence, with probability at least $1 - q^{-d}$, the graph $\boldsymbol{E}$ is a negative input for f. On the other hand, each of the sets S_i is contained in $\boldsymbol{E}$ with probability $\gamma^{|S_i|} = \gamma^{s+1}$. Thus, in this case,

$$t \geq \frac{1-q^{-d}}{r^{s+1}\gamma^{s+1}} \geq \left(\frac{q}{2d^2 \ln q}\right)^{\Omega(s)} \geq 2^{\Omega(s)} \geq q^{\Omega(d)} ,$$

where the third inequality holds for all $d \leq (q/\ln q)^{1/2}/2$.

We have proved that the function f can be t-simple only if $t \geq q^{\Omega(d)}$. By Theorem 9.13, this function cannot be computed by monotone circuits of size smaller than $q^{\Omega(d)}$. □

Exercises

9.1. The *independence number* $\alpha(\mathcal{F})$ of a family $\mathcal{F} \subseteq 2^X$ is defined as the maximum cardinality $|S|$ of a set $S \subseteq X$ which does not contain any member of $\mathcal{F}$. Prove that $\alpha(\mathcal{F}) = |X| - \tau(\mathcal{F})$.

9.2. Let T be a minimal blocking set of a family $\mathcal{F}$. Show that, for every $x \in T$, there exists an $A \in \mathcal{F}$ such that $T \cap A = \{x\}$.

9.3. Show that the solution to Proposition 9.2 is optimal: if $\mathcal{F}$ is an antichain, then at least $|b(\mathcal{F})|$ locks are also necessary.

9.4. Let $\mathcal{F}$ be an r-uniform family and suppose that $\tau(\mathcal{F} \setminus \{A\}) < \tau(\mathcal{F})$ for all $A \in \mathcal{F}$. Prove that $|\mathcal{F}| \leq \binom{r+\tau(\mathcal{F})-1}{r}$. *Hint*: Observe that, for each $A \in \mathcal{F}$, there is a set B of size $\tau(\mathcal{F})-1$ which is disjoint from A but intersects all other members of $\mathcal{F}$; apply the Bollobás theorem (Theorem 8.7).

9.5. Let $\mathcal{F}$ and $\mathcal{H}$ be antichains over some set X. Prove that:

(i) $\mathcal{H} = b(\mathcal{F})$ if and only if for every coloring of the points in X in Red and in Blue, *either* $\mathcal{F}$ has a Red set (i.e., all points in this set are red), *or* (exclusive) $\mathcal{H}$ has a Blue set.
(ii) $\mathcal{F} \succ \mathcal{H}$ if and only if $b(\mathcal{H}) \succ b(\mathcal{F})$.

9.6. Consider the following family $\mathcal{F}$. Take k disjoint sets $V_1, \ldots, V_k$ such that $|V_i| = i$ for $i = 1, \ldots, k$. The members of $\mathcal{F}$ are all the sets of the form $V_i \cup T$, where T is any set such that $|T| = k - i$ and $|T \cap V_j| = 1$ for all $j = i+1, \ldots, k$. Show that this family is self-dual, i.e., that $\mathcal{F} = b(\mathcal{F})$. (This construction is due to Erdős and Lovász.)

9.7. A pair of sets (A, B) *separates* a pair of elements (x, y) if $x \in A \setminus B$ and $y \in B \setminus A$. A family $\mathcal{F} = \{A_1, \ldots, A_m\}$ of subsets of $X = \{x_1, \ldots, x_n\}$ is a *complete separator* if every pair of elements in X is separated by at least one pair of sets in $\mathcal{F}$. Let $\mathcal{F}^*$ be the family of all non-empty sets $X_i := \{j : x_i \in$

$A_j\}$. Prove that $\mathcal{F}$ is a complete separator if and only if $\mathcal{F}^*$ is an antichain. *Hint*: $X_i \not\subseteq X_j$ means that there exists k such that $k \in X_i$ and $k \notin X_j$, i.e., that $x_i \in A_k$ and $x_j \notin A_k$.

9.8. Let $\mathcal{F}$ be a family of rank r. Show that then, for any $s \geq 1$, the family $\mathcal{F}$ has at most r^s minimal blocking sets of size s.

9.9. Prove that any decision tree for the function $f = \bigwedge_{i=1}^m \bigvee_{j=1}^m x_{ij}$ has depth m^2. *Hint*: Take an arbitrary decision tree for f and construct a path from the root by the following "adversary" rule. Suppose we have reached a node v labeled by x_{ij}. Then follow the outgoing edge marked by 1 if and only if *all* the variables x_{il} with $l \neq j$ were already tested before we reached the node v.

9.10. The *storage access function* is a boolean function $f(x,y)$ on $n+k$ variables $x = (x_0, \ldots, x_{n-1})$ and $y = (y_0, \ldots, y_{k-1})$ where $n = 2^k$, and is defined as follows: $f(x,y) := x_{int(y)}$, where $int(y) := \sum_{i=0}^{k-1} y_i 2^i$ is the integer whose binary representation is the vector y. Prove that f is a $(k+1)$-DNF function although some of its minterms have length 2^k. *Hint*: For the first claim observe that the value of f only depends on $k+1$ bits $y_0, \ldots, y_{k-1}$ and $x_{int(y)}$. For the lower bound, consider the monomial $x_0 x_1 \cdots x_{n-1}$ and show that it is a minterm of f.

9.11. A *partial* b*–*(n,k,λ) *design* is a family $\mathcal{F}$ of k-element subsets of $\{1, \ldots, n\}$ such that any b-element set is contained in at most λ of its members. We can associate with each such design $\mathcal{F}$ a monotone boolean function $f_{\mathcal{F}}$ such that $f_{\mathcal{F}}(S) = 1$ if and only if $S \supseteq F$ for at least one $F \in \mathcal{F}$. Assume that $\ln|\mathcal{F}| < k-1$ and that each element belongs to at most N members of $\mathcal{F}$. Use Theorem 9.13 to show that for every integer $a \geq 2$, every monotone circuit computing $f_{\mathcal{F}}$ has size at least

$$\ell := \min\left\{ \frac{1}{2}\left(\frac{k}{2b\ln|\mathcal{F}|}\right)^a, \quad \frac{|\mathcal{F}| - a\cdot N}{\lambda \cdot a^b}, \right\}.$$

Hint: Take $s = a$, $r = b$ and show that under this choice of parameters, the function $f_{\mathcal{F}}$ can be t-simple only if $t \geq \ell$. When doing this, note that the members of $\mathcal{F}$ are positive inputs for $f_{\mathcal{F}}$. To handle the case of negative inputs, take a random subset in which each element appears independently with probability $p = (1+\ln|\mathcal{F}|)/k$, and show that its complement can contain a member of $\mathcal{F}$ with probability at most $|\mathcal{F}|(1-p)^k \leq \mathrm{e}^{-1}$.

9.12. Derive Theorem 9.15 from the previous exercise. *Hint*: Observe that the family of all q^d graphs of polynomials of degree at most $d-1$ over $\mathbb{F}_q$ forms a partial b–(n,k,λ) design with parameters $n = q^2$, $k = q$ and $\lambda = q^{d-b}$.

9.13. Andreev (1987) has shown how, for any prime power $q \geq 2$ and $d \leq q$, to construct an explicit family $\mathcal{D}$ of subsets of $\{1, \ldots, n\}$ which, for every $b \leq d+1$, forms a partial b–(n,k,λ) design with parameters $n = q^3$, $k = q^2$, $\lambda = q^{2d+1-b}$ and $|\mathcal{D}| = q^{2d+1}$. Use Exercise 9.11 to show that the corresponding boolean function $f_{\mathcal{D}}$ requires monotone circuits of size exponential in $\Omega\left(n^{1/3-o(1)}\right)$.

9.14. (due to Berkowitz). A *k-threshold* is a monotone boolean function $T_k^n(x_1, \ldots, x_n)$ which outputs 1 if and only if the input vector $x = (x_1, \ldots, x_n)$ has *weight* at least k, i.e., if $|x| := x_1 + \cdots + x_n \geq k$. Show that

$$T_k^{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \overline{x}_i,$$

for all inputs $(x_1, \ldots, x_n)$ such that $x_1 + \cdots + x_n = k$.

9.15. A boolean function f is a *slice* function if there is some $0 \leq k \leq n$ such that for every input $x \in \{0,1\}^n$,

$$f(x) = \begin{cases} 0 & \text{if } |x| < k; \\ 1 & \text{if } |x| > k. \end{cases}$$

That is, f can be non-trivial only on inputs with exactly k ones; in this case we also say that f is the *k-slice* function. Use the previous exercise and the fact that the threshold function T_k^n has a monotone circuit of size $O(n^2)$ to prove that for such functions, using the negations cannot help much. Namely, prove that if a slice function f has a non-monotone circuit of size ℓ, then f can also be computed by a monotone circuit of size at most $\ell + O(n^3)$.

10. Density and Universality

In many applications (testing logical circuits, construction of k-wise independent random variables, etc.), vector sets $A \subseteq \{0,1\}^n$ with the following property play an important role:

> For any subset of k coordinates $S = \{i_1, \ldots, i_k\}$ the projection of A onto the indices in S contains all possible 2^k configurations.

Such sets are called (n,k)-*universal.* If the same holds not for all but only for *at least one* subset S of k indices, then A is called (n,k)-*dense.* The maximal number k, for which A is (n,k)-dense, is also known as the *Vapnik–Chervonenkis dimension* of A.

Given n and k, the problem is to find a universal (or dense) set A with as few vectors as possible. In this chapter we will discuss several approaches to its solution.

10.1 Dense sets

Given a vector $v = (v_1, \ldots, v_n)$, its *projection* onto a set of coordinates $S = \{i_1, \ldots, i_k\}$ is the vector $v\restriction_S := (v_{i_1}, \ldots, v_{i_k})$. The projection of a set of vectors $A \subseteq \{0,1\}^n$ onto S is the set of vectors $A\restriction_S := \{v\restriction_S : v \in A\}$. Thus, A is (n,k)-dense iff $A\restriction_S = \{0,1\}^k$ for at least one subset of k coordinates S.

It is clear that every (n,k)-dense set must contain at least 2^k vectors. On the other hand, if A is the set of all vectors in $\{0,1\}^n$ with less than k ones, then A has

$$H(n,k) := \sum_{i=0}^{k-1} \binom{n}{i}$$

vectors but is *not* (n,k)-dense. It turns out, however, that every larger set already *is* (n,k)-dense! This interesting fact, whose applications range from probability theory to computational learning theory, was discovered indepen-

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science. An EATCS Series, DOI 10.1007/978-3-642-17364-6_10,

dently by three sets of authors in remarkable simultaneity: Perles and Shelah (see Shelah 1972), Sauer (1972), and Vapnik and Chervonenkis (1971). No less remarkable is the range of contexts in which the results arose: logic, set theory, and probability theory.

Theorem 10.1. *If $A \subseteq \{0,1\}^n$ and $|A| > H(n,k)$ then A is (n,k)-dense.*

Proof. Induction on n and k. If $k = 1$ then A has at least two different vectors and hence is $(n,1)$-dense. For the induction step take an arbitrary set $A \subseteq \{0,1\}^n$ of size $|A| > H(n,k)$. Let B be the projection of A onto the first $n-1$ coordinates, and C be the set of all vectors u in $\{0,1\}^{n-1}$ for which *both* vectors $(u,0)$ and $(u,1)$ belong to A. A simple but crucial observation is that

$$|A| = |B| + |C|.$$

Now, if $|B| > H(n-1,k)$ then the set B is $(n-1,k)$-dense by induction, and hence the whole set A is also (n,k)-dense. If $|B| \leq H(n-1,k)$ then, using the identity $\binom{n}{i} - \binom{n-1}{i} = \binom{n-1}{i-1}$ (see Proposition 1.3) we obtain

$$\begin{aligned} |C| = |A| - |B| &> H(n,k) - H(n-1,k) \\ &= \sum_{i=0}^{k-1}\binom{n}{i} - \sum_{i=0}^{k-1}\binom{n-1}{i} = \sum_{i=0}^{k-2}\binom{n-1}{i} \\ &= H(n-1,k-1). \end{aligned}$$

By the induction hypothesis, the set C is $(n-1,k-1)$-dense, and since $C \times \{0,1\}$ lies in A, the whole set A is also (n,k)-dense. □

10.2 Hereditary sets

Alon (1983) and Frankl (1983) have independently made an intriguing observation that for results like Theorem 10.1, we can safely restrict our attention to sets with a very special structure.

A set $A \subseteq \{0,1\}^n$ is *hereditary* or *downwards closed* if $v \in A$ and $u \leq v$ implies $u \in A$. (Here, as usual, $u \leq v$ means that $u_i \leq v_i$ for all i.) Thus, being hereditary means that we can arbitrarily switch 1s to 0s, and the resulting vectors will still belong to the set.

For a set $S \subseteq \{1,\ldots,n\}$ of coordinates, let $t_S(A)$ denote the number of vectors in the projection $A\restriction_S$. If v is a vector and i is any of its coordinates, then the *i-th neighbor* of v is the vector $v_{i\to 0}$ obtained from v by switching its i-th bit to 0; if this bit is 0 then we let $v_{i\to 0} = v$.

Theorem 10.2. *For every subset A of the n-cube $\{0,1\}^n$ there is a hereditary subset B such that $|B| = |A|$ and $t_S(B) \leq t_S(A)$ for all sets S of coordinates.*

Before we prove this result, observe that it immediately implies Theorem 10.1: if B is hereditary and $|B| > H(n,k)$, then B must contain a vector v with at least k ones, and so, must contain all the 2^k vectors obtained from v by changing any subset of these ones to zeroes.

Proof. If A itself is hereditary, there is nothing to do. Otherwise, we have some "bad"coordinates, i.e., coordinates i such that $v_{i\to 0} \not\in A$ for some $v \in A$. To correct the situation, we will apply for each such bad coordinate i, the following transformation T_i. Take a vector $v \in A$ with $v_i = 1$, and see if $v_{i\to 0}$ belongs to A. If so, do nothing; otherwise, replace the vector v in A by $v_{i\to 0}$. Apply this transformation as long as possible, and let B denote the resulting set. It is clear that $|B| = |A|$. We also claim that $t_S(B) \le t_S(A)$ for every $S \subseteq \{1,\ldots,n\}$.

Indeed, if $i \not\in S$ then $t_S(B) = t_S(A)$, and we are done. Suppose that $i \in S$ and let $S' = S \setminus \{i\}$. Assume, for notational convenience, that i was the first coordinate, i.e., that $i = 1$. Now, if $t_S(B) \ge t_S(A) + 1$, this can happen only when A has two vectors $x = (1, u, w_1)$ and $y = (1, u, w_2)$ with $u \in \{0,1\}^{S'}$ and $w_1 \ne w_2$, and such that *exactly one* of them, say x, was altered by T_i. That is, the S-projection of B contains both vectors $(1,u)$ and $(0,u)$, whereas $(0,u)$ does not appear in the S-projection of A. But this is impossible because the fact that the other vector $y = (1, u, w_2)$ was *not* altered by T_i means that its i-th neighbor $(0, u, w_2)$ belongs to A, and hence vector $(0,u)$ must appear among the vectors in the S-projection of A. This contradiction proves that $t_S(B) \le t_S(A)$.

Thus, starting with A, we can apply the transformations T_i along all n coordinates $i = 1,\ldots,n$, and obtain the set $B = T_n(T_{n-1}(\cdots T_1(A)\cdots))$, which is hereditary, has the same number of vectors as the original set A and satisfies the condition $t_S(B) \le t_S(A)$ for all S. □

Frankl (1983) observed that this result also has other interesting consequences. For a set $A \subseteq \{0,1\}^n$, let $t_s(A) = \max t_S(A)$ over all $S \subseteq \{1,\ldots,n\}$ with $|S| = s$; hence, $t_n(A) = |A|$.

Theorem 10.3 (Bondy 1972). *If* $|A| \le n$ *then* $t_{n-1}(A) = |A|$.

Proof. We will give a direct proof of this result in Sect. 11.1; here we show that it is a consequence of Theorem 10.2.

By this theorem we may assume that A is hereditary. If A is empty, there is nothing to prove. Otherwise, A must contain the all-0 vector. Hence, at least one of n unit vectors

$$e_i = (0,\ldots,0,1,0,\ldots,0),$$

with the 1 in the i-th coordinate, does not belong to A. As A is hereditary, this implies that $|A| = t_n(A) = t_S(A)$ for $S = \{1,\ldots,n\} \setminus \{i\}$. □

Bollobás (see Lovász 1979, Problem 13.10) extended this result to larger sets.

Theorem 10.4. *If* $|A| \leq \lceil \frac{3}{2}n \rceil$ *then* $t_{n-1}(A) \geq |A| - 1$.

Proof. By Theorem 10.2 we may assume that A is hereditary. If there is an i such that $e_i \notin A$, then again $t_{n-1}(A) = |A|$, and we are done. Otherwise, A contains the all-0 vector and all unit vectors $e_1, \ldots, e_n$. Let A' be the set of all vectors in A with precisely two 1s. Each such vector covers only two of the unit vectors. Therefore, some e_i must remain uncovered, for otherwise we would have $|A| \geq 1 + n + \lceil n/2 \rceil > \lceil \frac{3}{2}n \rceil$. But this means that e_i is the only vector in A with 1 in the i-th coordinate, implying that for $S = \{1, \ldots, n\} \setminus \{i\}$, $t_S(A) = |A \setminus \{e_i\}| = |A| - 1$. □

Combining Theorem 10.2 with the deep Kruskal–Katona theorem about the shadows of arbitrary families of sets (see Theorem 10.16 below), Frankl (1983) derived the following general result, which is the best possible whenever t divides n (see Exercise 10.9). We state it without proof.

Theorem 10.5 (Frankl 1983). *If* $A \subseteq \{0,1\}^n$ *and* $|A| \leq \lceil n(2^t - 1)/t \rceil$, *then* $t_{n-1}(A) \geq |A| - 2^{t-1} + 1$.

The following result concerning the intersection of hereditary sets, due to Kleitman, has many generalizations and applications (see, for example, Exercise 10.8 and Theorem 7.7):

Theorem 10.6 (Kleitman 1966). *Let* A, B *be downwards closed subsets of* $\{0,1\}^n$. *Then*

$$|A \cap B| \geq \frac{|A| \cdot |B|}{2^n}.$$

Proof. Apply induction on n, the case $n = 0$ being trivial. For $\epsilon \in \{0,1\}$, set $c_\epsilon = |A_\epsilon|$ and $d_\epsilon = |B_\epsilon|$, where

$$A_\epsilon := \{(a_1, \ldots, a_{n-1}) \ : \ (a_1, \ldots, a_{n-1}, \epsilon) \in A\}$$

and

$$B_\epsilon := \{(b_1, \ldots, b_{n-1}) \ : \ (b_1, \ldots, b_{n-1}, \epsilon) \in B\}.$$

Then

$$\begin{aligned} |A \cap B| &= |A_0 \cap B_0| + |A_1 \cap B_1| \\ &\geq (c_0 d_0 + c_1 d_1)/2^{n-1} \quad \text{(by induction)} \\ &= (c_0 + c_1)(d_0 + d_1)/2^n + (c_0 - c_1)(d_0 - d_1)/2^n. \end{aligned}$$

Since sets A, B are downwards closed, we have $A_1 \subseteq A_0$ and $B_1 \subseteq B_0$, implying that $(c_0 - c_1)(d_0 - d_1) \geq 0$. Since $c_0 + c_1 = |A|$ and $d_0 + d_1 = |B|$, we are done. □

10.3 Matroids and approximation

Given a family $\mathcal{F}$ of subsets of some finite set X, called the *ground-set*, and a weight function assigning each element $x \in X$ a non-negative real number $w(x)$, the *optimization problem* for $\mathcal{F}$ is to find a member $A \in \mathcal{F}$ whose weight $w(A) = \sum_{x\in A} w(x)$ is maximal. For example, given a graph $G = (V, E)$ with non-negative weights on edges, we might wish to find a matching (a set of vertex-disjoint edges) of maximal weight. In this case $X = E$ is the set of edges, and members of $\mathcal{F}$ are matchings. As it happens in many other situations, the resulting family is *hereditary*, that is, $A \in \mathcal{F}$ and $B \subseteq A$ implies $B \in \mathcal{F}$.

In general, some optimization problems are extremely hard—the so-called "NP-hard problems." In such situations one is satisfied with an "approximative" solution, namely, with a member $A \in \mathcal{F}$ whose weight is at least $1/k$ times the weight of an optimal solution, for some real constant $k \geq 1$.

One of the simplest algorithms to solve an optimization problem is the *greedy algorithm.* It first sorts the elements $x_1, x_2, \ldots, x_n$ of X by weight, heaviest first. Then it starts with $A = \emptyset$ and in the i-th step adds the element x_i to the current set A if and only if the result still belongs to $\mathcal{F}$. A basic question is: for what families $\mathcal{F}$ can this trivial algorithm find a good enough solution?

Namely, say that a family $\mathcal{F}$ is *greedy k-approximative* if, for every weight function, the weight of the solution given by the greedy algorithm is at least $1/k$ times the weight of an optimal solution. Note that being greedy 1-approximative means that for such families the greedy algorithm always finds an optimal solution.

Given a real number $k \geq 1$, what families are greedy k-approximative?

In the case $k = 1$ (when greedy is optimal) a surprisingly tight answer was given by introducing a notion of "matroid." This notion was motivated by the following "exchange property" in linear spaces: If A, B are two sets of linearly independent vectors, and if $|B| > |A|$, then there is a vector $b \in B \setminus A$ such that the set $A \cup \{b\}$ is linearly independent.

Now let $\mathcal{F}$ be a family of subsets of some finite set X; we call members of $\mathcal{F}$ *independent sets.* A *k-matroid* is a hereditary family $\mathcal{F}$ satisfying the following *k-exchange property*: For every two independent sets $A, B \in \mathcal{F}$, if $|B| > k|A|$ then there exists $b \in B \setminus A$ such that* $A + b$ is independent (belongs to $\mathcal{F}$). Matroids are k-matroids for $k = 1$.

Matroids have several equivalent definitions. One of them is in terms of maximum independent sets. Let $\mathcal{F}$ be a family of subsets of X (whose members we again call independent sets), and $Y \subseteq X$. An independent set $A \in \mathcal{F}$ is a *maximum independent* subset of Y (or a *basis* of Y in $\mathcal{F}$) if $A \subseteq Y$ and $A + x \notin \mathcal{F}$ for all $x \in Y \setminus A$. A family is *k-balanced* if for every subset $Y \subseteq X$

* Here and in what follows, $A + b$ will stand for the set $A \cup \{b\}$.

and any two of its maximum independent subsets $A, B \subseteq Y$ we have that $|B| \leq k|A|$.

Lemma 10.7. *A hereditary family is k-balanced if and only if it is a k-matroid.*

Proof. ($\Leftarrow$) Let $Y \subseteq X$, and let $A, B \subseteq Y$ be two sets in $\mathcal{F}$ that are maximum independent subsets of Y. Suppose that $|B| > k|A|$. Then by the k-exchange property, we can add some element b of $B \setminus A$ to A and keep the result $A + b$ in $\mathcal{F}$. But since A and B are both subsets of Y, the set $A + b$ is also a subset of Y and thus A is not *maximum* independent in Y, a contradiction.

($\Rightarrow$) We will show that if $\mathcal{F}$ does *not* satisfy the k-exchange property, then it is not k-balanced. Let A and B be two independent sets such that $|B| > k|A|$ but no element of $B \setminus A$ can be added to A to get a result in $\mathcal{F}$. We let Y be $A \cup B$. Now A is a maximum independent set in Y, since we cannot add any of the other elements of Y to it. The set B may not be a maximum independent set in Y, but if it isn't there is some subset B' of Y that contains it and is maximum independent in Y. Since this set is at least as big as B, it is strictly bigger than $k|A|$ and we have a violation of the k-balancedness property. □

For $k = 1$, the ($\Leftarrow$) direction of the following theorem was proved by Rado (1942), and the ($\Rightarrow$) direction by Edmonds (1971).

Theorem 10.8. *A hereditary family is greedy k-approximative if and only if it is a k-matroid.*

Proof. ($\Leftarrow$) Let $\mathcal{F}$ be a k-matroid over some ground-set X. Fix an arbitrary weight function, and order the elements of the ground-set X according to their weight, $w(x_1) \geq w(x_2) \geq \ldots \geq w(x_n)$. Let A be the solution given by the greedy algorithm, and B an optimal solution. Our goal is to show that $w(B)/w(A) \leq k$.

Let $Y_i := \{x_1, \ldots, x_i\}$ be the set of the first i elements considered by the greedy algorithm. The main property of the greedy algorithm is given by the following simple claim.

Claim 10.9. For every i, the set $A \cap Y_i$ is a maximum independent subset of Y_i.

Proof. Suppose that the independent set $A \cap Y_i$ is not a maximum independent subset of Y_i. Then there must exist an element $x_j \in Y_i \setminus A$ (an element *not* chosen by the algorithm) such that the set $A \cap Y_i + x_j$ is independent. But then $A \cap Y_{j-1} + x_j$ (as a subset of an independent set) is also independent, and should have been chosen by the algorithm, a contradiction. □

Now let $A_i := A \cap Y_i$. Since $A_i \setminus A_{i-1}$ is either empty or is equal to $\{x_i\}$,

$$\begin{aligned} w(A) &= w(x_1)|A_1| + \sum_{i=2}^{n} w(x_i)(|A_i| - |A_{i-1}|) \\ &= \sum_{i=1}^{n-1}(w(x_i) - w(x_{i+1}))|A_i| + w(x_n)|A_n|\,. \end{aligned}$$

Similarly, letting $B_i := B \cap Y_i$, we get

$$w(B) = \sum_{i=1}^{n-1}(w(x_i) - w(x_{i+1}))|B_i| + w(x_n)|B_n|\,.$$

Using the inequality $(a+b)/(x+y) \le \max\{a/x, b/y\}$ we obtain that $w(B)/w(A)$ does not exceed $|B_i|/|A_i|$ for some i. By Claim 10.9, the set A_i is a maximum independent subset of Y_i. Since B_i is also a (not necessarily maximum) independent subset of Y_i, the k-balancedness property implies that $|B_i| \le k|A_i|$. Hence, $w(B)/w(A) \le |B_i|/|A_i| \le k$, as desired.

($\Rightarrow$) We will prove that if our family $\mathcal{F}$ fails to satisfy the k-exchange property, then there is some weight function on which the greedy algorithm fails to approximate an optimal solution by a factor of $1/k$.

Suppose there are two sets A and B in $\mathcal{F}$, with $|B| > k|A|$, such that no element of $B \setminus A$ can be added to A while keeping the result in $\mathcal{F}$. Let $m = |A|$. Take any two positive numbers a and b such that $0 < a - b \le 1/k$. Define the weight function as follows: elements in A have weight $m + a$, elements in $B \setminus A$ have weight $m + b$, and other elements have weight 0. Then the greedy algorithm tries elements of weight $m + a$ first, gets all m of them, but then is stuck because no element of weight $m + b$ fits; hence, the total score of the greedy algorithm is $m(m + a)$. But the optimum is at least the total weight $(m + b)|B| \ge (m + b)(km + 1)$ of elements in B. Thus, the greedy algorithm can $(1/k)$-approximate this particular optimization problem only if $(m + b)(km + 1) \le km(m + a)$, or equivalently, if $k(a - b) \ge 1 + b/m$. But this is impossible because $a - b \le 1/k$ and $b > 0$. □

When trying to show that a given family is a k-matroid, the following somewhat easier to verify property, suggested by Mestre (2006), is often useful. We say that a family $\mathcal{F}$ is *k-extendible* if for every sets $A \subset B \in \mathcal{F}$ and for every element $x \notin B$ the following holds: If the set $A + x$ is independent then the set $B + x$ can be made independent by removing from B at most k elements not in A, that is, there exists $Y \subseteq B \setminus A$ such that $|Y| \le k$ and the set $B \setminus Y + x$ is independent.

Lemma 10.10. *Every k-extendible hereditary family is a k-matroid.*

Proof. Given two independent sets A and B with $|B| > k|A|$, we need to find an element $z \in B \setminus A$ such that the set $A + z$ is independent. If $A \subset B$ then we are done since all subsets of B are independent. Suppose now that $A \not\subseteq B$. The idea is to pick an element $x \in A \setminus B$ and apply the k-extendibility property to the sets $C := A \cap B$ and $D := B$ to find a subset $Y \subseteq D \setminus C = B \setminus A$ with at most k elements such that the set $B' = B \setminus Y + x$ is independent. If A is still not a subset of B', then repeat the same procedure. Since, due to the condition $Y \subseteq B \setminus A$, at any step none of the already added elements of A are removed, after at most $|A \setminus B|$ steps we will obtain an independent set B' such that $A \subseteq B'$. From $|B| > k|A|$, we have that $|B \setminus A| > k|A \setminus B|$. Since in each step at most k elements of B are removed, at least one element $z \in B \setminus A$ must remain in B', that is, A is a *proper* subset of B'. But then the set $A + z$ is independent, because B' is such, and we are done. □

In the case of matroids ($k = 1$) we also have the converse.

Lemma 10.11. *Every matroid is* 1*-extendible.*

Proof. Let $\mathcal{F}$ be a matroid. Given sets $A \subset B \in \mathcal{F}$ and an element $x \notin B$ such that the set $A + x$ is independent, we need to find an element $y \in B \setminus A$ such that $B - y + x$ is independent. If necessary, we can repeatedly apply the matroid property to add elements of $B \setminus A$ to A until we get a subset A' such that $A \subseteq A' \subset B$, $A' + x \in \mathcal{F}$ and $|A' + x| = |B|$. Since $x \notin B$, this implies that $B \setminus A'$ consists of just one element y. But then $B - y + x = A' + x$ belongs to $\mathcal{F}$, as desired. □

It can be shown (see Exercise 10.12) that for $k \geq 2$ the converse of Lemma 10.10 does not hold, that is, not every k-matroid is k-extendible. Still, together with Theorem 10.8, Lemma 10.10 gives us a handy tool to show that some unrelated optimization problems can be approximated quite well by using the trivial greedy algorithm.

Example 10.12 (Maximum weight f-matching)*.* Given a graph $G = (V, E)$ with non-negative weights on edges and degree constraints $f : V \to \mathbb{N}$ for vertices, an f*-matching* is a set of edges M such that for all $v \in V$ the number $\deg_M(v)$ of edges in M incident to v is at most $f(v)$. The corresponding optimization problem is to find an f-matching of maximal weight.

In this case we have a family $\mathcal{F}$ whose ground-set is the set $X = E$ of edges of G and f-matchings are independent sets (members of $\mathcal{F}$). Note that $\mathcal{F}$ is already not a matroid when $f(v) = 1$ for all $v \in V$: if $A = \{a, b\}$ and $B = \{\{c, a\}, \{b, d\}\}$ are two matchings, then $|B| > |A|$ but no edge of B can be added to A. We claim that this family is 2-extendible, and hence, is a 2-matroid.

To show this, let $A + x$ and B be any two f-matchings, where $A \subset B$ and $x = \{u, v\}$ is an edge not in B. If $B + x$ is an f-matching, we are done. If not, then $\deg_B(u) = f(u)$ or $\deg_B(v) = f(v)$ (or both). But we know that $\deg_A(u) < f(u)$ and $\deg_A(v) < f(v)$, for otherwise $A + x$ would not be an f-matching. Thus, we can remove at most two edges of B not in A so that the resulting graph plus the edge x forms a f-matching.

Example 10.13 (Maximum weight traveling salesman problem). We are given a complete directed graph with non-negative weights on edges, and we must find a maximum weight Hamiltonian cycle, that is, a cycle that visits every vertex exactly once. This problem is very hard: it is a so-called "NP-hard" problem. On the other hand, using Theorem 10.8 and Lemma 10.10 we can show that the greedy algorithm can find a Hamiltonian cycle whose weight is at least one third of the maximum possible weight of a Hamiltonian cycle.

The ground-set X of our family $\mathcal{F}$ in this case consists of the directed edges of the complete graph. A set is independent if its edges form a collection of vertex-disjoint paths or a Hamiltonian cycle. It is enough to show that $\mathcal{F}$ is 3-extendible.

To show this, let $A+x$ and B be any two members of $\mathcal{F}$, where $A \subset B$ and $x = (u, v)$ is an edge not in B. First remove from B the edges (if any) out of u and into v. There can be at most two such edges, and neither of them can belong to A since otherwise $A + (u, v)$ would not belong to $\mathcal{F}$. If we add (u, v) to B then every vertex has in-degree and out-degree at most one. Hence, the only reason why the resulting set may not belong to $\mathcal{F}$ is that there may be a non-Hamiltonian cycle which uses (u, v). But then there must be an edge in the cycle, not in A, that we can remove to break it: if all edges, except for (u, v), of the cycle belong to A, then $A + (u, v)$ contains a non-Hamiltonian cycle and could not belong to $\mathcal{F}$. Therefore we need to remove at most three edges in total.

10.4 The Kruskal–Katona theorem

A *neighbor* of a binary vector v is a vector which can be obtained from v by flipping one of its 1-entries to 0. A *shadow* of a set $A \subseteq \{0, 1\}^n$ of vectors is the set $\partial(A)$ of all its neighbors. A set A is *k-regular* if every vector in A contains exactly k 1-entries. Note that in this case $\partial(A)$ is $(k-1)$-regular.

A basic question concerning shadows is the following one: What can one say about $|\partial(A)|$ in terms of the total number $|A|$ of vectors in a k-regular set A?

In general one cannot improve on the trivial upper bound $|\partial(A)| \leq k|A|$. But what about *lower* bounds? The question is non-trivial because one and the same vector with $k-1$ ones may be a neighbor of up to $n-k+1$ vectors in A. Easy counting shows that

$$|\partial(A)| \geq \frac{k}{n-k+1}|A| = \frac{|A|}{\binom{n}{k}}\binom{n}{k-1}.$$

This can be shown by estimating the number N of pairs (u, v) of vectors such that $v \in A$ and u is a neighbor of v. Since every $v \in A$ has exactly k neighbors, we have that $N = k|A|$. On the other hand, every vector u

with $k-1$ ones can be a neighbor of at most $n-k+1$ vectors of A. Hence, $k|A| = N \leq (n-k+1)|\partial(A)|$, and the desired lower bound on $|\partial(A)|$ follows.

Best possible lower bounds on $|\partial(A)|$ were obtained by Kruskal (1963) and Katona (1966). The idea, again, is to show that the minimum of $|\partial(A)|$ over all sets A with $|A| = m$ is achieved by sets of a very special structure, and use the Pascal identity for binomial coefficients $\binom{x}{k} = x(x-1)\cdots(x-k+1)/k!$: for every real number $x \geq k$

$$\binom{x}{k-1} + \binom{x}{k} = \binom{x+1}{k}. \tag{10.1}$$

In Proposition 1.3 we gave a combinatorial proof of this identity in the case when x is a natural number. The case when x is not necessarily an integer can be shown by a simple algebraic manipulation:

$$\begin{aligned}\binom{x}{k-1} + \binom{x}{k} &= \frac{x!}{(x-(k-1))!(k-1)!} + \frac{x!}{(x-k)!k!} \\ &= \frac{kx! + (x+1-k)x!}{(x+1-k)!k!} = \binom{x+1}{k}.\end{aligned}$$

The following lemma allows us to restrict our attention to sets with a very special structure. For a set of vectors $A \subseteq \{0,1\}^n$, let A_0 and A_1 denote the sets of vectors in A starting, respectively, with 0 and 1. Hence, $A = A_0 \cup A_1$. Let also e_i denote the vector in $\{0,1\}^n$ with exactly one 1-entry in the i-th position.

Proposition 10.14. *For every set $B \subseteq \{0,1\}^n$ there is a set $A \subseteq \{0,1\}^n$ of the same size such that $|\partial(B)| \geq |\partial(A)|$ and*

$$\partial(A_0) + e_1 \subseteq A_1\,. \tag{10.2}$$

That is, if we take a vector v in A with $v_1 = 0$, flip any of its 1s to 0 and at the same time flip its first bit to 1, then the obtained vector will again belong to A.

Proof. For $1 < j \leq n$, the *j-th shift* of B is the set $s_j(B)$ of vectors defined as follows. First, we include in $s_j(B)$ all vectors $v \in B_1$. For the vectors $v \in B_0$ we look whether $v_j = 1$. If yes, we include in $s_j(B)$ the vector $v \oplus e_1 \oplus e_j$ (obtained from vector v by flipping its 1-st and j-th bits), but only if this vector does not already belong to B; if $v \oplus e_1 \oplus e_j$ belongs to B, we include in $s_j(B)$ the vector v itself. This last requirement ensures that $|s_j(B)| = |B|$ for every $1 < j \leq n$. For example, if

$$B = \begin{array}{c} 1\,0\,1\,0 \\ 1\,1\,0\,1 \\ 0\,1\,1\,0 \\ 0\,1\,0\,1 \end{array} \qquad \text{then} \qquad s_2(B) = \begin{array}{c} 1\,0\,1\,0 \\ 1\,1\,0\,1 \\ 0\,1\,1\,0 \\ 1\,0\,0\,1 \end{array}$$

We claim that the shifting operation preserves the neighborhood. Namely, for every $1 < j \leq n$,

$$\partial(s_j(B)) \subseteq s_j(\partial(B))\,.$$

The following diagram sketches the proof idea:

$$\begin{tikzcd}
(0\ldots1\ldots1\ldots) \arrow[r, "\text{shift}"] \arrow[d, "\text{neighbor}"] & (1\ldots0\ldots1\ldots) \arrow[d, "\text{neighbor}"] \\
(0\ldots1\ldots0\ldots) \arrow[r, "\text{shift}"] & (1\ldots0\ldots0\ldots)
\end{tikzcd}$$

If we repeatedly apply the shift operators s_j, $j = 2, \ldots, n$ to B, the number of vectors containing 1 in the first position increases, so that after a finite number of applications the shifts must therefore cease to make any change. We have then obtained a new set A of the same size as B, with $s_j(A) = A$ for each $j \geq 2$, and with $|\partial(B)| \geq |\partial(A)|$. We claim that A satisfies (10.2).

To show this, take a vector $u \in \partial(A_0)$. Then $u + e_j$ belongs to A_0 for some $j \geq 2$, and hence, $u + e_1$ belongs to $s_j(A) = A$. □

We first state and prove a slightly weaker but much more handy version of the Kruskal–Katona theorem.

Theorem 10.15. *If $A \subseteq \{0,1\}^n$ is k-regular, and if*

$$|A| \geq \binom{x}{k} = x(x-1)\cdots(x-k+1)/k!$$

for some real number $x \geq k$, then

$$|\partial(A)| \geq \binom{x}{k-1}\,. \tag{10.3}$$

Note that this is the best possible: If $A \subseteq \{0,1\}^n$ is the set of all $\binom{n}{k}$ vectors with exactly k ones, then $|\partial(A)| = \binom{n}{k-1}$.

Proof (due to Lovász 1979). By Proposition 10.14, we can assume that A satisfies (10.2). Consider the set

$$A^0 := \{(0, w) \ : \ (1, w) \in A\}$$

obtained from A_1 by flipping the first bit from 1 to 0. Note that $|A^0| = |A_1|$. Observe also that

$$|\partial(A)| \geq |A^0| + |\partial(A^0)|\,. \tag{10.4}$$

Indeed, vectors in the set A^0 are neighbors of A by the definition of this set. Moreover, each neighbor of A^0 plus the unit vector e_1 is also a neighbor of A.

We now argue by double induction on k and $m = |A|$. For $k = 1$ and m arbitrary, (10.3) holds trivially.

For the induction step, we first use the fact that A has a special structure—namely, satisfies (10.2)—to show that $|A^0|$ cannot be smaller than $\binom{x-1}{k-1}$. To show this, assume the opposite. Then

$$|A_0| = |A| - |A_1| = |A| - |A^0| > \binom{x}{k} - \binom{x-1}{k-1} = \binom{x-1}{k},$$

and so, by induction, $|\partial(A_0)| \geq \binom{x-1}{k-1}$. But then (10.2) implies that

$$|A^0| = |A_1| \geq \binom{x-1}{k-1},$$

a contradiction. Hence, $|A^0| \geq \binom{x-1}{k-1}$.

Since A^0 is $(k-1)$-regular, the induction hypothesis yields $|\partial(A^0)| \geq \binom{x-1}{k-2}$. Together with (10.4) this implies

$$|\partial(A)| \geq |A^0| + |\partial(A^0)| \geq \binom{x-1}{k-1} + \binom{x-1}{k-2} = \binom{x}{k-1},$$

as desired. □

To state the Kruskal–Katona theorem in its original form, we write $m = |A|$ in *k-cascade* form:

$$m = \binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \cdots + \binom{a_s}{s} \tag{10.5}$$

where $a_k > a_{k-1} > \ldots > a_s \geq s \geq 1$ are integers. Such a representation of m can be obtained as follows. Let a_k be the maximal integer for which $\binom{a_k}{k} \leq m$. Then choose a_{k-1} as the largest integer for which $\binom{a_{k-1}}{k-1} \leq m - \binom{a_k}{k}$. If $a_{k-1} \geq a_k$, then we would have $m \geq \binom{a_k}{k} + \binom{a_k}{k-1} = \binom{1+a_k}{k}$, contradicting the maximality of a_k. Therefore $a_{k-1} < a_k$. Continuing this process we eventually reach a stage where the choice of a_s for some $s \geq 2$ actually gives an equality,

$$\binom{a_s}{s} = m - \binom{a_k}{k} - \binom{a_{k-1}}{k-1} - \cdots - \binom{a_{s+1}}{s+1},$$

or we get right down to choosing a_1 as the integer such that

$$\binom{a_1}{1} \leq m - \binom{a_k}{k} - \cdots - \binom{a_2}{2} < \binom{a_1+1}{1}$$

in which case we have

$$0 \leq m - \binom{a_k}{k} - \cdots - \binom{a_1}{1} < 1,$$

so that

$$m = \binom{a_k}{k} + \cdots + \binom{a_1}{1}.$$

It can be shown by induction (do this!) that the representation (10.5) is unique.

Theorem 10.16 (Kruskal–Katona Theorem)**.** *If $A \subseteq \{0,1\}^n$ is k-regular, and if*

$$|A| = \binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \cdots + \binom{a_s}{s}$$

then

$$|\partial(A)| \geq \binom{a_k}{k-1} + \binom{a_{k-1}}{k-2} + \cdots + \binom{a_s}{s-1}.$$

We leave the proof as an exercise. It is the same as that of Theorem 10.15 with $\binom{x}{k}$ and $\binom{x}{k-1}$ replaced by the corresponding sums of binomial coefficents.

The representation (10.5) of $m = |A|$ in the k-cascade form seems somewhat magical. To interpret this representation, let us consider the so-called *colexicographic order* (or *colex order*) of vectors in $\{0,1\}^n$. This order is defined by letting $u \prec v$ iff there is an i such that $u_i = 0$, $v_i = 1$ and $u_j = v_j$ for all $j > i$. Note that the only difference from the more standard *lexicographic* order is that we now scan the strings from right to left. For example, the colex order of all $\binom{5}{3} = 10$ vectors in $\{0,1\}^5$ with exactly 3 ones is (with the "smallest" vector on the top):

$$\begin{array}{c}
1\;1\;1\;0\;0\\
1\;1\;0\;1\;0\\
1\;0\;1\;1\;0\\
0\;1\;1\;1\;0\\
1\;1\;0\;0\;1\\
1\;0\;1\;0\;1\\
0\;1\;1\;0\;1\\
1\;0\;0\;1\;1\\
0\;1\;0\;1\;1\\
0\;0\;1\;1\;1
\end{array}$$

Let E^n_k denote the k-th slice of the binary n-cube, that is, the set of all vectors in $\{0,1\}^n$ with exactly k ones.

Proposition 10.17. *If the m-th vector in the colex order of E^n_k contains* 1*s in positions $a_1 + 1 < a_2 + 1 < \ldots < a_k + 1$ then*

$$m = \binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \cdots + \binom{a_1}{1}.$$

Proof. Let v be the m-th vector in the colex order of E^n_k. To reach v we must skip all vectors whose k-th 1 appears before position $a_k + 1$, and there are

$\binom{a_k}{k}$ of these. Some vectors with last (rightmost) 1 in position a_k may also precede v. These are the vectors whose first $k-1$ 1s precede position $a_{k-1}+1$, and there are $\binom{a_{k-1}}{k-1}$ of these. Arguing further in this way gives the result. □

By the same argument one can show that the shadow of the first $m = \sum_{i=1}^{k} \binom{a_i}{i}$ vectors in the colex order of E_k^n consists of the first $\partial_k(m) := \sum_{i=1}^{k} \binom{a_i}{i-1}$ vectors in the colex order of E_{k-1}^n. Thus, the Kruskal–Katona theorem says that the shadow of a family of m vectors in E_k^n is minimized by the set consisting of the first m vectors in the colex ordering on E_{k-1}^n. Furthermore, the size of the shadow is $\partial_k(m)$.

10.5 Universal sets

The (n,k)-density of a set of vectors means that its projection on *at least one* set of k coordinates gives the whole binary k-cube. We now consider a stronger property – (n,k)-universality – where we require that the same holds for *all* subsets of k coordinates.

Of course, the whole cube $\{0,1\}^n$ is (n,k)-universal for every $k \leq n$. This is the trivial case. Do there exist smaller universal sets? Note that 2^k is a trivial lower bound.

Using the probabilistic argument it can be shown that there *exist* (n,k)-universal sets of size only $k2^k \log n$ (see Theorem 3.2).

This result tells us only that small universal sets exist, but gives us no idea of how to construct them. In this section we will show how to construct explicit sets in $\{0,1\}^n$ which only have size n and are (n,k)-universal as long as $k2^k < \sqrt{n}$. The construction employs some nice combinatorial properties of so-called Paley graphs.

In this section we introduce one property of (bipartite) graphs which is equivalent to the universality property of 0-1 vectors. In the next section we will describe an explicit construction of such graphs based on the famous theorem of Weil (1948) regarding character sums.

By a bipartite graph with parts of size n we will mean a bipartite graph $G = (V_1, V_2, E)$ with $|V_1| = |V_2| = n$. We say that a node $y \in V_2$ is a *common neighbor* for a set of nodes $A \subseteq V_1$ if y is joined to *each* node of A. Dually, a node $y \in V_2$ is a *common non-neighbor* for a set of nodes $B \subseteq V_1$ if y is joined to *no* node of B. Given two disjoint subsets A and B of V_1, we denote by $v(A,B)$ the number of nodes in V_2 which are common neighbors for A, and at the same time are common non-neighbors for B. That is, $v(A,B)$ is the number of nodes in V_2 joined to each node of A and to no node of B.

Definition 10.18. A bipartite graph $G = (V_1, V_2, E)$ satisfies the *isolated neighbor condition for* k if $v(A,B) > 0$ for any two disjoint subsets $A, B \subseteq V_1$ such that $|A| + |B| = k$.

Such graphs immediately yield (n, k)-universal sets of 0-1 strings:

Proposition 10.19. *Let G be a bipartite graph with parts of size n and C be the set of columns of its incidence matrix. If G satisfies the isolated neighbor condition for k then C is (n, k)-universal.*

Proof. Let $G = (V_1, V_2, E)$ and $M = (m_{x,y})$ be the adjacency matrix of G. That is, M has n rows labeled by nodes x from V_1, n columns labeled by nodes y from V_2, and $m_{x,y} = 1$ if and only if $(x, y) \in E$.

Let $S = \{i_1, \ldots, i_k\}$ be an arbitrary subset of k rows of M and $v = (v_{i_1}, \ldots, v_{i_k})$ be an arbitrary (column) vector in $\{0,1\}^k$. Each row of M corresponds to a node in V_1. Let A be the set of nodes in V_1 corresponding to the 1-coordinates of v, and B be the set of nodes corresponding to the 0-coordinates of v. Since $|A| + |B| = |S| = k$ and our graph satisfies the isolated neighbor condition for k, there must be a node $y \in V_2$ which is joined to each node of A and to no node of B. But this means that the values of the y-th column of M at rows from S coincide with the corresponding values of the vector v, as desired. □

10.6 Paley graphs

Here we will show how to construct explicit bipartite graphs satisfying the isolated neighbor condition for k close to $\log n$.

A *bipartite Paley graph* is a bipartite graph $G_q = (V_1, V_2, E)$ with parts $V_1 = V_2 = \mathbb{F}_q$ for q odd prime congruent to 1 modulo 4; two nodes, $x \in V_1$ and $y \in V_2$, are joined by an edge if and only if $x - y$ is a non-zero square in $\mathbb{F}_q$, i.e., if $x - y = z^2 \bmod q$ for some $z \in \mathbb{F}_q$, $z \neq 0$. The condition $q \equiv 1 \bmod 4$ is only to ensure that -1 is a square in the field (see Exercise 10.7), so that the resulting graph is undirected.

Given two disjoint sets of nodes $A, B \subseteq V_1$, let $v(A, B)$, as before, denote the number of nodes in V_2 joined to each node of A and to no node of B. It turns out that for $|A| + |B| = k < (\log q)/3$, this number is very close to $q/2^k$, independent of what the sets A, B actually are.

Theorem 10.20. *Let $G_q = (V_1, V_2, E)$ be a bipartite Paley graph with $q \geq 9$, and A, B be disjoint sets of nodes in V_1 such that $|A| + |B| = k$. Then*

$$\left| v(A, B) - 2^{-k} q \right| \leq k\sqrt{q}. \tag{10.6}$$

In particular, $v(A, B) > 0$ as long as $k 2^k < \sqrt{q}$.

This result is a slight modification of a similar result of Bollobás and Thomason (1981) about general (non-bipartite) Paley graphs; essentially the same result was proved earlier by Graham and Spencer (1971). The proof is based on the theorem of Weil (1948) regarding character sums. Its special case states the following.

Let χ be the *quadratic residue character* in $\mathbb{F}_q$: $\chi(x) = x^{(q-1)/2}$. That is, $\chi(x) = 1$ if x is a non-zero square in $\mathbb{F}_q$, $\chi(x) = -1$ if x is non-square, and $\chi(0) = 0$. Also, $\chi(x \cdot y) = \chi(x) \cdot \chi(y)$.

Theorem 10.21 (Weil 1948)**.** *Let $f(t)$ be a polynomial over $\mathbb{F}_q$ which is not the square of another polynomial, and has precisely s distinct zeros. Then*

$$\left|\sum_{x\in\mathbb{F}_q} \chi\big(f(x)\big)\right| \le (s-1)\sqrt{q}.$$

We omit the proof of this important result. Weil's original proof relied heavily on several ideas from algebraic geometry. Since then other (but still complicated) proofs have been found; the interested reader can find the details in (Schmidt 1976).

With Weil's result, the above stated property of Paley graphs can be derived by easy computations.

Proof of Theorem 10.20. Recall that (x, y) is an edge in G_q if and only if $\chi(x - y) = 1$. Say that a node $x \in V_2$ is a *copy* of a node $y \in V_1$ if both these nodes correspond to the same element of $\mathbb{F}_q$; hence, each node of V_2 is a copy of precisely one node in V_1. Moreover, no x is joined to its copy y since then $\chi(x - y) = \chi(0) = 0$.

Let A' and B' be the set of all copies of nodes in A and, respectively, in B. Also let $U := V_2 \setminus (A' \cup B')$. Define

$$g(x) := \prod_{a\in A}\big(1 + \chi(x-a)\big)\prod_{b\in B}\big(1 - \chi(x-b)\big)$$

and observe that, for each node $x \in U$, $g(x)$ is non-zero if and only if x is joined to every node in A and to no node in B, in which case it is precisely 2^k. Hence,

$$\sum_{x\in U} g(x) = 2^k \cdot v^*(A, B), \tag{10.7}$$

where $v^*(A, B)$ is the number of those nodes in U which are joined to each node of A and to no node of B.

Expanding the expression for $g(x)$ and using the fact that $\chi(x \cdot y) = \chi(x) \cdot \chi(y)$, we obtain

$$g(x) = 1 + \sum_{C} (-1)^{|C\cap B|}\chi\big(f_C(x)\big),$$

where $f_C(x)$ denotes the polynomial $\prod_{c\in C}(x - c)$, and the sum is over all *non-empty* subsets C of $A \cup B$. By Weil's theorem,

$$\Big|\sum_{x\in\mathbb{F}_q}\chi\big(f_C(x)\big)\Big| \le (|C|-1)\sqrt{q}.$$

Hence,

$$\begin{aligned}\Big|\sum_{x\in\mathbb{F}_q} g(x)-q\Big| &\le \sum_{C}(|C|-1)\sqrt{q} = \sqrt{q}\sum_{s=2}^{k}\binom{k}{s}(s-1)\\ &= \sqrt{q}\big((k-2)2^{k-1}+1\big).\end{aligned}$$

Here the last equality follows from the identity $\sum_{s=1}^{k} s\binom{k}{s} = k2^{k-1}$ (see Exercise 1.5).

The summation above is over all nodes $x \in V_2 = \mathbb{F}_q$. However, for every node $x \in A' \cup B'$, $g(x) \le 2^{k-1}$, and the nodes of $A' \cup B'$ can contribute at most

$$\Big|\sum_{x\in A'\cup B'} g(x)\Big| \le k\cdot 2^{k-1}.$$

Therefore,

$$\Big|\sum_{x\in U} g(x)-q\Big| \le \sqrt{q}\big((k-2)2^{k-1}+1\big)+k\cdot 2^{k-1}.$$

Dividing both sides by 2^k and using (10.7), together with the obvious estimate $v(A,B)-v^*(A,B) \le |A'\cup B'| = k$, we conclude that

$$\Big|v(A,B)-2^{-k}q\Big| \le \frac{k\sqrt{q}}{2}-\sqrt{q}+\frac{\sqrt{q}}{2^k}+\frac{k}{2}+k, \tag{10.8}$$

which does not exceed $k\sqrt{q}$ as long as $q \ge 9$. □

Theorem 10.20 together with Proposition 10.19 give us, for infinitely many values of n, and for every k such that $k2^k < \sqrt{n}$, an explicit construction of (n,k)-universal sets of size n. In Sect. 17.4 we will show how to construct such sets of size $n^{O(k)}$ for arbitrary k using some elementary properties of linear codes.

10.7 Full graphs

We have seen that universal sets of 0-1 strings correspond to bipartite graphs satisfying the isolated neighbor condition. Let us now ask a slightly different question: how many vertices must a graph have in order to contain *every* k-vertex graph as an induced subgraph? Such graphs are called *k-full.* That is, given k, we are looking for graphs of small order (the order of a graph is the number of its vertices) which contain every graph of order k as an induced subgraph.

Note that if G is a k-full graph of order n then $\binom{n}{k}$ is at least the number of non-isomorphic graphs of order k, so

$$\binom{n}{k} \geq 2^{\binom{k}{2}}/k!$$

and thus

$$n \geq 2^{(k-1)/2}.$$

On the other hand, for every k it is possible to exhibit a k-full graph of order $n = 2^k$. This nice construction is due to Bollobás and Thomason (1981).

Let P_k be a graph of order $n = 2^k$ whose vertices are subsets of $\{1, \ldots, k\}$, and where two distinct vertices A and B are joined if and only if $|A \cap B|$ is even; if one of the vertices, say A, is an empty set then we join B to A if and only if $|B|$ is even. Note that the resulting graph is regular: each vertex has degree $2^{k-1} - 1$.

Theorem 10.22 (Bollobás–Thomason 1981)**.** *The graph P_k is k-full.*

Proof. Let G be a graph with vertex set $\{v_1, v_2, \ldots, v_k\}$. We claim that there are sets $A_1, A_2, \ldots, A_k$ uniquely determined by G, such that

$$A_i \subseteq \{1, \ldots, i\}, \quad i \in A_i,$$

and, for $i \neq j$,

$$|A_i \cap A_j| \text{ is even if and only if } v_i \text{ and } v_j \text{ are joined in } G.$$

Indeed, suppose we have already chosen the sets $A_1, A_2, \ldots, A_{j-1}$. Our goal is to choose the next set A_j which is properly joined to all the sets $A_1, A_2, \ldots, A_{j-1}$, that is, $|A_j \cap A_i|$ must be even precisely when v_j is joined to v_i in G. We will obtain A_j as the last set in a sequence $B_1 \subseteq B_2 \subseteq \ldots \subseteq B_{j-1} = A_j$, where, for each $1 \leq i < j$, B_i is a set properly joined to all sets $A_1, A_2, \ldots, A_i$.

As the first set B_1 we take either $\{j\}$ or $\{1, j\}$ depending on whether v_j is joined to v_1 or not. Having the sets $B_1, \ldots, B_{i-1}$ we want to choose a set B_i. If v_j is joined to v_i then we set $B_i = B_{i-1}$ or $B_i = B_{i-1} \cup \{i\}$ depending on whether $|B_{i-1} \cap A_i|$ is even or odd. If v_j is not joined to v_i then we act dually. Observe that our choice of whether i is in B_i will effect $|B_i \cap A_i|$ (since $i \in A_i$) but none of $|B_i \cap A_l|$, $l < i$ (since $A_l \subseteq \{1, \ldots, l\}$). After $j - 1$ steps we will obtain the desired set $B_{j-1} = A_j$. □

Exercises

10.1. Let $A \subseteq \{0,1\}^n$ be (n,k)-dense and suppose that no vector in A has more than r ones. Prove that some two vectors in A have at most $r-k$ ones in common.

10.2. (Alon 1986). Let A be a 0-1 matrix of 2^n rows and n columns, the i-th row being the binary representation of $i-1$ $(1 \leq i \leq 2^n)$. Show that for any choice of k distinct columns of A and any choice of k bits, there are exactly 2^{n-k} rows of A that have the j-th chosen bit in the j-th chosen column.

10.3. Let $A \subseteq \{0,1\}^n$, $|A| = n$. By induction on k prove that, for every $k = 1, 2, \ldots, n-1$, there exist k coordinates such that the projection of A onto these coordinates has more than k vectors. For $k = n-1$ this is the well-known Bondy's theorem (Theorem 11.1).

10.4. (Chandra et al. 1983). Prove the following (n,k)-universality criterion for the case $k = 2$. Given a set $A \subseteq \{0,1\}^n$ of $m = |A|$ vectors, look at it as an $m \times n$ matrix, whose rows are the vectors of A. Let $v_1, \ldots, v_n \in \{0,1\}^m$ be the *columns* of this matrix, and let $\overline{v}_1, \ldots, \overline{v}_n$ be their complements, i.e., $\overline{v}_i$ is obtained from v_i by switching all its bits to the opposite values. Prove that A is $(n,2)$-universal if and only if all the vectors $v_1, \ldots, v_n, \overline{v}_1, \ldots, \overline{v}_n$ are different and form an antichain in $\{0,1\}^m$, i.e., are mutually incomparable.

10.5. Let $A \subseteq \{0,1\}^n$, $|A| = m$. Look at A as an $m \times n$ matrix, and let $\mathcal{F}_A$ be the family of those subsets of $\{1, \ldots, m\}$, whose incidence vectors are columns of this matrix. Show that A is (n,k)-universal if and only if the family $\mathcal{F}_A$ is *k-independent* in the following sense: for every k distinct members $S_1, \ldots, S_k$ of $\mathcal{F}_A$ all 2^k intersections $\bigcap_{i=1}^k T_i$ are non-empty, where each T_i can be either S_i or its complement $\overline{S}_i$.

10.6. Show that the converse of Proposition 10.19 also holds: if the set of rows of the incidence matrix of a given bipartite graph is (n,k)-universal then the graph satisfies the isolated neighbor condition for k.

10.7. Let p be a prime with $p \equiv 1 \bmod 4$. Show that -1 is a square in the field $\mathbb{F}_p$. *Hint*: Let P be the product of all nonzero elements of $\mathbb{F}_p$. If -1 is not a square, then $x^2 = -1$ has no solutions; so, the set of all $p-1$ nonzero elements of $\mathbb{F}_p$ can be divided into $(p-1)/2$ pairs such that the product of the elements in each pair is -1; hence $P = 1$. On the other hand, for any $x \neq \pm 1$ there exists exactly one $y \neq x$ with $xy = 1$, so all the elements of $\mathbb{F}_p \setminus \{-1, 0, +1\}$ can be divided into pairs so that the product of elements in each pair is 1; hence, $P = -1$, a contradiction.

10.8. Recall that a set $A \subseteq \{0,1\}^n$ of vectors is downwards closed if $v \in A$ and $u \leq v$ implies $u \in A$. Similarly, say that a set is upwards closed if $v \in A$ and $u \geq v$ implies $u \in A$. Show that Kleitman's theorem (Theorem 10.6) implies the following: Let A, B be upwards closed and C downwards closed subsets of $\{0,1\}^n$. Then

$$|A \cap B| \geq \frac{|A| \cdot |B|}{2^n}$$

and

$$|A \cap C| \leq \frac{|A| \cdot |C|}{2^n}.$$

Hint: For the first inequality, apply Kleitman's theorem to the complements of A and B. For the second inequality, take $B := \{0,1\}^n \setminus C$, and apply the first inequality to the pair A, B to get

$$|A| - |A \cap C| = |A \cap B| \geq 2^{-n}|A|(2^n - |C|).$$

10.9. Show that the lower bound $t_{n-1}(A) \geq |A| - 2^{t-1} + 1$ given in Theorem 10.5 is the best possible whenever t divides n. *Hint*: Split $\{1, \ldots, n\}$ into n/t disjoint subsets $S_1, \ldots, S_{n/t}$ with $|S_i| = t$ and define A to be the set of all vectors $v \neq \mathbf{0}$ such that $v_j = 0$ for all $j \notin S_i$.

10.10. Let $\mathcal{F}$ be a matroid over a ground-set X, and $Y \subseteq X$. Recall that a maximum independent subset of Y is a member $A \in \mathcal{F}$ such that $A \subseteq Y$ and $A + x \notin \mathcal{F}$ for all $x \in Y \setminus A$. Use the exchange property of matroids to show that if $Z \subseteq Y$ then every maximal independent set in Z can be extended to a maximal independent set in Y.

10.11. Let $\mathcal{F}$ be a matroid over a ground-set X. By Lemma 10.7, we know that, for every subset $Y \subseteq X$, all independent subsets of Y have the same number of elements. This number $r(Y)$ is callled the *rank* of the set Y within $\mathcal{F}$. Use Exercise 10.10 to show that then the rank function is submodular: for every subsets $Y, Z \subseteq X$, $r(Y \cup Z) + r(Y \cap Z) \leq r(Y) + r(Z)$. *Hint*: $|B \cap Y| + |B \cap Z| = |B \cap (Y \cup Z)| + |B \cap (Y \cap Z)|$.

10.12. Let $X = Y \cup \{x\}$ where $|Y| = k+2$ and $x \notin Y$. Let $\mathcal{F}$ be a hereditary family whose only maximum independent sets are the set Y and all 2-element sets $\{x, y\}$ with $y \in Y$. Show that $\mathcal{F}$ is a k-matroid, but is not k-extendible.

10.13. Show that the intersection of k matroids is a k-matroid. *Hint*: Show that the intersection of k 1-extendible systems is k-extendible and use Lemma 10.10.

11. Witness Sets and Isolation

Given a set A of distinct 0-1 vectors and a vector u in A, how many bits of u must we know in order to distinguish it from the other vectors in A? Such a set of bits is a *witness* for the fact that $u \notin A \setminus \{u\}$. In this chapter we will give some basic estimates on the size of these witnesses. We will also consider a related problem of how to *isolate* an object within a given universum according to its weight. Finally, we will describe the so-called "dictator paradox" saying that, if the society fulfills some simple "democracy axioms," then there will always be an individual (a dictator?) whose options prevail against all options.

11.1 Bondy's theorem

Let $A \subseteq \{0,1\}^n$ be a set of m distinct 0-1 vectors of length n. A set $S \subseteq \{1, \ldots, n\}$ of coordinates is a *witness* for a vector u in A if for every other $v \in A$ there exists a coordinate in S on which u differs from v. We may also say that *exposing* the entries of u corresponding to S uniquely determines u among vectors in A. The minimum size of a witness for u in A is denoted by $\mathrm{w}_A(u)$ (or by $\mathrm{w}(u)$, if the underlying set A is clear from the context).

It is easy to show that every set of m vectors contains a vector whose witness has size at most $\log_2 m$ (see Exercise 11.2). On the other hand, it is obvious that $\mathrm{w}_A(u) \leq |A| - 1$ for any A and $u \in A$, and a simple example shows that this is tight: if A consists of the all-0 vector 0^n and the n vectors with precisely one 1, then $\mathrm{w}_A(0^n) = n$.

The following result, due to Bondy (1972), shows that if we take only $m \leq n$ vectors, then all the vectors will already have *one and the same* witness of size at most $m-1$. The *projection* of a vector $v = (v_1, \ldots, v_n)$ onto a set of coordinates $S = \{i_1, \ldots, i_k\}$ is the vector $v\restriction_S := (v_{i_1}, \ldots, v_{i_k})$. The projection of a set of vectors A is the set $A\restriction_S = \{v\restriction_S : v \in A\}$.

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_11,

Theorem 11.1 (Bondy 1972). *For every set A of* 0-1 *vectors there exists a set S of at most $|A| - 1$ coordinates such that all the vectors $\{v\restriction_S \ : \ v \in A\}$ are distinct.*

Proof. Suppose that A is a counterexample, that is, $|A\restriction_S| < |A|$ for every set S of at most $|A| - 1$ coordinates. Let S be a *maximal* set of coordinates for which $|A\restriction_S| \geq |S| + 1$. Since $|A\restriction_S| \leq |A| - 1$, at least two vectors $u \neq v \in A$ must coincide on S. Take a coordinate $i \not\in S$ on which these two vectors differ, and set $T := S \cup \{i\}$. Since the vectors u,v coincide on S but differ on T, the projection $A\restriction_T$ must have at least one more vector than $A\restriction_S$; hence,

$$|A\restriction_T| \geq |A\restriction_S| + 1 \geq |S| + 2 = |T| + 1,$$

a contradiction with the maximality of S. □

Given k, how large must a set A be in order to be sure that at least one of its vectors will have no witness of size $\leq k$? It is clear that any such set A must have more than 2^k vectors; this is a trivial lower bound. A trivial upper bound is 2^n. The following simple observation shows that much fewer vectors are enough.

Proposition 11.2. *In every set of more than $2^k\binom{n}{k}$* 0-1 *vectors of length n there is a vector which has no witness of size k.*

Proof. Let A be a set of 0-1 vectors of length n, and assume that every vector in it has a witness of size k. Then each vector $u \in A$ has its own set S_u of k coordinates on which this vector differs from all other vectors in A. That is, we can assign to each vector $u \in A$ its "pattern" – a set S_u of k bits and the projection $u\restriction_S$ of u onto this set – so that different vectors will receive different patterns, i.e., if $u \neq v$ then either $S_u \neq S_v$ or $S_u = S_v$ but u and v differ on some coordinate in S_u. There are $\binom{n}{k}$ possible subsets of k coordinates and, on each of these sets, vectors can take no more than 2^k possible values. Thus, there are at most $\binom{n}{k}2^k$ possible patterns and, since each vector in A must have its own pattern, we conclude that $|A| \leq \binom{n}{k}2^k$. □

11.2 Average witnesses

Since the worst-case witness sets may have to be large, it is natural to consider the *average* witness size:

$$\mathrm{w}_{\mathrm{ave}}(A) := \frac{1}{|A|}\sum_{u\in A} \mathrm{w}_A(u).$$

The same example, as in the previous section, shows that the gap between the worst-case witness size and the average witness size may be exponential: if A

is the set of $n+1$ vectors with at most one 1, then $\mathrm{w}_{\mathrm{ave}}(A) = 2n/(n+1) \leq 2$, but in the all-0 vector all n bits must be exposed.

How large can $\mathrm{w}_{\mathrm{ave}}(A)$ be as a function of $|A|$? The following result of Kushilevitz, Linial, Rabinovitch, and Saks (1996) says that the average witness size of *any* set does not exceed the square root of its size, and that this bound is almost optimal.

Theorem 11.3. *For every set A of m 0-1 vectors, $\mathrm{w}_{\mathrm{ave}}(A) \leq 2m^{1/2}$. On the other hand, for infinitely many numbers m, there exists a set A of m 0-1 vectors such that $\mathrm{w}_{\mathrm{ave}}(A) \geq \frac{1}{2\sqrt{2}} m^{1/2}$.*

Proof. Upper bound. Take an arbitrary set A of m vectors and order its vectors $u_1, u_2, \ldots, u_m$ by decreasing value of their smallest witness size: $\mathrm{w}(u_1) \geq \mathrm{w}(u_2) \geq \cdots \geq \mathrm{w}(u_m)$. Consider the sum of the first k largest values $\sum_{i=1}^{k} \mathrm{w}(u_i)$ for a value k soon to be set. Find a set T of at most $k-1$ coordinates as guaranteed by Bondy's theorem applied to the set $\{u_1, \ldots, u_k\}$ and expose the T-coordinates in *all* vectors of A. By the property of T, vectors $u_1, \ldots, u_k$ are already mutually distinguished. The T-coordinates of every vector u_j with $j > k$, distinguish u_j from all $u_1, \ldots, u_k$, except, perhaps, *one* u_i (because no two of the vectors $u_1, \ldots, u_k$ coincide on T). It is possible to expose a single additional bit in u_i to distinguish u_i from u_j. Apply this step for every u_j, $j > k$. Consequently, each of $u_1, \ldots, u_k$ is distinguished from every other vector in A. No more than $m-k$ additional bits get exposed in this process, so:

$$\sum_{i=1}^{k} \mathrm{w}(u_i) \leq k(k-1) + m - k = k^2 - 2k + m. \tag{11.1}$$

In particular, it follows that $\mathrm{w}(u_k) \leq k - 2 + m/k$.

Putting these two observations together we get

$$\begin{aligned}\sum_{i=1}^{m} \mathrm{w}(u_i) &= \sum_{i=1}^{k} \mathrm{w}(u_i) + \sum_{i=k+1}^{m} \mathrm{w}(u_i) \\ &\leq (k^2 - 2k + m) + (m-k)\left(k - 2 + \frac{m}{k}\right).\end{aligned}$$

Pick $k := m^{1/2}$; the above inequality then yields $\sum_{i=1}^{m} \mathrm{w}(u_i) \leq 2m^{3/2}$, which means that $\mathrm{w}_{\mathrm{ave}}(A) \leq 2m^{1/2}$, as desired.

Lower bound. We will explicitly construct a set $A \subseteq \{0,1\}^n$ which achieves the lower bound. Let p be a prime and consider a projective plane $\mathrm{PG}(2,p)$ of order p (see Sect. 12.4). Such a plane consists of $n = p^2 + p + 1$ points $P = \{1, \ldots, n\}$ and n subsets of points $L_1, \ldots, L_n \subseteq P$ (called *lines*) satisfying the following three conditions: (i) each line has exactly $p+1$ points; (ii) every two lines intersect in exactly one point, and (iii) exactly $p+1$ lines meet in one point.

We consider n-dimensional vectors where the coordinates correspond to points of P, and define $A \subseteq \{0,1\}^n$ to be the family of $m = 2n$ binary vectors, of which n are the incidence vectors of lines of $\mathrm{PG}(2,p)$, and another n are all unit vectors, i.e., incidence vectors of all singletons $\{i\}$, $i \in P$.

For a vector $u \in A$, corresponding to a line L, $\mathrm{w}(u) = 2$, since it suffices to expose the coordinates corresponding to any two points on L. Such a pair distinguishes u from all singletons, and since distinct lines share exactly one point, this pair also distinguishes u from the incidence vectors of other lines.

On the other hand, $\mathrm{w}(u) = p+2$ if $u = (0,\dots,0,1,0,\dots,0)$ corresponds to a singleton point $i \in P$. To distinguish u from the incidence vector of a line L containing i, a zero in u should be exposed in a coordinate that corresponds to a point on L other than i. There are $p+1$ lines, whose pairwise intersection is $\{i\}$, so to distinguish u from all of them, at least $p+1$ distinct 0-entries should be exposed. To distinguish u from other singletons, the 1-entry should be exposed as well (the alternative being to expose all $p^2 + p$ 0-entries).

Putting things together, we get

$$\mathrm{w}_{\mathrm{ave}}(A) = \frac{1}{|A|}\sum_{u\in A}\mathrm{w}(u) = \frac{1}{2n}(2n + (p+2)n) = \frac{p+4}{2} \geq \frac{n^{1/2}}{2\sqrt{2}}.$$

□

The next natural problem concerning 0-1 vectors is the following question about the *distribution* of their witness sizes:

> Given an integer t, $1 \leq t \leq m$, and a set of m vectors, how many of its vectors have a witness of size at least (or at most) t?

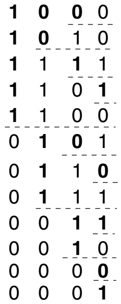
If we know nothing more about the set except for its size, the question turns out to be difficult. Still, Kushilevitz et al. (1996) have found the following interesting partial solutions (see also Exercise 11.3):

Lemma 11.4. *Let A be a set of m distinct 0-1 vectors. Then*

(a) *for any $t \leq m$ at most t of vectors in A have a minimal witness of size at least $t + m/t - 2$;*
(b) *for any $t \leq \sqrt{m}$ at least $t^2 - t$ of vectors in A have a witness of size at most $2t + \log_2 m$.*

Proof. The first claim (a) follows from the proof of the upper bound in Theorem 11.3: let k be the number of vectors $u \in A$ for which $\mathrm{w}(u) \geq t + m/t - 2$, and apply (11.1).

To prove the second claim (b), reorder the vectors in A as follows: split the vectors into two groups according to their first coordinate, and let the vectors of the smaller group (i.e., of the group containing at most half of the vectors) precede those in the larger. Expose the first coordinate in all vectors of the smaller group. Proceed recursively in the same manner on each group separately (by looking at next coordinates), and so on, until each group reduces to a single vector (see Fig. 11.1). Observe that:



1 **0** **0** 0
1 **0** 1 0
1 1 **1** 1
1 1 0 **1**
1 1 0 0
0 **1** **0** 1
0 **1** 1 **0**
0 **1** 1 1
0 0 **1** **1**
0 0 **1** 0
0 0 0 **0**
0 0 0 **1**

Fig. 11.1 Exposed bits are in boldface; a vector u *follows* vector v if u is below v.

(i) each vector is distinguished from all those following it (but not necessarily from those preceding it);
(ii) no vector has more than $\log_2 m$ bits exposed (since each time one bit is exposed in at most one-half of the vectors of a current group).

Let B be the set of the first t^2 vectors. Applying the first claim (a) to this set, we conclude that at most t of its vectors have a witness of size at least $2t$. Therefore, at least $t^2 - t$ of the vectors in B can be distinguished from other members of B at the cost of exposing at most $2t$ additional bits in each of them. We call these vectors *good.* By (i) and (ii), at the cost of exposing at most $\log_2 m$ bits, each good vector v is already distinguished from all the vectors in A following it. On the other hand, all the vectors preceding v belong to B, and hence, v is distinguished also from them by at most $2t$ additional bits. Thus, we have at least $t^2 - t$ good vectors v and for each of them, $\mathrm{w}_A(v) \leq 2t + \log_2 m$. □

11.3 The isolation lemma

Let X be some set of n points, and $\mathcal{F}$ be a family of subsets of X. Let us assign a weight $w(x)$ to each point $x \in X$ and let us define the weight of a set E to be $w(E) = \sum_{x \in E} w(x)$. It may happen that several sets of $\mathcal{F}$ will have the minimal weight. If this is not the case, i.e., if $\min_{E \in \mathcal{F}} w(E)$ is achieved by a unique $E \in \mathcal{F}$, then we say that w is *isolating for* $\mathcal{F}$.

The following lemma, due to K. Mulmuley, U. Vazirani, and V. Vazirani (1987), says that – independent of what our family $\mathcal{F}$ actually is – a randomly chosen w is isolating for $\mathcal{F}$ with large probability.

Lemma 11.5. *Let $\mathcal{F}$ be a family of subsets of an n-element set X. Let $\boldsymbol{w} : X \to \{1, \ldots, N\}$ be a random function, each $\boldsymbol{w}(x)$ independently and uniformly chosen over the range. Then*

$$\Pr[\boldsymbol{w} \textit{ is isolating for } \mathcal{F}] \geq 1 - \frac{n}{N}.$$

Proof (Spencer 1995). For a point $x \in X$, set

$$\alpha(x) = \min_{E \in \mathcal{F};\, x \notin E} \boldsymbol{w}(E) - \min_{E \in \mathcal{F};\, x \in E} \boldsymbol{w}(E \setminus \{x\}).$$

A crucial observation is that evaluation of $\alpha(x)$ does not require knowledge of $\boldsymbol{w}(x)$. As $\boldsymbol{w}(x)$ is selected uniformly from $\{1, \ldots, N\}$,

$$\Pr[\boldsymbol{w}(x) = \alpha(x)] \leq 1/N,$$

so that

$$\Pr[\boldsymbol{w}(x) = \alpha(x) \text{ for some } x \in X] \leq n/N.$$

But if $\boldsymbol{w}$ had two minimal sets $A, B \in \mathcal{F}$ and $x \in A \setminus B$, then

$$\min_{E \in \mathcal{F}; x \notin E} \boldsymbol{w}(E) = \boldsymbol{w}(B),$$

$$\min_{E \in \mathcal{F}; x \in E} \boldsymbol{w}(E \setminus \{x\}) = \boldsymbol{w}(A) - \boldsymbol{w}(x),$$

so $\boldsymbol{w}(x) = \alpha(x)$. Thus, if $\boldsymbol{w}$ is *not* isolating for $\mathcal{F}$ then $\boldsymbol{w}(x) = \alpha(x)$ for some $x \in X$, and we have already established that the last event can happen with probability at most n/N. □

11.4 Isolation in politics: the dictator paradox

One of the problems of politics involves averaging out individual preferences to reach decisions acceptable to society as a whole. In this section we will prove one isolation-type result due to Arrow (1950) which shows that, under some simple "democracy axioms" this is indeed a difficult task.

The simple process of voting can lead to surprisingly counterintuitive paradoxes. For example, if three people vote for three candidates, giving the rankings $x < y < z$, $y < z < x$, $z < x < y$, then a majority prefers y to x ($x < y$), x to z ($z < x$) but also z to y ($y < z$). In general, we have the following situation.

Suppose that $I = \{1, \ldots, n\}$ is a society consisting of a set of n individuals. These individuals are to be offered a choice among a set X of options, for example, by a referendum. We assume that each individual i has made her/his mind up about the relative worth of the options. We can describe this by a total order $<_i$ on X, for each $i \in I$, where $x <_i y$ means that the individual i prefers option y to option x. So, after a referendum we have a set $R = \{<_1, \ldots, <_n\}$ of total orders on X. A *social choice function* F takes such a set of total orders as input and comes up with a "social preference" on X, i.e., with some total order $<$ on X. Being total means, in particular, that the order $<$ is transitive: if $x < y$ and $y < z$ then $x < z$.

Given a social choice function F, a *dictator* is an individual $i_0 \in I$ such that for every referendum, the resulting social preference $<$ coincides with the preference $<_{i_0}$ of this individual. That is, for any given set of total orders $R = \{<_1, \ldots, <_n\}$, the social choice function will output the order $<_{i_0}$, independent of preferences $<_i$ made by other individuals $i \neq i_0$.

Arrow's theorem asserts that, if the choice function F fulfills some natural "democracy axioms," then there will *always* be a dictator! That is, once we fix some social choice function F, then there will be an individual (a dictator?) whose options prevail against all options.

Let us consider the following three natural *democracy axioms*:

(A1) If $x < y$ (in the social preference), then the same remains true if the individual preferences are changed in y's favor.

(A2) If $Y \subseteq X$ is a set of options and if during two referendums no individual changes his/her mind about the options within the set Y (i.e. no one changes his mind about $y < y'$ or $y' < y$ for y, y' that are both in Y), then the society also don't changes its mind about these options.

(A3) For any distinct options $x, y \in X$, there is some system of individual preferences for which the corresponding social preference has $x < y$. That is, it should be possible for society to prefer y to x if enough individuals do so.

Theorem 11.6. *If $|X| \geq 3$ then for every social choice function, satisfying the three democracy axioms above, there is a dictator.*

Proof. We follow the elegant argument from Cameron (1994). Suppose that we have a social choice function. If (x, y) is an ordered pair of distinct options, we say that a set J of individuals is (x, y)-*decisive* if, whenever all members of J prefer y to x, then so does the social order; formally, if $x <_i y$ for all $i \in J$, then $x < y$. Further, we say that J is *decisive* if it is (x, y)-decisive for some distinct $x, y \in X$.

Let J be a minimal decisive set. It follows from (A1)–(A3) that, for any distinct options $x, y \in X$, if every individual prefers y to x then so does the social order. Hence, $J \neq \emptyset$. Suppose that J is (x, y)-decisive, and let i_0 be a member of J.

Claim 11.7. $J = \{i_0\}$.

To prove the claim, suppose the opposite and let $J' := J \setminus \{i_0\}$ and $K := I \setminus J$. Let v be an option in X different from x and y (remember that $|X| \geq 3$). Consider the individual preferences $<_i, i \in I$ for which

$$\begin{array}{llll} x <_{i_0} y & <_{i_0} v & \\ v <_i & x <_i & y & \text{for all } i \in J' \\ y <_j & v <_j & x & \text{for all } j \in K \end{array}$$

Then $x < y$, since all members of the (x, y)-decisive set J think so, and $y < v$, since if $v < y$ then J' would be (v, y)-decisive (nobody outside J' thinks so), contradicting the minimality of J. Hence $x < v$. But then $\{i_0\}$ is (x, v)-decisive, since nobody else agrees with this order. By minimality of J, we have $J = \{i_0\}$, as desired.

Claim 11.8. i_0 is a dictator.

We have to prove that $\{i_0\}$ is (u, v)-decisive for any pair of different options $u \neq v$. The case when $u = x$ is covered by the (proof of) Claim 11.7, and we are left with two possible situations: either $v = x$ or neither $v = x$ nor $u = x$. The argument in both cases is similar.

Case 1: $u \neq x$ and $v \neq x$.

Consider individual preferences in which

$$\begin{array}{llll} u <_{i_0} x & <_{i_0} v & \\ v <_j \;\; u & <_j \;\; x & \text{for all } j \neq i_0 \end{array}$$

Then $u < x$ (because everybody thinks so) and $x < v$ (because i_0 thinks so and, by Claim 11.7, is (x, v)-decisive for any $v \neq x$); hence $u < v$, and $\{i_0\}$ is (u, v)-decisive because nobody else agrees with this order.

Case 2: $v = x$.

Take $z \notin \{u, x\}$ and consider individual preferences in which

$$\begin{array}{llll} u <_{i_0} z & <_{i_0} x & \\ z <_j \;\; x & <_j \;\; u & \text{for all } j \neq i_0 \end{array}$$

Then $u < z$ (because i_0 thinks so and both u, z differ from x) and $z < x$ (because everybody thinks so); hence $u < x$, and $\{i_0\}$ is (u, x)-decisive.

This completes the proof of the claim, and thus, the proof of the theorem. □

Exercises

11.1. Bondy's theorem (Theorem 11.1) implies that, if we take n binary vectors of length n, then all these vectors differ on some set of $n - 1$ bits. Does this hold for $n + 1$ vectors?

11.2. Prove that every set of m vectors contains a vector whose witness has size at most $\log_2 m$.

11.3. Generalize Lemma 11.4 as follows. Let A be a set of m 0-1 vectors. For an integer l, $1 \leq l \leq m$, let

$$f(m, l) = \min\{k \; : \; k \geq 1 \text{ and } k + m/k \geq l + 2\}.$$

Prove that:

(a) at most $f(m, l)$ vectors in A have a minimal witness of size at least l;
(b) for any $k \leq m$, at least $k - f(k, l - \log_2 m)$ vectors in A have witness of size at most l.

11.4. Lemma 11.5 isolates the unique set with the minimal weight. With what probability will there be a unique set with the *maximal* weight?

11.5. Prove that Lemma 11.5 also holds when the weight of a set is defined to be the *product* of the weights of its elements.

12. Designs

The use of combinatorial objects, called designs, originates from statistical applications. Let us assume that we wish to compare v varieties of wines. In order to make the testing procedure as fair as possible it is natural to require that:

(a) each participating person tastes the same number (say k) of varieties so that each person's opinion has the same weight;
(b) each pair of varieties is compared by the same number (say λ) of persons so that each variety gets the same treatment.

One possibility would be to let everyone taste all the varieties. But if v is large, this is very impractical (if not dangerous, as in the case of wines), and the comparisons become rather unreliable. Thus, we should try to design the experiment so that $k < v$.

Definition 12.1. Let $X = \{1, \ldots, v\}$ be a set of *points* (or *varieties*). A (v, k, λ) *design* over X is a collection $\mathcal{D}$ of distinct subsets of X (called *blocks*) such that the following properties are satisfied:

(1) each set in $\mathcal{D}$ contains exactly k points;
(2) every pair of distinct points is contained in exactly λ blocks.

The number of blocks is usually denoted by b. If we replace (2) by the following property:

(2') every t-element subset of X is contained in exactly λ blocks,

then the corresponding family is called a t–(v, k, λ) *design*. A *Steiner system* $S(t, k, v)$ is a t–(v, k, λ)-design with $\lambda = 1$. A design, in which $b = v$ (i.e., the number of blocks and points is the same) is often called *symmetric*.

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_12,

12.1 Regularity

In every design, every pair of points lies in the same number of blocks. It is easy to show that then the same also holds for every single point. A family of sets $\mathcal{F}$ is *r-regular* if every point lies in exactly r sets; r is the *replication number* of $\mathcal{F}$.

Theorem 12.2. *Let $\mathcal{D}$ be a (v, k, λ) design containing b blocks. Then $\mathcal{D}$ is r-regular with the replication number r satisfying the equations*

$$r(k-1) = \lambda(v-1). \tag{12.1}$$

and

$$bk = vr. \tag{12.2}$$

Proof. Let $a \in X$ be fixed and assume that a occurs in r_a blocks. We count in two ways the cardinality of the set

$$\{(x, B) \,:\, B \in \mathcal{D}; a, x \in B; x \neq a\}.$$

For each of the $v-1$ possibilities for x $(x \neq a)$ there are exactly λ blocks B containing both a and x. The cardinality of the set is therefore $(v-1)\lambda$. On the other hand, for each of the r_a blocks B containing a, the element $x \in B \setminus \{a\}$ can be chosen in $|B|-1 = k-1$ ways. Hence $(v-1)\lambda = r_a(k-1)$. This shows that r_a is independent of the choice of a and proves (12.1).

To prove the second claim we count in two ways the cardinality of the set

$$\{(x, B) \,:\, B \in \mathcal{D},\, x \in B\}.$$

For each $x \in X$ the block B can be chosen in r ways. On the other hand, for each of the b blocks B the element $x \in B$ can be chosen in k ways. Hence $vr = bk$, as desired. □

Thus, every design is an r-regular family with the parameter r satisfying both equations (12.1) and (12.2). It turns out that for regularity the second condition (12.2) is also sufficient. The proof presented here is due to David Billington (see Cameron 1994).

Theorem 12.3. *Let $k < v$ and $b \leq \binom{v}{k}$. If $bk = vr$ then there is an r-regular family $\mathcal{F}$ of k-subsets of $\{1, \ldots, v\}$ with $|\mathcal{F}| = b$.*

Proof. There is a simple way to make a k-uniform family $\mathcal{F}$ "more regular." (We have already used a similar argument in the proof of Theorem 10.2 to make a given set of binary vectors "more hereditary.")

Let r_x be the *replication number* of x, the number of sets of $\mathcal{F}$ which contain x. (In our previous notation this is the *degree* $d(x)$ of a point in the family $\mathcal{F}$. Here we follow the notation which is usual in the design theory.)

If $r_x > r_y$, then there must exist a $(k-1)$-set A, containing neither x nor y, such that $\{x\} \cup A \in \mathcal{F}$ and $\{y\} \cup A \notin \mathcal{F}$. Now form a new family $\mathcal{F}'$ by removing $\{x\} \cup A$ from $\mathcal{F}$ and including $\{y\} \cup A$ in its place. In the new family, $r'_x = r_x - 1$, $r'_y = r_y + 1$, and all other replication numbers are unaltered. Starting with *any* family of k-sets, we reach by this process a family in which all the replication numbers differ by at most 1 (an *almost regular* family), containing the same number of sets as the original family. By double counting, the average replication number is

$$\frac{1}{v}\sum r_x = \frac{1}{v}\sum_{A\in\mathcal{F}} |A| = \frac{bk}{v};$$

and an almost regular family whose average replication number is an integer must be regular. □

12.2 Finite linear spaces

Sometimes it is possible to show that a design has at least as many blocks as it has points. The well-known Fisher's Inequality (Theorem 7.5) implies that if $\mathcal{D}$ is a (v, k, λ) design then $|\mathcal{D}| \geq v$ (see Exercise 12.1). Many generalizations exist. For example, the Petrenjuk–Ray-Chaudhuri–Wilson Inequality (Petrenjuk 1968, Ray-Chaudhuri, and Wilson 1975) states that, if $\mathcal{D}$ is a $2s$–(v, k, λ) design with $v \geq k + s$ then $|\mathcal{D}| \geq \binom{v}{s}$. Both results can be obtained using the linear algebra method (cf. Exercise 7.6).

Some of these results, however, may be proved by direct double counting. Such, for example, is the argument due to Conway for the case of "finite linear spaces." (Do not confuse these linear spaces with those from Analysis.)

A (finite) *linear space* over a set X is a family $\mathcal{L}$ of its subsets, called *lines*, such that:

- every line contains at least two points, and
- any two points are on exactly one line.

Theorem 12.4 (De Bruijn–Erdős 1948). *If $\mathcal{L}$ is a linear space over X then $|\mathcal{L}| \geq |X|$, with equality iff any two lines share exactly one point.*

Proof (due to J. Conway). Let $b = |\mathcal{L}| \geq 2$ and $v = |X|$. For a point $x \in X$, let r_x, as above, be its replication number, i.e., the number of lines in $\mathcal{L}$ containing x. If $x \notin L$ then $r_x \geq |L|$ because there are $|L|$ lines joining x to the points on L. Suppose $b \leq v$. So, for $x \notin L$, we have

$$b(v - |L|) \geq v(b - r_x)\,.$$

Hence

$$b = \sum_{L\in\mathcal{L}} 1 = \sum_{L\in\mathcal{L}} \sum_{x:x\notin L} \frac{1}{v-|L|} \le \frac{b}{v} \sum_{L\in\mathcal{L}} \sum_{x:x\notin L} \frac{1}{b-r_x}$$
$$= \frac{b}{v} \sum_{x\in X} \sum_{L:x\notin L} \frac{1}{b-r_x} = \frac{b}{v} \sum_{x\in X} 1 = b,$$

and this implies that all inequalities are equalities so that $b = v$, and $r_x = |L|$ whenever $x \notin L$. □

There are several methods to construct (symmetric) designs. In the next two sections we will study two of them: one comes from "difference sets" in abelian groups, and the other from "finite geometries." The third important construction, which arises from Hadamard matrices, will be described in Chap. 14 (see Theorem 14.11).

12.3 Difference sets

Let $\mathbb{Z}_v$ be an additive Abelian group of integers modulo v. We can look at $\mathbb{Z}_v$ as the set of integers $\{0, 1, \ldots, v-1\}$ where the sum is modulo v.

Definition 12.5. Let $2 \le k < v$ and $\lambda \ge 1$. A (v, k, λ) *difference set* is a k-element subset $D = \{d_1, d_2, \ldots, d_k\} \subseteq \mathbb{Z}_v$ such that the collection of values $d_i - d_j$ $(i \ne j)$ contains every element in $\mathbb{Z}_v \setminus \{0\}$ exactly λ times.

Since the number of pairs (i, j) with $i \ne j$ equals $k(k-1)$ and these give each of the $v-1$ nonzero elements λ times as a difference, it follows that

$$\lambda(v-1) = k(k-1). \tag{12.3}$$

If D is a difference set, we call the set

$$a + D := \{a + d_1, a + d_2, \ldots, a + d_k\}$$

a *translate* of D. Notice that our assumption $k < v$ together with (12.3) implies that all the translates of a difference set are different. Indeed, if $a + D = D$ for some $a \ne 0$, then there is a permutation π of $\{1, \ldots, k\}$ so that $\pi(i) \ne i$ and $a + d_i = d_{\pi(i)}$ for all i. Hence, a can be expressed as a difference $d_{\pi(i)} - d_i$ in k ways; but $\lambda < k$ by (12.3) and our assumption that $k < v$.

Theorem 12.6. *If $D = \{d_1, d_2, \ldots, d_k\}$ is a (v, k, λ) difference set then the translates*

$$D, 1 + D, \ldots, (v-1) + D$$

are the blocks of a symmetric (v, k, λ) design.

Proof. We have v blocks over v points. Since, clearly, every one of the translates contains k points, it is sufficient to show that every pair of points is contained in exactly λ blocks. Let $x, y \in \mathbb{Z}_v$, $x \neq y$. Suppose that $x, y \in a+D$ for some $a \in \mathbb{Z}_v$. Then $x = a + d_i$ and $y = a + d_j$ for some pair $i \neq j$. Also, we have $d_i - d_j = x - y = d$. Now, there are exactly λ pairs $i \neq j$ such that $d_i - d_j = d$, and for each such pairs, there is exactly one a for which $x, y \in a + D$, namely, $a = x - d_i = y - d_j$. □

Let us now describe one construction of difference sets. *Squares* (or *quadratic residues*) in $\mathbb{Z}_v$ are the elements a^2 for $a \in \mathbb{Z}_v$.

Theorem 12.7. *If v is a prime power and $v \equiv 3 \bmod 4$, then the nonzero squares in $\mathbb{Z}_v$ form a (v, k, λ) difference set with $k = (v-1)/2$ and $\lambda = (v-3)/4$.*

The condition $v \equiv 3 \bmod 4$ is only used to ensure that -1 is not a square in $\mathbb{Z}_v$, i.e., that $-1 \not\equiv a^2 \bmod v$ for all $a \in \mathbb{Z}_v$. This fact follows from elementary group theory, and we omit its proof here.

Proof. Let D be the set of all nonzero squares, and $k = |D|$. First, observe that $k = (v-1)/2$. Indeed, the nonzero squares in $\mathbb{Z}_v$ are the elements a^2 for $a \in \mathbb{Z}_v \setminus \{0\}$. But for every such a the equation $x^2 = a^2$ has two different solutions $x = \pm a$. So, every pair $(+a, -a)$ gives rise to only one square. This means that exactly half of the nonzero elements in $\mathbb{Z}_v$ are squares, and hence $k = (v-1)/2$.

By the remark above, -1 is not a square in $\mathbb{Z}_v$. Hence, if S is the set of all nonzero squares then $-S = \{-s : s \in S\}$ is exactly the set of nonsquares. For any $s \in S$, the pair $(x, y) \in S \times S$ satisfies the equation $x - y = 1$ if and only if the pair $(sx, sy) \in S \times S$ satisfies the equation $sx - sy = s$, or equivalently, if and only if the pair $(sy, sx) \in S \times S$ satisfies the equation $sy - sx = -s$. This shows that all nonzero squares $s \in S$ and all nonsquares $-s \in -S$ have the same number λ of representations as a difference of two nonzero squares. We can compute λ from the equation (12.3), which gives $\lambda = k(k-1)/(v-1) = (v-3)/4$. □

12.4 Projective planes

Let $\mathcal{L} \subseteq 2^X$ be a linear space with $|\mathcal{L}| = b$ and $|X| = v$. By Theorem 12.4, $b \geq v$. In this section we will consider linear spaces with $b = v$ and with an additional requirement that every line has the same number, say $q + 1$, of points. Then $\mathcal{L}$ turns into a symmetric (v, k, λ) design with $\lambda = 1$ and $k = q + 1$. Such a design is known as a *projective plane of order q*. (The reason for taking the block size of the form $k = q + 1$ is that, for any prime power q, such a design has a very transparent construction using the Galois field $\mathbb{F}_q$; we will give this construction below.) By Theorem 12.2, we have $v = b = q^2 + q + 1$.

Projective planes have many applications. They are particularly useful to show that some bounds in Extremal Set Theory are optimal (cf., for example, Lemma 2.1 and Theorems 6.2, 11.3). Due to their importance, projective planes deserve a separate definition.

Definition 12.8. A *projective plane* of order q consists of a set X of q^2+q+1 elements called *points*, and a family $\mathcal{L}$ of subsets of X called *lines*, having the following properties:

- Every line has $q+1$ points.
- Every two points lie on a unique line.

The only possible projective plane of order $q=1$ is a triangle. For $q=2$, the unique projective plane of order q is the famous *Fano plane* (see Fig. 12.1).

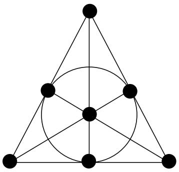


Fig. 12.1 The Fano plane with 7 lines and 3 points on a line

Additional properties of projective planes are summarized as follows:

Proposition 12.9. *A projective plane of order q has the properties:*

(i) *Any point lies on* $q+1$ *lines.*
(ii) *There are* q^2+q+1 *lines.*
(iii) *Any two lines meet in a unique point.*

Proof. (i) Take a point x. There are $q(q+1)$ points different from x; each line through x contains q further points, and there are no overlaps between these lines (apart from x). So, there must be $q+1$ lines through x.

(ii) Counting in two ways the pairs (x, L) with $x \in L$, we obtain $|\mathcal{L}| \cdot (q+1) = (q^2+q+1) \cdot (q+1)$, so $|\mathcal{L}| = q^2+q+1$.

(iii) Let L_1 and L_2 be lines, and x a point of L_1. Then the $q+1$ points of L_2 are joined to x by different lines; since there are only $q+1$ lines through x, they all meet L_2 in a point; in particular, L_1 meets L_2. □

A nice property of projective planes is their duality. Let $(X, \mathcal{L})$ be a projective plane of order q, and let $M = (m_{x,L})$ be its incidence matrix. That is, M is n by n 0-1 matrix, the rows and columns of which correspond to points and lines, and $m_{x,L} = 1$ iff $x \in L$. Each row and column of M has exactly $q+1$ ones, and any two rows and any two columns share exactly one 1.

12.4.1 The construction

The standard construction for projective planes of any prime order $q \geq 2$ is the following.

Let V be the set of all vectors (x_0, x_1, x_2) of elements of $\mathbb{F}_q$, where x_0, x_1, x_2 are not all zero. We identify the vectors that can be obtained from each other by multiplying by a nonzero element of $\mathbb{F}_q$, and call each such collection of vectors a *point*. That is, points of our plane are *sets*

$$[x_0, x_1, x_2] = \{(cx_0, cx_1, cx_2) \,:\, c \in \mathbb{F}_q,\ c \neq 0\}$$

of $q - 1$ vectors in V. There are $(q^3 - 1)/(q - 1) = q^2 + q + 1$ such sets, and hence, so many points. The *line* $L(a_0, a_1, a_2)$, where $(a_0, a_1, a_2) \in V$, is defined to be the set of all those points $[x_0, x_1, x_2]$ for which

$$a_0 x_0 + a_1 x_1 + a_2 x_2 = 0. \tag{12.4}$$

How many points does such a line $L(a_0, a_1, a_2)$ have?

Because $(a_0, a_1, a_2) \in V$, this vector has at least one nonzero component; say $a_0 \neq 0$. Therefore, the equation (12.4) has exactly $q^2 - 1$ solutions $(x_0, x_1, x_2) \in V$: for arbitrary x_1, x_2, not both zero, this equation uniquely determines x_0. Since each $[x_0, x_1, x_2]$ consists of $q - 1$ vectors, there are exactly $(q^2 - 1)/(q - 1) = q + 1$ points $[x_0, x_1, x_2]$ satisfying (12.4). In other words: there are exactly $q + 1$ points on each line. So, it remains to verify that *any two points lie on a unique line.*

To show this, let $[x_0, x_1, x_2]$ and $[y_0, y_1, y_2]$ be two distinct points. How many lines contain both these points? For each such line $L(a_0, a_1, a_2)$,

$$\begin{aligned} a_0 x_0 + a_1 x_1 + a_2 x_2 &= 0, \\ a_0 y_0 + a_1 y_1 + a_2 y_2 &= 0. \end{aligned}$$

Without loss of generality $x_0 \neq 0$. Then $a_0 = -a_1 x_1 / x_0 - a_2 x_2 / x_0$, and we can replace the second equation by

$$a_1 \left(y_1 - \frac{y_0}{x_0} x_1 \right) + a_2 \left(y_2 - \frac{y_0}{x_0} x_2 \right) = 0. \tag{12.5}$$

If

$$y_1 - \frac{y_0}{x_0} x_1 = y_2 - \frac{y_0}{x_0} x_2 = 0$$

then $(y_0, y_1, y_2) = (cx_0, cx_1, cx_2)$ with $c = y_0/x_0$, and hence, $[y_0, y_1, y_2] = [x_0, x_1, x_2]$, which is impossible since we consider distinct points. Therefore, at least one of them, say $y_1 - (y_0/x_0)x_1$, is nonzero. Then for arbitrary nonzero a_2, both a_1 and a_0 are uniquely determined by (12.5) and the first equation; and if (a_0, a_1, a_2) is a solution then (ca_0, ca_1, ca_2) for $c \neq 0$ are all the solutions. Consequently, every two different points $[x_0, x_1, x_2]$ and $[y_0, y_1, y_2]$ are contained in a unique line, as desired.

The constructed projective plane is usually denoted by $\mathrm{PG}(2,q)$.

12.4.2 Bruen's theorem

A *blocking set* in a projective plane is a set of points which intersects every line. The smallest (with respect to the set–theoretic inclusion) blocking sets are just the lines (show this!). This is why blocking sets containing a line are called *trivial*.

What can be said about the size of non-trivial blocking sets? Lines themselves have $q+1$ points, and these are trivial blocking sets. Can we find a non-trivial blocking set with, say $q+2$ or $q+3$ points? The fundamental result due to Bruen (1970) says that any non-trivial blocking set in a projective plane of order q must have at least $q+\sqrt{q}+1$ points, and this lower bound is tight when q is a square (that is, for square q blocking sets of this size exist). For the prime order q, Blokhuis (1994) improved Bruen's bound to $3(q+1)/2$ (which is also optimal).

Theorem 12.10 (Bruen 1970). *Let B be a non-trivial blocking set in a projective plane of order q. Then $|B| \geq q+\sqrt{q}+1$.*

This result captures a very interesting property of projective planes: if we take *any* set of at most $q+\sqrt{q}$ points, then either it contains a line or avoids a line (the third is impossible!).

Proof. Let $|B| = q+m$, and assume that $m \leq \sqrt{q}+1$. We will show that $|B| = q+\sqrt{q}+1$. Let l_i be the number of lines containing precisely i points of B. Counting lines, point-line pairs (x,L) with $x \in B \cap L$, and triples (x,y,L) with $x \neq y$ in $B \cap L$, we obtain

$$\sum_{i=1}^{m} l_i = q^2+q+1$$

$$\sum_{i=1}^{m} i \cdot l_i = |B|(q+1) \qquad \text{every point lies in } q+1 \text{ lines}$$

$$\sum_{i=1}^{m} i(i-1)l_i = |B|(|B|-1) \qquad \text{two points lie on exactly one line.}$$

Since $m \leq \sqrt{q}+1$, we have that $i-\sqrt{q}-1 \leq 0$ for all $i = 1,\ldots,m$. This, in particular, implies that

$$\begin{aligned}0 &\geq \sum_{i=1}^{m}(i-1)(i-\sqrt{q}-1)l_i \\ &= \sum_{i=1}^{m} i(i-1)l_i - (\sqrt{q}+1)\sum_{i=1}^{m} i\cdot l_i + (\sqrt{q}+1)\sum_{i=1}^{m} l_i \\ &= |B|(|B|-1) - (\sqrt{q}+1)|B|(q+1) + (\sqrt{q}+1)(q^2+q+1) \\ &= [|B| - (q+\sqrt{q}+1)]\cdot[|B| - (q\sqrt{q}+1)]\,.\end{aligned}$$

Since $|B| \leq q+\sqrt{q}+1$ (by our assumption), the second term is negative, implying that the first one cannot be negative, that is, $|B| \geq q+\sqrt{q}+1$, as desired. □

12.5 Resolvable designs

Suppose $\mathcal{D}$ is a (v, k, λ) design over a set X. A *parallel class* in $\mathcal{D}$ is a subset of disjoint blocks from $\mathcal{D}$ whose union is X. Observe that a parallel class contains v/k blocks, and every point of X appears in exactly one of these blocks. Moreover, by (12.1) and (12.2), we have

$$r = |\mathcal{D}|k/v = \lambda(v-1)/(k-1)$$

such classes, where r is the replication number of $\mathcal{D}$ (the number of blocks containing a given point). A partition of $\mathcal{D}$ into r parallel classes is called a *resolution*, and a design is said to be *resolvable* if it has at least one resolution.

Let us consider the following example from Anderson and Honkala (1997). We have a football league of $2n$ teams and each team plays exactly once against every other team. We wish to arrange the league schedule so that all the matches are played during $2n-1$ days, and on each day every team plays one match. Is this possible?

What we are looking for is a resolvable $(2n, 2, 1)$ design. For convenience, let our ground set (of teams) be $X = \{*, 1, \ldots, 2n-1\}$, where $*$ is some symbol different from $1, \ldots, 2n-1$. Since by (12.1), the replication number equals

$$r = \lambda(v-1)/(k-1) = 1\cdot(2n-1)/(2-1) = 2n-1,$$

we have to show how to partite the collection $\mathcal{D}$ of all 2-element subsets of X into $2n-1$ parallel classes $\mathcal{D}_1, \ldots, \mathcal{D}_{2n-1}$; the i-th class $\mathcal{D}_i$ gives us the set of matches played at the i-th day.

Define $\{i, *\} \in \mathcal{D}_i$ for all $i \in X \setminus \{*\}$, and $\{a, b\} \in \mathcal{D}_i$, if

$$a + b \equiv 2i \bmod 2n-1$$

for $a, b \in X \setminus \{*\}$. Since $2n-1$ is odd, each 2-element subset of X belongs to a unique $\mathcal{D}_i$; and the unique block in $\mathcal{D}_i$ containing an element $a \in X$ is $\{a, b\}$ where $b \equiv 2i - a \bmod 2n-1$ if $a \neq i$, and $\{i, *\}$ if $a = i$.

12.5.1 Affine planes

An *affine plane* $\mathrm{AG}(2, q)$ of order q is a $(q^2, q, 1)$ design. By (12.1), each point of this plane belongs to $r = (q^2-1)/(q-1) = q+1$ lines, and by (12.2), we have $b = vr/k = q^2 + q$ lines altogether. Put otherwise, an affine plane of order q has q^2 points and satisfies the following conditions:

- every line has q points;
- any two points lie on a unique line;
- any point lies on $q+1$ lines;
- there are $q^2 + q$ lines.

Hence, the main difference from projective planes is that now we can have "parallel" lines, i.e., lines which do not meet each other.

There are two basic constructions of affine planes.

Construction 1. An affine plane can be obtained from a projective plane by removing any one of its lines. Let $(X, \mathcal{L})$ be a projective plane of order q. Fix one of its lines $L_0 \in \mathcal{L}$ and consider the design $(X', \mathcal{L}')$ where $X' = X \setminus L_0$ and $\mathcal{L}' = \{L \setminus L_0 : L \in \mathcal{L},\ L \neq L_0\}$ (see Fig. 12.2). It is easy to verify (Exercise 12.11) that the obtained design $(X', \mathcal{L}')$ is an affine plane of order q. The line L_0 is called the *line at infinity*. For each line $L' \in \mathcal{L}'$ of the affine plane there is a unique point $x \in L_0$ such that $L' \cup \{x\} \in \mathcal{L}$; this point is called the *infinite point* of L'.

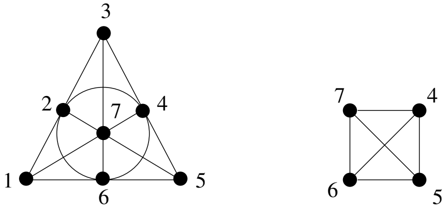


Fig. 12.2 Construction 1 applied to the Fano plane; $L_0 = \{1, 2, 3\}$ is the removed line (the line at infinity).

Construction 2. Let q be a prime power, and consider the set of points $X = \mathbb{F}_q \times \mathbb{F}_q$. Let $\mathcal{D}$ be the set of all blocks of the form

$$L(a, b) := \{(x, y) \in X \,:\, y = ax + b\}$$

and

$$L(c) := \{(c, y) \,:\, y \in \mathbb{F}_q\},$$

where $a, b, c \in \mathbb{F}_q$. We will show that $\mathcal{D}$ is a $(q^2, q, 1)$ design. Clearly, there are q^2 points and each block contains exactly q of them. Hence, we only need to show that every pair of points $(x_1, y_1), (x_2, y_2) \in \mathbb{F}_q \times \mathbb{F}_q$ is contained in a unique block. If $x_1 = x_2$ then the unique block containing this pair is $L(x_1)$. If $x_1 \neq x_2$ then the system of equations $y_1 = ax_1 + b$, $y_2 = ax_2 + b$ has a unique solution (a, b); hence, the unique block containing that pair is $L(a, b)$, and we are done.

It can be shown that affine planes are resolvable designs (see Exercise 12.12): the parallel classes are $\{L \setminus \{x\} : x \in L\}$ for $x \in L_0$, in Construction 1, and $\{L(c) : c \in \mathbb{F}_q\}$ together with $\{L(a, b) : b \in \mathbb{F}_q\}$ for all $a \in \mathbb{F}_q$, in Construction 2.

Exercises

12.1. Let $\mathcal{D} = \{A_1, \ldots, A_m\}$ be a (v, k, λ) design over some set X with $|X| = v$. Use Fisher's inequality (Theorem 7.5) to show that $|\mathcal{D}| \geq v$. *Hint*: Consider the sets $S_x = \{i : x \in A_i\}$ and show that $|S_x \cap S_y| = \lambda$ for all $x \neq y \in X$.

12.2. Let $\mathcal{D} \subseteq 2^X$ be a (v, k, λ) design with b blocks, and let r be its replication number (i.e., each element occurs in r blocks). Prove that its complement $\overline{\mathcal{D}} := \{X \setminus B : B \in \mathcal{D}\}$ is a $(v, v-k, b-2r+\lambda)$ design provided that $b - 2r + \lambda > 0$. *Hint*: A pair of elements $x \neq y$ is contained in $X \setminus B$ if and only if B contains neither x nor y. The number of blocks of $\mathcal{D}$ containing neither x nor y is $b - 2r + \lambda$ by the principle of inclusion and exclusion.

12.3. Show that the number b of blocks in a t–(v, k, λ) design is given by $b = \lambda\binom{v}{t}/\binom{k}{t}$. *Hint*: Count in two ways the number of pairs (T, B) where T is a t-element set of points and B is a block.

12.4. Construct a projective plane $\mathrm{PG}(2, q)$ of order $q = 3$.

12.5. Show that, in a projective plane of order q, its lines are the only blocking sets of size $q + 1$. *Hint*: See Exercise 7.7.

12.6. From the previous example we know that no set of q points intersects all the lines. We can ask the dual question: are there sets of size q that intersect every *non-trivial* blocking set? Show that there are no such sets and that the only sets of size $q + 1$ that intersect every blocking set are the lines. *Sketch*: (due to Blokhuis): Take any set S of q points, and start with a line L_0 not intersecting this set. Delete one (suitable) point $x \in L_0$ from that line, and add a point $y_i \in L_i \setminus S$ on every other line L_i through the deleted point x. The resulting set $(L_0 \setminus \{x\}) \cup \{y_1, \ldots, y_q\}$ is blocking and is disjoint from S. This blocking set might be trivial (i.e., contain a line), but it can be made non-trivial by deleting some unnecessary point of the line we started with.

12.7. (due to Bruen). Let S be a nontrivial blocking set in a projective plane of order q. Show that:

(i) $|S| \leq q^2 - \sqrt{q}$, and
(ii) no line contains more than $|S| - q$ points of S.

Hint: To (i): observe that the complement of S is also a nontrivial blocking set, and apply Bruen's theorem. To (ii): take a line L and a point $x \in L \setminus S$; there are q other lines through x and these lines intersect S; argue that then $|L \cap S| + q \leq |S|$.

12.8. Let S be a set of $q + 2$ points in a projective plane of order q. Prove that every line, that meets S, meets it twice.

12.9. Let S be a set of points in a projective plane of order q. Suppose that no three points of S are colinear (i.e., lie on a line). Prove that then $|S| \leq q+1$ if q is odd, and $|S| \leq q + 2$ if q is even. *Hint*: Fix a point $x \in S$; for each other point $y \in S$ the pair x, y lies on one of $q + 1$ lines (containing x), and these lines must be different for different y. This proves the odd case. For the even case show that, if $|S| = q + 2$ then every line that meets S, meets it twice.

12.10. Color some q points of a projective plane of order q in red, and the rest in blue. Prove that, for any two different sets $A \neq B$ of red points, there is a set C of q blue points such that $A \cup C$ is a blocking set but $B \cup C$ avoids at least one line. *Hint*: Take a point $x \in A \setminus B$, and show that some two lines L_1 and L_2 meet in the (red) point x and have no more red points; take $C = L_1 \setminus \{x\}$.

12.11. Take a projective plane of order q, i.e., a design satisfying the conditions (P1)–(P5), and apply the first construction from Sect. 12.5.1 to it. Show that the resulting design satisfies the conditions (A1)–(A4).

12.12. Consider a $(q^2, q, 1)$ design, i.e., an affine plane of order q. Show that this design is resolvable. More generally, let a *parallel class* be a set of mutually disjoint lines, and show the following:

(i) each parallel class contains q lines;
(ii) there are $q + 1$ such classes;
(iii) any two lines from different classes meet in a point;
(iv) lines of each parallel class cover the whole point set.

Hint: Each parallel class contains exactly one line through any point; so, the $q + 1$ lines through a point x contain representatives of all the classes.

Part III
The Linear Algebra Method

13. The Basic Method

The general framework for the *linear algebra method* in combinatorics is the following: if we want to come up with an upper bound on the size of a set of objects, associate them with elements in a vector space V of relatively low dimension, and show that these elements are linearly independent; hence, we cannot have more objects in our set than the dimension of V.

13.1 The linear algebra background

A *field* is a set $\mathbb{F}$ closed under addition, subtraction, multiplication and division by nonzero element. By addition and multiplication, we mean commutative and associative operations which obey distributive laws. The additive identity is called zero, and the multiplicative identity is called unity. Examples of fields are reals $\mathbb{R}$, rationals $\mathbb{Q}$, and integers modulo a prime p. We will be mostly concerned with finite fields. The cardinality of a finite field must be a power of a prime and all finite fields with the same number of elements are isomorphic. Thus, for each prime power q there is essentially one field $\mathbb{F}$ with $|\mathbb{F}| = q$. This field is usually denoted as $GF(q)$ or $\mathbb{F}_q$.

A *linear space* (or *vector space*) V over a field $\mathbb{F}$ is an additive Abelian group $(V, +, \mathbf{0})$ closed under (left) multiplication by elements of $\mathbb{F}$ (called *scalars*). It is required that this multiplication is distributive with respect to addition in both V and $\mathbb{F}$, and associative with respect to multiplication in $\mathbb{F}$. Elements of V are called *vectors* or *points*. Standard examples of vector spaces are subsets $V \subseteq \mathbb{F}^n$ of vectors closed under the component-wise addition $u + v = (u_1 + v_1, \ldots, u_n + v_n)$ and multiplication by scalars $\lambda v = (\lambda v_1, \ldots, \lambda v_n)$, $\lambda \in \mathbb{F}$.

A *linear combination* of the vectors $v_1, \ldots, v_m$ is a vector of the form $\lambda_1 v_1 + \ldots + \lambda_m v_m$ with $\lambda_i \in \mathbb{F}$. A *subspace* of V is a nonempty subset of V, closed under linear combinations. The *span* of $v_1, \ldots, v_m$, denoted by span$\{v_1, \ldots, v_m\}$, is the set of all linear combinations of these vectors. A

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_13,

vector u *depends* on the vectors $v_1, \ldots, v_m$ if $u \in \mathrm{span}\{v_1, \ldots, v_m\}$. The vectors $v_1, \ldots, v_m$ are *linearly independent* if none of them is dependent on the rest. Equivalently, the (in)dependence can be defined as follows.

A *linear relation* among the vectors $v_1, \ldots, v_m$ is a linear combination that gives the zero vector:

$$\lambda_1 v_1 + \ldots + \lambda_m v_m = \mathbf{0}.$$

This relation is nontrivial if $\lambda_i \neq 0$ for at least one i. It is easy to see that the vectors $v_1, \ldots, v_m$ are linearly independent if and only if no nontrivial relation exists between them. A *basis* of V is a set of independent vectors which spans V. A fundamental fact in linear algebra says that *any two bases of V have the same cardinality*; this number is called the *dimension*, $\dim V$, of V.

A further basic fact is the so-called *linear algebra bound* (see any standard linear algebra book for the proof):

Proposition 13.1. *If $v_1, \ldots, v_k$ are linearly independent vectors in a vector space of dimension m then $k \leq m$.*

An important operation in vector spaces is the scalar product of two vectors. Given two vectors $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$, their *scalar product* $\langle u, v \rangle$ (also called *inner product* and denoted $u \cdot v$) is defined by:

$$\langle u, v \rangle = u^\top \cdot v := u_1 v_1 + \cdots + u_n v_n.$$

Vectors u and v are *orthogonal* if $\langle u, v \rangle = 0$; in this case one also writes $u \perp v$. If $U \subseteq V$ is a subspace of V then the *dual* (or *orthogonal complement*) is the subspace

$$U^\perp = \{v \in V \ : \ \langle u, v \rangle = 0 \text{ for all } u \in U\}.$$

The following useful equality connects the dimensions of two orthogonal subspaces.

Proposition 13.2. *Let V be a finite dimensional linear space, and $U \subseteq V$ be a subspace. Then* $\dim U + \dim U^\perp = \dim V$.

A consequence of this is that, for every linear subspace $U \subseteq \mathbb{R}^n$ and every vector $x \in U$, there are uniquely defined vectors $u \in U$ and $w \in U^\perp$ such that $x = u + w$. The vector u is then called the *projection* of x onto U.

If $A = (a_{ij})$ is an m-by-n matrix over some field $\mathbb{F}$ and x is a vector in $\mathbb{F}^m$, then $x^\top \cdot A$ is the vector in $\mathbb{F}^n$ whose j-th coordinate is the scalar product of x with the j-th *column* of A. Thus, the rows of A are linearly independent if and only if $x^\top \cdot A \neq \mathbf{0}$ for all $x \neq \mathbf{0}$. Similarly, if $y \in \mathbb{F}^n$, then $A \cdot y$ is the vector in $\mathbb{F}^m$ whose i-th coordinate is the scalar product of y with the i-th *row* of A.

The *column rank* of a matrix A is the dimension of the vector space spanned by its columns. The *row rank* of A is the dimension of the vector space spanned by its rows. One of the first nontrivial results in matrix theory asserts that the *row and column ranks are equal*; this common value is the rank of A, denoted by $\mathrm{rk}(A)$. There are several equivalent definitions of the rank of an m-by-n matrix $A = (a_{ij})$ over a given field $\mathbb{F}$:

(a) $\mathrm{rk}(A)$ is the smallest r such that $A = B \cdot C$ for some m-by-r matrix B and r-by-n matrix C;
(b) $\mathrm{rk}(A)$ is the smallest r such that A can be written as a sum of r rank-1 matrices;
(c) $\mathrm{rk}(A)$ is the smallest r such that A is a matrix of scalar products of vectors in $\mathbb{F}^r$: there exist vectors $u^1, \ldots, u^m$ and $v^1, \ldots, v^n$ in $\mathbb{F}^r$ such that $a_{ij} = \langle u^i, v^j \rangle$.

Usually, the underlying field $\mathbb{F}$ will be clear from the context. If the field still needs to be specified, we will write $\mathrm{rk}_{\mathbb{F}}(A)$ instead of $\mathrm{rk}(A)$.

The following inequalities hold for the rank:

$$\mathrm{rk}(A) - \mathrm{rk}(B) \le \mathrm{rk}(A + B) \le \mathrm{rk}(A) + \mathrm{rk}(B); \tag{13.1}$$

$$\mathrm{rk}(A) + \mathrm{rk}(B) - n \le \mathrm{rk}(AB) \le \min\{\mathrm{rk}(A), \mathrm{rk}(B)\}\ , \tag{13.2}$$

if A is an m-by-n and B an n-by-k matrix.

The *determinant*, $\det(A)$, of an $n \times n$ matrix $A = (a_{ij})$ is the sum of $n!$ signed products $\pm a_{1i_1} a_{2i_2} \cdots a_{ni_n}$, where $(i_1, i_2, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$, the sign being $+1$ or -1, depending on whether the number of *inversions* of $(i_1, i_2, \ldots, i_n)$ is even or odd; an inversion occurs when $i_r > i_s$ but $r < s$. It can be shown (do this!) that $\det(A) \neq 0$ implies $\mathrm{rk}(A) = n$.

If $x = (x_1, \ldots, x_n)$ denotes the vector of indeterminates, and b is a vector in $\mathbb{F}^m$, then the matrix equation $A \cdot x = b$ is a concise form of writing a system of m linear equations in variables $x_1, \ldots, x_n$:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \qquad (i = 1, \ldots, m).$$

We have the following useful criterion for such a system being solvable. Let $a_1, \ldots, a_n \in \mathbb{F}^m$ denote the columns of A. Observe that $A \cdot x = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$. It follows that the set $\{A \cdot x \,:\, x \in \mathbb{F}^n\}$ is a *columns space* of A, i.e., the set of all vectors spanned by the columns of A. The system $A \cdot x = b$ is thus solvable if and only if $b \in \mathrm{span}\{a_1, \ldots, a_n\}$ or, equivalently, if and only if $\mathrm{rk}(A) = \mathrm{rk}([A|b])$, where $[A|b]$ denotes the $m \times (n+1)$ matrix obtained by adding the column b to A. A system $A \cdot x = b$ is *homogeneous* if $b = \mathbf{0}$. The set of solutions of $A \cdot x = \mathbf{0}$ is clearly a subspace (of all vectors that are orthogonal to all the rows of A) and, by Proposition 13.2, its dimension is $n - \mathrm{rk}(A)$. We summarize this important result:

Proposition 13.3. *Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Then the set of solutions of the system of linear equations $A \cdot x = \mathbf{0}$ is a linear subspace of dimension $n - \mathrm{rk}(A)$ of the space $\mathbb{F}^n$.*

This subspace is called the *kernel* of A.

The *norm* (or *length*) of a vector $v = (v_1, \dots, v_n)$ in $\mathbb{R}^n$ is the number

$$\|v\| := \langle v, v\rangle^{1/2} = \Big(\sum_{i=1}^{n} v_i^2\Big)^{1/2}.$$

The following basic inequality, known as the *Cauchy–Schwarz inequality*, estimates the scalar product of two vectors in terms of their norms (we have already used it in previous sections; now we will prove it):

Proposition 13.4. *For any real vectors $u, v \in \mathbb{R}^n$,*

$$|\,\langle u, v\rangle\,| \le \|u\| \cdot \|v\|$$

with an equality iff u and v are linearly dependent.

When expressed explicitly, this inequality turns to:

$$\Big(\sum_{i=1}^{n} u_i v_i\Big)^2 \le \Big(\sum_{i=1}^{n} u_i^2\Big)\Big(\sum_{i=1}^{n} v_i^2\Big). \tag{13.3}$$

Proof. We may assume that $u \ne \mathbf{0}$. For any constant $\lambda \in \mathbb{R}$ we have

$$\begin{aligned} 0 \le \langle \lambda u - v, \lambda u - v\rangle &= \langle \lambda u, \lambda u - v\rangle - \langle v, \lambda u - v\rangle \\ &= \lambda^2\langle u, u\rangle - 2\lambda\langle u, v\rangle + \langle v, v\rangle. \end{aligned}$$

Substituting $\lambda = \frac{\langle u,v\rangle}{\langle u,u\rangle}$ we get

$$0 \le \frac{\langle u, v\rangle^2}{\langle u, u\rangle^2}\langle u, u\rangle - 2\frac{\langle u, v\rangle^2}{\langle u, u\rangle} + \langle v, v\rangle = \langle v, v\rangle - \frac{\langle u, v\rangle^2}{\langle u, u\rangle}$$

Rearranging the last inequality, we get $\langle u, v\rangle^2 \le \langle u, u\rangle\langle v, v\rangle = \|u\|^2 \cdot \|v\|^2$. □

A scalar λ is an *eigenvalue* of a square real matrix A if the equation $Ax = \lambda x$ has a solution $x \in \mathbb{R}^n$, $x \ne \mathbf{0}$, which is the case iff the *characteristic polynomial* $p_A(z) = \det(A - zI)$ has λ as a root; here, I is a unit matrix with 1s on the diagonal, and 0s elsewhere. A non-zero x with $Ax = \lambda x$ is called an *eigenvector* corresponding to the eigenvalue λ. Since p_A has degree n, we can have at most n (complex) eigenvalues. If the matrix A is symmetric, that is, $A^\top = A$, then all its eigenvalues are real numbers.

The following are standard facts about the eigenvalues of a real symmetric $n \times n$ matrix $A = (a_{ij})$:

1. A has exactly n (not necessarily distinct) real eigenvalues $\lambda_1 \ge \ldots \ge \lambda_n$.
2. There exists a set of n eigenvectors $x_1, \dots, x_n$, one for each eigenvalue, that are normalized and mutually orthogonal, that is, $\|x_i\| = 1$ and $\langle x_i, x_j\rangle = 0$ over the reals. Hence, $x_1, \dots, x_n$ form an *orthonormal* basis of $\mathbb{R}^n$.

3. The rank of A is equal to the number of its nonzero eigenvalues, including multiplicities: $\mathrm{rk}(A) = |\{i : \lambda_i \neq 0\}|$.
4. The sum of all eigenvalues $\sum_{i=1}^n \lambda_i$ is equal to the trace $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$.
5. The product $\prod_{i=1}^n \lambda_i$ of all eigenvalues is equal to $\det(A)$.
6. Perron–Frobenius theorem: If $A = (a_{ij})$ is a real $n \times n$ matrix with nonnegative entries $a_{ij} \geq 0$ and irreducible, then there is a real eigenvalue r of A such that
$$\min_i \sum_j a_{ij} \leq r \leq \max_i \sum_j a_{ij}$$
and any other eigenvalue λ satisfies $|\lambda| \leq r$. A matrix *is reducible* if there is a subset $I \subseteq [n]$ such that $a_{ij} = 0$ for all $i \in I$ and $j \notin I$. In particular, an adjacency matrix of a graph is irreducible iff the graph is connected.

There is also a general formula to compute eigenvalues explicitly. A *weighted average* of a sequence $x = (x_1, \ldots, x_n)$ of numbers is a number

$$\sum_{i=1}^n a_i x_i \qquad \text{with} \qquad \sum_{i=1}^n a_i = 1\,.$$

We will use the following easy fact (prove it!): For every sequence x, any weighted average of x is $\geq \min_i x_i$ and is $\leq \max_i x_i$.

Theorem 13.5. *The k-th largest eigenvalue of a symmetric $n \times n$ matrix A is*

$$\lambda_k = \max_{\dim U = k} \min_{x \in U} \frac{x^\top A x}{x^\top x} \tag{13.4}$$

$$= \min_{\dim U = k-1} \max_{x \perp U} \frac{x^\top A x}{x^\top x}\,. \tag{13.5}$$

Here, the maximum/minimum is over all subspaces U of a given dimension, and over all nonzero vectors x in the respective subspace. In particular, (13.5) yields:

$$\lambda_1 = \max_{x \neq \mathbf{0}} \frac{x^\top A x}{x^\top x} = \max_{\|x\|=1} x^\top A x \tag{13.6}$$

and

$$\lambda_2 = \max_{x \perp \mathbf{1}} \frac{x^\top A x}{x^\top x} = \max_{x \perp \mathbf{1}, \|x\|=1} x^\top A x\,, \tag{13.7}$$

where $\mathbf{1}$ is the all-1 vector, and the second equality follows since we can replace x by $x/\|x\|$, since the first maximum is over all nonzero vectors x.

Proof. We only prove the first equality (13.4)—the proof of the second one is analogous. First of all, note that the quantity (known as the *Rayleigh quotient*)

$$f_A(x) = \frac{x^\top A x}{x^\top x}$$

is invariant under replacing x by any nonzero multiple cx. Therefore, we can assume that x is a unit vector, that is, $\|x\| = 1$ and hence $x^\top x = 1$ (by replacing $x \mapsto cx$ with $c = 1/\|x\|$, if necessary).

Consider an orthonormal basis of eigenvectors $u^1, \ldots, u^n$. Any vector x in $\mathbb{R}^n$ can be written as $x = \sum_{i=1}^n a_i u^i$, and the expression $x^\top Ax$ reduces to

$$x^\top Ax = \Big(\sum_{i=1}^n a_i u^i\Big)^\top A\Big(\sum_{i=1}^n a_i u^i\Big) = \sum_{i,j=1}^n \langle u^i, \lambda_j u^j\rangle = \sum_{i=1}^n a_i^2 \lambda_i\,,$$

where the last equality follows because the scalar product of vectors u^i and u^j is 1 if $i = j$, and is 0 otherwise. By a similar argument, we have that $x^\top x = \sum_{i=1}^n a_i^2$, and for unit vector x we get $\sum_{i=1}^n a_i^2 = 1$. Thus, for each unit vector x, the expression $f_A(x)$ can be interpreted as a weighted average of the eigenvalues.

Now consider the subspace $U \subseteq \mathbb{R}^n$ generated by the first k eigenvectors $u^1, \ldots, u^k$. For any unit vector $x \in U$, we get $x^\top Ax = \sum_{i=1}^k a_i^2 \lambda_i$ and $\sum_{i=1}^k a_i^2 = 1$. The weighted average $f_A(x)$ of the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_k$ is at least the smallest of the first k eigenvalues, so $\min_{x\in U} f_A(x) \geq \lambda_k$ holds for this special k-dimensional subspace U.

On the other hand, consider any subspace U of dimension k, and a subspace V of dimension $n-k+1$ generated by the last $n-k+1$ eigenvectors $u^k, u^{k+1} \ldots, u^n$. These two subspaces must have nontrivial intersection, that is, there must exist a nonzero vector $z \in U \cap V$. By normalization, we can assume that $z = \sum_{j=k}^n b_j u^j$ is a unit vector, $z^\top z = \sum_{j=k}^n b_j^2 = 1$ and we obtain $f_A(z) = \sum_{j=k}^n b_j^2 \lambda_j \leq \lambda_k$, since $f_A(z)$ is a weighted average of the last $n-k+1$ eigenvalues and the largest of these eigenvalues is λ_k. Consequently, $\min_{x\in U} f_A(x) \leq f_A(z) \leq \lambda_k$ holds for any k-dimensional subspace U. □

The *spectral norm* of a matrix A is defined as

$$\|A\| := \max_{x\neq\mathbf{0}} \frac{\|Ax\|}{\|x\|}\,.$$

The name "spectral norm" comes from the fact that

$$\|A\| = \text{ square root of the largest eigenvalue of } A^\top A.$$

This holds because $x^\top(A^\top A)x = \langle Ax, Ax\rangle = \|Ax\|^2$. The *Frobenius norm* of A is just the Euclidean norm

$$\|A\|_{\mathrm{F}} := \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2}$$

of the corresponding vector of length n^2. The following fact relates these two norms with the rank over the reals.

Proposition 13.6. *For every real matrix A,*

$$\frac{\|A\|_{\mathrm{F}}}{\sqrt{\mathrm{rk}(A)}} \leq \|A\| \leq \|A\|_{\mathrm{F}} \,.$$

Proof. Observe that $\|A\|_{\mathrm{F}}^2$ is equal to the trace, that is, the sum of diagonal elements of the matrix $B = A^{\top}A$. On the other hand, the trace of any real matrix is equal to the sum of its eigenvalues. Hence, $\|A\|_{\mathrm{F}}^2 = \sum_{i=1}^{n} \lambda_i$ where $\lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues of B. Since B has only $\mathrm{rk}(B) = \mathrm{rk}(A) = r$ non-zero eigenvalues, and since all eigenvalues of B are nonnegative, the largest eigenvalue λ_1 is bounded by $\|A\|_{\mathrm{F}}^2/r \leq \lambda_1 \leq \|A\|_{\mathrm{F}}^2$. It remains to use the fact mentioned above that $\|A\| = \sqrt{\lambda_1}$. □

Let us now see how the linear algebra argument works in concrete situations.

13.2 Graph decompositions

A bipartite clique is a bipartite complete graph $K_{A,B} = (A \cup B, E)$ with $A \cap B = \emptyset$ and $E = A \times B$.

Let $f(n)$ be the smallest number t such that the complete graph K_n on n vertices $1, 2, \ldots, n$ can be decomposed into t edge-disjoint bipartite cliques. It is not difficult to see that $f(n) \leq n - 1$. Indeed, it is enough to remove the vertices $1, 2, \ldots, n-1$ one-by-one, together with their incident edges. This gives us a decomposition of K_n into edge-disjoint stars, that is, bipartite cliques K_{A_i,B_i} with $A_i = \{i\}$ and $B_i = \{i+1, \ldots, n\}$, $i = 1, \ldots, n-1$.

This is, however, just one special decomposition and does not exclude better ones. Still, a classical result of Graham and Pollak (1971) says that the trivial decomposition is in fact the best one! This can be shown using linear algebra.

Theorem 13.7. *The edges of K_n cannot be decomposed into fewer than $n-1$ edge-disjoint biartite cliques.*

Proof (due to Trevberg 1982). We consider a more general question: What is the smallest number t such that the sum of products

$$S(x) := \sum_{1 \leq i < j \leq n} x_i x_j$$

in indeterminates $x = (x_1, \ldots, x_n)$ can be written as the sum

$$S(x) = \sum_{i=1}^{t} \Big(\sum_{j \in A_i} x_j\Big) \cdot \Big(\sum_{j \in B_i} x_j\Big) = \sum_{i=1}^{t} L_i(x) \cdot R_i(x)$$

of products-of-sums with $A_i \cap B_i = \emptyset$ for all $i = 1, \ldots, t$? To answer this question, set $T(x) := \sum_{i=1}^n x_i^2$ and observe that

$$\Bigl(\sum_{i=1}^n x_i\Bigr)^2 = \sum_{i=1}^n x_i^2 + 2\sum_{i<j} x_i x_j = T(x) + 2S(x)\,,$$

and hence,

$$T(x) = \Bigl(\sum_{i=1}^n x_i\Bigr)^2 - 2S(x) = \Bigl(\sum_{i=1}^n x_i\Bigr)^2 - 2 \cdot \sum_{i=1}^t L_i(x) \cdot R_i(x)\,. \qquad (13.8)$$

Consider now a homogeneous system of $t+1$ linear equations over $\mathbb{R}$:

$$L_1(x) = 0\,,\ \ldots\,,\ L_t(x) = 0\,,\ x_1 + \cdots + x_n = 0$$

and assume that $t \leq n-2$. Then the system has more variables than equations, implying that it must have a solution $x \in \mathbb{R}^n$ with $x \neq \mathbf{0}$. From $\sum_{i=1}^n x_i = 0$ and $L_i(x) = 0$ for all $i = 1, \ldots, t$ it follows that, for this vector x, the right-hand side of (13.8) must be equal to 0. But the left-hand side is not equal to 0, since $x \neq \mathbf{0}$ implies $T(x) = \sum_{i=1}^n x_i^2 \neq 0$. Thus, our assumption that $t \leq n-2$ has led to a contradiction. □

13.3 Inclusion matrices

A celebrated result, due to Razborov (1987), says that the majority function cannot be computed by constant depth circuits of polynomial size, even if we allow unbounded fanin And, Or and Parity functions as gates. This result was obtained in two steps:

(i) show that functions, computable by small circuits, can be approximated by low degree polynomials, and
(ii) prove that the majority function is hard to approximate by such polynomials.

The proof of (i) is probabilistic, and we will present it later (see Lemma 18.11). The proof of (ii) employs the linear algebra argument, and we present it below.

The *k-threshold function* is a boolean function $T_k^n(x_1, \ldots, x_n)$ which outputs 1 if and only if at least k of the bits in the input vector are 1. A boolean function $g(x_1, \ldots, x_n)$ is a polynomial of degree d over $\mathbb{F}_2$ if it can be written as a sum modulo 2 of products of at most d variables.

Lemma 13.8 (Razborov 1987). *Let $n/2 \leq k \leq n$. Every polynomial of degree at most $2k - n - 1$ over $\mathbb{F}_2$ differs from the k-threshold function on at least $\binom{n}{k}$ inputs.*

Proof (due to Lovász–Shmoys–Tardos 1995). Let g be a polynomial of degree $d \leq 2k - n - 1$ over $\mathbb{F}_2$ and let U denote the set of all vectors where it differs from T_k^n. Let A denote the set of all 0-1 vectors of length n containing exactly k ones. By our choice of d, the coordinate-wise And $a \wedge b$ of any two vectors $a, b \in A$ must contain at least $d + 1$ ones.

Consider the 0-1 matrix $M = (m_{a,u})$ whose rows are indexed by the members of A, columns are indexed by the members of U, and $m_{a,u} = 1$ if and only if $a \geq u$. For two vectors a and b we denote by $a \wedge b$ the coordinate-wise And of these vectors. Our goal is to prove that the columns of M span the whole linear space; since the dimension of this space is $|A| = \binom{n}{k}$, this will mean that we must have $|U| \geq \binom{n}{k}$ columns.

The fact that the columns of M span the whole linear space follows directly from the following claim saying that every unit vector lies in the span.

Claim 13.9. Let $a \in A$ and $U_a = \{u \in U : m_{a,u} = 1\}$. Then, for every $b \in A$,

$$\sum_{u \in U_a} m_{b,u} = \begin{cases} 1 & \text{if } b = a; \\ 0 & \text{if } b \neq a. \end{cases}$$

Proof. By the definition of U_a, we have (all sums are over $\mathbb{F}_2$):

$$\sum_{u \in U_a} m_{b,u} = \sum_{\substack{u \in U \\ u \leq a \wedge b}} 1 = \sum_{x \leq a \wedge b} (T_k^n(x) + g(x)) = \sum_{x \leq a \wedge b} T_k^n(x) + \sum_{x \leq a \wedge b} g(x).$$

The second term of this last expression is 0, since $a \wedge b$ has at least $d+1$ ones (Exercise 13.10). The first term is also 0 except if $a = b$.

This completes the proof of the claim, and thus, the proof of the lemma. □

13.4 Disjointness matrices

Let $k \leq n$ be natural numbers, and X be a set of n elements. A *k-disjointness matrix* over X is a 0-1 matrix $D = D(n,k)$ whose rows and columns are labeled by subsets of X of size at most k; the entry $D_{A,B}$ in the A-th row and B-th column is defined by:

$$D_{A,B} = \begin{cases} 0 & \text{if } A \cap B \neq \emptyset, \\ 1 & \text{if } A \cap B = \emptyset. \end{cases}$$

This matrix plays an important role in computational complexity. Its importance stems from the fact that it has full rank over $\mathbb{F}_2$, i.e., all its $\sum_{i=0}^{k} \binom{n}{i}$ rows are linearly independent.

Theorem 13.10. *The k-disjointness matrix $D = D(n,k)$ has full rank over $\mathbb{F}_2$, that is,*

$$\mathrm{rk}_{\mathbb{F}_2}(D) = \sum_{i=0}^{k} \binom{n}{i}.$$

There are several proofs of this result. Usually, it is derived from more general facts about Möbius inversion or general intersection matrices. Here we present one particularly simple and direct proof due to Razborov (1987).

Proof. Let $N = \sum_{i=0}^{k} \binom{n}{i}$. We must show that the rows of D are linearly independent over $\mathbb{F}_2$, i.e., that for any nonzero vector $\lambda = (\lambda_{I_1}, \lambda_{I_2}, \ldots, \lambda_{I_N})$ in $\mathbb{F}_2^N$ we have $\lambda^\top \cdot D \neq 0$. For this, consider the following polynomial:

$$f(x_1, \ldots, x_n) := \sum_{|I| \le k} \lambda_I \prod_{i \in I} x_i.$$

Since $\lambda \neq 0$, at least one of the coefficients λ_I is nonzero, and we can find some I_0 such that $\lambda_{I_0} \neq 0$ and I_0 is *maximal* in that $\lambda_I = 0$ for all $I \supset I_0$. Assume w.l.o.g. that $I_0 = \{1, \ldots, t\}$, and make in the polynomial f the substitution $x_i := 1$ for all $i \notin I_0$. After this substitution has been made, a *nonzero* polynomial over the first t variables $x_1, \ldots, x_t$ remains such that the term $x_1 x_2 \cdots x_t$ is left untouched (here we use the maximality of I_0). Hence, after the substitution we obtain a polynomial which is 1 for some assignment $(a_1, \ldots, a_t)$ to its variables. But this means that the polynomial f itself takes the value 1 on the assignment $b = (a_1, \ldots, a_t, 1, \ldots, 1)$. Hence,

$$1 = f(b) = \sum_{|I| \le k} \lambda_I \prod_{i \in I} b_i.$$

Let $J_0 := \{i : a_i = 0\}$. Then $|J_0| \le k$ and, moreover, $\prod_{i \in I} b_i = 1$ if and only if $I \cap J_0 = \emptyset$, which is equivalent to $D_{I,J_0} = 1$. Thus,

$$\sum_{|I| \le k} \lambda_I D_{I,J_0} = 1,$$

meaning that the J_0-th coordinate of the vector $\lambda^\top \cdot D$ is nonzero. □

In order to apply the linear algebra method, in many situations it is particularly useful to associate sets not to their incidence vectors but to some (multivariate) polynomials $f(x_1, \ldots, x_n)$ and show that these polynomials are linearly independent as members of the corresponding functions space. This idea has found many applications. All these applications are based on the following simple and powerful lemma connecting algebra to linear algebra.

Lemma 13.11 (Independence Criterion)**.** *For $i = 1, \ldots, m$ let $f_i : \Omega \to \mathbb{F}$ be functions and $v_i \in \Omega$ elements such that*

(a) *$f_i(v_i) \neq 0$ for all $1 \le i \le m$;*
(b) *$f_i(v_j) = 0$ for all $1 \le j < i \le m$.*

Then $f_1, \dots, f_m$ are linearly independent members of the space $\mathbb{F}^{\Omega}$.

Proof. By contradiction: Suppose there is a nontrivial linear relation

$$\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_m f_m = 0$$

between the f_i's. Take the largest i for which $\lambda_i \neq 0$. Substitute v_i for the variables. By the assumption, all but the i-th term vanish. What remains is $\lambda_i f_i(v_i) = 0$, which implies $\lambda_i = 0$ because $f_i(v_i) \neq 0$, a contradiction. □

13.5 Two-distance sets

Our first illustration of the independence criterion prepares for some surprisingly powerful applications, which we will consider in Sects. 13.6 and 13.7.

Let $a_1, \dots, a_m$ be points in the n-dimensional Euclidean space $\mathbb{R}^n$. If the pairwise distances of the a_i are all equal then $m \leq n + 1$. (Show this!) But what happens if we relax the condition and require only that the pairwise distances between the a_i take *two* values? Such a set is called *two-distance* set.

We shall see that then m is about $n^2/2$. Indeed, it is easy to construct a two-distance set in $\mathbb{R}^n$ with $\binom{n}{2}$ points (Exercise 13.13). On the other hand, we have the following upper bound.

Theorem 13.12 (Larman–Rogers–Seidel 1977)**.** *Every two-distance set in $\mathbb{R}^n$ has at most $\binom{n}{2} + 3n + 2$ points.*

Proof. Let $a_1, \dots, a_m$ be a two-distance set of distinct points in $\mathbb{R}^n$. The distance between two points x, y in $\mathbb{R}^n$ is $\|x - y\|$. Since for our set of points $a_1, \dots, a_m$ this distance can take only one of two values d_1 or d_2, none of which is zero (why?), it is natural to associate with each point a_i the following polynomial in n real variables $x \in \mathbb{R}^n$:

$$f_i(x) := (\|x - a_i\|^2 - d_1^2) \cdot (\|x - a_i\|^2 - d_2^2).$$

Then $f_i(a_i) = (d_1 d_2)^2 \neq 0$, but $f_i(a_j) = 0$ for every $j \neq i$. By Lemma 13.11, these polynomials are linearly independent (as members of the space of all functions $f : \mathbb{R}^n \to \mathbb{R}$). What is a vector space in which they reside? It is easy to see that every such polynomial is an appropriate linear combination of the following polynomials

$$\Big(\sum_{i=1}^{n} x_i^2\Big)^2, \ \Big(\sum_{i=1}^{n} x_i^2\Big)x_j, \ x_i x_j, \ x_i, \ 1, \quad \text{for } i, j = 1, \dots, n;$$

their number is $1 + n + \big(\binom{n}{2} + n\big) + n + 1 = \binom{n}{2} + 3n + 2$. Thus, the polynomials $f_1, \dots, f_m$ belong to a linear space of dimension at most $\binom{n}{2} + 3n + 2$. As they are linearly independent, their number m cannot exceed the dimension, completing the proof of the theorem. □

We can rewrite the upper bound in Theorem 13.12 as

$$m \leq \binom{n}{2} + 3n + 2 = \binom{n+2}{2} + n + 1.$$

A significant improvement was achieved by Blokhuis (1981) who showed that the second term $n+1$ here is redundant. His trick was to show that the polynomials $f_1, \ldots, f_m$ *together* with the polynomials $x_1, \ldots, x_n, 1$ are linearly independent. This nice idea was later employed to derive more impressing results (cf. Exercise 13.17).

13.6 Sets with few intersection sizes

In this section we demonstrate how the polynomial technique can be used to obtain far reaching extensions of Fisher's inequality (see Theorem 7.5).

Let $\mathcal{F}$ be a family of subsets of some n-element set, and let $L \subseteq \{0, 1, \ldots\}$ be a finite set of integers. We say that $\mathcal{F}$ is *L-intersecting* if $|A \cap B| \in L$ for every pair A, B of distinct members of $\mathcal{F}$.

Suppose we know only the size of L. What can then be said about the number of sets in $\mathcal{F}$? Fisher's inequality tells us that $|\mathcal{F}| \leq n$ when $|L| = 1$. In the case of uniform families, the celebrated result of Ray-Chaudhuri and Wilson (1975) gives the upper bound $|\mathcal{F}| \leq \binom{n}{|L|}$. The non-uniform version of this result was proved by Frankl and Wilson (1981).

Theorem 13.13 (Frankl–Wilson 1981). *If $\mathcal{F}$ is an L-intersecting family of subsets of a set of n elements, then $|\mathcal{F}| \leq \sum_{i=0}^{|L|} \binom{n}{i}$.*

Both these results are best possible: for $L = \{0, 1, \ldots, s-1\}$ one can take the family of all subsets of an n-element set with s elements (with at most s elements, respectively).

The original proof of these theorems used the machinery of higher incidence matrices. Fortunately, these results now admit conceptually simpler proofs using linear spaces of multivariate polynomials.

Proof of Theorem 13.13 (due to Babai 1988). Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ where $|A_1| \leq \ldots \leq |A_m|$. Let $L = \{l_1, \ldots, l_s\}$ be the set of all possible intersection sizes. That is, for every $i \neq j$ there is a k such that $|A_i \cap A_j| = l_k$. With each set A_i we associate its incidence vector $v_i = (v_{i1}, \ldots, v_{in})$, where $v_{ij} = 1$ if $j \in A_i$; otherwise $v_{ij} = 0$. For $x, y \in \mathbb{R}^n$, let (as before) $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$ denote their standard scalar product. Clearly, $\langle v_i, v_j \rangle = |A_i \cap A_j|$.

For $i = 1, \ldots, m$, let us define the polynomial f_i in n variables by

$$f_i(x) = \prod_{k\,:\,l_k<|A_i|} (\langle v_i, x\rangle - l_k) \qquad (x \in \mathbb{R}^n).$$

Observe that $f_i(v_j) = 0$ for all $1 \le j < i \le m$, and $f_i(v_i) \ne 0$ for all $1 \le i \le m$. By Lemma 13.11, the polynomials $f_1, \ldots, f_m$ are linearly independent over $\mathbb{R}$. What is a small vector space in which these polynomials can reside? The f_i's are polynomials of degree at most s, but we can do better. The domain being $\{0,1\}^n$ implies that $x_i^2 = x_i$ for each variable x_i. Thus, pure monomials of degree $\le s$ form a basis (where a pure or multilinear monomial has at most one occurrence of each variable), and we have only $\sum_{i=0}^{s} \binom{n}{i}$ of them. □

Using essentially the same argument, we can also prove the following "modular" version of this theorem (we leave the proof as Exercise 13.16). Write $r \in L \bmod p$ if $r = l \bmod p$ for at least one $l \in L$.

Theorem 13.14 (Deza–Frankl–Singhi 1983). *Let L be a set of integers and p be a prime number. Assume $\mathcal{F} = \{A_1, \ldots, A_m\}$ is a family of subsets of a set of n elements such that*

(a) $|A_i| \notin L \bmod p$ *for all* $1 \le i \le m$*;*
(b) $|A_i \cap A_j| \in L \bmod p$ *for all* $1 \le j < i \le m$.

Then $|\mathcal{F}| \le \sum_{i=0}^{|L|} \binom{n}{i}$.

These theorems and their modifications have found many striking applications in combinatorics and geometry. An excellent exposition is given in the book by Babai and Frankl (1992).

13.7 Constructive Ramsey graphs

Roughly, the main idea of these applications is the following. If we identify the members of our family $\mathcal{F}$ with vertices and join two members if and only if their intersection has a particular size, then the theorems above ensure that the graph cannot have a large clique or a large independent set (or both). To demonstrate the idea, we use it to construct so-called Ramsey graphs.

Recall that a *clique* of size t in a graph is a set of t of its vertices, each pair of which is joined by an edge. Similarly, an *independent set* of size t is a set of t vertices with no edge between them. A graph is a *Ramsey graph* (with respect to t) if it has no clique and no independent set of size t.

Given t, we are interested in the largest possible number n for which such a graph (on n vertices) exists. The existence of Ramsey graphs of size $n = 2^{t/2}$ is known: this was proved by Erdős (1947) using the *probabilistic method* (see Theorem 4.17). The theorem states only the mere *existence* of a graph, and gives no way to find it.

For many years, only an easy construction of a Ramsey graph of size $n = (t-1)^2$ was known: take the disjoint union of $t - 1$ cliques of size

$t-1$ each. The first non-trivial construction of Ramsey graphs on $n = \Omega(t^3)$ vertices was given by Zsigmond Nagy in 1972 (see Exercise 13.15).

Substantial progress in that direction was made by Frankl (1977) who was able to construct Ramsey graphs of super-polynomial size

$$n = t^{\Omega(\ln t/\ln\ln t)}.$$

Subsequently, a simpler proof was found by Frankl and Wilson (1981), using their result about families with one missing intersection size modulo a prime power. To be self-contained, here we present a slightly weaker version whose proof relies only on Theorems 13.13 and 13.14.

The desired graph is defined as follows. Let p be a prime number, and $v = p^3$. Let G_p be a graph whose vertices are subsets of $\{1,\ldots,v\}$ of cardinality p^2-1, and where two vertices A and B are joined by an edge if and only if

$$|A \cap B| \neq -1 \bmod p\,.$$

Theorem 13.15 (Frankl 1977, Frankl–Wilson 1981)**.** *The graph G_p has $n = \binom{v}{p^2-1}$ vertices and has neither a clique nor an independent set on more than $t = \sum_{i=0}^{p-1}\binom{v}{i}$ vertices.*

Proof. If $A_1,\ldots,A_r$ is a clique in G_p then $|A_i \cap A_j| \neq -1 \bmod p$ for every $1 \le i < j \le r$, implying that $|A_i \cap A_j| \in L \bmod p$ for $L = \{0,1,\ldots,p-2\}$. On the other hand, each of the sets A_i has size $p^2-1 = -1 \bmod p$, and hence, $|A_i| \not\in L \bmod p$. Theorem 13.14 implies that in this case $r \le t$.

Now suppose that the sets $A_1,\ldots,A_r$ form an independent set in G_p. Then $|A_i \cap A_j| \in L$, for every $1 \le i < j \le r$, where $L = \{p-1, 2p-1, \ldots, p^2-p-1\}$. Theorem 13.13 again yields that in this case $r \le t$. □

To get the desired lower bound $n \ge t^{\Omega(\ln t/\ln\ln t)}$, we just have to select p appropriately. The exact computation is somewhat tedious, and we leave it as an exercise. We only sketch the way these computations should proceed. By the density of primes, there is always a prime between N and $2N$, for any positive integer N; so we may pretend that p is an integer rather than a prime. Since $v = p^3$, $t = \sum_{i=0}^{p-1}\binom{v}{i} \le p^{O(p)}$ whereas $n = \binom{v}{p^2-1} \ge p^{\Omega(p^2)}$, which is at least $t^{\Omega(\ln t/\ln\ln t)}$ for $p = \Omega(\ln t/\ln\ln t)$.

13.8 Zero-patterns of polynomials

Let $\mathbf{f} = \{f_i(x_1,\ldots,x_n) : i = 1,\ldots,m\}$ be a sequence of polynomials over some field $\mathbb{F}$. A subset $S \subseteq [m]$ is a *zero-pattern* of $\mathbf{f}$ if there exists a vector, a *witness* for this zero-pattern, $v \in \mathbb{F}^n$ such that $S = \{i : f_i(v) \neq 0\}$. Let $Z_{\mathbb{F}}(\mathbf{f})$ denote the number of distinct zero-patterns of $\mathbf{f}$ as v ranges over $\mathbb{F}^n$.

The following upper bound was proved by Rónyai, Babai, and Ganapathy (2001).

Theorem 13.16. *Let d_i denote the degree of f_i, and $D = \sum_{i=1}^{m} d_i$. Then*

$$Z_{\mathbb{F}}(\mathbf{f}) \le \binom{n+D}{n}.$$

Proof. Assume that $\mathbf{f}$ has M different zero-patterns, and let $v_1, \ldots, v_M$ be witnesses to these zero-patterns. Let $S_i = \{k : f_k(v_i) \neq 0\}$ be a zero-pattern witnessed by the i-th vector v_i, and consider the polynomials $g_i := \prod_{k \in S_i} f_k$. We claim that these polynomials are linearly independent over $\mathbb{F}$. This claim completes the proof of the theorem since each g_i has degree at most D and the dimension of the space of polynomials of degree at most D is exactly $\binom{n+D}{D}$ (see Exercise 13.22).

To prove the claim, it is enough to note that $g_i(v_j) \neq 0$ if and only if $S_i \subseteq S_j$. Indeed, $g_i(v_j) \neq 0$ iff $f_k(v_j) \neq 0$ for all $k \in S_i$ iff $f_k(v_i) \neq 0$ implies $f_k(v_j) \neq 0$ iff $S_i \subseteq S_j$.

Assume now, for the sake of contradiction, that a nontrivial linear relation $\sum_{i=1}^{M} \lambda_i g_i = 0$ exists ($\lambda_i \in \mathbb{F}$). Let j be a subscript such that $|S_j|$ is minimal among the S_i with $\lambda_i \neq 0$. Substitute v_j in the relation. While $\lambda_j g_j(v_j) \neq 0$, we have $\lambda_i g_i(v_j) = 0$ for all $i \neq j$, a contradiction. □

If all polynomials $f_1, \ldots, f_m$ are of degree at most d, and if we have $m \ge n$ polynomials, then the upper bound $Z_{\mathbb{F}}(\mathbf{f}) \le \binom{md+n}{n}$ can be improved to

$$Z_{\mathbb{F}}(\mathbf{f}) \le \binom{md}{n} < \left(\frac{emd}{n}\right)^n. \tag{13.9}$$

The assumption $m \ge n$ is justified by the observation that for $m \le n$, the trivial upper bound $Z_{\mathbb{F}}(\mathbf{f}) \le 2^m$ can be attained even for $d = 1$, over every field: just take $f_i = x_i$.

It is also shown by Rónyai, Babai, and Ganapathy (2001) that, for $m \ge nd$, $d \ge 1$ and any sufficiently large field $\mathbb{F}$ (including infinite fields), the upper bound (13.9) is almost optimal: There exists a constant $\epsilon > 0$ and a sequence $\mathbf{f}$ of m polynomials of degree at most d in n variables such that

$$Z_{\mathbb{F}}(\mathbf{f}) \ge \left(\frac{\epsilon md}{n}\right)^n.$$

Exercises

13.1. Prove the Pythagoras theorem: if the vectors x, y are orthogonal, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

13.2. Let $x \in \mathbb{F}_2^n$ be a nonzero vector. Show that it is orthogonal to exactly half of vectors in $\mathbb{F}_2^n$. *Hint*: Take an i for which $x_i = 1$ and split the space $\mathbb{F}_2^n$ into 2^{n-1} pairs y, y' that differ only in their i-th coordinate. For each of these pairs, $\langle x, y\rangle \neq \langle x, y'\rangle$.

13.3. Let $\mathcal{F}$ be a family of subsets of an n-element set such that: (i) every set of $\mathcal{F}$ has an *even* number of elements, and (ii) each pair of sets shares an *even* number of elements. Construct such a family with at least $2^{\lfloor n/2\rfloor}$ sets.

13.4. (Babai–Frankl 1992). Show that the upper bound $2^{\lfloor n/2\rfloor}$ in the previous exercise cannot be improved. *Hint*: Let S be the set of incidence vectors of all sets in $\mathcal{F}$, and let U the span of this set (over $\mathbb{F}_2$). Argue that the rules (i) and (ii) imply that U is a subspace of $U^\perp$, and apply Proposition 13.2.

13.5. Prove the following "Oddtown Theorem" (see Babai and Frankl (1992) for the explanation of this name). Let $\mathcal{F}$ be a family of subsets of an n-element set such that: (i) every set of $\mathcal{F}$ has an *odd* number of elements, and (ii) each pair of sets share an *even* number of elements. Prove that then $|\mathcal{F}| \leq n$. Compare this with Exercise 13.3. *Hint*: The incidence vectors of sets in $\mathcal{F}$ are linearly independent over $\mathbb{F}_2$.

13.6. The *Hamming distance* between two vectors of the same length is just the number of positions in which these two strings differ. Show that the Euclidean distance between any two 0-1 vectors is the square root of their Hamming distance.

13.7. Show that the pairwise orthogonality of $(+1, -1)$-vectors implies their linear independence (over the reals).

13.8. Using the Cauchy–Schwarz inequality show that if $u = (u_1, \ldots, u_n)$ is a vector in $\mathbb{R}^n$ then $|u| \leq \sqrt{n} \cdot \|u\|$, where $|u| := |u_1| + \ldots + |u_n|$ and $|u_i|$ is the absolute value of u_i. *Hint*: Take a vector $v = (v_1, \ldots, v_n)$ with $v_i = 1$ if $u_i > 0$ and $v_i = -1$, otherwise. Observe that $|u| = \langle u, v\rangle$ and $\|v\| = \sqrt{n}$.

13.9. Let $f(x_1, \ldots, x_n)$ be a polynomial over $\mathbb{F}_2$ of degree $d < n$ which is not identically 1. Show that then $f(v) = 0$ for at least one nonzero vector v with at most $d + 1$ ones.

13.10. Let $h = \prod_{i\in S} x_i$ be a monomial of degree $d = |S| \leq n-1$, and let a be a 0-1 vector with at least $d+1$ ones. Show that then, over $\mathbb{F}_2$, $\sum_{b\leq a} h(b) = 0$. *Hint*: There are only two possibilities: either $a_i = 1$ for all $i \in S$, or not.

13.11. Suppose $A \subseteq \mathbb{Z}_3^n$ has the property that for all distinct vectors $a, b \in A$, there is a coordinate $i \in [n]$ such that $a_i - b_i = 1$ (subtraction in $\mathbb{Z}_3 = \{0, 1, 2\}$). Show that $|A| \leq 2^n$. *Hint*: Consider polynomials $f_a(x) = \prod_{i=1}^n (x_i - a_i - 1)$ over $\mathbb{Z}_3$.

13.12. (Babai et al. 1991). Let $\mathbb{F}$ be a field, $H_1, \ldots, H_m \subseteq \mathbb{F}$ and $H = H_1 \times \cdots \times H_m$. Prove that, for any function $f : H \to \mathbb{F}$ there exists a polynomial $\tilde{f}$ in m variables over $\mathbb{F}$ such that:

(i) $\tilde{f}$ has degree $< |H_i|$ in its i-th variable, and
(ii) $\tilde{f}$, restricted to H, agrees with f.

Is such a polynomial unique? *Hint*: Associate with each vector $u = (u_1, \ldots, u_m)$ in H a polynomial

$$g_u(x) = \prod_{i=1}^{m} \prod_{h \in H_i \setminus \{u_i\}} (x_i - h)$$

and show that every function $f : H \to \mathbb{F}$ is a linear combination of the g_u's, restricted to H.

13.13. Construct a two-distance set in $\mathbb{R}^n$ of size $\binom{n}{2}$. *Hint*: What about 0-1 vectors with two 1s in each?

13.14. Prove the following generalization of Theorem 13.12 for *s-distance sets.* Let $a_1, \ldots, a_m$ be vectors in $\mathbb{R}^n$ and suppose that the pairwise distances between them take at most s values. Prove that $m \leq \binom{n+s+1}{s}$. *Hint*: Let $d_1, \ldots, d_s$ be the distances permitted, and consider the polynomials $f_i(x) = \prod_{i=1}^{s}(\|x - a_i\|^2 - d_i^2)$. To estimate the dimension of the subspace containing all of them, expand the norm-square expression in each factor, replace the sum $\sum_{i=1}^{n} x_i^2$ by a new variable z, and multiply the constant terms by a new variable t. Observe that then each f_i becomes a homogeneous polynomial of degree s in $n + 2$ variables $x_1, \ldots, x_n, z, t$, and apply Proposition 1.5.

13.15. (Nagy 1972). Let G be a graph whose vertices are 3-element subsets of $\{1, \ldots, t\}$, and where two vertices A and B are joined by an edge if and only if $|A \cap B| = 1$. Use Exercise 13.5 and Fisher's inequality to show that this graph has neither a clique nor an independent set of size $t + 1$.

13.16. Write down a complete proof of Theorem 13.14. *Hint*: Work in the finite field $\mathbb{F}_p$ instead of that of real numbers $\mathbb{R}$. This time, due to the condition $|A_i| \not\in L \bmod p$, we can take $f_i(x) = \prod_{l \in L}(\langle v_i, x\rangle - l)$, i.e., we do not need the condition $l < |A_i|$.

13.17. (Ray-Chaudhuri–Wilson 1975). Prove the following uniform version of Theorem 13.13: if $A_1, \ldots, A_m$ is a k-uniform L-intersecting family of subsets of an n-element set, then $m \leq \binom{n}{s}$, where $s = |L|$. *Sketch*: (Alon–Babai–Suzuki 1991): Start as in the proof of Theorem 13.13 and define the same polynomials $f_1, \ldots, f_m$ of degree at most s. Associate with each subset I of $\{1, \ldots, n\}$ of cardinality $|I| \leq s - 1$ the following polynomial of degree at most s:

$$g_I(x) = \Big(\big(\sum_{j=1}^{n} x_j\big) - k\Big) \prod_{i \in I} x_i,$$

and observe that for any subset $S \subseteq \{1, \ldots, n\}$, $g_I(S) \neq 0$ if and only if $|S| \neq k$ and $S \supseteq I$. Use this property to show that the polynomials g_I *together* with the polynomials f_i are linearly independent. For this, assume

$$\sum_{i=1}^{m} \lambda_i f_i + \sum_{|I| \leq s-1} \mu_I g_I = 0$$

for some $\lambda_i, \mu_I \in \mathbb{R}$. Substitute A_j's for the variables in this equation to show that $\lambda_j = 0$ for every $j = 1, \ldots, m$. What remains is a relation among the g_I. To show that this relation must be also trivial, assume the opposite and re-write this relation as $\mu_1 g_{I_1} + \cdots + \mu_t g_{I_t} = 0$ with all $\mu_i \neq 0$ and $|I_1| \geq |I_j|$ for all $j > 1$. Substitute the first set I_1 for the variables and observe that all but the first function vanish.

13.18. Let $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ be subsets of an n-element set such that $|A_i \cap B_i|$ is odd for all $1 \leq i \leq m$, and $|A_i \cap B_j|$ is even for all $1 \leq i < j \leq m$. Show that then $m \leq n$.

13.19. (Frankl–Wilson 1981). Let p be a prime, and $n = 4p - 1$. Consider the graph $G = (V, E)$ whose vertex set V consists of all 0-1 vectors of length n with precisely $2p - 1$ ones each; two vectors are adjacent if and only if the Euclidean distance between them is $\sqrt{2p}$. Show that G has no independent set of size larger than $\sum_{i=0}^{p-1} \binom{n}{i}$. *Hint*: Use Exercise 13.6 to show that two vectors from V are adjacent in G precisely when they share $p - 1$ ones in common, and apply Theorem 13.14.

Comment: This construction was used by Frankl and Wilson (1981) to resolve an old problem proposed by H. Hadwiger in 1944: how many colors do we need in order to color the points of the n-dimensional Euclidean space $\mathbb{R}^n$ so that each monochromatic set of points misses some distance? A set is said to *miss distance* d if no two of its points are at distance d apart from each other. Larman and Rogers (1972) proved that Hadwiger's problem reduces to the estimating the minimum number of colors $\chi(n)$ necessary to color the points of $\mathbb{R}^n$ such that pairs of points of unit distance are colored differently. The graph G we just constructed shows that $\chi(n) \geq 2^{\Omega(n)}$ (see the next exercise). Kahn and Kalai (1993) used a similar construction to disprove another 60 years old and widely believed conjecture of K. Borsuk (1933) that every set of diameter one in n-dimensional real space R^n can be partitioned in at most $n + 1$ disjoint pieces of smaller diameter. Kahn and Kalai presented an infinite sequence of examples where the minimum number of pieces grew as an exponential function of $\sqrt{n}$, rather than just as a linear function $n + 1$, as conjectured. The interested reader can find these surprising solutions in the book of Babai and Frankl (1992).

13.20. The *unit distance graph* on $\mathbb{R}^n$ has the infinite set $\mathbb{R}^n$ as its vertex set, and two points are adjacent if their (Euclidean) distance is 1. Let $\chi(n)$ be the minimum number of colors necessary to color the points of the Euclidean space $\mathbb{R}^n$ such that pairs of points of unit distance are colored differently. Use the graph from the previous exercise to show that $\chi(n) \geq 2^{\Omega(n)}$. *Hint*: Observe that $\chi(G) \geq |V|/\alpha(G)$ and replace each 0-1 vector v by the vector ϵv, where $\epsilon = 1/\sqrt{2p}$. How does this change the distance?

13.21. Let $x_1, \ldots, x_n$ be real numbers, and $\sigma : [n] \to [n]$ a permutation of $[n] = \{1, \ldots, n\}$. Show that then $\sum_{i=1}^{n} x_i \cdot x_{\sigma(i)} \leq \sum_{i=1}^{n} x_i^2$. *Hint*: Use the Cauchy–Schwarz inequality.

13.22. Use Proposition 1.3 and Exercise 1.7 to show that the number of distinct monomials of degree at most d is $\binom{n+d}{d}$.

14. Orthogonality and Rank Arguments

Linear independence is one of the most basic concepts in linear algebra. No less important are the concepts of orthogonality and rank. In this chapter we consider some combinatorial applications of these two concepts.

14.1 Orthogonal coding

Linear independence is not the only way to obtain good upper bounds. If the members of a family $\mathcal{F}$ can be *injectively* associated with the elements of $\mathbb{F}_q^m$, then $|\mathcal{F}| \leq q^m$. If we are lucky, the associated "code-vectors" will be orthogonal to some subspace of dimension d, which (due to Proposition 13.2) immediately improves our bound to $|\mathcal{F}| \leq q^{m-d}$. We demonstrate this idea by the following result. Recall that two vectors u, v are *orthogonal* if their scalar product is zero, $\langle u, v\rangle = 0$.

Given two families $\mathcal{A}$ and $\mathcal{B}$ of subsets of an n-element set, satisfying some conditions, we are interested in how large $|\mathcal{A}| \cdot |\mathcal{B}|$ can be. If we know nothing about the families, then this number can be as large as 2^{2n}. If we know that both families are monotone increasing (or monotone decreasing), Kleitman's theorem (Theorem 10.6) gives a non-trivial upper bound:

$$|\mathcal{A}| \cdot |\mathcal{B}| \leq 2^n \cdot |\mathcal{A} \cap \mathcal{B}|.$$

If we know that all the intersections $A \cap B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$, have the same size modulo 2, then we can get an even better bound.

Theorem 14.1 (Ahlswede–El Gamal–Pang 1984)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be two families of subsets of an n-element set with the property that $|A \cap B|$ is even for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then $|\mathcal{A}| \cdot |\mathcal{B}| \leq 2^n$.*

Proof (due to Delsarte and Piret 1985). With each subset of X associate its incidence vector, and look at these vectors as elements of the n-dimensional

An EATCS Series, DOI 10.1007/978-3-642-17364-6_14,

vector space $\mathbb{F}_2^n$. Let U and V be the sets of incidence vectors of $\mathcal{A}$ and $\mathcal{B}$, respectively. Fix a vector $v_0 \in V$ and let $V_0 := \{v_0 + v \ : \ v \in V\}$. Moreover, let U' and V_0' be the subspaces spanned by U and V_0, respectively. Then

$$|\mathcal{A}| \cdot |\mathcal{B}| = |U| \cdot |V| = |U| \cdot |V_0| \leq |U'| \cdot |V_0'| \leq 2^{\dim U' + \dim V_0'}. \tag{14.1}$$

The key point is that (in $\mathbb{F}_2$) $\langle u, w\rangle = 0$ for all $u \in U$ and $w \in V_0$. This (with $w = v_0 + v$, $v \in V$) follows from the fact that $\langle u, v\rangle$ is exactly the parity of points in the intersection of the corresponding sets, and from our assumption that all these intersections have the same parity: $\langle u, w\rangle = \langle u, v_0\rangle + \langle u, v\rangle = 0$.

Therefore, by Proposition 13.2, we obtain that $\dim U' \leq n - \dim V_0'$. Putting this estimate in (14.1) we get the desired upper bound $|\mathcal{A}| \cdot |\mathcal{B}| \leq 2^n$. □

The same holds also with "even" replaced by "odd." In this case the bound is slightly better: $|\mathcal{A}| \cdot |\mathcal{B}| \leq 2^{n-1}$ (see Exercise 14.1).

14.2 Balanced pairs

Given a family $A_1, \ldots, A_m$ of distinct sets, a *balanced pair* in it is a pair of disjoint non-empty subsets of indices $I, J \subseteq [m]$ such that

$$\bigcup_{i\in I} A_i = \bigcup_{j\in J} A_j \quad \text{and} \quad \bigcap_{i\in I} A_i = \bigcap_{j \in J} A_j. \tag{14.2}$$

Theorem 14.2 (Lindstrom 1993). *Every family of $m \geq n+2$ distinct subsets of an n-element set contains a balanced pair.*

Proof. With each subset A of $\{1, \ldots, n\}$ we can associate the incidence vector $v = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$ of the pair $(A, \overline{A})$ in the usual way: $x_i = 1$ iff $i \in A$, and $y_i := 1 - x_i$. These vectors belong to the vector space V (over $\mathbb{R}$) of all vectors v for which $x_1 + y_1 = \cdots = x_n + y_n$.

Claim 14.3. The dimension of V is $n + 1$.

To prove the claim, observe that for any vector $v = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$ in V, the knowledge of $n+1$ coordinates $x_1, \ldots, x_n, y_1$ is enough to reconstruct the whole vector v; namely $y_i = x_1 + y_1 - x_i$. So, our space V is the set of solutions $v \in \mathbb{R}^{2n}$ of the system of linear equations $M \cdot v = 0$, where M is the $(n-1) \times (2n)$ matrix

$$\begin{pmatrix} 1 & 1 & -1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 0 & 0 & \cdots & -1 & -1 \end{pmatrix}$$

By Proposition 13.3, $\dim V = 2n - \mathrm{rk}(M) = 2n - (n-1) = n+1$, as desired.

Now let $v_i = (v_{i,1}, \ldots, v_{i,2n})$ be the vector corresponding to the i-th set A_i, $i = 1, \ldots, m$. By the assumption, the vectors $v_1, \ldots, v_m$ are distinct and all belong to the subspace V. Since $m \geq n+2 > n+1 = \dim V$, there must be a nontrivial linear relation between these vectors, which we can write as

$$\sum_{i \in I} \alpha_i v_i = \sum_{j \in J} \beta_j v_j,$$

where I and J are non-empty, $I \cap J = \emptyset$, and $\alpha_i, \beta_j > 0$ for all $i \in I$ and $j \in J$. When interpreted in combinatorial terms, the equality means that the *sets* of nonzero coordinates of the vectors on both sides must be the same, implying that $\cup_{i \in I} A_i = \cup_{j \in J} A_j$ and $\cup_{i \in I} \overline{A_i} = \cup_{j \in J} \overline{A_j}$. Using the identity $\overline{A} \cup \overline{B} = \overline{A \cap B}$, the last equality is equivalent to $\cap_{i \in I} A_i = \cap_{j \in J} A_j$. □

With a similar argument one can prove the following useful fact.

Proposition 14.4. *Among any $n+2$ distinct vectors in $\mathbb{R}^n$ there must be two whose scalar product is non-negative.*

Proof. Suppose $v_1, \ldots, v_{n+2} \in \mathbb{R}^n$, but $\langle v_i, v_j \rangle < 0$ for all $i \neq j$. Let $\widehat{v}_i := (v_i, 1)$, that is, append a 1 to each vector. Since the number $n+2$ of these vectors exceeds the dimension $n+1$ of the vector space $\mathbb{R}^{n+1}$ they lie in, the vectors must be linearly dependent. Choose coefficients α_i, not all zero, so that $\sum_i \alpha_i \widehat{v}_i = 0$. In other words, the α_i satisfy

$$\sum_{i=1}^{n+2} \alpha_i v_i = \mathbf{0} \qquad \text{and} \qquad \sum_{i=1}^{n+2} \alpha_i = 0\,.$$

Since the α_i are not all zero, some of them are positive and some are negative. Let

$$P = \{i \;:\; \alpha_i > 0\} \qquad \text{and} \qquad N = \{i \;:\; \alpha_i < 0\}\,.$$

Note that P and N are disjoint and both nonempty. Consider the vector

$$y := \sum_{i \in P} \alpha_i v_i = \sum_{j \in N} -\alpha_j v_j\,.$$

Then

$$0 \leq \langle y, y \rangle = \Big(\sum_{i \in P} \alpha_i v_i\Big)\Big(\sum_{j \in N} -\alpha_j v_j\Big) = \sum_{i \in P} \sum_{j \in N} -\alpha_i \alpha_j \langle v_i, v_j \rangle\,.$$

But the right-hand side is < 0, because $-\alpha_i \alpha_j > 0$ and $\langle v_i, v_j \rangle < 0$, a contradiction. □

14.3 Hadamard matrices

A Hadamard matrix is a square $n \times n$ matrix H with entries in $\{-1, +1\}$ and with row vectors mutually orthogonal over the reals (and hence with column vectors mutually orthogonal). Such matrices have many interesting properties and arise in many applications.

Lemma 14.5 (Lindsey's Lemma). *The absolute value of the sum of all entries in any $a \times b$ submatrix of an $n \times n$ Hadamard matrix does not exceed $\sqrt{abn}$.*

Proof (due to Babai, Frankl and Simon 1986). Let H be an $n \times n$ Hadamard matrix, and A one of its $a \times b$ submatrices. Assume for simplicity that A consists of its first a rows and b columns. Let α be the sum of all entries of A. We want to prove that $\alpha \leq \sqrt{abn}$.

Let $v_1, \ldots, v_a$ be the first a rows of H, and $y = \sum_{i=1}^{a} v_i$. If we take the vector $x = (1^b 0^{n-b})$, then $\alpha^2 = \langle x, y\rangle^2 \leq \|x\|^2 \|y\|^2 = b \cdot \|y\|^2$. On the other hand, the conditions $\langle v_i, v_i\rangle = n$ and $\langle v_i, v_j\rangle = 0$ for all $i \neq j$ imply that $\|y\|^2 = \sum_{i,j=1}^{a} \langle v_i, v_j\rangle = \sum_{i=1}^{a} \langle v_i, v_i\rangle = an$. Thus, $\alpha^2 \leq b \cdot \|y\|^2 = abn$, as desired. □

Another interesting property of every Hadamard matrix is that each of its $k \times n$ submatrices maps each(!) nonzero vector $x \in \mathbb{R}^k$ to a vector with at least n/k nonzero entries.

Lemma 14.6 (Alon 1990a). *Every non-trivial linear combination of any k rows of a Hadamard matrix has at least n/k nonzero entries.*

Proof. Let A be a $k \times n$ submatrix of an $n \times n$ Hadamard matrix with rows $a^i = (a_{i1}, \ldots, a_{in})$, $i = 1, \ldots, k$. Let $y = x^\top A$ for some nonzero vector $x = (x_1, \ldots, x_k)$ in $\mathbb{R}^k$, $S = \{i : y_i \neq 0\}$ and $s = |S|$. We have to show that $s \geq n/k$.

Assume, without loss of generality, that $|x_1| = \max_{1 \leq i \leq k} |x_i|$. Since the vectors $a^1, \ldots, a^k$ are mutually orthogonal, we have

$$\begin{aligned} k x_1^2 n &\geq \sum_{i=1}^{k} x_i^2 n = \sum_{i=1}^{k} \langle x_i a^i, x_i a^i\rangle = \left\langle \sum_{i=1}^{k} x_i a^i, \sum_{i=1}^{k} x_i a^i \right\rangle = \langle y, y\rangle \\ &= \sum_{j=1}^{n} y_j^2 = \sum_{j\in S} |y_j|^2 = \frac{1}{s}\Big(\sum_{j\in S} 1\Big)\Big(\sum_{j\in S} |y_j|^2\Big) \geq \frac{1}{s}\Big(\sum_{j\in S} |y_j|\Big)^2, \end{aligned}$$

where the last inequality follows from the Cauchy–Schwarz inequality (13.3). On the other hand, since a^1 is orthogonal to all the vectors $a^2, \ldots, a^k$,

$$\begin{aligned}\sum_{j=1}^{n}|y_j| \geq \sum_{j=1}^{n} y_j a_{1j} &= \sum_{j=1}^{n}\sum_{i=1}^{k} x_i a_{ij} a_{1j} \\ &= \sum_{i=1}^{k} x_i \sum_{j=1}^{n} a_{ij} a_{1j} = \sum_{i=1}^{k} x_i \langle a^i, a^1\rangle = x_1\langle a^1, a^1\rangle = x_1 \cdot n.\end{aligned}$$

Substituting this estimate into the previous one we obtain $s \geq n/k$, as desired. □

If H is an $n \times n$ Hadamard matrix, then $H^\top H = nI_n$ where I_n is the $n \times n$ identity matrix. Hence, all n eigenvalues of $H^\top H$ are equal to n, implying that H has spectral norm $\|H\| = \sqrt{n}$. Combining this with Proposition 13.6, we can show that all large enough submatrices of H have large rank over the reals.

Lemma 14.7. *Let H be an $n \times n$ Hadamard matrix, and A one of its $a \times b$ submatrices. Then* $\mathrm{rk}(A) \geq ab/n$.

Proof. Since A is a submatrix of H, we have that $\|A\| \leq \|H\|$. So, by Proposition 13.6, we obtain

$$\mathrm{rk}(A) \geq \frac{\|A\|_{\mathrm{F}}^2}{\|A\|^2} \geq \frac{\|A\|_{\mathrm{F}}^2}{\|H\|^2} = \frac{ab}{n},$$

where the last equality follows because $\|A\|_{\mathrm{F}}^2$ is precisely the number of entries in A. □

We can always reduce the rank of a real-valued matrix by changing some of its entries. The *rigidity* of a matrix M is a function $R_M(r)$ equal to the minimum number of entries of M that one needs to change in order to reduce the rank to r or less.

Matrix rigidity is an important measure in boolean circuit complexity. In particular, an *explicit* boolean $n \times n$ matrix M having rigidity $R_M(\epsilon n) \geq n^{1+\delta}$ over $\mathbb{F}_2$ for some constants $\epsilon, \delta > 0$ would give us the first super-linear lower bound on log-depth linear circuits computing the linear transformation $y = Mx$. This was shown by Valiant (1977).

Due to its importance, the rigidity of Hadamard matrices merits particular attention. For an $n \times n$ Hadamard matrix H, Pudlák, Razborov, and Savický (1988) proved that $R_H(r) \geq n^2/r^3 \log r$. It can be also shown that Alon's lemma yields $R_H(r) \geq n^2/r^2$ (Exercise 14.6). Kashin and Razborov (1998) improved this to $R_H(r) \geq n^2/256r$. De Wolf (2006) later re-derived this bound using a spectral argument, with a better constant.

Theorem 14.8. *Let H be an $n \times n$ Hadamard matrix. If $r \leq n/2$ then $R_H(r) \geq n^2/4r$.*

The condition $r \leq n/2$ is important here. If H is symmetric then its eigenvalues are all $\pm\sqrt{n}$, so we can reduce the rank to $n/2$ by adding or subtracting the diagonal matrix $\sqrt{n}I_n$. This shows that $R_H(n/2) \leq n$.

Proof (due to Ronald de Wolf 2006). Let R be the minimum number of changes that brought the rank of H down to r. By a simple averaging argument, we can find $2r$ rows of H that contain a total of at most $2rR/n$ changes. If $n \leq 2rR/n$, then $R \geq n^2/2r$ and we are done. Hence, we can assume that $n - 2rR/n > 0$. Consider the $n - 2rR/n$ columns that contain no changes in the above set of rows. We thus get a $2r \times (n - 2rR/n)$ submatrix B that contains no changes and hence is a submatrix of H. By definition of R, this submatrix must have rank at most r. Applying Lemma 14.7, we get $r \geq \mathrm{rk}(B) \geq 2r(n-2rR/n)/n$. Rearranging this inequality, we get $R \geq n^2/4r$. □

We can multiply any rows and columns of a Hadamard matrix by -1 to obtain other Hadamard matrices. In particular, starting from an arbitrary Hadamard matrix, we can reduce it to the form where the first row or the first column (or both) consist entirely of 1s. In this case the matrix is called *normalized.* Such matrices have additional structural properties.

Theorem 14.9. *If H is a Hadamard matrix of order n and its first row consists entirely of* 1*s, then every other row has $n/2$ positive and $n/2$ negative entries. If $n > 2$ then any two rows other than the first have exactly $n/4$* 1*s in common.*

Proof. The first statement immediately follows from the fact that the scalar product of any row with the first row is 0.

To prove the second statement, let u and v be two rows other than the first, and let a (resp. b) be the number of places where they both have 1s (resp. -1s). Because u has the same number $n/2$ of 1s and -1s, we get the following picture:

$$\begin{array}{ccccc}
u & +1+1\ldots+1 & +1+1\ldots+1 & -1-1\ldots-1 & -1-1\ldots-1 \\
v & +1+1\ldots+1 & -1-1\ldots-1 & +1+1\ldots+1 & -1-1\ldots-1 \\
 & & & & \\
 & a & n/2-a & n/2-b & b
\end{array}$$

Since the total number of $+1$'s in v is $n/2$, we have $a + (n/2 - b) = n/2$, and hence, $a = b$. The orthogonality of u and v then implies that $a - (n/2 - a) - (n/2 - b) + b = 0$, i.e., that $a = n/4$. □

Let H be a Hadamard matrix of order n. Take all the rows of H and $-H$, and change all -1's to 0. This way we obtain a set of $2n$ binary vectors of length n called the *Hadamard code C_n*.

Theorem 14.10. *Every two codewords in C_n differ in at least $n/2$ coordinates.*

Proof. Take any $x, y \in C_n$, $x \neq y$. If these two vectors have been obtained from the i-th rows of H and $-H$ respectively, then they disagree in all n coordinates. Otherwise, there are two different rows u and v in H such that x is obtained (by changing -1s to 0s) from u or $-u$, and y from v or $-v$. In all cases, x and y differ in $n/2$ coordinates, because $\pm u$ and $\pm v$ are orthogonal. □

Hadamard matrices can also be used to construct combinatorial designs with good parameters. Recall that a (v, k, λ) *design* is a k-uniform family of subsets (also called *blocks*) of a v-element set such that every pair of distinct points is contained in exactly λ of these subsets; if the number of blocks is the same as the number v of points, then the design is *symmetric* (see Chap. 12).

By Theorem 14.9, we have that, if there is a Hadamard matrix of order n, then $n = 2$ or n is divisible by 4. It is conjectured that Hadamard matrices exist for *all* orders that are divisible by 4.

Theorem 14.11. *Every Hadamard matrix of order* $4n$ *gives a symmetric* $(4n-1, 2n-1, n-1)$ *design.*

Proof. Let H be a Hadamard matrix of order $4n$, and assume that it is normalized, i.e., the first row and the first column consist entirely of 1s. Form a $(4n-1) \times (4n-1)$ 0-1 matrix M by deleting the first column and the first row in H, and changing -1s to 0s. This is the incidence matrix of a symmetric $(4n-1, 2n-1, n-1)$ design, because by Theorem 14.9, each row of M has $2n-1$ ones and any two columns of M have exactly $n-1$ ones in common. □

14.4 Matrix rank and Ramsey graphs

A matrix $A = (a_{ij})$ is *lower co-triangular* if $a_{ii} = 0$ and $a_{ij} \neq 0$ for all $1 \leq j < i \leq n$. That is, such a matrix has zeroes on the diagonal and nonzero entries below the diagonal; the entries above the diagonal may be arbitrary.

Lemma 14.12. *Let* p *be a prime number, and* A *an* $n \times n$ *lower co-triangular matrix over* $\mathbb{F}_p$ *of rank* r. *Then*

$$n \leq \binom{r+p-2}{p-1} + 1 \leq (r+p)^{p-1}.$$

Proof. Let $r = \mathrm{rk}_{\mathbb{F}_p}(A)$ and $A = B \cdot C$ be the corresponding decomposition of A. For $i = 1, \ldots, n$ consider the polynomials $f_i(x) = 1 - g_i(x)^{p-1}$ in r variables $x = (x_1, \ldots, x_r)$ over $\mathbb{F}_p$, where $g_i(x)$ is the scalar product of x with the i-th row of B. Let $c_1, \ldots, c_n$ be the columns of C. Then $g_i(c_i) = 0$ and $g_i(c_j) \neq 0$ for every $i > j$. Since p is a prime, Fermat's Little Theorem (see Exercise 1.15) implies that $a^{p-1} = 1$ for every $a \neq 0$ in $\mathbb{F}_p$. Hence, $f_i(c_i) \neq 0$

and $g_i(c_j) = 0$ for every $i > j$. By Lemma 13.11, the polynomials $f_1, \ldots, f_n$ are linear independent elements of a vector space V of all polynomials over $\mathbb{F}_p$ of degree $p-1$, all of whose monomials $\prod_{i=1}^r x_i^{t_i}$ satisfy $\sum_{i=1}^r t_i = p-1$ and $t_i \geq 0$. By Proposition 1.5, the number of such monomials is $\binom{r+(p-1)-1}{p-1}$. Since the polynomials can also have a constant term (which accounts for the "+1" in the final equation), we have that

$$n \leq \dim V \leq \binom{r+p-2}{p-1} + 1 \leq (r+p)^{p-1}. \qquad \square$$

Let R be a ring and $A = (a_{ij})$ an $n \times n$ matrix with entries from R. The rank $\mathrm{rk}_R(A)$ of A over R is defined as the minimum number r for which there exists an $n \times r$ matrix B and an $r \times n$ matrix C over R such that $A = B \cdot C$; if all entries of A are zeroes then $\mathrm{rk}_R(A) = 0$. If $R = \mathbb{F}$ is a field, then $\mathrm{rk}_R(A)$ is the usual rank over $\mathbb{F}$, that is, the largest number of linear independent rows.

By Lemma 14.12, lower co-triangular matrices over $R = \mathbb{Z}_m$ have large rank, if m is a prime number. But what about $R = \mathbb{Z}_m$ for non-prime m, say, for $m = 6$? In this case R is no longer a field—it is just a ring (division is not defined). Still one can extend the notion of rank also to rings.

Let R be a ring and $A = (a_{ij})$ an $n \times n$ matrix with entries from R. The rank $\mathrm{rk}_R(A)$ of A over R is defined as the minimum number r for which there exists an $n \times r$ matrix B and an $r \times n$ matrix C over R such that $A = B \cdot C$; if all entries of A are zeroes then $\mathrm{rk}_R(A) = 0$. If $R = \mathbb{F}$ is a field, then $\mathrm{rk}_R(A)$ is the usual rank over $\mathbb{F}$, that is, the largest number of linear independent rows.

It turns out that explicit low rank matrices over the ring $R = \mathbb{Z}_6$ of integers modulo 6 would give us explicit graphs with good Ramsey properties, that is, graphs without any large clique or large independent set.

Let $A = (a_{ij})$ be an $n \times n$ lower co-triangular matrix over $\mathbb{Z}_6$. Associate with A the graph $G_A = (V, E)$ with $V = \{1, \ldots, n\}$, where two vertices $i > j$ are adjacent iff a_{ij} is odd.

Lemma 14.13 (Grolmusz 2000). *If $r = \mathrm{rk}_{\mathbb{Z}_6}(A)$ then the graph G_A contains neither a clique on $r+2$ vertices nor an independent set of size $\binom{r+1}{2} + 2$.*

Proof. It is clear that $\mathrm{rk}_{\mathbb{F}_p}(A) \leq r$ for $p \in \{2, 3\}$. Let $S \subseteq V$ be a clique in G_A of size $|S| = s$, and $B = (b_{ij})$ be the corresponding $s \times s$ submatrix of A; hence, $b_{ii} = 0$ and $b_{ij} \in \{1, 3, 5\}$ for all $i > j$. Then $B \bmod 2$ is a lower co-triangular matrix over $\mathbb{F}_2$, and Lemma 14.12 (with $p = 2$) implies that $|S| \leq r + 1$.

Now let $T \subseteq V$ be an independent set in G_A of size $|T| = t$, and $C = (c_{ij})$ be the corresponding $t \times t$ submatrix of A; hence, $c_{ii} = 0$ and $c_{ij} \in \{2, 4\}$ for all $i > j$. Then $C \bmod 3$ is a lower co-triangular matrix over $\mathbb{F}_3$, and Lemma 14.12 (with $p = 3$) implies that $|T| \leq \binom{r+1}{2} + 1$. $\square$

In Sect. 13.7 (Theorem 13.15) we have shown how to construct explicit n-vertex graphs with no clique or independent set larger than

$$t := 2^{c\sqrt{\ln n \ln\ln n}}$$

for an absolute constant c. Grolmusz (2000) constructed a co-triangular $n \times n$ matrix A over $R = \mathbb{Z}_6$ with $\mathrm{rk}_{\mathbb{Z}_6}(A) \leq t$. Together with Lemma 14.13, this gives an alternative construction of a graph G_A with no clique or independent set larger than t.

14.5 Lower bounds for boolean formulas

Boolean *formulas* (or De Morgan formulas) are defined inductively as follows:

- Every boolean variable x_i and its negation $\overline{x}_i$ is a formula of size 1 (these formulas are called *leaves*).
- If F_1 and F_2 are formulas of size l_1 and l_2, then both $F_1 \wedge F_2$ and $F_1 \vee F_2$ are formulas of size $l_1 + l_2$.

Note that the size of F is exactly the number of leaves in F.

Often one uses an equivalent definition of a formula as a circuit with And, Or, and Not gates, whose underlying graph is a tree. That is, now negation is allowed not only at the leaves. But using De Morgan rules $\neg(x \vee y) = \neg x \wedge \neg y$ and $\neg(x \wedge y) = \neg x \vee \neg y$ one can move all negations to leaves without increasing the formula size.

Given a boolean function f, how it can be shown that it is hard, i.e., that it cannot be computed by a formula of small size? Easy counting shows that almost all boolean functions in n variables require formulas of size exponential in n. Still, for a *concrete* boolean function f, the largest remains the lower bound $n^{3-o(1)}$ proved by Håstad (1993).

The main difficulty here is that we allow negated variables $\overline{x}_i$ as leaves. It is therefore natural to look at what happens if we forbid this and require that our formulas are *monotone* in that they do not have negated leaves. Of course, not every boolean function $f(x_1, \ldots, x_n)$ can be computed by such a formula – the function itself must be also *monotone*: if $f(x_1, \ldots, x_n) = 1$ and $x_i \leq y_i$ for all i, then $f(y_1, \ldots, y_n) = 1$. Under this restriction progress is substantial: we are able to prove that some explicit monotone functions require monotone formulas of super-polynomial size.

14.5.1 Reduction to set-covering

Let A and B be two disjoint subsets of $\{0, 1\}^n$. A boolean formula F *separates* A and B if $F(a) = 1$ for all $a \in A$ and $F(b) = 0$ for all $b \in B$. A *rectangle* is a subset $R \subseteq A \times B$ of the form $R = S \times T$ for some $S \subseteq A$ and $T \subseteq B$. A rectangle is *monochromatic* if there exist an $\epsilon \in \{0,1\}$ and a position $i \in [n]$ such that $a_i = \epsilon$ and $b_i = 1 - \epsilon$ for all $a \in S$ and $b \in T$. That is, the rectangle $R = S \times T$ is monochromatic if S and T can be separated by a single variable x_i or by its negation $\overline{x}_i$. If we have a stronger condition that $a_i = 1$ and $b_i = 0$ for all $a \in S$ and $b \in T$ (i.e. if we do not allow negations $\overline{x}_i$) then the rectangle is *monotone monochromatic*.

The following simple lemma reduces the (computational) problem of proving a lower bound on the size of a formulas separating a pair A, B to a (combinatorial) problem of proving a lower bound on the number of mutually disjoint rectangles covering the Cartesian product $A \times B$.

Lemma 14.14 (Rychkov 1985). *If A and B can be separated by a (monotone) formula of size t then the set $A \times B$ can be covered by t mutually disjoint (monotone) monochromatic rectangles.*

Proof. Let F be an optimal formula which separates the pair A, B, i.e., $A \subseteq F^{-1}(1)$ and $B \subseteq F^{-1}(0)$. Let $t = size(F)$. We argue by induction on t.

Base case. If $size(F) = 1$ then F is just a single variable x_i or its negation. In that case the Cartesian product $A \times B$ is a monochromatic rectangle itself, and we are done.

Induction step. Assume that the theorem holds for all formulas smaller than F, and suppose that $F = F_1 \wedge F_2$ (the case $F = F_1 \vee F_2$ is similar). Let $t_i = size(F_i)$, hence $t = t_1 + t_2$. Define $B_1 := \{b \in B : F_1(b) = 0\}$ and $B_2 := B \setminus B_1$. Notice that F_i separates A and B_i for $i = 1, 2$. Applying the induction hypothesis to the subformula F_i yields that the product $A \times B_i$ can be covered by t_i mutually disjoint monochromatic rectangles, for both $i = 1, 2$. Since $A \times B_1$ and $A \times B_2$ form a partition of $A \times B$, we have that the set $A \times B$ can be covered by $t_1 + t_2 = t$ monochromatic rectangles, as desired. □

We can use Rychkov's lemma to derive the well-known lower bound due to Khrapchenko (1971). Given two disjoint subsets A and B of $\{0,1\}^n$, define the set

$$A \otimes B = \{(a,b) : a \in A \text{ and } b \in B \text{ and } a \sim b\},$$

where $a \sim b$ means that inputs a and b differ on exactly one bit. Intuitively, if $A \otimes B$ is large, then every formula separating A and B should be large, since the formula must distinguish many pairs of adjacent inputs. Just how large the formulas must be says the following theorem. Viewing $A \otimes B$ as the set of edges of a bipartite graph with parts A and B, it states that the size of any formula separating A and B must be at least the product of the average degrees of these two parts.

Theorem 14.15 (Khrapchenko 1971). *Every formula separating a pair of non-empty disjoint subsets $A, B \subseteq \{0,1\}^n$ must have size at least*

$$\frac{|A \otimes B|^2}{|A| \cdot |B|}.$$

Proof. The main property of the set $A \otimes B$ is accumulated in the following

Claim 14.16. No monochromatic $s \times t$ rectangle can cover more than $\sqrt{st}$ elements of $A \otimes B$.

To prove the claim, let $S \times T$ be a monochromatic $s \times t$ subrectangle of $A \times B$. Since the rectangle is monochromatic, each element of S differs from each element in T in one particular position j, whereas (a, b) is in $A \otimes B$ only if a and b differ in exactly one position. Hence, for any given $a \in S$, the only possible $b \in T$ for which $a \sim b$ is one which differs from a exactly in position j. As a result, we have that $S \times T$ can cover at most $\min\{|S|, |T|\} = \min\{s, t\} \leq \sqrt{st}$ entries of $A \otimes B$.

Now suppose we have a decomposition of $A \times B$ into r monochromatic rectangles of dimensions $s_i \times t_i$, $i = 1, \ldots, r$. Let c_i be the number of elements of $A \otimes B$ in the i-th of these rectangles. By Claim 14.16, we know that $c_i^2 \leq a_i b_i$. Since the rectangles are disjoint and cover the whole rectangle $A \times B$, we also have that $|A \otimes B| = \sum_{i=1}^r c_i$ and $|A \times B| = \sum_{i=1}^r a_i b_i$. Applying the Cauchy–Schwarz inequality $(\sum x_i y_i)^2 \leq (\sum x_i^2) \cdot (\sum y_i^2)$ with $x_i = c_i$ and $y_i = 1$, we obtain

$$|A \otimes B|^2 = \left(\sum_{i=1}^r c_i\right)^2 \leq r \sum_{i=1}^r c_i^2 \leq r \cdot \sum_{i=1}^r a_i b_i = r \cdot |A \times B|\,. \qquad \square$$

Khrapchenko's theorem can be used to show that some explicit boolean functions require formulas of quadratic size. Consider, for example, the parity function $f = x_1 \oplus \cdots \oplus x_n$. Taking $A = f^{-1}(1)$ and $B = f^{-1}(0)$ we see that $|A \otimes B| = n|A| = n|B|$, and hence, f requires formulas of size at least n^2.

Unfortunately, this (quadratic) lower bound is the best that we can achieve using this theorem (Exercise 14.12).

14.5.2 The rank lower bound

In order to apply Ryckov's lemma, we must be able to show that, for some explicit disjoint subsets $A, B \subseteq \{0,1\}^n$, the rectangle $A \times B$ cannot be decomposed into few disjoint monochromatic rectangles. In general, this is a very difficult task: no explicit pair A, B requiring, say, n^3 monochromatic rectangles is known.

Fortunately, in the *monotone* case—when rectangles in a decomposition are required to be monotone—the situation is much better: here we can prove even super-polynomial lower bounds of the form $n^{\Omega(\log n)}$. And this can be done using rank arguments.

Fix an arbitrary field $\mathbb{F}$. Given a monotone boolean function f, we can associate with every pair of subsets $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$ a matrix $M : A \times B \to \mathbb{F}$. If $R \subseteq A \times B$ is a set of its entries, then we denote by M_R the matrix which is obtained from the matrix M by changing to 0 all its entries outside R. Define

$$\mu(M) := \frac{\mathrm{rk}(M)}{\max_R \mathrm{rk}(M_R)} \tag{14.3}$$

where the maximum is over all monotone monochromatic rectangles $R \subseteq A \times B$. For a monotone boolean function f, let $\mu(f)$ be the maximum of $\mu(M)$ over all pairs A, B separated by f and all matrices $M : A \times B \to \mathbb{F}$.

Lemma 14.17 (Razborov 1990)**.** *Any monotone formula computing a monotone boolean function f must have size at least $\mu(f)$.*

Proof. Suppose that f can be computed by a monotone formula of size t. Take an arbitrary pair A, B of sets separated by f, and an arbitrary matrix $M : A \times B \to \mathbb{F}$. By Rychkov's lemma we know that the rectangle $A \times B$ can be covered by at most t mutually disjoint monotone monochromatic rectangles $R_1, \ldots, R_t$. Let $M_i = M_{R_i}$ be the matrix corresponding to R_i. Since the rectangles are mutually disjoint, we have that $M = \sum_{i=1}^{t} M_i$. The sub-additivity of the rank (see (13.1)) implies that

$$\mathrm{rk}(M) = \mathrm{rk}\Big(\sum_{i=1}^{t} M_i\Big) \le \sum_{i=1}^{t} \mathrm{rk}(M_i) \le t \cdot \max_i \mathrm{rk}(M_i)\,. \qquad \square$$

It is clear that Lemma 14.17 also holds for *non-monotone* formulas if we allow non-monotone monochromatic rectangles. However, Razborov (1992) has proved that in this case the result is useless: for any boolean function f in n variables, the fraction on the right-hand side of (14.3) does not exceed $O(n)$. Fortunately, in the monotone case, Lemma 14.17 *can* give non-trivial lower bounds.

Let us consider bipartite graphs $G = (V_1, V_2, E)$ with $|V_1| = |V_2| = n$. With any such graph we can associate a monotone boolean function $f_{G,k}$ as follows. The function has $2n$ variables, one for each node of G, and accepts a set of nodes $X \subseteq V_1 \cup V_2$ if and only if X contains some subset $S \subseteq V_1$ of size at most k, together with the set of its *common neighbors*

$$N(S) := \{j \in V_2 \,:\, (i,j) \in E \text{ for all } i \in S\}.$$

That is, $f_{G,k}$ is the Or of all $\sum_{i=0}^{k} \binom{n}{i}$ monomials $\bigwedge_{i \in S \cup N(S)} x_i$ where $S \subseteq V_1$ and $|S| \le k$. By $\overline{N}(S)$ we will denote the set of all *common non-neighbors* of S, that is,

$$\overline{N}(S) := \{j \in V_2 \,:\, (i,j) \in E \text{ for no } i \in S\}.$$

By its definition, every function $f_{G,k}$ can be computed by a monotone formula of size at most $2n \sum_{i=0}^{k} \binom{n}{i}$. It turns out that, for graphs satisfying the isolated neighbor condition (see Definition 10.18), this trivial formula is almost optimal.

Recall that a bipartite graph $G = (V_1, V_2, E)$ satisfies the *isolated neighbor condition for k* if for any two disjoint subsets $S, T \subseteq V_1$ such that $|S| + |T| = k$, there is a node $v \in V_2$ which is a common neighbor of all the nodes in S and is isolated from all the nodes in T, i.e., if $N(S) \cap \overline{N}(T) \neq \emptyset$.

Lemma 14.18 (Gál 1998)**.** *If G satisfies the isolated neighbor condition for $2k$, then the function $f_{G,k}$ does not have a monotone DeMorgan formula of size smaller than $\sum_{i=0}^{k} \binom{n}{i}$.*

Proof. Associate with each subset $S \subseteq V_1$ of $|S| \leq k$ vertices the following two vectors a_S and b_S in $\{0,1\}^{2n}$:

$$a_S(i) = 1 \text{ if and only if } i \in S \cup N(S)$$

and

$$b_S(i) = 0 \text{ if and only if } i \in S \cup \overline{N}(S)\,.$$

Let A denote the set of all vectors a_S, and B the set of all vectors b_S where S ranges over all subsets of V_1 of size at most k.

By the definition of the function $f = f_{G,k}$, we have that $f(x) = 1$ if and only if $x \geq a_S$ for some S. Hence, $f(a) = 1$ for all $a \in A$. We claim that $f(b) = 0$ for all $b \in B$. To show this, we use the fact that the graph G satisfies the isolated neighbor condition for $2k$. This condition implies that, for any two subsets $S, T \subseteq V_1$ of size at most k,

$$S \cap T = \emptyset \text{ if and only if } N(S) \cap \overline{N}(T) \neq \emptyset. \tag{14.4}$$

Now take an arbitrary vector b_T in B. To show that $f(b_T) = 0$, it is enough to show that, for every $a_S \in A$ there is a position i such that $b_T(i) = 0$ and $a_S(i) = 1$. If $S \cap T \neq \emptyset$ then every position i in the intersection $S \cap T$ has this property. If $S \cap T = \emptyset$ then (14.4) implies that $N(S) \cap \overline{N}(T) \neq \emptyset$, and again, every position i in the intersection $N(S) \cap \overline{N}(T)$ has this property.

Thus, we have shown that the function f separates the pair A, B. Now define the matrix $M : A \times B \to \mathbb{F}_2$ by

$$M[a_S, b_T] = 1 \text{ if and only if } S \cap T = \emptyset.$$

This is a disjointness matrix $D(n,k)$, considered in Sect. 13.4, and we already know (see Theorem 13.10) that it has full rank over $\mathbb{F}_2$:

$$\mathrm{rk}_{\mathbb{F}_2}(M) = \sum_{i=0}^{k} \binom{n}{i}.$$

Thus, by Lemma 14.17, it remains to show that $\mathrm{rk}_{\mathbb{F}_2}(R) \leq 1$ for every monotone monochromatic rectangle $R \subseteq A \times B$.

Since R is monotone monochromatic, there must exist a position i such that $v_S(i) = 1$ and $u_T(i) = 0$ for all $(v_S, v_T) \in R$. If $i \in V_1$ then the corresponding entry of the intersection matrix M is 0 because then $S \cap T \neq \emptyset$; if $i \in V_2$ then this entry is 1 because then $N(S) \cap \overline{N}(T) \neq \emptyset$, and by (14.4),

$S \cap T = \emptyset$. Thus, depending on whether $i \in V_1$ or $i \in V_2$, the matrix M_R is either the all-0 matrix or a matrix consisting of 1s in all entries in R and 0s elsewhere; in this last case, M_R is a matrix of rank 1. □

Explicit bipartite graphs, satisfying the isolated neighbor condition for $k = \Omega(\log n)$ are known. Such are, for example, Paley graphs G_n constructed in Sect. 10.6. By Lemma 14.18, the corresponding boolean function $f_{G_n,k}$ requires monotone formula of size at least $\binom{n}{k} = n^{\Omega(\log n)}$.

Exercises

14.1. Let $\mathcal{A}$ and $\mathcal{B}$ be families of subsets of an n-element set with the property that $|A \cap B|$ is *odd* for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Prove that then $|\mathcal{A}| \cdot |\mathcal{B}| \leq 2^{n-1}$. *Hint*: Replace A in the proof of Theorem 14.1 by a larger set $A' = A \cup A_0$ where $A_0 = \{u_0 + u \ : \ u \in A\}$ for some fixed $u_0 \in A$. Show that $A \cap A_0 = \emptyset$, and argue as in that proof with A' instead of A.

14.2. Define the matrices H_{2m}, $m = 1, 2, 4, 8, \ldots$, inductively as follows:

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad H_{2m} = \begin{pmatrix} H_m & H_m \\ H_m & -H_m \end{pmatrix}.$$

Show that these matrices are Hadamard.

14.3. (Due to Gatis Midrijanis) Let H_n be a matrix from the previous exercise. Give a *direct* proof that its rigidity is $R_{H_n}(r) \geq n^2/4r$ for all $1 \leq r \leq n/2$. *Hint*: Divide H_n uniformly into $(n/2r)^2$ submatrices of size $2r \times 2r$, and observe that each of them is $\pm H_{2r}$.

14.4. Let $n = 2^m$. The $n \times n$ Sylvester ±1-matrix $S_n = (s_{xy})$ by labeling the rows and columns by m-bit vectors $x, y \in \mathbb{F}_2^m$ and letting $s_{xy} = (-1)^{\langle x,y \rangle}$, where the scalar product is over $\mathbb{F}_2$. Show that S_n is a Hadamard matrix.

14.5. Show that Alon's lemma (Lemma 14.6) is sharp, at least whenever k divides n and there exists a $k \times k$ Hadamard matrix H. *Hint*: Take n/k copies of H.

14.6. Use Lemma 14.6 to show that if $t > (1 - 1/r)n$, then every $r \times t$ sub-matrix H' of an $n \times n$ Hadamard matrix H has rank r (over the reals).

14.7. Take an $n \times n$ matrix over some field $\mathbb{F}$ and suppose that all of its rows are different. Prove the following: if some column is linearly dependent on the others then after the deletion of this column, all the rows in the resulting n by $n - 1$ matrix are still different.

14.8. (Babai–Frankl 1992). Give a linear algebra proof of Bondy's theorem (Theorem 11.1): In any $n \times n$ 0-1 matrix M with all rows being different, there is a column, after deletion of which the resulting n by $n-1$ matrix still has no equal rows. *Hint*: Consider two cases depending on what is the determinant of M. If $\det(M) = 0$ then some column is linearly dependent on the others, and we are in the situation of the previous exercise. If $\det(M) \neq 0$ then take a row v_i of M with the minimal number of 1s, and expand the determinant by this row. Conclude that for some j, the term $v_{ij} \cdot \det M_{ij} \neq 0$, where M_{ij} is the $(n-1)$ by $(n-1)$ minor obtained by deleting the i-th row and the j-th column. Hence, $v_{ij} = 1$ and no two rows of the minor are identical. Use this to prove that deleting the j-th column from the whole matrix M leaves no equal rows.

14.9. For $n \geq 1$, $d \geq 0$, $n \equiv d \bmod 2$, let $K(n,d)$ denote the minimal cardinality of a family V of ± 1 vectors of length n, such that for any ± 1 vector w of length n, there is a $v \in V$ such that the value of the scalar product $\langle v, w\rangle$ (over the reals) lies between $-d$ and d. Prove that:

(i) $K(n,0) \leq n$ (Knuth 1986). *Hint*: Consider ± 1 vectors $v_0, v_1, \ldots, v_n$ of length n, where the i-th vector v_i has first i coordinates equal to -1 and the rest equal to $+1$; hence, v_0 has no -1's at all whereas v_n consists entirely of -1's. Observe that $\langle w, v_0\rangle = -\langle w, v_n\rangle$, while $\langle w, v_i\rangle = \langle w, v_{i+1}\rangle \pm 2$ for each $i = 0, 1, \ldots, n-1$.

(ii) $K(n,d) \leq \lceil n/(d+1)\rceil$ (Alon et al. 1988). *Hint*: Consider the same vectors as before, and select only the vectors $u_j := v_{j\cdot(d+1)+1}$ for $j = 0, 1, \ldots, r$; $r = \lceil n/(d+1)\rceil - 1$. Observe that for any ± 1 vector w and any j, $0 \leq j < r$, $\langle w, u_j\rangle = \langle w, u_{j+1}\rangle \pm (2d+2)$. Note: Alon et al. (1988) have also proved that this upper bound is tight.

14.10. (Alon et al. 1988). Let V be the set of all ± 1 vectors of length n. A vector is *even* if it has an even number of -1's, otherwise it is *odd*. Let $f(x_1, \ldots, x_n)$ be a multilinear polynomial of degree less than $n/2$ over the reals, i.e.,

$$f = \sum_{|S|<n/2} \alpha_S \prod_{i\in S} x_i,$$

where $\alpha_S \in \mathbb{R}$. Suppose that $f(v) = 0$ for every even vector $v \in V$. Prove that then $f \equiv 0$, i.e., $\alpha_S = 0$ for all S. Does the same hold if $f(v) = 0$ for every odd vector $v \in V$? *Hint*: By the hypothesis, for every even subset $T \subseteq N$ we have

$$\sum_{|S|<n/2} \alpha_S (-1)^{|S\cap T|} = 0.$$

It thus suffices (why?) to show that the rows of the matrix

$$A = \left\{(-1)^{|S\cap T|} \;:\; |T| \text{ even and } |S| < n/2\right\}$$

are linearly independent (over the reals). For this, show that the matrix $M = A^T A$ has non-zero determinant. The (S_1, S_2)-th entry of M is the sum

$$\sum_T (-1)^{|S_1\cap T|+|S_2\cap T|} = \sum_T (-1)^{|(S_1\oplus S_2)\cap T|}$$

over all even T. If $S_1 = S_2$, then this sum is 2^{n-1}. If $S_1 \neq S_2$, then $0 < |S_1 \oplus S_2| < n$; use this to show that in this case the sum is 0.

14.11. Let $n = 2m+1$, and consider the majority function $\mathrm{MAJ}_n(x_1, \ldots, x_n)$, which outputs 1 iff $x_1 + \ldots + x_n \geq m + 1$. The best known upper bound $O(n^{4.57})$ for the formula size of MAJ_n is due to Valiant. Use Khrapchenko's theorem to show that this function requires formulas of size $\Omega(n^2)$. *Hint*: Take $A = \{a : |a| = m + 1\}$ and $B = \{b : |b| = m\}$.

14.12. Show that Khrapchenko's theorem cannot yield larger than quadratic lower bounds. *Hint*: Each vector in $\{0, 1\}^n$ has only n neighbors.

14.13. Research problem: It is not known if the converse of Rychkov's lemma (Lemma 14.14) holds. Suppose that $A \times B$ can be covered by t mutually disjoint rectangles. Does there then exist a formula which separates A, B and has size at most t^c for some absolute constant c?

14.14. Let $V \subseteq \mathbb{F}_2^n$ be a linear space and $y \in \mathbb{F}_2^n$ be a vector. Assume that $y \notin V^\perp$. Show that then $v \cdot y = 0$ for precisely one-half of the vectors v in V. *Hint*: Split V into V_0 and V_1 according to whether $v \cdot y = 0$ or $v \cdot y = 1$. Take $x \in V$ such that $x \cdot y = 1$; hence, $x \in V_1$. Show that then $x + V_0 \subseteq V_1$, $x + V_1 \subseteq V_0$, $|x + V_0| = |V_0|$ and $|x + V_1| = |V_1|$.

14.15. The *general disjointness matrix* D_n is a $2^n \times 2^n$ 0-1 matrix whose rows and columns are labeled by the subsets of an n-element set, and the (A, B)-th entry is 1 if and only if $A \cap B = \emptyset$. Prove that this matrix has full rank, i.e., that $\mathrm{rk}(D_n) = 2^n$. *Hint*: Use the induction on n together with the following recursive construction of D_n:

$$D_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \qquad D_n = \begin{pmatrix} D_{n-1} & D_{n-1} \\ D_{n-1} & 0 \end{pmatrix}$$

14.16. The *intersection matrix* Q_n is a $(2^n - 1) \times (2^n - 1)$ 0-1 matrix whose rows and columns are labeled by the *non-empty* subsets of an n-element set, and the (A, B)-th entry is 1 if and only if $A \cap B \neq \emptyset$. Prove that this matrix also has full rank over any field. *Hint*: If we subtract D_n from the all-1 matrix I_n then we get a matrix Q_n with one additional null column and row. Combine this fact with (13.1) and the previous exercise.

15. Eigenvalues and Graph Expansion

A very important class of sparse graphs consists of *expander* graphs. Among other things, they are the model for a good network. They are also used to derandomize algorithms as well as to construct good error-correcting codes. Basically, an expander has the property that every subset of its vertices has a large set of neighbors. This particularly implies that any pair of vertices is connected by a short path.

In general, it is difficult to decide whether a given graph is a good expander: One must test whether *all* subsets of vertices have many neighbors. Fortunately, linear algebra can help us in this situation. Namely, it turns out that a graph is a good expander if it has large *spectral gap*, that is, if the difference between the first and the second largest eigenvalues of its adjacency matrix is large.

15.1 Expander graphs

For a graph $G = (V, E)$ and a vertex $u \in V$, let $\Gamma(u)$ denote the set of neighbors of u, that is, the set of all vertices adjacent to u. For a subset $S \subseteq V$, its neighborhood is defined as the set

$$\Gamma(S) = \{v \in V \setminus S : \ v \text{ is adjacent to some vertex } u \in S\}$$

of all proper neighbors of S.

An (n, d, c)-*expander* is a d-regular graph $G = (V, E)$ on n vertices such that every subset $S \subseteq V$ with $|S| \leq n/2$ is connected by edges of G to at least $c|S|$ vertices outside the set S, that is,

$$|\Gamma(S)| \geq c|S| \text{ for all } S \subseteq V \text{ with } |S| \leq n/2.$$

The smaller the degree d and the larger the expansion constant $c > 0$ is, the better the expander we have.

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science. An EATCS Series, DOI 10.1007/978-3-642-17364-6_15,

To see that a small degree expander makes a good network, suppose we want to route a message from a vertex x to another vertex y in an (n, d, c)-expander G. Every vertex has degree d. By the property of an expander, there are at least $(1 + d)(1 + c)$ vertices at distance ≤ 2 from x. Working outwards from there, there will be at least $(1+d)(1+c)^k$ vertices at distance $\leq k + 1$ from x. We can continue expanding from x until the reachable set V_x of vertices has $|V_x| > n/2$ vertices. The vertex y may not be among them. But if we expand from y in the same way, we eventually obtain a set V_y of $|V_y| > n/2$ vertices reachable from y. Since both sets V_x and V_y have more than $n/2$ vertices, they must overlap. The overlap contains vertices on a path from x to y.

In this way, we have shown that any two vertices x and y are connected by a path of length at most $2(k + 1)$, as long as

$$k > \log_{1+c} \frac{n}{2(1 + d)} \,.$$

If $c > 0$ is a constant, then any two vertices are connected by a path of length *logarithmic* in the total number n of vertices.

15.2 Spectral gap and the expansion

The basic question about expanders is: If A is the adjacency matrix of a graph G, what properties of A ensure that G is a good expander? All graphs considered in this chapter are undirected. Recall that the *adjacency matrix* of a graph G on vertices $\{1, 2, \ldots, n\}$ is an $n \times n$ 0-1 matrix $A = (a_{ij})$ with $a_{ij} = 1$ iff i and j are adjacent in G. Note that A is *symmetric*, that is, $a_{ij} = a_{ji}$ for all i, j.

Of course, there is a trivial *combinatorial* property of A ensuring that G is an (n, d, c)-expander: For every subset $S \subseteq [n]$ of $|S| \leq n/2$ rows, at least $c|S|$ columns outside S must have at least one 1 in these rows. This answer is, however, not satisfactory because the property must hold for *all* subsets S and it is difficult to test all these $2^{\Omega(n)}$ possibilities. What we would like to have is an *algebraic* condition on A ensuring good expansion of G. It turns out that this property is captured by the two largest eigenvalues of A.

By eigenvalues of a graph $G = ([n], E)$ we will mean the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ of its adjacency matrix $A = (a_{ij})$ with $a_{ij} = 1$ iff i and j are adjacent in G. Since A has zeroes on the diagonal, its trace is equal to 0. Hence, we always have that $\lambda_1 + \cdots + \lambda_n = 0$.

Example 15.1. The complete graph K_n has an adjacency matrix equal to $A = J - I$, where J is the all-1 matrix and I is the identity matrix. The rank of J is 1, i.e. there is one nonzero eigenvalue equal to n, with an eigenvector $\mathbf{1} = (1, 1, ..., 1)$. All the remaining eigenvalues are 0. Subtracting the identity shifts all eigenvalues by -1, because $Ax = (J - I)x = Jx - x$. Therefore the eigenvalues of K_n are $\lambda_1 = n - 1$ and $\lambda_2 = \ldots = \lambda_n = -1$

If G is d-regular, then $\mathbf{1} = (1, 1, ..., 1)$ is an eigenvector. We get $A\mathbf{1} = d\mathbf{1}$, and hence d is an eigenvalue. It is easy to show that no eigenvalue can be larger than d (see Exercises 15.7 and 15.8).

Interestingly, already the difference $d - \lambda_2$ (known as the *spectral gap*) between the degree d of a d-regular graph and its second-largest eigenvalue λ_2 gives us a lot of information about the expansion properties of the corresponding graphs. The larger this difference is, the better expansion properties the graph has.

For two not necessarily disjoint subsets $S, T \subseteq V$ of vertices, let $e(S, T)$ denote the number of edges of G with one endpoint in S and the other in T. If $S \cap T = \emptyset$, then $0 \leq e(S, T) \leq |S| \cdot |T|$. In this case, $e(S, T)$ is the number of "crossing edges" between the sets S and T. Also, let $e(S)$ denote the number of edges, both endpoints of which lie in S.

Lemma 15.2 (Expander Crossing Lemma)**.** *Let $G = (V, E)$ be a d-regular graph on n vertices $V = \{1, \ldots, n\}$, and let $\lambda = \lambda_2$ be the second largest eigenvalue of its adjacency matrix A. Then, for every partition $V = S \cup T$,*

$$e(S, T) \geq \frac{(d - \lambda)|S| \cdot |T|}{n}\,. \tag{15.1}$$

Proof. By Lemma 13.5, we know that $\lambda = \lambda_2$ is the maximum of the Rayleigh quotient $x^\top Ax/\|x\|^2$ over all vectors x such that $\langle x, \mathbf{1}\rangle = 0$. So, to get a lower bound on λ we can plug any vector $x \perp \mathbf{1}$ into the Rayleigh quotient. For this purpose, we take the following vector x related to our partition $V = S \cup T$. Let $s = |S|$ and $t = |T| = n - s$. We can assume w.l.o.g. that $S = \{1, 2, \ldots, s\}$ and $T = \{s + 1, \ldots, n\}$. Consider the vector $x = (x_1, \ldots, x_n)$ with $x_i = -t$ for $i \in S$ and $x_i = s$ for $i \in T$. That is,

$$x = (\overbrace{-t, -t, \ldots, -t}^{|S| \text{ times}}, \overbrace{s, s, \ldots, s}^{|T| \text{ times}})\,.$$

Then

$$\langle x, \mathbf{1}\rangle = \sum_{i \in S}(-t) + \sum_{i \in T} s = s(-t) + ts = 0$$

Also

$$\|x\|^2 = \langle x, x\rangle = \sum_{i \in S}(-t)^2 + \sum_{i \in T} s^2 = st^2 + ts^2 = st(s + t) = stn\,.$$

By (13.7),

$$x^\top Ax \leq \lambda\|x\|^2 = \lambda stn. \tag{15.2}$$

Our goal now is to show that $x^\top Ax = stdn - e(S, T)n^2$, from which the desired lower bound on $e(S, T)$ follows.

Letting $\Gamma(i)$ denote the set of all neighbors of vertex i, we have

$$\begin{aligned} x^\top Ax &= \sum_{i=1}^n x_i\Big(\sum_{j\in\Gamma(i)} x_j\Big) = 2\sum_{\{i,j\}\in E} x_ix_j \\ &= 2t^2\cdot e(S) + 2s^2\cdot e(T) - 2st\cdot e(S,T)\,. \end{aligned} \tag{15.3}$$

To eliminate $e(S)$ and $e(T)$, observe that the sum of degrees $\sum_{i\in S} d_i$ of vertices in S is equal to $d|S| = ds$, since the graph is d-regular, and is equal to $2e(S)+e(S,T)$, since each edge with both endpoints in S is counted twice. This gives $2e(S) = ds - e(S,T)$. Similarly, $2e(T) = dt - e(S,T)$. Substituting this in (15.3) we obtain

$$\begin{aligned} x^\top Ax &= (ds - e(S,T))t^2 + (dt - e(S,T))s^2 - 2st\cdot e(S,T) \\ &= (dst^2 + dts^2) - (s^2 + 2st + t^2)\cdot e(S,T) \\ &= std(s+t) - (s+t)^2\cdot e(S,T) \\ &= stdn - n^2\cdot e(S,T)\,. \end{aligned}$$

Together with (15.2), this implies

$$e(S,T) \geq \frac{\lambda stn - stdn}{n^2} = \frac{(d-\lambda)st}{n}\,. \qquad \square$$

In particular, this lemma implies that if $\lambda < d$, then for any partition of the vertices, there is a "crossing edge" going from one part to another. That is, $\lambda < d$ implies that the graph is connected. In fact, the converse holds as well.

Proposition 15.3. *A d-regular graph with second-largest eigenvalue λ_2 is connected if and only if $\lambda_2 < d$.*

Proof. Let G be connected and d-regular. Suppose on the contrary that there is a vector $x \in \mathbb{R}^n$, $x \neq \mathbf{0}$ such that $\langle x, \mathbf{1}\rangle = 0$ and $Ax = dx$. Let x_i be the smallest and x_j the largest entry of x. Since $\langle x, \mathbf{1}\rangle = 0$ and $x \neq \mathbf{0}$, we have that $x_i < 0$ and $x_j > 0$. Take $c := -1/x_i$ and consider the vector $y := \mathbf{1} + cx$. Then $y \geq \mathbf{0}$. Moreover, $y_i = 1 + cx_i = 1 - 1 = 0$ and $y_j = 1 + cx_j = 1 - x_j/x_i > 1$. But

$$Ay = A\mathbf{1} + cAx = d\mathbf{1} + cdx = dy\,,$$

and hence, $A^ty = d^ty$ for all t. In particular,

$$A^t[i,1]y_1 + \cdots + A^t[i,j]y_j + \cdots + A^t[i,n]y_n = dy_i = 0\,.$$

Together with $y \geq \mathbf{0}$ and $y_j > 0$, this implies that $A^t[i,j] = 0$ for any t, that is, there is no walk joining i and j, a contradiction with the connectedness of G. □

An important consequence of the Expander Crossing Lemma is that every d-regular graph whose second-largest eigenvalue is strictly smaller than d, is a good enough expander.

Theorem 15.4. *If $\lambda = \lambda_2$ is the second-largest eigenvalue of the incidence matrix of a d-regular graph G on n vertices, then G is an (n, d, c)-expander for $c = (d - \lambda)/2d$.*

Proof. Let S be a subset of $|S| \leq n/2$ vertices, and let $\Gamma(S)$ be the set of all neighbors of S in the complement $\overline{S}$ of S. An edge can lie between S and $\overline{S}$ only if one its endpoint belongs to S and the other to $\Gamma(S)$. Since every vertex in $\Gamma(S)$ has at most d neighbors in S, this implies that $e(S, \overline{S}) \leq d|\Gamma(S)|$. Together with Lemma 15.2 this implies

$$|\Gamma(S)| \geq \frac{(d - \lambda)|S|(n - |S|)}{dn} \geq \frac{d - \lambda}{2d},$$

where the last inequality follows since $n - |S| \geq n/2$. □

For not necessarily disjoint subsets of vertices S and T, we have a slightly worse lower bound on $e(S, T)$.

Lemma 15.5 (Expander Mixing Lemma)**.** *If G is a d-regular graph on n vertices and $\lambda = \lambda_2$ is the second-largest eigenvalue of its adjacency matrix then, for every two subsets S and T of vertices,*

$$\left| e(S, T) - \frac{d|S| \cdot |T|}{n} \right| \leq \lambda \sqrt{|S| \cdot |T|}\,.$$

The left-hand side measures the deviation between two quantities: one is $e(S, T)$, the number of edges between the two sets; the other is the expected number of edges between S and T in a random graph of edge density d/n, namely $d|S||T|/n$. A small λ (or large spectral gap) implies that this deviation (or discrepancy as it is sometimes called) is small, so the graph behaves like a random graph in this sense!

Proof. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of the adjacency matrix A of G, and let $x^1, \ldots, x^n$ be the corresponding orthonormal basis; here x^1 is $\frac{1}{\sqrt{n}}$ times the all-1 vector $\mathbf{1}$. Let v_S and v_T be the characteristic vectors of S and T. Expand these two vectors as linear combinations

$$v_S = \sum_{i=1}^{n} a_i x^i \quad \text{and} \quad v_T = \sum_{i=1}^{n} b_i x^i$$

of the basis vectors. Since the x^i are orthonormal eigenvectors,

$$e(S, T) = v_S^\top A v_T = \Big(\sum_{i=1}^{n} a_i x^i \Big)^\top A \Big(\sum_{i=1}^{n} b_i x^i \Big) = \sum_{i=1}^{n} \lambda_i a_i b_i\,. \tag{15.4}$$

Since the graph G is d-regular, we have $\lambda_1 = d$. The first two coefficients a_1 and b_1 are scalar products of $x^1 = \frac{1}{\sqrt{n}}\mathbf{1}$ with v_S and v_T; hence, $a_1 = |S|/\sqrt{n}$ and $b_1 = |T|/\sqrt{n}$. Thus, the first term $\lambda_1 a_1 b_1$ in the sum (15.4) is precisely $d|S||T|/n$. Since $\lambda = \lambda_2$ is the second largest eigenvalue, the absolute value of the sum of the remaining $n-1$ terms in this sum does not exceed $\lambda \sum_{i=2}^{n} |a_i b_i|$ which, by the Cauchy–Schwarz inequality, does not exceed

$$\lambda\|a\| \cdot \|b\| = \lambda\|v_S\| \cdot \|v_T\| = \lambda\sqrt{|S| \cdot |T|}\,. \qquad \square$$

There are several *explicit* constructions of constant degree expanders with a constant expansion factor $c > 0$. We mention just two of them:

1. Vertices are pairs of integers $x \in \mathbb{Z}_m = \{0, 1, \ldots, m-1\}$, and each vertex (x, y) is connected to four vertices $(x+y, y), (x-y, y), (x, y+y), (x, x-y)$, where all operations are modulo m.
2. Vertices are elements of the field $\mathbb{Z}_p$ (for p a prime number), and every vertex $x \neq 0$ is connected to three vertices $x+1, x-1, x^{-1}$, where again, $x \pm 1$ is computed modulo p and x^{-1} is the multiplicative inverse of x in $\mathbb{Z}_p$. The vertex $x = 0$ is connected to 0, 1 and $p-1$.

15.2.1 Ramanujan graphs

The second-largest-eigenvalue λ of d-regular graphs lies roughly between $\sqrt{d}$ and d. More precisely, it is known that the second eigenvalue is always at least $2\sqrt{d-1} - o(1)$. Graphs achieving this lower bound are called Ramanujan graphs. That is, a *Ramanujan graph* is a d-regular graph whose second-largest eigenvalue λ satisfies $\lambda \leq 2\sqrt{d-1}$.

Example 15.1 shows that the complete graph K_n is a Ramanujan graph for any n. The problem, however is, that K_n has huge degree. In applications we need an explicit sequence G_n, $n = 1, 2, \ldots$ of graphs such that infinitely many of them have some (fixed) constant degree d and are Ramanujan (or "nearly" Ramanujan) graphs.

Explicit constructions of $(p+1)$-regular Ramanujan graphs on n vertices for every prime $p \equiv 1 \bmod 4$ and infinitely many values of n were given in Margulis (1973), Lubotzky, Phillips and Sarnak (1988); these were later extended to the case where p is an arbitrary prime power in Morgenstern (1994) and Jordan and Livné (1997).

Most important in these (rather non-trivial) constructions is that the degree $p+1$ of constructed n-vertex graphs is *constant*, it does not grow with n. On the other hand, Ramanujan n-vertex graphs of degree about $\sqrt{n}$ can be constructed quite easily.

In Sect. 2.2 we described explicit n-vertex graphs with no 4-cycles and almost maximal number $\Omega(n^{3/2})$ of edges (see Exercise 2.5). Now we will show that these graphs are Ramanujan graphs. Later, in Sect. 25.4, we will combine this fact with the expander mixing lemma to prove an important result in extremal number theory about sum-product sets.

Let $n = p(p-1)$, where p is a prime number. The vertices of our graph G are pairs (a,b) of elements of a finite field $\mathbb{Z}_p$ with $a \neq 0$, and two vertices (a,b) and (c,d) are joined by an edge iff $ac = b+d$ (all operations modulo p). We have already shown that this graph is $(p-1)$-regular, that is, every vertex has $p-1$ incident edges (some edges may be loops). We have also shown that any two vertices (a,b) and (c,d) have

(i) no common neighbor, if $a = c$ or $b = d$;
(ii) exactly one common neighbor, if $a \neq c$ and $b \neq d$.

Now we will show that, for any prime number $p \geq 5$, the graph G is a Ramanujan graph. Let $\lambda = \lambda(G)$ be the second-largest eigenvalue of the adjacency matrix of G.

Lemma 15.6. $|\lambda| < \sqrt{3p}$.

Proof. Let M be the adjacency matrix of G. The (u,v)-entry of M^2 is the number of walks from u to v of length 2. If $u = v$, this number is the degree $p-1$, while if $u \neq v$, with $u = (a,b)$ and $v = (c,d)$, then properties (i) and (ii) tell us that this number is 1 if $a \neq c$ and $b \neq d$, and is 0 otherwise. It follows that

$$M^2 = J + (p-2)I - E\,, \tag{15.5}$$

where J is the all-1 matrix, I is the identity matrix, and E is the "error matrix," the adjacency matrix of the graph G_E whose vertex set is the same as that of G, and in which two vertices (a,b) and (c,d) are connected by an edge if $a = c$ or $b = d$. It is easy to see that G_E is a $(2p-3)$-regular graph.

Since G is regular and its adjacency matrix M is symmetric, we know that the all-1 vector is an eigenvector of M and all other eigenvectors are orthogonal to it. It is easy to check that G is connected and not bipartite, so that the eigenvalue $p-1$ has multiplicity 1, and for any other eigenvalue θ we have $|\theta| < p-1$.

Given such an eigenvalue θ, let x be a corresponding eigenvector. Then by equation (15.5),

$$\theta^2 x = (p-2)x - Ex\,,$$

since Jx is the all-0 vector. Therefore $p-2-\theta^2$ is an eigenvalue of E.

Now, the degree $2p-3$ of G_E is an upper bound on the absolute value of every eigenvalue of E (see Exercise 15.8). It follows that

$$p - 2 - \theta^2 \geq -2p + 3$$

which implies $|\theta| < \sqrt{3p}$, as desired. □

15.3 Expanders and derandomization

Random algorithms use random bits (results of coin-flips) during the computation and are allowed to produce a wrong answer with some small probability. Such algorithms are usually much faster than known deterministic algorithms. But we must pay for this: we must expect errors and it is time consuming to produce random bits. It turns out that expander graphs can help to decrease the error-probability as well as to reduce the number of required random bits.

Suppose we have a boolean function $f : \{0,1\}^n \to \{0,1\}$ and a probabilistic polynomial time algorithm $\mathcal{A}$ that approximates f in the sense that, for random $r \in \{0,1\}^m$ we have:

$$\Pr_r\left[\mathcal{A}(x,r) \neq f(x)\right] \leq \frac{1}{4} \qquad \text{for every } x \in \{0,1\}^n. \tag{15.6}$$

We could reduce the error to 4^{-t} by running the algorithm $2t+1$ times and taking the majority of its outputs as the result. But this requires $(2t+1)m$ coin tosses. We want to reduce errors while using a small number of coin tosses. (A general procedure, when we reduce the number of random bits by modifying a probabilistic algorithm, is called *derandomization*.)

Take a d-regular graph $G = (V,E)$ on $|V| = 2^m$ vertices, and let $\lambda = \lambda_2$ be the second-largest eigenvalue of its adjacency matrix. Let us consider the following algorithm $\mathcal{B}$ that uses only m coin tosses. For a given input x, it picks a vertex $v \in V$ uniformly at random, and outputs

$$\mathcal{B}(x,v) := \text{Majority}_{u\in\Gamma(v)}\mathcal{A}(x,v)\,.$$

Claim 15.7. For every $x \in \{0,1\}^n$,

$$\Pr_v\left[\mathcal{B}(x,v) \neq f(x)\right] \leq 4\left(\frac{\lambda}{d}\right)^2.$$

Proof. Fix an input x. Let $S = \{v \in V\colon \mathcal{B}(x,v) \neq f(x)\}$ be the set of vertices on which $\mathcal{B}$ errs, and $T = \{v \in V\colon \mathcal{A}(x,v) \neq f(x)\}$ be the set of vertices on which $\mathcal{A}$ errs. Observe that every vertex $u \in S$ must be adjacent to at least $d/2$ vertices $v \in T$, implying that $e(S,T) \geq d|S|/2$. Moreover, $|T| \leq |V|/4 = n/4$, by (15.6). The Expander Mixing Lemma yields:

$$\begin{aligned}
e(S,T) - \frac{d|S|\cdot|T|}{n} &\leq \lambda\sqrt{|S|\cdot|T|}\\
\frac{d|S|}{2} - \frac{d|S|}{4} &\leq \lambda\sqrt{|S|n/4}\\
\frac{d|S|}{4} &\leq \lambda\sqrt{|S|n/4}
\end{aligned}$$

from which

$$\Pr_v[\mathcal{B}(x,v) \neq f(x)] = \frac{|S|}{n} \leq 4\left(\frac{\lambda}{d}\right)^2 .$$

follows. □

So, taking Ramanujan graphs the error probability can be reduced to $4(2/\sqrt{d})^2$ without any increase of the number of random bits!

We will present yet another application of expander graphs to reduce the number of random bits in Sect. 23.3. More applications of expanders as well as their constructions can be found in a beautiful survey paper by Hoory, Linial and Wigderson (2006).

Exercises

15.1 (Unique neighbors)**.** Let $G = (V, E)$ and $S \subseteq V$. A *unique neighbor* of S is a vertex in $\Gamma(S)$ connected by an edge to only one vertex in S. Suppose that G is an (n, d, c)-expander. Show that then every subset S of size $|S| \leq n/2$ has at least $(c - d/2)|S|$ unique neighbors. *Hint*: Let $T \subseteq \Gamma(S)$ be the set of non-unique neighbors and count the number of edges between S and T in two ways.

15.2. Let A be a square symmetric matrix, and λ one of its eigenvalues. Show that, for every integer $k \geq 1$, λ^k is an eigenvalue of A^k.

15.3. Let G be a d-regular graph on n vertices, and A its adjacency matrix. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of A. Show that the eigenvalues of the adjacency matrix of the complement graph $\overline{G}$ are $n - 1 - d$ and $-1 - \lambda_i$ for $i = 2, \ldots, n$. *Hint*: The adjacency matrix of $\overline{G}$ is $J - I - A$. If vector x is orthogonal to $\mathbf{1}$, then $Jx = \mathbf{0}$.

15.4. Let G be a *bipartite* d-regular graph on n vertices, and A its adjacency matrix. Show that $-d$ is also an eigenvalue of A. *Hint*: If G is bipartite with parts of size p and q with $p + q = n$, then

$$A = \begin{bmatrix} \mathbf{0} & B \\ B^\top & \mathbf{0} \end{bmatrix}$$

for a $p \times q$ matrix B. Take the vector

$$x = (\underbrace{1, \ldots, 1}_{p}, \underbrace{-1, \ldots, -1}_{q}) .$$

15.5. Let $d > 1$ be a constant, and A be the adjacency matrix of a d-regular graph on n vertices. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of A; hence, $\lambda_1 = d$. Let $\lambda = \max_{i \neq 1} |\lambda_i|$. Show that $\lambda \geq (1 - o(1))\sqrt{d}$ as $n \to \infty$. *Hint*: Use the fact that $\lambda_1 + \ldots + \lambda_n$ is the trace of A to estimate the trace of A^2.

15.6. Let A be a square 0-1 matrix with exactly d ones in each row and in each column. Show that then $x^\top Ax \leq d$ holds for every vector $x \in \mathbb{R}^n$ with $\|x\| = 1$. *Hint*: The Birkhoff–Von Neumann theorem and Exercise 13.21.

15.7. Let G be a graph, and A its adjacency matrix. Let δ be the average degree and $\varDelta$ the maximum degree of G. Let λ_{max} be the largest eigenvalue of A. Show that $\delta \leq \lambda_{\max} \leq \varDelta$. *Hint*: Let $\lambda = \lambda_{max}$, take a vector x with $Ax = \lambda x$ and give an upper bound on $|\lambda| \max_i |x_i|$. For the lower bound apply Theorem 13.5 with $x = \mathbf{1}$.

15.8. Use the previous exercise to show that $\lambda_1 = d$ is the largest eigenvalue of the adjacency matrix of every d-regular graph.

15.9. Show that a scalar λ is an eigenvalue for A with eigenvector x if and only if $|x^\top Ax| = \|Ax\| \cdot \|x\|$.

15.10. Let $A = (a_{ij})$ be an upper triangular matrix, that is, $a_{ij} = 0$ for all $i > j$. Show that then the diagonal elements $a_{11}, a_{22}, \ldots, a_{nn}$ are the eigenvalues of A. *Hint*: Recall that eigenvalues of A are the roots of the characteristic polynomial $p_A(z) = \det(A - zI)$ of A, where I is a unit matrix with 1s on the diagonal, and 0s elsewhere.

15.11. Let A be a symmetric $n \times n$ 0-1 matrix with 1s on the diagonal, and let $|A|$ be the total number of 1-entries in A. Show that then $\mathrm{rk}(A) \geq n^2/|A|$. *Hint*: Consider the trace of A^2, and recall that $\mathrm{rk}(A)$ is the number of (not necessarily distinct) nonzero eigenvalues of A.

15.12. Let A be a real $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $x_1, \ldots, x_n$. Let $I \subseteq [n]$ be such that all eigenvalues λ_i with $i \in I$ are distinct.

1. Show that the vectors x_i with $i \in I$ are linearly independent.
2. Let A be symmetric. Show that the eigenvectors corresponding to different eigenvalues are mutually orthogonal, that is, $\langle x_i, x_j \rangle = 0$ for all $i \neq j \in I$. *Hint*: Show that $y^\top Ax = x^\top Ay$ holds for all vectors x, y.

15.13. Let $G = (X \cup Y, E)$ be the point-line incidence graph of the projective plane $\mathrm{PG}(2, q)$ (see Sect. 12.4). That is, vertices in X correspond to points, vertices in Y to lines of $\mathrm{PG}(2, q)$, and a point x is joined to a line y by an edge iff $x \in y$. The graph has $|X| = |Y| = n := q^2 + q + 1$ vertices on each side, and each vertex has degree $q + 1$. The most important property of this graph is that it contains no copies of $K_{2,2}$, that is, it contains no cycle on four vertices. Show that every subset $S \subseteq X$ has at least $(1 - q/|S|)n$ neighbors in Y. *Hint*: Corrádi's lemma, Lemma 2.1).

16. The Polynomial Method

This method is based on various extensions of the following basic fact about univariate (single-variable) polynomials —known as the “factor theorem”—to the case of multivariate polynomials, that is, polynomials on many variables:

(i) Every nonzero polynomial of degree d has at most d roots.
(ii) For every set S of points there exists a nonzero polynomial f of degree at most $|S|$ such that $f(x) = 0$ for all $x \in S$.

Thus, to obtain an *upper* bound on the size of a given set S, it is enough to exhibit a nonzero low-degree polynomial that vanishes on S; conversely, to *lower* bound the size of S, it is enough to show that the only low-degree polynomial that vanishes on S is the zero polynomial.

16.1 DeMillo–Lipton–Schwartz–Zippel lemma

Let $x_1, \ldots, x_n$ be variables. A *monomial* of degree t is a product $x_1^{t_1} x_2^{t_2} \cdots x_n^{t_n}$ with integers $t_i \geq 0$ such that $t_1 + t_2 + \cdots + t_n = t$. Hence, constant 1 is the only monomial of degree 0. For a fixed field $\mathbb{F}$, let $\mathbb{F}[x_1, \ldots, x_n]$ denote the ring of all multivariate polynomials over $\mathbb{F}$. Each such polynomial is a linear combination of monomials with coefficients taken from $\mathbb{F}$. The *degree*, $\deg(f)$, of f is the maximum degree among its monomials with a nonzero coefficient. A (multivariate) polynomial is *homogeneous* if all its monomials with nonzero coefficients have the same degree. A polynomial f *vanishes* on a subset $E \subseteq \mathbb{F}^n$ if $f(x) = 0$ for all $x \in E$. A point $x \in \mathbb{F}^n$ with $f(x) = 0$ is a *root* of f. A polynomial $f(x)$ is the *zero polynomial* if all its coefficients are 0.

The following lemma extends the second claim (ii) of the factor theorem to multivariate polynomials.

Lemma 16.1. *Given a set $E \subseteq \mathbb{F}^n$ of size $|E| < \binom{n+d}{d}$, there exists a nonzero polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ of degree at most d that vanishes on E.*

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_16,

Proof. Let V_d be the vector space of polynomials in $\mathbb{F}[x_1, \ldots, x_n]$ of degree at most d. It is not difficult to show (see Exercise 13.22) that V_d has dimension $\binom{n+d}{n}$. On the other hand, the vector space $\mathbb{F}^E$ of all functions $g : E \to \mathbb{F}$ has dimension $|E| < \binom{n+d}{d}$. Hence the evaluation map $f \mapsto (f(a))_{a \in E}$ from V_d to $\mathbb{F}^E$ is non-injective. So, at least two polynomials f_1 and f_2 in V_d are mapped to the same string in $\mathbb{F}^E$. But then the polynomial $f = f_1 - f_2$ belongs to V_d and is mapped to the all-0 string, meaning that f vanishes on E. □

The first claim (i) of the factor theorem can be also extended to the case of multivariate polynomials by analyzing the behavior of polynomials on lines.

Lemma 16.2. *Every nonzero polynomial $f(x_1, \ldots, x_n)$ of degree d over a finite field with q elements has at most dq^{n-1} roots.*

Proof (due to Dvir 2009 and Moshkovitz 2010). Let us assume $n \geq 2$, $1 \leq d \leq q$, where $q = |\mathbb{F}|$. The proof is by reduction to the case $n = 1$. Write $f = g + h$, where g is homogeneous of degree d, and h contains only monomials of degree strictly smaller than d. Since f is a nonzero polynomial, $g(w) \neq 0$ for some $w \in \mathbb{F}^n$, $w \neq \mathbf{0}$. Associate with each vector $u \in \mathbb{F}^n$ the line

$$L_u = \{u + tw \colon t \in \mathbb{F}\}$$

in direction w through u. Then $L_u \cap L_v = \emptyset$ as long as $v \notin L_u$. Since $w \neq \mathbf{0}$, each line L_u contains $|L_u| = q$ points. Hence, we can partition $\mathbb{F}^n$ into $q^n/q = q^{n-1}$ lines. It remains therefore to show that the number of zeros of f on each of the lines L_u is at most d.

To show this, observe that, for every $u \in \mathbb{F}^n$, the function $p_u(t) = f(u+tw)$ is a univariate polynomial in t of degree at most d. Moreover, this polynomial is not identically zero, because the coefficient of t^d in $p_u(t)$ is $g(w) \neq 0$. Thus, $p_u(t)$ can have at most d roots, implying that the polynomial f can vanish on at most d points of the line L_u Since we have only q^{n-1} lines in a partition of $\mathbb{F}^n$, the total number of roots of f cannot exceed dq^{n-1}, as claimed. □

A more general result was proved by DeMillo and Lipton (1978), Zippel (1979) and Schwartz (1980). The lemma bounds the probability that a nonzero multivariate polynomial will have roots at randomly selected test points.

Lemma 16.3 (DeMillo–Lipton–Schwartz–Zippel lemma)**.** *For every set $S \subseteq \mathbb{F}$ of $|S| \geq d$ field elements, every nonzero polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ of degree d can have at most $d|S|^{n-1}$ roots in S^n.*

Proof. Suppose that f is a nonzero polynomial. We proceed by induction on n, the number of variables of f. The statement is true for $n = 1$ since the number of roots of f does not exceed its degree. Now let $n \geq 2$ and arrange f according to the powers of x_n:

$$f = f_0 + f_1 x_n + f_2 x_n^2 + \cdots + f_t x_n^t$$

where $f_0, \dots, f_t$ are polynomials of the $n-1$ variables $x_1, \dots, x_{n-1}$, the term f_t is not identically 0, and $t \leq d$. Our goal is to estimate for how many of the points $(a, b) \in S^{n-1} \times S$, $f(a, b) = 0$. We distinguish between two cases:

Case 1. $f_t(a) = 0$. Since f_t is nonzero and has total degree $\leq d-t$, we have by the induction hypothesis, that it can vanish on at most $(d-t)|S|^{n-2}$ points in S^{n-1}. Therefore, in this case, there are at most $(d-t)|S|^{n-1}$ points $(a, b) \in S^{n-1} \times S$ for which $f(a, b) = 0$ and $f_t(a) = 0$.

Case 2. $f_t(a) \neq 0$. For every (fixed) point $a \in S^{n-1}$ for which $f_t(a) \neq 0$, the polynomial $f(a, x_n)$ is a polynomial in one variable of degree t, and it is not identically zero. Therefore it has at most t roots. Since there are at most $|S|^{n-1}$ such points a, the number of points $(a, b) \in S^{n-1} \times S$ for which $f(a, b) = 0$ and $f_t(a) \neq 0$, does not exceed $t \cdot |S|^{n-1}$.

Thus, there are at most $(d-t)|S|^{n-1} + t \cdot |S|^{n-1} = d \cdot |S|^{n-1}$ points $(a, b) \in S^n$ for which $f(a, b) = 0$. □

The DeMillo–Lipton–Schwartz–Zippel lemma can be used to design efficient *randomized* algorithms, that is, algorithms that are allowed to flip a coin during their computation.

The fundamental question of identity testing is: given a polynomial f of degree d on n variables, how hard is it to tell whether or not the polynomial is identically equal to zero? Note that we can only evaluate the polynomial at points of our choice, and do not have access to the coefficients of the polynomial. It is not hard to see that a deterministic algorithm that can only evaluate the polynomial, could need as many as $(d+1)^n$ points in total.

The basic idea of what is known as a *randomized algorithm* is that we write *random* numbers in place of the variables and compute the value of the polynomial. Now if the value computed is not zero then we know the answer: $f \neq 0$. But what happens if we get zero? Well, we just hit a root of f and try again. Another root? Try once more. After a number of runs we are tired and would like to stop and conclude the answer is $f = 0$. How big will the error probability of such a decision be?

To answer this question, it is helpful to reformulate Lemma 16.3 in probabilistic terms.

Lemma 16.4. *Suppose that $f(x_1, \dots, x_n)$ is a nonzero polynomial of degree d over a field $\mathbb{F}$ and $S \subseteq \mathbb{F}$ is a non-empty subset of the field elements. Then*

$$\Pr\left[f(r_1, \dots, r_n) = 0\right] \leq \frac{d}{|S|},$$

where $r_1, \dots, r_n$ are random elements selected uniformly and independently from S.

Thus, if we take $|S| = 2d$ elements of the field then, assuming $f \neq 0$, the algorithm will *not* discover this in one iteration with probability at most $1/2$. With 100 experiments repeated independently of each other, the probability that this occurs *every* time is at most 2^{-100}. So, if the algorithm does not prove that $f \neq 0$, we can be pretty certain that actually $f = 0$. Not 100% certain, but if we lose the bet, we would know that an experiment that had only two possible outcomes ended with the one that had probability 2^{-100}. This should compensate for our trouble: we found a needle in a haystack!

As our next example, consider the following situation. We have two friends, Alice and Bob. Alice maintains a large database of information. Bob maintains a second copy of the database. Periodically, they must compare their databases for consistency. Because the transmission between Alice and Bob is expensive, they would like to discover the presence of inconsistency without transmitting the entire database between them. Denote Alice's data by the sequence $a = a_0 \cdots a_{n-1}$ and Bob's data by the sequence $b = b_0 \cdots b_{n-1}$ where $a_i, b_i \in \{0, 1\}$. It is clear that any deterministic consistency check that transmits fewer than n bits will fail (just because an adversary can modify the unsent bits). Using randomness it is possible to design a strategy that detects an inconsistency with high probability (at least $1 - n^{-1}$) while transmitting many fewer than n bits, namely only $O(\log n)$ bits.

Think of the strings a and b as (strings of coefficients of) univariate polynomials over the field $\mathbb{F}_p$ where p is a prime such that $n^2 < p < 2n^2$ (theorems regarding the density of primes guarantee the existence of such p). That is, consider polynomials

$$\begin{aligned} A(x) &= a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} \pmod{p}, \\ B(x) &= b_0 + b_1 x + \ldots + b_{n-1} x^{n-1} \pmod{p}. \end{aligned}$$

In order to detect whether $a = b$, Alice and Bob use the following strategy:

> Alice picks uniformly at random a number $\boldsymbol{r}$ in $\mathbb{F}$ and sends to Bob the numbers $\boldsymbol{r}$ and $A(\boldsymbol{r})$. Bob responds with 1 if $A(\boldsymbol{r}) = B(\boldsymbol{r})$ and with 0 otherwise. The number of bits transmitted is $1 + 2\log p = O(\log n)$.

If $a = b$ then $A(\boldsymbol{r}) = B(\boldsymbol{r})$ for all $\boldsymbol{r}$, so the output is always 1. If $a \neq b$ we have two distinct polynomials $A(x)$ and $B(x)$ of degree at most $n - 1$. By Lemma 16.4, the probability of error is

$$\Pr\left[A(\boldsymbol{r}) = B(\boldsymbol{r})\right] \leq \frac{n-1}{|\mathbb{F}|} = \frac{n-1}{p} \leq \frac{1}{n}\,.$$

16.2 Solution of Kakeya's problem in finite fields

A famous unsolved problem in mathematics is the Kakeya conjecture in geometric measure theory. This conjecture is descended from the following question asked in 1917 by Japanese mathematician Soichi Kakeya: What is the smallest set in the plane in which one can rotate a needle around completely? He likened this to a samurai turning his lance around in a small toilet. For

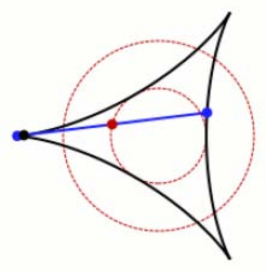


Fig. 16.1 As the middle (flexible) point moves around the smaller circle, the needle rotates through 360°.

instance, one can rotate a unit needle inside a unit disk, which has area $\pi/4$. By using a deltoid one requires only $\pi/8$ area (see Fig. 16.1).

The Kakeya conjecture in more dimensions states that any subset of $\mathbb{R}^n$ that contains a unit line segment in every direction has Hausdorff dimension equal to n. This conjecture remains open in dimensions three and higher, and gets more difficult as the dimension increases.

To approach this question, Wolff (1999) proposed a simpler *finite* field analogue of the Kakeya conjecture. If $\mathbb{F}^n$ is a vector space over a finite field $\mathbb{F}$, define a Kakeya set to be a subset $K \subseteq \mathbb{F}^n$ which contains a line in every direction, namely for any $v \in \mathbb{F}^n$ there exists a vector $w \in \mathbb{F}^n$ such that the line $\{w + tv \ : \ t \in \mathbb{F}\}$ is contained in K; here, vector w is the origin and vector v the direction of the line. The finite field Kakeya conjecture stated that there exists a constant $c > 0$ depending only on the dimension n such that every Kakeya set $K \subseteq \mathbb{F}^n$ has cardinality $|K| \geq c|\mathbb{F}|^n$.

This finite field version of the conjecture has had a significant influence on the subject, in particular inspiring work on the sum-product phenomenon in finite fields, which has since proved to have many applications in number theory and computer science. Modulo minor technicalities, the progress on the finite field Kakeya conjecture was, however, essentially the same as that of the original "Euclidean" Kakeya conjecture.

Recently Dvir (2009) used a surprisingly simple application of the polynomial method to prove the finite field Kakeya conjecture.

Lemma 16.5. *Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial of degree at most $q - 1$ over a finite field with $q = |\mathbb{F}|$ elements. If f vanishes on a Kakeya set K, then f is the zero polynomial.*

Proof. The argument is similar to that in the proof of Lemma 16.2. Suppose for a contradiction that f is nonzero. We can write $f = \sum_{i=0}^{d} f_i$, where $0 \leq d \leq q - 1$ is the degree of f and f_i is the i-th homogeneous component; thus f_d is nonzero. Since f vanishes on K, d cannot be zero. Hence, f_d is a nonzero polynomial.

Let $v \in \mathbb{F}^n \setminus \{\mathbf{0}\}$ be an arbitrary direction. As K is a Kakeya set, K contains a line $\{w + tv \,:\, t \in \mathbb{F}\}$ for some $w \in \mathbb{F}^n$, thus $f(w + tv) = 0$ for all $t \in \mathbb{F}$. The left-hand side is a polynomial $g_{w,v}(t)$ in t of degree at most $q - 1$, and must be the zero polynomial by the factor theorem, that is, all its coefficients are zero. In particular, the coefficient of t^d, which is $f_d(v)$, must be zero. Since v was arbitrary, it follows that the polynomial $f_d(x)$ vanishes on all points in $\mathbb{F}^n$. But since $dq^{n-1} \le (q-1)q^{n-1} < q^n$, Lemma 16.2 implies that f_d must be a zero polynomial. □

Theorem 16.6 (Dvir 2009). *Let $K \subset \mathbb{F}^n$ be a Kakeya set. Then*

$$|K| \ge \binom{|F| + n - 1}{n} \ge \frac{|\mathbb{F}|^n}{n!}\,.$$

Proof. Let $q = |\mathbb{F}|$ and suppose that $|K| < \binom{n+q-1}{n}$. Then, by Lemma 16.1, there exists a *nonzero* polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ of degree at most $q - 1$ that vanishes on K, which contradicts Lemma 16.5. □

16.3 Combinatorial Nullstellensatz

The following special case of Hilbert's Nullstellensatz has found numerous applications in combinatorics.

Theorem 16.7 (Nullstellensatz). *Let $f \in \mathbb{F}[x_1, \ldots, x_n]$, and let $S_1, \ldots, S_n$ be nonempty subsets of $\mathbb{F}$. If $f(x) = 0$ for all $x \in S_1 \times \cdots \times S_n$, then there are polynomials $h_1, \ldots, h_n \in \mathbb{F}[x_1, \ldots, x_n]$ such that* $\deg(h_i) \le \deg(f) - |S_i|$ *and*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} h_i(x_1, \ldots, x_n) \prod_{s \in S_i} (x_i - s)\,.$$

Proof (due to Alon 1999). Define $d_i = |S_i| - 1$ for all i, and consider polynomials

$$g_i(x_i) = \prod_{s \in S_i} (x_i - s) = x_i^{d_i+1} - \sum_{j=0}^{d_i} a_{ij} x_i^j.$$

Observe that if $x_i \in S_i$ then $g_i(x_i) = 0$, that is,

$$x_i^{d_i+1} = \sum_{j=0}^{d_i} a_{ij} x_i^j\,. \tag{16.1}$$

Let $\overline{f}$ be the polynomial obtained by writing f as a linear combination of monomials and replacing, repeatedly, each occurrence of $x_i^{t_i}$ $(1 \le i \le n)$, where $t_i > d_i$, by a linear combination of smaller powers of x_i, using the relations (16.1). The resulting polynomial $\overline{f}$ is clearly of degree at most d_i in

x_i for each $1 \leq i \leq n$, and is obtained from f by subtracting from it products of the form $h_i g_i$, where $\deg(h_i) \leq \deg(f) - \deg(g_i) = \deg(f) - |S_i|$. So,

$$\overline{f}(x) = f(x) - \sum_{i=1}^{n} h_i(x) g_i(x_i)\,. \tag{16.2}$$

Moreover, $\overline{f}(x) = f(x)$ for all $x \in S_1 \times \cdots \times S_n$, since the relations (16.1) hold for these values of x. Since, by our assumption, $f(x) = 0$ for all these values, we obtain that $\overline{f}(x) = 0$ for all $x \in S_1 \times \cdots \times S_n$ as well, and Exercise 16.5 implies that $\overline{f}(x) = 0$ for all $x \in \mathbb{F}^n$. Together with (16.2), this implies that $f = \sum_{i=1}^{n} h_i g_i$, as desired. □

Using the Nullstellensatz we can derive the following generalization of the DeMillo–Lipton–Schwartz–Zippel lemma.

Theorem 16.8 (Combinatorial Nullstellensatz)**.** *Let $f(x_1, \ldots, x_n)$ be a polynomial of degree d over a field $\mathbb{F}$. Suppose that the coefficient of the monomial $x_1^{t_1} \cdots x_n^{t_n}$ in f is nonzero and $t_1 + \cdots + t_n = d$. If $S_1, \ldots, S_n$ are finite subsets of $\mathbb{F}$ with $|S_i| \geq t_i + 1$, then there are exists a point x in $S_1 \times \cdots \times S_n$ for which $f(x) \neq 0$.*

Proof. We may assume that $|S_i| = t_i + 1$ for all i. Suppose the result is false, and define $g_i(x_i) = \prod_{s \in S_i}(x_i - s)$. Let $h_1, \ldots, h_n$ be the polynomials guaranteed by the Nullstellensatz. Hence,

$$\deg(h_i) \leq \deg(f) - \deg(g_i) = \deg(f) - (t_i + 1) \tag{16.3}$$

and $f(x) = \sum_{i=1}^{n} h_i(x) g_i(x_i)$, that is,

$$f(x) = \sum_{i=1}^{n} x_i^{t_i+1} h_i(x) + (\text{terms of degree} < \deg(f))\,.$$

By assumption, the coefficient of $\prod_{i=1}^{n} x_i^{t_i}$ on the left-hand side is nonzero, while it is impossible to have such a monomial on the right-hand side, a contradiction. □

In a similar vein is the following result.

Theorem 16.9 (Chevalley–Warning)**.** *Let p be a prime, and $f_1, \ldots, f_m$ polynomials in $\mathbb{F}_p[x_1, \ldots, x_n]$. If $\sum_{i=1}^{m} \deg(f_i) < n$ then the number of common zeros of $f_1, \ldots, f_m$ is divisible by p. In particular, if there is one common zero, then there is another one.*

Although this theorem can be derived from the Combinatorial Nullstellensatz, we give a slightly more direct proof due to Alon (1995).

Proof. By Fermat's Little Theorem (see Exercise 1.15), $a^{p-1} \equiv 1 \bmod p$ for all $a \in \mathbb{F}_p$, $a \neq 0$. Hence, the number N of common zeros of $f_1, \ldots, f_m$ satisfies

$$N = \sum_{x_1,\ldots,x_n \in \mathbb{F}_p} \prod_{j=1}^{m} \left(1 - f_j(x_1, \ldots, x_n)^{p-1}\right) . \quad (\text{in } \mathbb{F}_p) \tag{16.4}$$

By expanding the right-hand side we get a linear combination of monomials of the form

$$\prod_{i=1}^{n} x_i^{t_i} \quad \text{with} \quad \sum_{i=1}^{n} t_i \leq (p-1) \sum_{j=1}^{m} \deg(f_j) < (p-1)n.$$

Hence, in each such monomial there is an i with $t_i < p - 1$. But then (see Exercise 16.1)

$$\sum_{x_i \in \mathbb{F}_p} x_i^{t_i} = 0 \quad (\text{in } \mathbb{F}_p),$$

implying that the contribution of each monomial to the sum (16.4) is 0 modulo p, completing the proof of the theorem. □

We illustrate the potential applications of the Combinatorial Nullstellensatz on several examples.

16.3.1 The permanent lemma

Let $A = (a_{i,j})$ be an $n \times n$ matrix over a field $\mathbb{F}$. The *permanent* $\mathrm{Per}(A)$ of A is the sum

$$\mathrm{Per}(A) = \sum_{(i_1,i_2,\ldots,i_n)} a_{1,i_1} a_{2,i_2} \cdots a_{n,i_n}$$

of $n!$ products, where $(i_1, i_2, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$.

Theorem 16.10. *If Per$(A) \neq 0$, then for any vector $b \in \mathbb{F}^n$, there is a subset of columns of A whose sum differs from b in all coordinates.*

This is just a special case of the following lemma for all $S_i = \{0, 1\}$.

Lemma 16.11 (Permanent Lemma)**.** *Let $b \in \mathbb{F}^n$ and $S_1, \ldots, S_n$ be subsets of $\mathbb{F}$, each of cardinality at least 2. If Per$(A) \neq 0$, then there exists a vector $x \in S_1 \times \cdots \times S_n$ such that Ax differs from b in all coordinates.*

Proof. The polynomial

$$f = \prod_{i=1}^{n} \big(\sum_{j=1}^{n} a_{ij} x_j - b_i \big)$$

is of degree n and the coefficient of $x_1 x_2 \cdots x_n$ in it is $\mathrm{Per}(A) \neq 0$. The result now follows directly from the Combinatorial Nullstellensatz with all $t_i = 1$. □

16.3.2 Covering cube by affine hyperplanes

An *affine hyperplane* is a set of vectors $H = \{x \in \mathbb{R}^n \colon \langle a, x\rangle = b\}$ with $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. How many such hyperplanes do we need to cover $\{0,1\}^n$? If we have no further restrictions on the covering, then just two hyperplanes $H_0 = \{x \in \mathbb{R}^n \colon \langle e_1, x\rangle = 0\}$ and $H_1 = \{x \in \mathbb{R}^n \colon \langle e_1, x\rangle = 1\}$ are enough, where $e_1 = (1, 0, \ldots, 0)$ is the first unit vector. But what if we, say, require that the all-0 vector $\mathbf{0}$ remains uncovered? In this case n hyperplanes $H_i = \{x \in \mathbb{R}^n \colon \langle e_i, x\rangle = 1\}$, $i = 1, \ldots, n$ are still enough. It turns out that this is already optimal.

Theorem 16.12. *Suppose that the hyperplanes $H_1, H_2, \ldots, H_m$ in $\mathbb{R}^n$ avoid $\mathbf{0}$, but otherwise cover all $2^n - 1$ vertices of the cube $\{0,1\}^n$. Then $m \geq n$.*

Proof. As $\mathbf{0}$ is not contained in H_i, we have that each hyperplane is of the form $H_i = \{x \in \mathbb{R}^n \colon \langle a^i, x\rangle = 1\}$ for some $a^i \in \mathbb{R}^n$. Assume that $m < n$, and consider the polynomial

$$f(x) = \prod_{i=1}^{m} (1 - \langle a^i, x\rangle) - \prod_{i=1}^{n} (1 - x_i)\,.$$

The degree of this polynomial is clearly n (since we assumed that $m < n$) and the coefficient at $x_1 \cdots x_n$ is $(-1)^{n+1} \neq 0$. When applied with $S_i = \{0,1\}$ and $t_i = 1$, the Combinatorial Nullstellensatz implies that there must be a point $x \in \{0,1\}^n$ for which $f(x) \neq 0$. We have $x \neq \mathbf{0}$, as $f(\mathbf{0}) = 1 - 1 = 0$. But then $\langle a^i, x\rangle = 1$ for some i (as x is covered by some H_i), implying that f vanishes at this point, a contradiction. □

16.3.3 Regular subgraphs

A graph is *p-regular* if all its vertices have degree p. The following sufficient condition for a graph to contain a regular subgraph was derived by Alon, Friedland and Kalai (1984) using the Combinatorial Nullstellensatz.

Theorem 16.13. *Let $G = (V, E)$ be a graph. Assume that G has no self-loops but multiple edges are allowed. Let p be a prime number. If G has average degree bigger than $2p-2$ and maximum degree at most $2p-1$, then G contains a spanning p-regular subgraph.*

Proof. Associate each edge e of G with a variable x_e and consider the polynomial

$$f = \prod_{v \in V} \left[1 - \left(\sum_{e \in E} a_{v,e} x_e \right)^{p-1} \right] - \prod_{e \in E} (1 - x_e)$$

over $GF(p)$, where $a_{v,e} = 1$ if $v \in e$ and $a_{v,e} = 0$ otherwise. Note that $\deg(f) = |E|$, since the degree of the first product is at most $(p-1)|V| < |E|$, by the assumption on the average degree $2|E|/|V|$ of G. Moreover, the coefficient of $\prod_{e\in E} x_e$ in f is $(-1)^{|E|+1} \neq 0$.

We can therefore apply the Combinatorial Nullstellensatz with $S_e = \{0,1\}$ and $t_e = 1$ for all $e \in E$ and obtain a $(0,1)$-vector $x = (x_e \colon e \in E)$ such that $f(x) \neq 0$. Consider the spanning subgraph H consisting of all edges $e \in E$ for which $x_e = 1$. As $f(\mathbf{0}) = 0$, we conclude that $x \neq \mathbf{0}$ and H is non-empty. The second summand in $f(x)$ is therefore zero and it follows from Fermat's Little Theorem (Exercise 1.15) that $\sum_{e\in E} a_{v,e}x_e \equiv 0 \bmod p$ for every vertex v. Therefore, in the subgraph H all degrees are divisible by p, and since the maximum degree is smaller than $2p$, all positive degrees are precisely p, as needed. □

16.3.4 Sum-sets

The Cauchy-Davenport Theorem, which has numerous applications in Additive Number Theory, is the following. Given two sets A and B of elements of some field $\mathbb{F}$, their *sum-set* is the set $A + B = \{a + b \colon a \in A, b \in B\}$.

Theorem 16.14 (Cauchy–Davenport)**.** *If p is a prime, and A, B are two non-empty subsets of $\mathbb{Z}_p$, then*

$$|A + B| \geq \min\{p,\, |A| + |B| - 1\}\,.$$

We will see in Sect. 25.3.1 that this theorem is just a special case of Kneser's theorem. Here we show how to derive this theorem from the Combinatorial Nullstellensatz.

Proof. If $|A| + |B| > p$ the result is trivial, since in this case for every $x \in \mathbb{Z}_p$ the two sets A and $x - B$ intersect, implying that $A + B = \mathbb{Z}_p$. Assume, therefore, that $|A| + |B| \leq p$ and suppose that $|A + B| \leq |A| + |B| - 2$. Let C be a subset of $\mathbb{Z}_p$ satisfying $A + B \subseteq C$ and $|C| = |A| + |B| - 2$. Define a polynomial $f(x, y) = \prod_{c\in C}(x + y - c)$ and observe that by the definition of C,

$$f(a, b) = 0 \text{ for all } (a, b) \in A \times B. \tag{16.5}$$

Put $t_1 = |A| - 1$, $t_2 = |B| - 1$ and note that the coefficient of $x^{t_1}y^{t_2}$ in f is the binomial coefficient

$$\binom{t_1 + t_2}{t_1} = \binom{|A| + |B| - 2}{|A| - 1}$$

which is nonzero in $\mathbb{Z}_p$, since $|A| + |B| - 2 < p$ (see Exercise 1.14). We can therefore apply the Combinatorial Nullstellensatz (with $n = 2$, $S_1 = A$ and $S_2 = B$) and obtain a poinr $(a, b) \in A \times B$ for which $f(a, b) \neq 0$, contradicting (16.5) and completing the proof. □

The Cauchy–Davenport theorem was extended in many ways. Let us mention one important result in that direction. For a subset $A \subseteq \mathbb{Z}_p$ and a natural number $1 \leq k \leq |A|$, let $s_k(A)$ be the number of elements $b \in \mathbb{Z}_p$ that can be represented as a sum $b = a_1 + \cdots + a_k$ of k *distinct* elements of A.

Theorem 16.15 (Dias da Silva and Hamidoune 1994)**.** *If p is a prime then, for every subset $A \subseteq \mathbb{Z}_p$ and every $1 \leq k \leq |A|$,*

$$s_k(A) \geq \min\{p, k|A| - k^2 + 1\}\,.$$

16.3.5 Zero-sum sets

Using the pigeonhole principle, one can show that any sequence $a_1, \ldots, a_n$ of n integers contains a non-empty consecutive subsequence $a_i, \ldots, a_{i+m}$ whose sum is divisible by n.

To show this, make n pigeonholes labeled from 0 up to $n-1$ and place the n sequences

$$(a_1), (a_1, a_2), \ldots, (a_1, a_2, \ldots, a_n)$$

into the pigeonholes corresponding to the remainder when the sum is divided by n. If any of these sequences is in the pigeonhole 0 then the sum of its numbers is divisible by n. If not, then the n sequences are in the other $n-1$ pigeonholes. By the pigeonhole principle some two of them, $(a_1, a_2, \ldots, a_r)$ and $(a_1, a_2, \ldots, a_s)$ with $r < s$, must lie in the same pigeonhole, meaning that the sum $a_{r+1} + a_{r+2} + \cdots + a_s$ is divisible by n.

A question of a similar flavor is the following one. Given a natural number n, what is the smallest N such that any sequence of N integers contains a subsequence of n (not necessarily consecutive) numbers whose sum is divisible by n? That is, this time we want to find a subsequence of a given length n. The sequence $0^{n-1}1^{n-1}$ of $n-1$ copies of 0 and $n-1$ copies of 1 shows that $N \geq 2n-1$. It turns out that this lower bound is also an upper bound for the sequence length N.

Theorem 16.16 (Erdős–Ginzburg–Ziv 1961)**.** *Any sequence of $2n-1$ integers contains a subsequence of cardinality n, the sum of whose elements is divisible by n.*

There are several different proofs of this theorem – the interested reader can find them, as well as some interesting extensions of this result to higher dimensions, in the paper of Alon and Dubiner (1993). The original proof was based on the Cauchy–Davenport theorem.

First proof of Theorem 16.16. We will first prove the theorem only for the case when $n = p$ is a prime number, and then show how the general case reduces to it.

Let $a_1 \leq a_2 \leq \ldots \leq a_{2p-1}$ be integers. If $a_i = a_{i+p-1}$ for some $i \leq p-1$, then $a_i + a_{i+1} + \cdots + a_{i+p-1} = pa_i = 0$ (in $\mathbb{Z}_p$) and the desired result follows. Otherwise, define $A_i := \{a_i, a_{i+p-1}\}$ for $i = 1, \ldots, p-1$. By repeated application of the Cauchy-Davenport theorem, we conclude that

$$\begin{aligned} |A_1 + A_2 + \cdots + A_{p-1}| &\geq \min\{p, |A_2 + \cdots + A_{p-1}| + 1\} \\ &\geq \min\{p, |A_3 + \cdots + A_{p-1}| + 2\} \\ &\cdots \\ &\geq \min\{p, |A_{p-1}| + p - 2\} = p\,, \end{aligned}$$

and hence, every number from $\mathbb{Z}_p$ is a sum of precisely $p-1$ of the first $2p-2$ elements of our sequence. In particular, the number $-a_{2p-1}$ is such a sum, supplying the required p-element subset whose sum is 0 in $\mathbb{Z}_p$.

The general case may be proved by induction on the number of primes in the prime factorization of n. Put $n = pm$ where p is a prime, and let $a_1, \ldots, a_{2n-1}$ be the given sequence. By the result for the prime case, each subset of $2p-1$ members of the sequence contains a p-element subset whose sum is 0 modulo p. Therefore, we can find pairwise disjoint p-element subsets $I_1, \ldots, I_\ell$ of $\{1, \ldots, 2n-1\}$, where

$$\sum_{j \in I_i} a_j \equiv 0 \bmod p$$

for each $i = 1, \ldots, \ell$. Moreover, $\ell \geq 2m-1$ since otherwise the number of left elements would still be $2pm - 1 - (2m-2)p = 2p - 1$, and we could choose the next subset $I_{\ell+1}$. Now define a sequence $b_1, \ldots, b_{2m-1}$ where

$$b_i = \sum_{j \in I_i} \frac{a_j}{p}$$

(recall that each of these sums is divisible by p). By the induction hypothesis this new sequence has a subset $\{b_i : i \in J\}$ of $|J| = m$ elements whose sum is divisible by m, and the union of the corresponding sets $\{a_j : j \in I_i\}$ with $i \in J$ supplies the desired n-element subset of our original sequence, whose sum is divisible by $n = pm$. □

Theorem 16.16 can also be derived using the Chevalley–Warning theorem about the zeroes of multivariate polynomials (see Theorem 16.9).

Second proof of Theorem 16.16 (Alon 1995). We will prove the theorem only for a prime $n = p$; the general case reduces to it (see the first proof of Theorem 16.16).

Let $a_1, a_2, \ldots, a_{2p-1}$ be integers, and consider the following system of two polynomials in $2p-1$ variables of degree $p-1$ over $\mathbb{F}_p$:

$$\sum_{i=1}^{2p-1} a_i x_i^{p-1} = 0, \quad \sum_{i=1}^{2p-1} x_i^{p-1} = 0\,.$$

Since $2(p-1) < 2p-1$ and $x_1 = x_2 = \cdots = x_{2p-1} = 0$ is a common solution, Theorem 16.9 implies the existence of a nontrivial common solution $(y_1, \ldots, y_{2p-1})$. Since p is a prime, Fermat's Little Theorem (see Exercise 1.15) tells us that $x^{p-1} = 1$ in $\mathbb{F}_p$ for every $x \in \mathbb{F}_p$, $x \neq 0$. So, if we take $I = \{i\colon y_i \neq 0\}$ then the first equation ensures that $\sum_{i \in I} a_i = 0$, while the second ensures that $|I| \equiv 0 \bmod p$, and hence, that $|I| = p$ because $|I| \leq 2p-1$. This completes the proof of the theorem for prime n. □

Exercises

16.1. Let $p \geq 2$ be a prime and consider the field $\mathbb{F}_p$. From Algebra we know that there is an $a \in \mathbb{F}_p$ such that $\mathbb{F}_p \setminus \{0\} = \{a^i\colon i = 0, 1, \ldots, p-2\}$. Use this fact together with Fermat's Little Theorem to prove that, for every $t \leq p-2$, $\sum_{x \in \mathbb{F}_p} x^t = 0$ in $\mathbb{F}_p$. *Hint*: $\sum_{i=0}^{n} z^i = (z^{n+1} - 1)/(z-1)$.

16.2 (Low local degree polynomials)**.** Let f be a nonzero polynomial in n variables over a field $\mathbb{F}$. Suppose that the maximum exponent of each of its variables in f does not exceed d. (Hence, the degree of f may be up to dn which makes the bound given by Lemma 16.3 trivial, if $|S| \leq n$.) Show that we still have the following upper bound: For any subset $S \subseteq \mathbb{F}$ of size $|S| \geq d$,

$$|\{x \in S^n\colon f(x) \neq 0\}| \geq (|S| - d)^n .$$

Hint: Argue by induction on n. In the induction step take a point $(a_1, \ldots, a_n) \in \mathbb{F}^n$ on which $f(a_1, \ldots, a_n) \neq 0$, and consider two polynomials:

$$\begin{aligned} f_0(x_1, \ldots, x_{n-1}) &:= f(x_1, \ldots, x_{n-1}, a_n) \\ f_1(x_n) &:= f(a_1, \ldots, a_{n-1}, x_n). \end{aligned}$$

16.3. Show that the bound in Exercise 16.2 is the best possible. *Hint*: Consider the polynomial $f(x_1, \ldots, x_n) = \prod_{i=1}^{d}(x_1 - i) \cdots \prod_{i=1}^{d}(x_n - i)$.

16.4. Use Exercise 16.2 to show the following: If S a subset of $\mathbb{F}$ that has $d+1$ elements, then any nonzero polynomial of local degree d has a nonzero point in S^n.

16.5. Prove the following "granulated" version of the result established in Exercise 16.2. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial, and let t_i be the maximum degree of x_i in f. Let $S_i \subseteq \mathbb{F}$ with $|S_i| \geq t_i$. If $f(x) = 0$ for all n-tuples $x \in S_1 \times \cdots \times S_n$, then $f(x) = 0$ for all $x \in \mathbb{F}^n$.

16.6. Let $f(x_1, \ldots, x_n)$ be a multivariate polynomial over a field $\mathbb{F}$ with the *degree sequence* $(d_1, \ldots, d_n)$, which is defined as follows: let d_1 be the maximum exponent of x_1 in f, and let $f_1(x_2, \ldots, x_n)$ be the coefficient of $x_1^{d_1}$ in f; then, let d_2 be the maximum exponent of x_2 in f_1, and $f_2(x_3, \ldots, x_n)$ be the coefficient of $x_2^{d_2}$ in f_1; and so on. Suppose that f is not the zero polynomial, and let $S_1, \dots, S_n \subseteq \mathbb{F}$ be arbitrary subsets. For $\boldsymbol{r}_i \in S_i$ chosen independently and uniformly at random, show that

$$\Pr\left[f(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n) = 0\right] \le \frac{d_1}{|S_1|} + \frac{d_2}{|S_2|} + \cdots + \frac{d_n}{|S_n|}.$$

16.7. (R. Freivalds 1977). Suppose that somebody gives us three $n \times n$ matrices A, B, C with real entries and claims that $C = A \cdot B$. We are too busy to verify this claim exactly and do the following. We take a random vector $\boldsymbol{r}$ of length n whose entries are integers chosen uniformly from the interval $\{0, 1, \dots, N-1\}$, and check whether $A \cdot (B \cdot \boldsymbol{r}) = C \cdot \boldsymbol{r}$. If this is true we accept the claim, otherwise we reject it. How large must N be set to make the probability of false acceptance smaller than $1/100$? *Hint*: Consider the matrix $X = A \cdot B - C$. If $C \neq A \cdot B$ then X has a row $x \neq 0$. Take a scalar product of this row with the random vector $\boldsymbol{r}$, observe that $\Pr\left[X \cdot \boldsymbol{r} = 0\right] \le \Pr\left[x \cdot \boldsymbol{r} = 0\right]$, and apply Exercise 16.2.

17. Combinatorics of Codes

In this chapter we will discuss some extremal properties of error-correcting codes. We will also use expander graphs to construct very easily decodable codes.

17.1 Error-correcting codes

Error-correcting codes enable one to store or transmit a message in such a way that one can later recover the message even if it is partially corrupted, that is, up to some number t of bits are flipped by noise. Messages are strings of some fixed length k over some alphabet A. In order to be able to recover corrupted messages, we encode our messages by strings of length $n > k$ over A (or some other alphabet). That is, we take a particular subset $C \subseteq A^n$ (called a *code*) and assign to each message w its own codeword $x = x_w \in C$.

During the transmission some bits may be flipped (by noise, by an adversary or whatever). So, the receiver is presented with a corrupted version $x' \in A^n$ of the original codeword $x \in C$. The receiver knows the set of all codewords C, as well as the encoding algorithm. He also knows that the received vector x' can differ from the original x codeword in at most t bits. What conditions must the code C fulfill in order that we may recover the original codeword x?

Here the notion of Hamming distance comes into the play. The *Hamming distance* $\mathrm{dist}(x, y)$ between two strings x and y of the same length is just the number of positions in which these two strings differ. The *minimum distance*, $\mathrm{dist}(C)$, of a subset $C \subseteq A^n$ is the minimal Hamming distance $\mathrm{dist}(C)$ between any pair of distinct strings in C.

The key observation is that, if $\mathrm{dist}(C) \geq 2t + 1$, then the receiver can (at least in principle) reconstruct the original codeword x from the received, possibly corrupted vector x'. The point is that x is the *only* codeword in the *Hamming ball* $\mathrm{B}_t(x') = \{y \in A^n : \mathrm{dist}(x', y) \leq t\}$ of radius t around the

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_17,

corrupted vector x': were another codeword $y \in C$ also to lie in $\mathrm{B}_t(x')$ then we would have that $\mathrm{dist}(x, y) \leq \mathrm{dist}(x, x') + \mathrm{dist}(x', y) \leq 2t$, contradicting $\mathrm{dist}(C) \geq 2t + 1$.

Thus, the larger $\mathrm{dist}(C)$ is, the better, the more errors we can correct. Another important parameter of codes is their size, that is, the total number of codewords the sender has in his disposal. The larger $|C|$ is, the better, the more distinct messages can be encoded.

Of course, if each two codewords must differ in many positions, then we cannot have many codewords. In particular, we have the following general upper bound.

Theorem 17.1 (Singleton bound). *If $C \subseteq A^n$ and $d = \mathrm{dist}(C)$ then*

$$|C| \leq |A|^{n-d+1} .$$

Proof. Clearly, all codewords in C are distinct. If we delete the first $d - 1$ letters of each codeword, then all resulting codewords must still be pairwise different, since all original codewords in C have Hamming distance at least d from each other. Thus the size of the code remains unchanged. The newly obtained codewords each have length $n - (d - 1) = n - d + 1$ and thus there can be at most $|A|^{n-d+1}$ of them. □

We now construct codes achieving this upper bound. These codes were proposed by Reed and Solomon (1960). Let $k \leq n \leq q$, where q is a power of a prime. As our alphabet we take a field $A := \mathbb{F}_q$ with $|A| = q$ elements. Fix n distinct elements $\alpha_1, \ldots, \alpha_n$ of $\mathbb{F}_q$; here we need that $q \geq n$. Messages are strings $w = (w_1, \ldots, w_k)$ of elements of $\mathbb{F}_q$. We identify each such message with the polynomial $p_w(z) = w_1 + w_2 z + \cdots + w_k z^{k-1}$ of degree at most $k - 1$ over $\mathbb{F}_q$. The codeword of the message w is then the string $x_w = \big(p_w(\alpha_1), \ldots, p_w(\alpha_n)\big)$ of the evaluation of the polynomial $p_w(z)$ at all n fixed points. Let $C = \{x_w : w \in A^k\}$ be the resulting code.

Since, by the factor theorem, no two distinct polynomials of degree at most $k-1$ can agree on k or more points, we have that $\mathrm{dist}(C) \geq n-k+1$. In this way we have constructed a code $C \subseteq A^n$ of minimum distance $d = n - k + 1$ and size $|C| = |A|^k = |A|^{n-d+1}$.

Thus, Reed–Solomon codes meet the Singleton bound. The drawback of these codes is the condition that $q \geq n$. The problem is that we need each coordinate of a codeword to correspond to a distinct element of $\mathbb{F}_q$.

There are different ways to reduce the alphabet size. A trivial one is just to encode all q elements of $\mathbb{F}_q$ by binary strings of length $\lceil \log_2 q \rceil$. Then one obtains a *binary* code $C \subseteq \{0, 1\}^n$ with codeword length $n = q\lceil \log_2 q \rceil$, size $|C| \geq q^k$ and minimum distance $\mathrm{dist}(C) \geq n - k + 1$. Although binary codes with better parameters are known, the binary Reed–Solomon codes are one of the most commonly used codes in practical applications. In particular, they are used to store information, music, and video on compact discs (CDs) and digital video discs (DVDs).

Another way to reduce the field size q is to use multivariate polynomials. In particular, using *bivariate* polynomials, q only needs to be $\sqrt{n}$. For this, let the message length be $k = (t/2+1)^2$ and look at messages $w \in \mathbb{F}_q^k$ as $(t/2+1) \times (t/2+1)$ matrices $w = (w_{ij})$ over $\mathbb{F}_q$. Each such message determines a bivariate polynomial

$$p_w(y,z) = \sum_{i,j=0}^{t/2} w_{ij} y^i z^j$$

of degree at most t over $\mathbb{F}_q$. The codeword of the message w is then the string $x_w = \big(p_w(\alpha,\beta) \,:\, \alpha,\beta \in \mathbb{F}_q\big)$ of the values of the polynomial p_w on all points of $\mathbb{F}_q \times \mathbb{F}_q$. Thus we only require that $q^2 \geq n$, instead of $q \geq n$. As before, the resulting code $C = \{x_w \,:\, w \in \mathbb{F}_q^k\}$ has $|C| = q^k$ codewords. Further, since the polynomials are of degree at most t, the DeMillo–Lipton–Schwartz–Zippel lemma implies that $\mathrm{dist}(C) \geq (1-t/q)n$.

Extending this idea to multivariate polynomials, we arrive at Reed–Muller codes. In this case we identify messages with polynomials of degree at most t in m variables $z_1,\ldots,z_m$. The codeword length in this case is $n = q^m$. The number of codewords is the number of monomials of degree at most ℓ in m variables which, by Exercise 13.22, is equal to $\binom{m+t}{t}$. Finally, the DeMillo–Lipton–Schwartz–Zippel lemma implies that $\mathrm{dist}(C) \geq (1-t/q)n$.

A *binary* Reed–Muller code $C \subseteq \{0,1\}^n$ corresponds to the case when $q = 2$ and $t = 1$. That is, messages are *multilinear* polynomials in m variables over $\mathbb{F}_2$. In this case we have that $|C| = \binom{m+1}{1} = m+1$ codewords of length $n = 2^m$, and the minimum distance is at least $n/2$. Note that these codes have exactly the same parameters as Hadamard codes constructed in Sect. 14.3 (see Theorem 14.10).

17.2 Bounds on code size

If the minimum distance d is given, how large can $|C|$ then be? To answer this question, let $\mathrm{Vol}(n,r)$ be the number of vectors in the Hamming ball $\mathrm{B}_r(\mathbf{0})$ of radius r around the all-0 vector. Since this ball consists of all vectors with at most r ones, we have that

$$2^{n\cdot H(r/n) - O(\log n)} \leq \mathrm{Vol}(n,r) = \sum_{i=0}^{r} \binom{n}{i} \leq 2^{n \cdot H(r/n)}\,,$$

where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function; the estimates in terms of this function are proved in Exercises 1.16 and 1.17.

Theorem 17.2 (Gilbert–Varshamov and Hamming bounds)**.** *Codes $C \subseteq \{0,1\}^n$ of minimal distance d and size*

$$|C| \geq \frac{2^n}{\mathrm{Vol}(n,d)} \tag{17.1}$$

exist, and for any such code with $d \geq 2t+1$*, we have that*

$$|C| \leq \frac{2^n}{\mathrm{Vol}(n,t)}\,. \tag{17.2}$$

The lower bound (17.1) is known as the *Gilbert–Varshamov bound*, and the upper bound (17.2) as the *Hamming bound.*

Proof. The lower bound can be obtained by a simple algorithm: Pick any vector $x \in \{0,1\}^n$, include it in C, remove the whole ball $\mathrm{B}_d(x)$ around this vector, and do the same in the set of the remaining vectors.

To show the upper bound, let $C \subseteq \{0,1\}^n$ be a code of minimum distance $d \geq 2t+1$. Draw a ball of radius t around each of the vectors in C. If the balls around some two codewords were to intersect, then the Hamming distance between these two codewords would be at most $2t < d$, a contradiction. Therefore, the balls are disjoint, and their number is limited by the overall volume divided by the volume of each ball. □

More upper bounds on the size of codes can be obtained using linear algebra. For this, we first embed the Hamming spaces (binary cubes) into Euclidean spaces over the reals. We embed a vector $v \in \{0,1\}^n$ as the vector $x_v \in \mathbb{R}^n$ by changing its coordinates by the rules $0 \mapsto +1$ and $1 \mapsto -1$. Thus, x_v is a ± 1 vector.

The following relations between the Hamming and Euclidean distances are easy to verify (do this!).

Proposition 17.3. *For* $u, v \in \{0,1\}^n$ *with Hamming distance* $\mathrm{dist}(u,v) = d$*, we have that* $\langle x_u, x_v \rangle = n - 2d$*,* $\|x_u\|^2 = n$ *and* $\|x_u - x_v\|^2 = 4d$*.*

Our first goal is to prove a geometric fact: In n dimensions there exist at most $2n$ vectors that pairwise subtend an angle of at least $\pi/2$ at the origin (that is, their pairwise scalar products are ≤ 0).

Lemma 17.4 (Obtuse angles)**.**

(i) *Let* $x_1, \ldots, x_m$ *be vectors in* $\mathbb{R}^n$*, and* α *a positive number. If* $\|x_i\| = 1$ *and* $\langle x_i, x_j \rangle \leq -\alpha$ *for all* $i \neq j$*, then* $m \leq 1 + \alpha^{-1}$*.*

(ii) *If* $y, x_1, \ldots, x_m \in \mathbb{R}^n$ *satisfy* $\langle x_i, x_j \rangle \leq 0$ *for all* $i \neq j$*, while* $\langle y, x_i \rangle > 0$ *for all* i*, then* $m \leq n$*.*

(iii) *If* $x_1, \ldots, x_m \in \mathbb{R}^n$ *are nonzero and satisfy* $\langle x_i, x_j \rangle \leq 0$ *for all* $i \neq j$*, then* $m \leq 2n$*.*

Proof. (i) Let $z = x_1 + \cdots + x_m$. On the one hand, we have $\langle z, z \rangle \geq 0$. On the other hand, we have

$$\begin{aligned}\langle z, z\rangle &= \sum_{i=1}^{m}\langle x_i, x_i\rangle + \sum_{i\neq j\in[m]}^{m}\langle x_i, x_j\rangle\\ &\leq m\cdot 1 + m(m-1)\cdot(-\alpha)\\ &= m(1-\alpha(m-1))\,.\end{aligned}$$

Putting the two inequalities together, we have $1-\alpha(m-1)\geq 0$, implying $m\leq 1+1/\alpha$.

(ii) Assume for the sake of contradiction that $m\geq n+1$. Then there must exist a linearly dependent set of vectors among the x_i's. Specifically, there exist disjoint sets $S, T\subseteq[m]$ and positive λ_i, for $i\in S\cup T$, such that $\sum_{i\in S}\lambda_i x_i = \sum_{j\in T}\lambda_j x_j$. It is not necessary that both S and T be non-empty, but at least one, say S, is non-empty. Let $z=\sum_{i\in S}\lambda_i x_i = \sum_{j\in T}\lambda_j x_j$. Our analysis divides into two cases depending on whether or not $z=\mathbf{0}$.

If $z\neq\mathbf{0}$, then we obtain a contradiction as follows:

$$0<\langle z, z\rangle = \Big\langle \sum_{i\in S}\lambda_i x_i, \sum_{j\in T}\lambda_j x_j\Big\rangle = \sum_{i\in S}\sum_{j\in T}\lambda_i\lambda_j\langle x_i, x_j\rangle\leq 0\,,$$

where the last inequality uses the fact that S and T are disjoint and so $\langle x_i, x_j\rangle\leq 0$ for every $i\in S$ and $j\in T$.

If $z=\mathbf{0}$, then we use the existence of the vector y to obtain a contradiction as follows:

$$0=\langle y,\mathbf{0}\rangle = \langle y, z\rangle = \Big\langle y, \sum_{i\in S}\lambda_i x_i\Big\rangle = \sum_{i\in S}\lambda_i\langle y, x_i\rangle > 0\,.$$

The last inequality is strict since $S\neq\emptyset$, $\lambda_i>0$ and $\langle y, x_i\rangle>0$.

(iii) Pick a vector y such that $\langle y, x_i\rangle\neq 0$ for all $i\in[m]$. At least half of the vectors x_i must have a positive scalar product with either y or $-y$. Assume w.l.o.g. that $x_1,\dots,x_{\lceil m/2\rceil}$ have a positive scalar product with y. Applying part (ii) to these vectors and y, we get $\lceil m/2\rceil\leq n$. □

Theorem 17.5 (Plotkin bound). *If $C\subseteq\{0,1\}^n$ and $d=\mathrm{dist}(C)$, then*

$$|C|\leq d2^{n-2d+2}\,.$$

Proof. We first prove the following claim: If $d\geq n/2$ then $|C|\leq 2n$. Let $c_1,\dots,c_m$ be the codewords of C, and let $x_1,\dots,x_m$ be their ±1 versions. Since $d\geq n/2$, Proposition 17.3 implies that $\langle x_i, x_j\rangle\leq n-2d\leq 0$ for all $i\neq j$, and Lemma 17.4(iii) implies that $m\leq 2n$, as claimed.

To derive the theorem from this claim, write $n=2d+k$. By restricting C to the most commonly occurring pattern in the first k coordinates and deleting these coordinates, we get a set $C'\subseteq\{0,1\}^{2d}$ of size $|C'|\geq|C|/2^k$ and minimum distance d. By the claim we just proved, this implies

$$|C| \le |C'| \cdot 2^k \le 2(2d) \cdot 2^k \le d2^{k+2} = d2^{n-2d+2} . \qquad \square$$

One of the best upper bounds is the following one.

Theorem 17.6 (Johnson bound). *If* $C \subseteq \{0,1\}^n$ *and* $\mathrm{dist}(C) \ge \delta n$ *then*

$$|C| \le \frac{n2^n}{\mathrm{Vol}(n, \tau n - 1)} \le 2^{(1-H(\tau))n} ,$$

where $\tau = (1 - \sqrt{1-2\delta})/2$.

Proof. For a set of strings $C \subseteq \{0,1\}^n$ and a positive integer t, let $\deg_t(C)$ denote the maximum number of vectors from C in a Hamming ball of radius t. If $\mathrm{dist}(C) \ge 2t+1$, then clearly $\deg_t(C) \le 1$. Hence, the following claim generalizes the Hamming bound (17.2).

Claim 17.7. For every subset $C \subseteq \{0,1\}^n$, and every positive integer t, we have that

$$|C| \le \frac{2^n \cdot \deg_t(C)}{\mathrm{Vol}(n,t)} .$$

Proof of Claim 17.7. If we consider the balls of radius t around the strings in C, then any string in $\{0,1\}^n$ is considered at most $\deg_t(C)$ times. Thus the sum of volumes of these balls is at most $2^n \cdot \deg_t(C)$. □

Now let $c_1, \dots, c_m$ be the codewords of a code of minimum distance $d = \delta n$ that are within a Hamming ball of radius $t = \tau n - 1$ from some string $b \in \{0,1\}^n$. By the claim above, it is enough to show that $m \le n$.

For $i \in [m]$, let x_i be the ±1 version of c_i scaled (i.e., multiplied) by $1/\sqrt{n}$, and let y be the ±1 version of b scaled by $1/\sqrt{n}$. We scale the vectors just to achieve $\|x_i\| = 1$ and $\|y\| = 1$. By Proposition 17.3, we have that $\langle x_i, x_j \rangle \le 1 - 2\delta$ for all $i \ne j$, and $\langle y, x_i \rangle > 1 - 2\tau$ for all i. Note the syntactic similarity of these conditions to those in Part (ii) of Lemma 17.4. In fact we can reduce our problem to exactly this case. We will just shift our origin to a new vector $v := \alpha y$ so that from this vector, the vectors x_i mutually subtend an angle at least $\pi/2$.

Claim 17.8. There exists an $0 \le \alpha < 1$ such that for $v := \alpha y$ we have that $\langle x_i - v, x_j - v \rangle \le 0$ and $\langle x_i - v, y - v \rangle > 0$ for all $i \ne j$.

Together with Lemma 17.4(ii), this claim gives the desired upper bound $m \le n$. So, it remains to prove the claim.

Proof of Claim 17.8. We will not specify α yet, only that it will lie in the interval $0 \le \alpha < 1$. Since $\langle x_i, x_j \rangle \le 1 - 2\delta$ and $\langle y, x_i \rangle > 1 - 2\tau$, for such an α we have that

$$\langle x_i - \alpha y, x_j - \alpha y \rangle \le 1 - 2\delta - 2\alpha(1 - 2\tau) + \alpha^2 = (1-\alpha)^2 + 4\alpha\tau - 2\delta .$$

The right-hand side is minimized at $\alpha := 1 - 2\tau$, and for this setting, it is equal to $4\tau - 4\tau^2 - 2\delta$. Since $\tau = (1 - \sqrt{1 - 2\delta})/2$ (by our choice), we have that $(1 - 2\tau)^2 = 1 - 2\delta$, which in turn implies $2\delta = 1 - (1 - 2\tau)^2 = 4\tau - 4\tau^2$, and hence, $4\tau - 4\tau^2 - 2\delta = 0$. We conclude that for this setting $\langle x_i - \alpha y, x_j - \alpha y\rangle \leq 0$, as desired. Since $\langle y, x_i\rangle > 1 - 2\tau$, for the same setting $\alpha = 1 - 2\tau$, we also have that

$$\langle x_i - \alpha y, (1-\alpha)y\rangle = (1-\alpha)\langle x_i, y\rangle - \alpha(1-\alpha)\|y\|^2 = (1-\alpha)[\langle x_i, y\rangle - \alpha] > 0\,,$$

which yields the other part of the claim. □

17.3 Linear codes

If $C \subseteq \{0,1\}^n$ forms a linear subspace over $\mathbb{F}_2$, then C is called a *linear code.* Being a linear subspace over $\mathbb{F}_2$ just means that $x \oplus y \in C$ for all $x, y \in C$. It can easily be verified (do this!) that the Reed–Solomon and Reed-Muller codes we constructed above using polynomials are linear.

If C has dimension k, then $|C| = 2^k$ and the code C can be described using its *generator matrix.* This is a $k \times n$ matrix G whose rows form a basis of C; hence, $C = \{u^\top G \,:\, u \in \mathbb{F}_2^k\}$. Dually, the *parity-check matrix* of C is the generator matrix H of the dual code

$$C^\perp = \{y \in \mathbb{F}_2^n \,:\, \langle x, y\rangle = 0 \text{ for all } x \in C\}\,.$$

That is, H is an $(n-k) \times n$ matrix H such that $C = \{x \,:\, Hx = \mathbf{0}\}$.

A general scenario is then as follows. Messages we want to send to our friend are vectors u in $\mathbb{F}_2^k$. We encode such a message as a vector $x = u^\top G$ and send it. Our friend receives some vector x' which may differ from x on up to t bits. He then searches for the unique vector x such that $Hx = \mathbf{0}$ and $\mathrm{dist}(x, x') \leq t$. In fact, it is enough to search for a unique vector $a \in \mathrm{B}_t(\mathbf{0})$ for which $Ha = Hx'$; then $x = x' \oplus a$ because $H(x' \oplus a) = \mathbf{0}$. This decoding procedure is known as *syndrome decoding*, because a gives us the locations of the errors.

The first important property of a linear code C (not shared by arbitrary codes) is the following fact. By a *weight* $|x|$ of a vector x we mean the number of its nonzero coordinates.

Proposition 17.9. *Every linear code has minimum distance equal to the minimum weight of its nonzero codewords.*

Proof. Let $C \subseteq \{0,1\}^n$ be a linear code, and let $w(C)$ be the minimum weight of a nonzero codeword. Take vectors $x \neq y$ and $z \neq \mathbf{0}$ in C such that $\mathrm{dist}(x, y) = \mathrm{dist}(C)$ and $|z| = w(C)$. Then $\mathrm{dist}(C) = |x \oplus y| \geq w(C)$, since $x \oplus y$ belongs to C. On the other hand, we have that $w(C) = |z| = \mathrm{dist}(z, \mathbf{0}) \geq \mathrm{dist}(C)$, since vector $\mathbf{0}$ belongs to C. □

The next important property of linear codes is that their minimum distance is related to linear independence of the columns of their parity-check matrices.

Theorem 17.10. *Let C be a linear code with parity-check matrix H. Then the minimum distance of C is d if and only if every set of $d-1$ columns of H are linearly independent but some d columns are linearly dependent.*

Proof. We already know that the minimum distance of C is equal to the smallest of the weights of the nonzero codewords. On the other hand, $Hx = \mathbf{0}$ for a nonzero vector $x \neq \mathbf{0}$ means that the columns of H corresponding to the 1-positions of x are linearly dependent. Thus, if $d = \mathrm{dist}(C)$ then some d columns of H must be linearly dependent (since C contains a codeword of weight d), and no $d-1$ columns can be linearly dependent, for otherwise C would contain a codeword of weight smaller than d. □

This fact can be used to show that linear codes with minimum distance d and about $2^n/n^d$ codewords exist.

Theorem 17.11. *A linear code $C \subseteq \{0,1\}^n$ of dimension k and minimum distance d exists provided that*

$$\sum_{i=0}^{d-2} \binom{n-1}{i} < 2^{n-k}. \tag{17.3}$$

That is, linear codes of size $|C| \geq 2^n/\mathrm{Vol}(n-1, d-2)$ exist. This is almost the same bound as for arbitrary codes given in (17.1).

Proof. We shall construct an $(n-k) \times n$ matrix H over $\mathbb{F}_2$ with the property that no $d-1$ columns are linearly dependent. Put $r = n-k$. Choose the first column of H to be any nonzero r-tuple in $\mathbb{F}_2^r$ Then choose the second column to be any nonzero r-tuple different from the first. Continue choosing successive columns so that each new column is not a linear combination of any $d-2$ or fewer previous columns. When we come to try to choose the i-th column, those r-tuples not available to us will be the $N(i) = \sum_{j=0}^{d-2} \binom{i-1}{j}$ linear combinations of $d-2$ or fewer columns from the $i-1$ columns already chosen. Not all of these linear combinations need be distinct vectors, but even in the worst case, where they are distinct, provided $N(i)$ is less than the total number 2^r of all r-tuples, then an i-th column can be added to the matrix. Thus, since (17.3) holds, we will reach a matrix H having n columns, as required. □

We are not going to sink into different constructions of explicit codes: there are so many of them. We just mention that so-called BCH codes, constructed by Bose, Chaudhuri and Hocquenghem, are linear codes with quite good parameters: for every integers $m \geq 3$ and $t < 2^{m-1}$ there is an explicit linear code $C \subseteq \{0,1\}^n$ with $n = 2^m - 1$, $|C| \geq 2^{n-mt}$ and $\mathrm{dist}(C) \geq 2t+1$. See the book by MacWilliams and Sloane (1977) for more constructions.

17.4 Universal sets from linear codes

Recall that a set of 0-1 strings $A \subseteq \{0,1\}^n$ is (n,k)-*universal* if, for any subset $S \subseteq [n]$ of $|S| = k$ coordinates, the projection of A onto the indices in S contains all possible 2^k configurations.

In Sect. 10.5 we have shown how to construct explicit (n,k)-universal sets of size about n, when $k \leq (\log n)/3$. This construction was based on Paley graphs. Here we will show how to construct such sets of size $n^{O(k)}$ for arbitrary k. The construction is based on some elementary properties of linear codes.

We already know that the minimal distance of C coincides with the minimum weight of (i.e., the number of 1s in) a nonzero vector from C. This simple property of linear codes implies that their duals are universal.

Proposition 17.12. *If C is a linear code of length n and its dual $C^{\perp}$ has minimal distance at least $k+1$ then the code itself is (n,k)-universal.*

Since a parity-check matrix of C is a generator matrix of $C^{\perp}$, an equivalent condition on C is that any k columns of its generator matrix must be linearly independent.

Proof. Take a set $S \subseteq \{1,\ldots,n\}$ with $|S| = k$. The set of all projections of vectors in C onto S forms a linear subspace in $\{0,1\}^{|S|}$ of dimension k. If this subspace were proper then, by Proposition 13.2, some nonzero vector x, whose support $\{i : x_i \neq 0\}$ lies in S, would belong to $C^{\perp}$, implying that $\mathrm{dist}(C^{\perp}) \leq |S| = k$, which contradicts our assumption. □

It is known (see, for example, MacWilliams and Sloane (1977)) that the dual of a binary BCH code of minimal distance k has only $O(n^{\lfloor k/2 \rfloor})$ vectors. By Proposition 17.12, these codes give us explicit (n,k)-universal sets consisting of only so many vectors.

One of the best known explicit constructions of (n,k)-universal sets of size only $2^{O(k^4)} \log n$ is due to Alon (1986a). His construction is based on a Justesen-type code constructed by Friedman (1984).

17.5 Spanning diameter

So far, we have used *algebraic* properties of linear spaces to derive some results in combinatorics. But these spaces themselves have some interesting *combinatorial* properties as well.

Let A be a set of vectors in $\{0,1\}^n$ and consider its span over the field $\mathbb{F}_2$. Each vector in $\mathrm{span}\, A$ is a linear combination of some vectors from A. Some of these combinations may be short, but some may be long. Given A, we are interested in the smallest number k such that every vector from $\mathrm{span}\, A$ is a sum (over $\mathbb{F}_2$) of at most k vectors of A; we call this k the *spanning diameter* of A.

Of course, the answer depends on how large the span is, compared with the set itself. It is easy to show that if $|A| > |\operatorname{span} A|/2$, then the spanning diameter of A is at most 2 (see Exercise 17.3). But what if A is a smaller fraction of span A, say, an α-fraction for some $\alpha > 1/4$? It turns out that then the spanning diameter does not exceed 4. In general, we have the following upper bound on k.

Theorem 17.13. *Let $A \subseteq \{0,1\}^n$. If $|A| \geq \alpha \cdot |\operatorname{span} A|$ for some $0 < \alpha \leq 1$, then every vector from* span A *is a sum of at most k vectors from A, where k is the maximal number satisfying the inequality*

$$k - \lfloor \log_2 k \rfloor - 1 \leq \log_2(1/\alpha) \tag{17.4}$$

Theorem 17.13 can be derived from known bounds on the covering radius of binary linear codes (see, for example, Cohen et al. 1997; Theorem 8.1.21). Here we present a direct argument due to Pavel Pudlák.

Proof. Take a maximal set of vectors $a^1, a^2, \ldots, a^k$ in A such that the vector

$$v = a^1 + a^2 + \cdots + a^k, \tag{17.5}$$

cannot be represented as a sum of fewer than k vectors from A. (Here and throughout the proof all the sums are over $\mathbb{F}_2$.) Our goal is to show that then k must satisfy (17.4). Since $\alpha \leq 1$, the cases $k = 1$ and $k = 2$ are trivial. So, assume that $k \geq 3$.

We will need a lower bound on the size of distance-3 codes. Such codes can be obtained by shortening the Hamming code (see, for example, MacWilliams and Sloane (1977)); Exercise 17.6 sketches a way to do this.

Claim 17.14. There exists a set $C \subseteq \mathbb{F}_2^k$ such that any two vectors of C differ in at least 3 coordinates and $\log_2 |C| \geq k - \lfloor \log_2 k \rfloor - 1$.

Fix such a set C, and let B be the set of all those vectors b from span A which can be represented in the form $b = c_1 a^1 + c_2 a^2 + \cdots + c_k a^k$ for $c = (c_1, \ldots, c_k) \in C$. The key point is that all the translates

$$b + A := \{b + a \,:\, a \in A\},$$

with $b \in B$, are mutually disjoint.

Claim 17.15. For every pair b, b' of distinct vectors from B, the sets $b + A$ and $b' + A$ are disjoint.

Proof of Claim 17.15. Suppose not. Then for some $a, a' \in A$ we have $b + a = b' + a'$, and hence, $a + a' = b + b'$. Let c, c' be the vectors from C for which $b = c_1 a^1 + \cdots + c_k a^k$ and $b' = c'_1 a^1 + \cdots + c'_k a^k$. Then

$$a + a' = b + b' = (c_1 + c_1')a^1 + (c_2 + c_2')a^2 + \cdots + (c_k + c_k')a^k.$$

Since vectors c and c' differ in at least three coordinates, we have on the right-hand side the sum of at least three vectors, say $a^{i_1} + \cdots + a^{i_l}$, with $l \geq 3$. But then in the equation (17.5) we can replace these three (or more) vectors $a^{i_1}, \ldots, a^{i_l}$ by two vectors a, a', which contradicts the minimality of k. □

The same argument also implies that no two distinct vectors $c, c' \in C$ can lead to one and the same vector $b \in B$, that is, $c \neq c' \in C$ implies $\sum_i c_i a^i \neq \sum_i c_i' a^i$. This means that $|B| = |C|$.

This, together with Claim 17.15, implies

$$|A| \cdot |C| = |A| \cdot |B| = \sum_{b \in B} |b + A| = |\bigcup_{b \in B} (b + A)| \leq |\mathrm{span}\, A|.$$

Hence, $\log_2 |C| \leq \log_2(1/\alpha)$ which, together with Claim 17.14, yields the desired upper bound (17.4) on k. □

17.6 Expander codes

If $C \subseteq \{0,1\}^n$ is a linear code with a $k \times n$ generator matrix G, then the encoding of messages $w \in \{0,1\}^k$ is very easy: just encode w by the codeword $x = w^\top G$. However, the decoding—that is, given a vector $y \in \{0,1\}^n$ find a codeword $x \in C$ closest to y—is in general linear codes a very difficult problem (it is "NP-hard").

We now show how using expander graphs one can construct linear codes for which decoding is almost trivial—it can be done in linear time! Moreover, if the expansion of the graph is good enough then the resulting codes achieve very good rate $(\log_2 |C|)/n$ and minimal distance (both these parameters are then absolute positive constants).

Let $G = (L \cup R, E)$ be a bipartite graph with $|L| = n$, $|R| = m$ and $E \subseteq L \times R$. Each such graph defines a linear code $C \subseteq \{0,1\}^n$ as follows. Associate with each vertex $u \in L$ a boolean variable x_u. Given a vector $x \in \{0,1\}^n$, say that a vertex $v \in R$ is *satisfied* by this vector if

$$\sum_{u \in \Gamma(v)} x_u \bmod 2 = 0\,,$$

where $\Gamma(v) = \{u \in L\,:\, uv \in E\}$ is the set of all neighbors of v on the left side (see Fig. 17.1). The code defined by the graph G is the set of vectors

$$C = \{x \in \{0,1\}^n \,:\, \text{all vertices in } R \text{ are satisfied by } x\}\,.$$

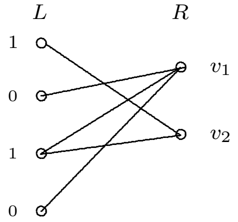


Fig. 17.1 Vertex v_2 is satisfied whereas v_1 is not satisfied by the vector $x = (1010)$.

That is, C is just the set of all solutions of m linear equations in n variables. Therefore, C is linear and $|C| \geq 2^{n-m}$.

Let $\operatorname{dist}(C)$ be the minimal Hamming distance between two different vectors in C. A graph $G = (L \cup R, E)$ is *left d-regular* if each vertex in L has degree d. Such a graph is an *(α, c)-expander* if every subset $I \subseteq L$ with $|I| \leq \alpha n$ has $|\Gamma(I)| > c|I|$ neighbors on the right side.

Lemma 17.16. *If $C \subseteq \{0,1\}^n$ is a code of a left d-regular (α, c)-expander with $c > d/2$, then*

$$\operatorname{dist}(C) > \alpha n\,.$$

Proof. Assume that $\operatorname{dist}(C) \leq \alpha n$. Then C must contain a vector x with at most αn ones. Hence, if we take the set $I = \{u \in L \,:\, x_u = 1\}$, then $|I| \leq \operatorname{dist}(C) \leq \alpha n$. Since G is an $(\alpha, d/2)$-expander, this implies $|\Gamma(I)| > d|I|/2$.

We claim that there must exist a vertex $v_0 \in \Gamma(I)$ with *exactly one* neighbor in I, that is, $|\Gamma(v_0) \cap I| = 1$. Indeed, otherwise every vertex $v \in \Gamma(I)$ would have at least two neighbors in I. Therefore the number of edges leaving I would be at least $2 \cdot \Gamma(I) > 2 \cdot (d|I|/2) = d|I|$, contradicting the left d-regularity of G.

Since $x_u = 0$ for all $u \notin I$, this implies that *exactly one* of the bits x_u of x with $u \in \Gamma(v_0)$ is equal to 1. So, $\sum_{u \in \Gamma(v_0)} x_u = 1$, and the vertex v_0 cannot be satisfied by the vector x, a contradiction with $x \in C$. □

By Lemma 17.16, expander codes can correct relatively many errors, up to $\alpha n/2$. Much more important, however, is that the decoding algorithm for such codes is very efficient. The decoding problem is the following one: given a vector $y \in \{0,1\}^n$ of Hamming distance $\leq \alpha n/2$ from some (unknown) codeword $x \in C$, find this codeword x. The decoding algorithm for expander codes is amazingly simple:

> *While there exists a variable such that most of its neighbors are not satisfied by the current vector, flip it.*

Lemma 17.17 (Sipser–Spielman 1996)**.** *If C is a code of a left d-regular (α, c)-expander with $c > \frac{3}{4}d$, then the algorithm solves the decoding problem in a linear number of steps.*

Proof. Let $y \in \{0,1\}^n$ be a vector of Hamming distance $\leq \alpha n/2$ from some (unknown) codeword $x \in C$. Our goal is to find this codeword x. Let

$$I = \{u \in L \,:\, y_u \neq x_u\}$$

be the set of errors in y. If I is empty, we are done. Otherwise, assume that $|I| \leq \alpha n$. We need this assumption to guarantee the expansion, and we will prove later that this assumption holds throughout the running of the algorithm.

Partition the set $\Gamma(I) = S \cup U$ into the set S of neighbors satisfied by y and the set U of neighbors not satisfied by y. Since $c > 3d/4$, we have that

$$|U| + |S| = |\Gamma(I)| > \tfrac{3}{4}d|I|\,. \tag{17.6}$$

Now, count the edges between I and $\Gamma(I)$. At least $|U|$ of these edges must leave U. Moreover, at least $2|S|$ of them must leave S because every vertex $v \in S$ must have at least two neighbors in I: If v had only one such neighbor, then y would not satisfy the vertex v since $y \neq x$, x satisfies v and y coincides with x outside I. Since the total number of edges between I and $\Gamma(I)$ is $d|I|$, this implies $|U| + 2|S| \leq d|I|$. Combining this with (17.6) we get that

$$d|I| - |U| \geq 2|S| > 2\left(\tfrac{3}{4}d|I| - |U|\right)$$

and therefore

$$|U| > \tfrac{1}{2}d|I|\,. \tag{17.7}$$

So, more than $d|I|/2$ neighbors of the $|I|$ vertices in I are unsatisfied. Therefore there is a variable in I that has more than $d/2$ unsatisfied neighbors. We have therefore shown the following claim:

> *If $I \neq \emptyset$ and $|I| \leq \alpha n$ then there is a variable with $> d/2$ unsatisfied neighbors.*

This implies that as long as there are errors *and* $|I| \leq \alpha n$ holds, some variable will be flipped by the algorithm. Since we flip a vertex with more unsatisfied neighbors than satisfied ones, $|U|$ decreases with every step (flipping x_u can only affect the satisfiability of neighbors of u). We deduce that if the distance $|I|$ of the actual vector y from x does not exceed $\alpha n/2$ throughout the run of the algorithm, then the algorithm will halt with the codeword x after a linear number of iterations.

To show that $|I|$ can never exceed αn, recall that $|I| \leq \alpha n/2$, and hence,

$$|U| \leq |\Gamma(I)| \leq \tfrac{1}{2}\alpha dn \tag{17.8}$$

hold in the beginning. Moreover, $|U|$ decreases after each iteration. Hence, if at some step we had that $|I| > \alpha n$, then (17.7) would imply $|U| > \alpha dn/2$, contradicting (17.8). □

In general, every linear code $C \subseteq \{0,1\}^n$ is defined by its *parity-check* matrix H such that $x \in C$ iff $Hx = \mathbf{0}$. Note that, if C is a code defined by a bipartite graph G, then H is just the transpose of the adjacency matrix of G. If G is left d-regular, then every row of H has exactly d ones. If G is an (α, c)-expander, then every subset I of $|I| \leq \alpha n$ columns of H has ones in at least $c|I|$ rows. The decoding algorithm above is, given a vector $y \in \{0,1\}^n$ such that $Hy \neq \mathbf{0}$, to flip its i-th bit provided that vector $H(y \oplus e_i)$ has fewer ones than vector Hy.

17.7 Expansion of random graphs

Explicit constructions of bipartite left d-regular (α, c)-expanders with $\alpha = \Omega(1)$ and $c > 3d/4$ are known. These constructions are however too involved to be presented here. Instead of that, we will show that *random* bipartite left-regular graphs have good expansion properties.

Let $d \geq 3$ be a constant. We construct a random bipartite left d-regular $n \times n$ graph $G_{n,d} = (L \cup R, E)$ as follows: For each vertex $u \in L$ choose its d neighbors independently at random, each with the same probability $1/n$. The graph obtained may have multi-edges, that is, some pairs of vertices may be joined by several edges.

Theorem 17.18. *For every constant $d \geq 3$, there is a constant $\alpha > 0$ such that for all sufficiently large n, the graph $G_{n,d}$ is an $(\alpha, d-2)$ expander with probability at least $1/2$.*

Proof. Set (with foresight) $\alpha := 1/(\mathrm{e}^3 d^4)$. Fix any $s \leq \alpha n$, and take any set $S \subseteq L$ of size $|S| = s$. We want to upper bound the probability that S does not expand by $d-2$. This means that the ds neighbors (including multiplicities) of the vertices in S hit fewer than $(d-2)s$ distinct vertices on the right side, that is, some $2s$ of these ds neighbors land on previously picked vertices. Each neighbor lands on a previously picked vertex with probability at most ds/n, so

$$\Pr\left[S \text{ does not expand by } (d-2)\right] \leq \binom{ds}{2s}\left(\frac{ds}{n}\right)^{2s}.$$

By the union bound, the probability that at least one subset S of size s does not expand by $(d-2)$ is at most

$$\binom{n}{s}\binom{ds}{2s}\left(\frac{ds}{n}\right)^{2s} \leq \left(\frac{\mathrm{e}n}{s}\right)^{s}\left(\frac{\mathrm{e}ds}{2s}\right)^{2s}\left(\frac{ds}{n}\right)^{2s} \leq \left(\frac{\mathrm{e}^3 d^4}{4n}\right)^{s} \leq \left(\frac{1}{4}\right)^{s},$$

by the choice of α. Thus, the probability that some set S of size $|S| \leq \alpha n$ does not expand by $(d-2)$ does not exceed

$$\sum_{s=1}^{\alpha n} 4^{-s} < \sum_{s=1}^{\infty} 4^{-s} < \frac{1}{2}\,.$$

Hence, the graph $G_{n,d}$ is an $(\alpha, d-2)$ expander with probability at least 1/2. □

Exercises

17.1. Prove the following stronger version of Proposition 17.12. Let C be a linear code of length n and minimal distance at least $k+1$ and let $C^{\perp}$ be its dual. Then for every subset S of $l \leq k$ coordinates, every 0-1 string of length l appears as a projection of $C^{\perp}$ onto S the same number of times. *Hint*: Take a matrix whose rows form a basis of $C^{\perp}$, observe that every k columns of this matrix are linearly independent and use Proposition 13.3.

17.2. Let $V \subseteq \mathbb{F}_2^n$ be a subspace of dimension d. Show that $|V| = 2^d$.

17.3. Let $A \subseteq \mathbb{F}_2^n$ and suppose that $|A| > |\mathrm{span}\, A|/2$. Prove that every vector in $\mathrm{span}\, A$ is the sum of at most 2 vectors from A. *Hint*: Show that, for every $v \in \mathrm{span}\, A$, the set $A \cap (v + A)$ has at least one vector.

17.4. Theorem 17.13 gives an *upper* bound on the spanning diameter of sets A in terms of their density $\alpha = |A|/|\mathrm{span}\, A|$. Show that for infinitely many values of k, the bound (17.4) is optimal, that is, exhibit sets A whose spanning diameter is the maximal number satisfying (17.4). *Hint*: Consider the set consisting of the all-0 vector and k vectors with precisely one 1; its density is $\alpha = (k+1)/2^k$.

17.5. (*Hamming code*). Let r be a positive integer, and let $k = 2^r - 1$. Consider the $r \times k$ matrix H whose columns are all the distinct nonzero vectors of $\{0,1\}^r$. Let $C \subseteq \mathbb{F}_2^k$ be the set of vectors, each of which is orthogonal (over $\mathbb{F}_2$) to all the rows of H. Prove that C is a linear code of minimal distance 3 and has precisely 2^{k-r} code words. *Hint*: Show that no vector of weight 1 or 2 can be orthogonal to all the rows of H, and use Proposition 17.9.

17.6. Prove Claim 17.14. *Hint*: If k has the form $k = 2^r - 1$, then we can take C to be a Hamming code (see previous exercise). Otherwise, take r such that $k = 2^r + x$ for some integer $0 \leq x < 2^r - 1$, and let C be a Hamming code of length $K = 2^{r+1} - 1$. By fixing the last $K - k$ of coordinates to appropriate constants, it is possible to obtain from C a set of vectors $C' \subseteq \{0,1\}^k$ of size $|C'| \geq |C|/2^{K-k} = 2^{k-r-1}$, such that any two of its vectors still differ in at least 3 coordinates. The code C' obtained may be not linear, but we do not require that.

17.7. Prove that among any $2^{k-1} + 1$ vectors in $\mathbb{F}_2^n$ some k of them must be linearly independent. *Hint*: Take a maximal subset of linearly independent vectors and form all possible sums (over $\mathbb{F}_2$).

Part IV
The Probabilistic Method

18. Linearity of Expectation

Let $X_1, \ldots, X_n$ be random variables, and $X = c_1X_1 + \cdots + c_nX_n$. Linearity of expectation states that

$$\mathrm{E}[X] = c_1\mathrm{E}[X_1] + \cdots + c_n\mathrm{E}[X_n].$$

The power of this principle comes from there being no restrictions on the dependence or independence of the X_i's. In applications we often use the fact that there must be a point in the probability space for which $X \geq \mathrm{E}[X]$ and a point for which $X \leq \mathrm{E}[X]$. This principle (known as the *pigeonhole property* of the expectation) is used in most arguments.

18.1 Hamilton paths in tournaments

A *tournament* is an oriented graph $T = (V, E)$ such that $(x, x) \notin E$ for all $x \in V$, and for any two vertices $x \neq y$ exactly one of (x, y) and (y, x) belongs to E. The vertices are players, each pair of which participates in a single match, and $(x, y) \in E$ if and only if x beats y. Given such a tournament, a *Hamiltonian path* in it is defined as a permutation $(x_1, x_2, \ldots, x_n)$ of players such that, for every i, x_i beats x_{i+1}.

It is easy to show (see Exercise 18.5) that every tournament contains a Hamiltonian path. On the other hand, there are tournaments with only one Hamiltonian path (the path itself). Are there tournaments with many Hamiltonian paths? The existence of such "rich" tournaments was proved by T. Szele in 1943. His proof is considered to be the first application of the probabilistic method in combinatorics.

Theorem 18.1 (Szele 1943)**.** *There is a tournament T with n players and at least $n!/2^{n-1}$ Hamiltonian paths.*

Proof. Take a random tournament $\boldsymbol{T}$ (where the outcome of each game is determined by the flip of fair coin), and let X be the number of Hamiltonian

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_18,

paths in it. For each permutation $\pi = (x_1, x_2, \ldots, x_n)$ of players, let X_π denote the indicator random variable for the event "π is a Hamiltonian path in $\boldsymbol{T}$." Then $X = \sum X_\pi$, the summation being over all $n!$ permutations π. For a given π, $\mathrm{E}[X_\pi] = 2^{-(n-1)}$, since that is the probability that the $n-1$ games x_i versus x_{i+1} all have the desired outcome. By the linearity of expectation,

$$\mathrm{E}[X] = \sum_\pi \mathrm{E}[X_\pi] = n!2^{-(n-1)}.$$

Since (by the pigeonhole property of the expectation) a random variable cannot always be smaller than its expectation, at least one tournament must have at least $\mathrm{E}[X]$ Hamiltonian paths. □

In the same paper, Szele also established an upper bound $O(n!/2^{3n/4})$ on the maximal possible number of Hamiltonian paths in any tournament with n players. Based on the solution of the well-known conjecture of H. Minc about the permanent of 0-1 matrices (found by Bregman in 1973), Alon (1990b) has essentially improved this upper bound to

$$cn^{3/2}n!/2^{n-1}\,, \tag{18.1}$$

where c is a positive constant independent of n.

18.2 Sum-free sets

Suppose we are given a finite set of nonzero integers, and are asked to mark an as large as possible subset of them under the restriction that the sum of any two marked integers cannot be marked. It turns out that (independent of what the given integers actually are!) we can always mark at least one-third of them.

A subset B of an additive group is called *sum-free* if $x + y \not\in B$ for all $x, y \in B$ ($x = y$ is allowed). For example, the set of all odd integers is sum-free, and the subset $B = \{n+1, n+2, , \ldots, 2n\}$ is a sum-free subset of $A = \{1, \ldots, 2n\}$. We are interested in the case when A is an *arbitrary* set of numbers: can we also then choose large sum-free subsets?

Theorem 18.2 (Erdős 1965)**.** *Let $A \subseteq \mathbb{Z}$ be a set of N nonzero integers. Then there is a sum-free subset B of A with $|B| > N/3$.*

Proof. Let $p = 3k + 2$ be a prime, which satisfies $p > 2\max_{a\in A} |a|$. Such a prime exists by Dirichlet's prime number theorem, stating that for any two positive co-prime integers a and d, there are infinitely many primes of the form $a + nd$, where $n \geq 0$. In other words: there are infinitely many primes which are congruent to a modulo d.

Write $S = \{k+1, k+2, \ldots, k+(k+1)\}$, and observe that S is a sum-free subset of the group $\mathbb{Z}_p$ (the integers modulo p), because, by the choice of p, the sum of any two numbers from S, taken modulo p, does not belong to S. Indeed, the sum $(k+1)+(k+1) = 2k+2 > 2k+1$ is too large, whereas the sum $(2k+1)+(2k+1) = 4k+2 = k \bmod p = k < k+1$ is too small.

We choose a subset of A as follows. Pick a random element $\boldsymbol{t} \in \mathbb{Z}_p \setminus \{0\}$, and let

$$A_{\boldsymbol{t}} = \{a \in A \, : \, a\boldsymbol{t} \bmod p \in S\}\,.$$

Note that $A_{\boldsymbol{t}}$ is sum-free, because for any $a, b \in A_{\boldsymbol{t}}$, the residues of $a\boldsymbol{t}$ and $b\boldsymbol{t}$ modulo p belong to S (by definition of $A_{\boldsymbol{t}}$) whereas the residue of $(a+b)\boldsymbol{t} = a\boldsymbol{t}+b\boldsymbol{t}$ cannot belong to S, by sum-freeness of S. It remains to show that A_t is large for some t. To do this, observe that for any fixed $a \neq 0$, as t ranges over all numbers $1, 2, \ldots, p-1$, the residues of $a \cdot t$ modulo p range over all nonzero elements of $\mathbb{Z}_p$. Thus, $\Pr[a\boldsymbol{t} \bmod p \in S] = |S|/(p-1) = (k+1)/(3k+1) > 1/3$, for every $a \in A$. By the linearity of expectation, we have that

$$\mathrm{E}[|A_{\boldsymbol{t}}|] = \sum_{a \in A} \Pr[a \in A_{\boldsymbol{t}}] = \sum_{a \in A} \Pr[a\boldsymbol{t} \bmod p \in S] > \tfrac{1}{3}|A|\,,$$

By the pigeonhole property of expectation, there is a value of t for which $|A_t| > |A|/3$. □

It is not clear what is the largest constant that works in place of 1/3 in the previous theorem. It is only known (see Alon and Kleitman 1990) that it must be smaller than 12/29.

18.3 Dominating sets

A *dominating set* of vertices in a graph $G = (V, E)$ is a set $S \subseteq V$ such that every vertex of G belongs to S or has a neighbor in S.

Theorem 18.3 (Alon 1990c). *If $G = (V, E)$ is an n-vertex graph with minimum degree $d > 1$, then G has a dominating set with at most $n\frac{1+\ln(d+1)}{d+1}$ vertices.*

Proof. Form a random vertex subset $\boldsymbol{S} \subseteq V$ by including each vertex independently with probability $p := \ln(d+1)/(d+1)$. Given $\boldsymbol{S}$, let $\boldsymbol{T}$ be the set of vertices outside $\boldsymbol{S}$ having no neighbor in $\boldsymbol{S}$; adding $\boldsymbol{T}$ to $\boldsymbol{S}$ yields a dominating set. So, it remains to estimate the expected size of this union. Since each vertex appears in $\boldsymbol{S}$ with probability p, $\mathrm{E}[|\boldsymbol{S}|] = np$.

The random variable $|\boldsymbol{T}|$ is the sum $\sum_{v \in V} X_v$ of n indicator variables X_v for whether individual vertices v belong to $\boldsymbol{T}$. We have $X_v = 1$ if and only if v and its neighbors all fail to be in $\boldsymbol{S}$, the probability of which is bounded by $(1-p)^{d+1}$, since v has degree at least d. Hence, $\mathrm{E}\left[|\boldsymbol{T}|\right] = \sum_{v\in V} \mathrm{E}\left[X_v\right] \leq n(1-p)^{d+1}$. As $(1-p)^{d+1} \leq \mathrm{e}^{-p(d+1)}$, we have

$$\mathrm{E}\left[|\boldsymbol{S}\cup\boldsymbol{T}|\right] \leq np + n\mathrm{e}^{-p(d+1)} = n\frac{1+\ln(d+1)}{d+1}\,.$$

By the pigeonhole property of the expectation, there must be some S for which $S\cup T$ is a dominating set of size no larger than this. □

18.4 The independence number

The independence number $\alpha(G)$ of a graph G is the maximum number of vertices with no edges between them. The following result is due to Caro (unpublished) and Wei (1981).

Theorem 18.4. *Let G be a graph on n vertices and let d_i denote the degree of the i-th vertex. Then*

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i+1}\,. \tag{18.2}$$

Proof. (Alon–Spencer 1992). Let $V = \{1,\ldots,n\}$ and let $\pi : V \to V$ be a random permutation taking its values uniformly and independently with probability $1/n!$. This permutation corresponds to a random ordering of vertices in V. Let A_i be the event that all neighbors j of i in G are greater than i in the ordering, i.e., that $\pi(j) > \pi(i)$ for all d_i neighbors j of i. There are $\binom{n}{d_i+1}$ possibilities to choose a (d_i+1)-element set $S \subseteq V$ of possible π-images of i and all its d_i neighbors. After that there are $(|S|-1)! = d_i!$ possibilities to arrange the π-images of neighbors of i within S (the place of $\pi(i)$ is fixed – it must come first), and $(n-|S|)! = (n-d_i-1)!$ possibilities to arrange the vertices outside S. Thus,

$$\Pr\left[A_i\right] = \binom{n}{d_i+1}\frac{d_i!(n-d_i-1)!}{n!} = \frac{1}{d_i+1}\,.$$

Let $\boldsymbol{U}$ be the set of those vertices i for which A_i holds. By linearity of expectation

$$\mathrm{E}\left[|\boldsymbol{U}|\right] = \sum_{i=1}^{n}\Pr\left[A_i\right] = \sum_{i=1}^{n} 1/(d_i+1)\,.$$

Thus, for some specific ordering, $|U| \geq \sum_{i=1}^{n} 1/(d_i+1)$. Now let $\{i,j\}$ be an edge of G. Then either $\pi(i) < \pi(j)$ or $\pi(j) < \pi(i)$. In the first case $j \notin U$, and in the second case $i \notin U$. That is, U is an independent set. □

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?

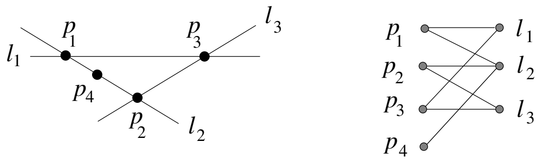


Fig. 18.1 We have four points and three lines. The number of incidences (edges in the point-line incidence graph on the right) is 7.

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 $\mathrm{E}\left[\mathrm{cr}(H)\right] \leq p^4x$. 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

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

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 $a\in\{0,1\}^n$,

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

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 $\Pr\left[\boldsymbol{h}_j(a)=1\right]\leq 1/2$ for all $j=1,\ldots,r$ (consult Exercise 18.11 for this conclusion). Hence,

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

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

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

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 $A = \{1, \dots, n\}$ 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 $f_1, \dots, f_t$ 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 $(f_1(x), \dots, f_t(x))$ 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 $B = \{1, \ldots, b\}$, 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*

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

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

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

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 $X_1, \ldots, X_k$ 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, $(x_1, \ldots, x_i, \ldots, x_k) \in T_i$ implies that $(x_1, \ldots, x_i', \ldots, x_k) \in T_i$ 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:

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

The *discrepancy* of f is the maximum $\operatorname{disc}(f) = \max_T \operatorname{disc}_T(f)$ 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 $\operatorname{disc}(f)$ is very hard to estimate. Fortunately, the discrepancy can be bounded from above using the following more tractable measure.

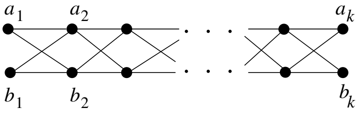


Fig. 18.2 A cube

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

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

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 $\mathrm{E}_D\left[f(D)\right]$ instead of $\mathrm{E}\left[f(D)\right]$.

Example 18.16. The difference between the measures $\operatorname{disc}(f)$ 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 $\operatorname{disc}_T(f)$ is just the sum of all entries in T divided by $|X|$. Thus, to determine $\operatorname{disc}(f)$

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 $1 \leq s, t \leq 2$.

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

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

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

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 $\operatorname{disc}(f) \leq \mathcal{E}(f)^{1/4}$. 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 $h : X \to \{-1, 1\}$, $\mathcal{E}(h) \geq (\mathrm{E}_{\boldsymbol{x}}[h(\boldsymbol{x})])^4$.

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

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

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

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

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

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

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

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

$$|\mathrm{E}_{\boldsymbol{x}}\left[f(\boldsymbol{x})\cdot g(\boldsymbol{x})\right]| \geq \mathrm{disc}_T(f)$$

and we can take $h(x) := f(x)\cdot g(x)$. Then $\left|\mathrm{E}_{\boldsymbol{x}}\left[h(\boldsymbol{x})\right]\right| \geq \mathrm{disc}(f)$. Moreover, $\mathcal{E}(h) = \mathcal{E}(f)$ because g_1 is constant on the rows and g_2 is constant on the columns so the product $g(D) = \prod_{x\in D} g(x)$ cancels to 1. □

This completes the proof of Theorem 18.17 in case $k = 2$. To extend it for arbitrary k, just repeat the argument k times. □

Say that a $(0, 1)$ matrix A is *odd* if the number of its all-1 rows is odd. Note that, if the matrix has only two columns, then it is odd iff the scalar (or inner) product of these columns over $GF(2)$ is 1. By this reason, a boolean function, detecting whether a given matrix is odd, is called the "generalized inner product" function. We will assume that input matrices have n rows and k columns.

That is, the *generalized inner product function* $\mathrm{GIP}(x)$ is a boolean function in kn variables, arranged in an $n \times k$ matrix $x = (x_{ij})$, and is defined by:

$$\mathrm{GIP}(x) = \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{k} x_{ij}\,.$$

Since we want our function to have range $\{-1, 1\}$, we will consider the function

$$f(x) = (-1)^{\mathrm{GIP}(x)} = \prod_{i=1}^{n}(-1)^{x_{i1}x_{i2}\cdots x_{ik}}\,. \tag{18.10}$$

Theorem 18.20. *For the* ± 1 *version* $f(x)$ *of the generalized inner product function we have that*

$$\mathcal{E}(f) = \left(1 - \frac{1}{2^k}\right)^n. \tag{18.11}$$

Proof. In our case, the function f is a mapping $f : X_1\times X_2\times\cdots X_k \to \{-1, 1\}$, where the elements of each set X_j are column vectors of length n. Hence, a cube D in our case is specified by two $n \times k$ $(0, 1)$ matrices $A = (a_{ij})$ and $B = (b_{ij})$. The cube D consists of all 2^k $n \times k$ matrices, the j-th column in each of which is either the j-th column of A or the j-th column of B. By (18.10), we have that

$$\begin{aligned} f(D) &= \prod_{x\in D} f(x) = \prod_{x\in D}\prod_{i=1}^{n}(-1)^{x_{i1}x_{i2}\cdots x_{ik}} \qquad\qquad \text{with } x_{ij}\in\{a_{ij}, b_{ij}\} \\ &= \prod_{i=1}^{n}\prod_{x\in D}(-1)^{x_{i1}x_{i2}\cdots x_{ik}} \\ &= \prod_{i=1}^{n}(-1)^{(a_{i1}+b_{i1})(a_{i2}+b_{i2})\cdots(a_{ik}+b_{ik})}\,. \end{aligned}$$

Note that the exponent $(a_{i1}+b_{i1})(a_{i2}+b_{i2})\cdots(a_{ik}+b_{ik})$ is even if $a_{ij}=b_{ij}$ for at least one $1\leq j\leq k$, and is equal to 1 in the unique case when $a_{ij}\neq b_{ij}$ for all $j=1,\ldots,k$, that is, when the i-th row of B is complementary to the i-th row of A. Thus,

$$f(D)=-1 \text{ iff the number of complementary rows in } A \text{ and } B \text{ is odd.}$$

Now, $\mathcal{E}(f)$ is the average of the above quantity over all choices of matrices A and B. We fix the matrix A and show that the expectation over all matrices B is precisely the right-hand side of (18.11). Let $A_1,\ldots,A_n$ be the rows of A and $B_1,\ldots,B_n$ be the rows of B. Then $f(D)=\prod_{i=1}^n g(B_i)$, where

$$g(B_i):=(-1)^{(a_{i1}+b_{i1})(a_{i2}+b_{i2})\cdots(a_{ik}+b_{ik})}=\begin{cases}+1 & \text{if } B_i\neq A_i\oplus\mathbf{1},\\ -1 & \text{if } B_i=A_i\oplus\mathbf{1}.\end{cases}$$

Thus, for every fixed matrix A, we obtain that

$$\begin{aligned}\mathrm{E}_B\Big[\prod_{i=1}^n g(B_i)\Big] &= \prod_{i=1}^n \mathrm{E}_{B_i}[g(B_i)] && \text{by (18.9)}\\ &= \prod_{i=1}^n \frac{1}{2^k}\sum_{B_i} g(B_i)\\ &= \prod_{i=1}^n \frac{1}{2^k}\left(2^k-1\right)\\ &= \left(1-\frac{1}{2^k}\right)^n. && \square\end{aligned}$$

18.11 Large deviation inequalities

A simple, but one of the most basic inequalities concerning the expectation of random variables states that a non-negative random variable X can take values much larger than $\mathrm{E}[X]$ with only small probability.

Theorem 18.21 (Markov's Inequality). *If X is a non-negative random variable then, for every real number $a>0$,*

$$\Pr[X\geq a]\leq\frac{\mathrm{E}[X]}{a}, \quad \textit{that is,} \quad \Pr[X\geq a\cdot\mathrm{E}[X]]\leq\frac{1}{a}.$$

Proof.

$$\mathrm{E}[X]=\sum_i i\cdot\Pr[X=i]\geq\sum_{i\geq a} a\cdot\Pr[X=i]=a\cdot\Pr[X\geq a]. \qquad \square$$

Intuitively, when $a \leq \mathrm{E}[X]$ the inequality is trivial. For $a > \mathrm{E}[X]$, it means the larger a is relative to the mean, the harder it is to have $X \geq a$. In particular, if $A_1, \ldots, A_n$ is a sequence of events, then Markov's inequality and the linearity of expectation (of their indicator random variables) implies that

$$\Pr[\text{fewer than } k \text{ events hold}] \geq 1 - \frac{\sum_{i=1}^n \Pr[A_i]}{k}.$$

In Markov's inequality, X can be an *arbitrary* non-negative random variable. In applications, however, X is often a sum of independent random variables. In these cases, Markov's inequality can be substantially sharpened. The main observation (due to Sergei Bernstein) is that, if X is a random variable and $t > 0$, then Markov's inequality yields

$$\Pr[X \geq a] = \Pr[\mathrm{e}^{tX} \geq \mathrm{e}^{ta}] \leq \mathrm{E}[\mathrm{e}^{tX}] \cdot \mathrm{e}^{-ta}. \tag{18.12}$$

There are many resulting inequalities known under a common name "Chernoff's inequalities." We mention just one of them.

Theorem 18.22 (Chernoff's Inequality). *Let $X_1, \ldots, X_n$ be independent random variables taking their values in the interval* $[0, 1]$. *Let $X = X_1 + \cdots + X_n$ and $\mu = \mathrm{E}[X]$. Then, for every real number $a > 0$, both* $\Pr[X \geq \mu + a]$ *and* $\Pr[X \leq \mu - a]$ *are at most* $\mathrm{e}^{-a^2/2n}$.

Note that the variables X_i need not be 0-1 variables: they can take arbitrary real values in the interval $[0, 1]$. Important restriction, however, is that these variables must be independent.

Proof. Consider random variables $Y_i = X_i - \mathrm{E}[X_i]$. Then $\mathrm{E}[Y_i] = 0$ and for their sum $Y = Y_1 + \cdots + Y_n$ we have that $Y = \sum_{i=1}^n X_i - \sum_{i=1}^n \mathrm{E}[X_i] = X - \mu$. Using (18.12) we have for every $t > 0$,

$$\begin{aligned}\Pr[X \geq \mu + a] &= \Pr[Y \geq a] \leq \mathrm{e}^{-ta}\, \mathrm{E}\left[\mathrm{e}^{tY}\right] = \mathrm{e}^{-ta}\, \mathrm{E}\left[\mathrm{e}^{\sum_{i=1}^n tY_i}\right] \\ &= \mathrm{e}^{-ta} \mathrm{E}\Big[\prod_{i=1}^n \mathrm{e}^{tY_i}\Big] = \mathrm{e}^{-ta} \prod_{i=1}^n \mathrm{E}\left[\mathrm{e}^{tY_i}\right],\end{aligned} \tag{18.13}$$

where in the last equality we used the independence of random variables Y_i, and hence, also of random variables e^{tY_i}.

In order to estimate $\mathrm{E}[\mathrm{e}^{tY_i}]$ from above, consider the function $f(x) = \mathrm{e}^{tx}$ and its derivatives. Since $t > 0$, the second derivative $f''(x)$ is positive, meaning that $f(x)$ is convex. Let $c + dx$ be a line through the points $(-1, f(-1))$ and $(1, f(1))$. Then $c - d = f(-1) = \mathrm{e}^{-t}$ and $c + d = f(1) = \mathrm{e}^{t}$, from which

$$c = \frac{\mathrm{e}^t + \mathrm{e}^{-t}}{2} \quad \text{and} \quad d = \frac{\mathrm{e}^t - \mathrm{e}^{-t}}{2}$$

follows. Since $f(x)$ is convex, all values $f(x)$ with $x \in [-1, 1]$ must lie below the line $c + dx$, that is, $\mathrm{e}^{tx} = f(x) \leq c + dx$ for all $x \in [-1, 1]$. Since $\mathrm{E}[Y_i] = 0$, we obtain

$$\mathrm{E}\left[\mathrm{e}^{tY_i}\right] \leq \mathrm{E}\left[c + dY_i\right] = c + d \cdot \mathrm{E}\left[Y_i\right] = c = \frac{1}{2}\left(\mathrm{e}^t + \mathrm{e}^{-t}\right) .$$

Using the Taylor series $\mathrm{e}^x = \sum_{k=0}^{\infty} x^k/k!$ we get

$$\begin{aligned} \mathrm{E}\left[\mathrm{e}^{tY_i}\right] &\leq \frac{1}{2}\Big(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots\Big) \\ &\quad + \frac{1}{2}\Big(1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} - \cdots\Big) \\ &= 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots + \frac{t^{2k}}{(2k)!} + \cdots \\ &\leq 1 + \frac{t^2}{2^1 \cdot 1!} + \frac{t^4}{2^2 \cdot 2!} + \cdots + \frac{t^{2k}}{2^k \cdot k!} + \cdots \qquad \text{since } 2^k \cdot k! \leq (2k)! \\ &= 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!} + \cdots \qquad \text{for } x = t^2/2 \\ &= \mathrm{e}^{t^2/2} . \end{aligned}$$

Together with (18.13) this gives the upper bound

$$\Pr\left[X \geq \mu + a\right] \leq \mathrm{e}^{-ta + t^2 n/2} .$$

The desired upper bound $\Pr\left[X \geq \mu + a\right] \leq \mathrm{e}^{-a^2/(2n)}$ now follows by taking $t = a/n$.

To prove the second inequality $\Pr\left[X \leq \mu - a\right] \leq \mathrm{e}^{-a^2/2n}$, it is enough to consider the random variable $X' := -X$. Then $X \leq \mu - a$ if and only if $X' \geq \mu' + a$, where $\mu' = \mathrm{E}\left[X'\right] = -\mu$. □

For sums of uniformly distributed ±1 random variables we have the following bounds. Let $Y = Y_1 + \cdots + Y_n$, where $\Pr\left[Y_i = +1\right] = \Pr\left[Y_i = -1\right] = 1/2$ and the Y_i are mutually independent. Then for any $a > 0$, both $\Pr\left[Y > a\right]$ and $\Pr\left[Y < -a\right]$ are smaller than $\mathrm{e}^{-a^2/2n}$.

Using Jensen's inequality to upper bound $\mathrm{E}\left[\mathrm{e}^{tX_i}\right]$, the following more general inequality can be derived: If $X_1, \ldots, X_n$ are mutually independent random variables with $|X_i| \leq c_i$ and $\mathrm{E}\left[X_i\right] = 0$, then

$$\Pr\left[X_1 + \cdots + X_n > a\right] < \exp\left(- \frac{a^2}{2(c_1^2 + \cdots + c_n^2)}\right) .$$

For sums of independent 0-1 random variables, the proof of Theorem 18.22 yields somewhat tighter bounds. Let $X = X_1 + \cdots + X_n$ be the sum of independent 0-1 random variables with $\Pr\left[X_i = 1\right] = p_i$. Let $\mu = \mathrm{E}\left[X\right] = p_1 + \cdots + p_n$. Since each X_i can only take values 0 or 1, the random variable e^{tX_i} can also take only values 1 or e^t. Hence, setting $a = (1 + \delta)\mu$ and

$t = \ln(1+\delta)$ in (18.12) and using the estimate

$$\mathrm{E}[(1+\delta)^{X_i}] = p_i \cdot (1+\delta) + (1-p_i)\cdot 1 = 1 + \delta p_i \leq \mathrm{e}^{\delta p_i}$$

we obtain that

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq \mathrm{e}^{\delta\mu}(1+\delta)^{-(1+\delta)\mu}\,.$$

Depending on how large the parameter δ is, one obtains different estimates. For example, if $\delta > 2\mathrm{e}-1$ then $(1+\delta)^{1+\delta} \geq (2\mathrm{e})^{1+\delta} \geq 2^{1+\delta}\mathrm{e}^{\delta}$, and we obtain that

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq 2^{-(1+\delta)\mu}$$

in this case. If $0 < \delta < 1$, then simple calculus yields

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq \mathrm{e}^{-\delta^2\mu/3}\,.$$

Similarly, $\Pr\left[X \leq (1-\delta)\mu\right] \leq \mathrm{e}^{-\delta^2\mu/2}$ holds for all $\delta > 0$.

Exercises

18.1. We have n letters going to n different persons and n envelopes with their addresses. We insert each letter into an envelope independently from each other at random (several letters may go in the same envelope). What is the expected number of correct matches? (Answer: $E = 1$.)

18.2. There are k people in a lift at the ground floor. Each wants to get off at a random floor of one of the n upper floors. What is the expected number of lift stops? *Hint*: Consider the indicator random variables X_i for the events that at least one person is off at the i-th floor, and apply the linearity of expectation. Answer: $E = n(1-(1-1/n)^k)$.

18.3. Let Ω be a uniform sample space, and let $X : \Omega \to \{0, 1, \ldots, M\}$ be a random variable with the expectation $\mu = M - a$ for some a. Prove that then, for any $1 \leq b \leq M$, $\Pr\left[X \geq M - b\right] \geq (b-a)/b$. *Hint*: Let B be the set of those points $\omega \in \Omega$ for which $X(\omega) < M - b$. Then $\Pr\left[B\right]\cdot(M-b) + \Pr\left[\overline{B}\right]\cdot M \geq M - a$, or $\Pr\left[B\right] \leq a/b$.

18.4. Let $\boldsymbol{T}$ be a random tournament chosen uniformly among all tournaments on n players. Then, by Szele's theorem, the expected number μ of Hamiltonian paths in it is $n!2^{-(n-1)}$. Use the argument of the previous exercise and Alon's upper bound (18.1) to prove that, with probability at least $\Omega(n^{-3/2})$, $\boldsymbol{T}$ has at least $n!2^{-n}$ Hamiltonian paths. *Hint*: Let $M = \Delta\cdot\mu$ with $\Delta = cn^{3/2}$, and take $a = M(1-1/\Delta)$, $b = M(1-1/2\Delta)$.

18.5. (Rédei 1934) Prove that every tournament contains a Hamiltonian path. *Hint*: Every path missing at least one player can be prolonged by adding him to that path.

18.6. Design an algorithm which, given an n-vertex graph $G = (V, E)$ of minimal degree d, constructs a dominating set $S \subseteq V$ of size $|S| \leq n\frac{1+\ln(d+1)}{d+1}$, whose existence is guaranteed by Theorem 18.3. *Hint*: For $S \subseteq V$, let $D(S)$ be the set of vertices dominated by S (i.e., $v \in D(S)$ if either $v \in S$ or v is joined by an edge with some vertex in S). Let $N(S) = V \setminus D(S)$. First show that, given $S \subseteq V$, there exists a vertex in $V \setminus S$ which dominates at least $|N(S)|(d+1)/n$ vertices in $N(S)$. Now construct the desired set S by iteratively adding a vertex with the maximum number of neighbors undominated by the vertices already chosen. Prove that at most $n/(d+1)$ vertices remain undominated after $n\ln(d+1)/(d+1)$ steps, such that adding them yields a dominating set of size at most $n\frac{1+\ln(d+1)}{d+1}$.

18.7. Show that Theorem 18.4 implies Turán's theorem: if a graph G has n vertices and $nk/2$ edges, then $\alpha(G) \geq n/(k+1)$ (see Exercise 4.8). *Hint*: Use the Cauchy–Schwarz inequality $\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right)$ with $a_i = (d_i+1)^{1/2}$ and $b_i = 1/a_i$.

18.8. Prove the Lieberher-Specker result for 2-satisfiable formulas: if F is a 2-satisfiable formula then at least γ-fraction of its clauses are simultaneously satisfiable, where $\gamma = (\sqrt{5}-1)/2$. *Sketch*: (Yannakakis 1994): Define the probability of a literal y to be satisfied to be: a if y occurs in a unary clause, and $1/2$ otherwise. Observe that then the probability that a clause C is satisfied is a if C is a unary clause, and at least $1 - a^2$ otherwise (at worst, a clause will be a disjunction of two literals whose negations appear as unary clauses); verify that $a = 1 - a^2$ for $a = \gamma$.

18.9. Prove that for any And-Or formula there is an input which satisfies at least half of its clauses. Is this bound optimal?

18.10. Given a graph $G = (V, E)$, define the And-Or formula

$$F_G = \bigwedge_{\{i,j\} \notin E} (\overline{x}_i \vee \overline{x}_j).$$

Each assignment $v = (v_1, \ldots, v_n) \in \{0,1\}^n$ can be interpreted as an incidence vector of the set of vertices $S_v = \{i : v_i = 1\}$. Show that S_v is a clique in G if and only if v satisfies the formula F_G.

18.11. Let $\boldsymbol{u}$ be a random vector uniformly distributed over $\mathbb{F}_2^n$, and let $u, v \in \mathbb{F}_2^n$ be any two distinct vectors with $v \neq 0$. The scalar product of u and v over $\mathbb{F}_2$ is the sum $\langle u, v\rangle = \sum_{i=1}^n u_i v_i \bmod 2$. Show that: $\Pr[\langle \boldsymbol{u}, v\rangle = 1] = 1/2$ for every $v \neq 0$, and $\Pr[\langle \boldsymbol{u}, v\rangle = \langle \boldsymbol{u}, w\rangle] = 1/2$ for every $v \neq w$. *Hint*: Exercise 13.2.

18.12. Recall that the *length* (or the *norm*) $\|v\|$ of a vector $v \in \mathbb{R}^n$ is the square root of the scalar product $\langle v, v\rangle$. Prove that for any vectors $v_1, \ldots, v_n$ in $\{+1, -1\}^n$ there are scalars $\epsilon_1, \ldots, \epsilon_n \in \{+1, -1\}$ such that

$$\|\epsilon_1 v_1 + \cdots + \epsilon_n v_n\| \leq n.$$

Hint: Choose the ϵ_i's independently at random to be $+1$ or -1 with probability $1/2$, and use the linearity of expectation to evaluate the expected length of the vector $\sum \epsilon_i v_i$ by computing the square of that quantity. When doing this, use the fact that ϵ_i and ϵ_j are independent and therefore $\mathrm{E}[\epsilon_i \cdot \epsilon_j] = \mathrm{E}[\epsilon_i] \cdot \mathrm{E}[\epsilon_j] = 0$.

18.13. Prove the following generalization of the previous result. Let $v_1, \ldots, v_n$ be vectors in $\{+1, -1\}^n$; $p_1 \ldots, p_n$ be real numbers in $[0,1]$, and set $w = p_1 v_1 + \cdots + p_n v_n$. Then there exist $\epsilon_1, \ldots, \epsilon_n \in \{0,1\}$ such that, setting $v = \epsilon_1 v_1 + \cdots + \epsilon_n v_n$, we have $\|w - v\| \leq n/2$. *Hint*: Pick the ϵ_i's independently with $\Pr[\epsilon_i = 1] = p_i$ and $\Pr[\epsilon_i = 0] = 1 - p_i$. Consider a random variable $X = \|w - v\|^2$, and prove that $\mathrm{E}[X] \leq n^2/4$.

18.14. Theorem 3.2 says that there exist (n,k)-universal sets of 0-1 vectors of size at most $r = k2^k \log n$. Give an alternative proof of this result using the linearity of expectation. *Hint*: Choose a random set $\boldsymbol{A}$ uniformly and independently from the family of all r-element subsets of $\{0,1\}^n$; the probability of one particular subset to be chosen is hence $\binom{2^n}{r}^{-1}$. For a set S of k coordinates and a vector $u \in \{0,1\}^k$, let $X_{S,u}$ denote the indicator random variable for the event $u \notin \mathrm{proj}_S(\boldsymbol{A})$, and let X be the sum of these random variables over all S and u. Show that for $r = k2^k \ln n$, $\mathrm{E}[X] < 1$.

19. The Lovász Sieve

Assume that we have a family of "bad" events. How can we make sure that there is some non-zero probability that none of the bad events will happen? By the union bound $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$, this probability is non-zero if the sum of probabilities of all these bad events is smaller than 1. In one sense this is best possible: when bad events are pairwise disjoint, the condition cannot be weakened. If we know that the bad events are independent, we can get a much better bound, by multiplying all the probabilities that each single bad event does not happen. This will work as long as each bad event has probability smaller than 1. But this will immediately fail, if at least two of the bad events are not independent.

In such cases—when there is some relatively small amount of dependence between events—one can use a powerful generalization of the union bound, known as the *Lovász Local Lemma.*

19.1 The Lovász Local Lemma

An event A is *mutually independent* of a collection of events if conditioning on any sub-collection $B_1, \ldots, B_m$ of these events does not affect the probability of A, that is,

$$\Pr[A \mid C_1 \cdots C_m] = \Pr[A]$$

for all $C_i \in \{B_i, \overline{B}_i\}$, $i = 1, \ldots, m$. Note that A might be independent of *each* of the events $B_1, \ldots, B_m$, but not be mutually independent of them. To see this, consider flipping a fair coin twice and the three events: B_1, B_2, A, where B_i is the event that the i-th flip is a head and A is the event that both flips are the same. Then A is independent of B_1 and of B_2 but $\Pr[A \mid B_1 B_2] = 1$.

Let $A_1, \ldots, A_n$ be events. A graph $G = (V, E)$ on the set of vertices $V = \{1, \ldots, n\}$ is said to be a *dependency graph* if, for all i, A_i is mutually independent of all the events A_j such that j is *not* adjacent to i in G, i.e.,

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_19,

for which $\{i, j\} \notin E$. We emphasize that A_i must not only be independent of each such A_j individually but also must be independent of any boolean combination of the A_j's. Such a graph G may be not uniquely defined, but we will not care about this. We will only be interested in the smallest possible degree of such a graph, which we call the *degree of dependence* of the events $A_1, \ldots, A_n$.

The following fact is known as the *Lovász Local Lemma.*

Lemma 19.1 (Erdős–Lovász 1975)**.** *Let* $A_1, \ldots, A_n$ *be events with* $\Pr[A_i] \leq p$ *for all* i*, and let* d *be the degree of their dependence. If* $\mathrm{e}p(d+1) \leq 1$ *then* $\Pr[\overline{A}_1\overline{A}_2 \cdots \overline{A}_n] > 0$*.*

As in the original proof of Erdős and Lovász, we will prove the lemma under the slightly stronger condition $4pd \leq 1$, and later show that the lemma remains true under weaker condition $\mathrm{e}p(d+1) \leq 1$, as well.

In the proof we will use two properties of the conditional probability which follow fairly easily from its definition as $\Pr[A \mid B] = \Pr[AB]/\Pr[B]$:

$$\Pr[A \mid BC] = \frac{\Pr[AB \mid C]}{\Pr[B \mid C]} \tag{19.1}$$

and

$$\Pr[A \mid BC] \cdot \Pr[B \mid C] \cdot \Pr[C] = \Pr[ABC]\ . \tag{19.2}$$

Proof (Spencer 1995). Fix a dependency graph G of our events of degree d. We prove by induction on m that for any m events (calling them $A_1, \ldots, A_m$ for convenience only)

$$\Pr[A_1 \mid \overline{A}_2 \cdots \overline{A}_m] \leq 2p.$$

For $m = 1$ this is obvious. Let $2, \ldots, k$ be the vertices from $\{2, \ldots, m\}$ which are adjacent to 1 in the dependency graph G. Using the identity (19.1), we can write

$$\Pr[A_1 \mid \overline{A}_2 \cdots \overline{A}_m] = \frac{\Pr[A_1\overline{A}_2 \cdots \overline{A}_k \mid \overline{A}_{k+1} \cdots \overline{A}_m]}{\Pr[\overline{A}_2 \cdots \overline{A}_k \mid \overline{A}_{k+1} \cdots \overline{A}_m]}. \tag{19.3}$$

We bound the numerator

$$\begin{aligned}\Pr[A_1\overline{A}_2 \cdots \overline{A}_k \mid \overline{A}_{k+1} \cdots \overline{A}_m] &\leq \Pr[A_1 \mid \overline{A}_{k+1} \cdots \overline{A}_m] \\ &= \Pr[A_1] \leq p\end{aligned}$$

since A_1 is mutually independent of $A_{k+1}, \ldots, A_m$. The denominator, on the other hand, can be bounded by the induction hypothesis

$$\begin{aligned}\Pr[\overline{A}_2 \cdots \overline{A}_k \mid \overline{A}_{k+1} \cdots \overline{A}_m] &= 1 - \Pr[A_2 \cup \cdots \cup A_k \mid \overline{A}_{k+1} \cdots \overline{A}_m] \\ &\geq 1 - \sum_{i=2}^{k} \Pr[A_i \mid \overline{A}_{k+1} \cdots \overline{A}_m] \\ &\geq 1 - 2p(k-1) \geq 1/2,\end{aligned}$$

because $k - 1 \leq d$ and $2pd \leq 1/2$. Thus

$$\Pr[A_1 \mid \overline{A}_2 \cdots \overline{A}_m] \leq p/(1/2) = 2p,$$

completing the induction. Finally, by (19.2),

$$\Pr[\overline{A}_1 \cdots \overline{A}_n] = \prod_{i=1}^{n} \Pr[\overline{A}_i \mid \overline{A}_1 \cdots \overline{A}_{i-1}] \geq (1-2p)^n > 0.$$

□

When the events A_i are not symmetric (i.e., when their probabilities might be very different) a more general form of the Lovász sieve is appropriate. This generalization is due to Spencer (1977).

Lemma 19.2. *Let $G = (V, E)$ be a dependency graph of events $A_1, \ldots, A_n$. Suppose there exist real numbers $x_1, \ldots, x_n$, $0 \leq x_i < 1$, so that, for all i,*

$$\Pr[A_i] \leq x_i \cdot \prod_{\{i,j\} \in E} (1 - x_j).$$

Then

$$\Pr[\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n] \geq \prod_{i=1}^{n} (1 - x_i).$$

In particular, with positive probability no event A_i holds.

Proof. The induction hypothesis of the earlier proof is replaced by

$$\Pr[A_1 \mid \overline{A}_2 \cdots \overline{A}_m] \leq x_1,$$

and, using the same identity (19.1), the denominator of (19.3) is set equal to

$$\prod_{j=2}^{k} \Pr[\overline{A}_j \mid \overline{A}_{j+1} \cdots \overline{A}_m],$$

which by the induction hypothesis, is at least

$$\prod_{j=2}^{k} (1 - x_j) = \prod_{\{1,j\} \in E} (1 - x_j)\,.$$ □

That, in the symmetric case, Lemma 19.1 holds with condition $4pd \leq 1$ replaced by $\mathrm{e}p(d+1) \leq 1$ follows from Lemma 19.2 by taking all $x_i = 1/(d+1)$ and observing that

$$x_i(1-x_i)^d = \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d > \frac{1}{d+1} \cdot \frac{1}{\mathrm{e}} \geq p\,,$$

by assumption of Lemma 19.1 and the fact that $(1 - 1/(d+1))^d > 1/\mathrm{e}$. This last fact is equivalent to $1 + 1/d < \mathrm{e}^{1/d}$, and holds because the Taylor series of $\mathrm{e}^{1/d}$ is (see (1.2)):

$$\mathrm{e}^{1/d} = 1 + \frac{1}{d} + \frac{1}{2!}\left(\frac{1}{d}\right)^2 + \frac{1}{3!}\left(\frac{1}{d}\right)^3 + \cdots\,.$$

The Lovász sieve works well when we have "much independence" between the events. In a similar vein, there is also an estimate, due to Razborov (1988), which works well if the events are "almost k-wise independent."

Let $A_1, \ldots, A_n$ be events, each of which appears with the same probability $\Pr[A_i] = p$. If *all* these events are mutually independent, then

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] = 1 - \Pr\left[\bigcap_{i=1}^{n} \overline{A}_i\right] = 1 - (1-p)^n \geq 1 - \mathrm{e}^{-pn}\,.$$

The mutual independence is a very strong requirement. It turns out that a reasonable estimate can be obtained also in the case when $\Pr\left[\bigcap_{i\in I} A_i\right]$ is only "near" to $p^{|I|}$ for the sets I of size up to some number k; in this case the events $A_1, \ldots, A_n$ are also called *almost k-wise independent.*

Lemma 19.3 (Razborov 1988)**.** *Let $n > 2k$ be any natural numbers, let $0 < p, \delta < 1$, and let $A_1, \ldots, A_n$ be events such that, for every subset $I \subseteq \{1, \ldots, n\}$ of size at most k,*

$$\left|\Pr\left[\bigcap_{i\in I} A_i\right] - p^{|I|}\right| \leq \delta.$$

Then

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \geq 1 - \mathrm{e}^{-pn} - \binom{n}{k+1}(\delta k + p^k).$$

Note that if the events are k-wise independent, then $\delta = 0$ and the obtained estimate worse by an additive term $\binom{n}{k+1}p^k$ than that for mutual independence.

Proof. Let us first consider the case where k is even. Let $B_1, \ldots, B_n$ be independent events, each having the success probability p. Applying the Bonferroni inequalities to $\Pr[\bigcup_{i=1}^{n} A_i]$ and $\Pr[\bigcup_{i=1}^{n} B_i]$ (see Exercise 1.37), we obtain that

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \geq \sum_{\nu=1}^{k}(-1)^{\nu+1} \sum_{|I|=\nu} \Pr\left[\bigcap_{i\in I} A_i\right] \tag{19.4}$$

and

$$\Pr\left[\bigcup_{i=1}^{n} B_i\right] \leq \sum_{\nu=1}^{k}(-1)^{\nu+1} \sum_{|I|=\nu} p^{|I|} + \sum_{|I|=k+1} p^{k+1}. \tag{19.5}$$

The assumption of the lemma that $A_1, \dots, A_n$ are almost k-wise independent implies that the right-hand side in (19.4) is at least

$$\sum_{\nu=1}^{k}(-1)^{\nu+1} \sum_{|I|=\nu} p^{|I|} - \delta k\binom{n}{k}. \tag{19.6}$$

On the other hand, the independence of $B_1, \dots, B_n$ implies that

$$\Pr\left[\bigcup_{i=1}^{n} B_i\right] = 1 - (1-p)^n \geq 1 - \mathrm{e}^{-pn}. \tag{19.7}$$

Combining (19.4), (19.5), (19.6) and (19.7) yields

$$\begin{aligned}\Pr\left[\bigcup_{i=1}^{n} A_i\right] &\geq 1 - \mathrm{e}^{-pn} - \delta k\binom{n}{k} - p^{k+1}\binom{n}{k+1} \\ &\geq 1 - \mathrm{e}^{-pn} - \binom{n}{k+1}(\delta k + p^{k+1}).\end{aligned}$$

In the case where k is odd, we use the above argument with $k-1$ substituted for k. □

19.2 Disjoint cycles

By a *digraph* we will mean a directed graph without parallel edges. Such a graph is *k-regular* if every vertex has exactly k outgoing edges.

Theorem 19.4. *Every k-regular digraph has a collection of $\lfloor k/(3\ln k)\rfloor$ vertex-disjoint cycles.*

Proof. Let $G = (V, E)$ be a k-regular digraph. Set $r := \lfloor k/(3\ln k)\rfloor$, and color the vertices uniformly at random using colors $\{1, \dots, r\}$. That is, each vertex v gets a particular color independently and with the same probability $1/r$. Let A_v be the event that v does not have any out-neighbor of the same color as v. (An out-neighbor of v is the second endpoint of an edge leaving v.) We need only to show that $\Pr[\bigcap_{v\in V}\overline{A_v}] > 0$.

Since each vertex has k out-neighbors, we have that

$$\Pr[A_v] = \left(1 - \frac{1}{r}\right)^k < \mathrm{e}^{-k/r} \leq \mathrm{e}^{-3\ln k} = k^{-3}\,.$$

For a vertex v, let $N(v)$ be the set consisting of v and all its k out-neighbors. Then A_v is mutually independent of the events in $\{A_u \,:\, N(u) \cap N(v) = \emptyset\}$. Since this set contains at most $(k+1)^2$ events, the degree of dependence of the events A_v is $d \leq (k+1)^2$. Hence, to apply the Lovász Local Lemma we only need that $4k^{-3}(k+1)^2 \leq 1$, which is true for $k \geq 6$. For $k < 6$ the theorem is trivially true since then $r = 1$. □

Alon, McDiarmid and Molloy (1996) proved that, in fact, $\Omega(k^2)$ vertex-disjoint cycles exist and conjectured that at least $\binom{k+1}{2}$ cycles should exist.

19.3 Colorings

A striking feature of the Lovász sieve is the lack of conditions on the total number n of events – only the degree of their dependence is important. This is particularly useful when dealing with large families whose members share not too many points in common. Let us demonstrate this with several typical examples.

First, let us consider 2-colorings of hypergraphs. Recall that a family of sets $\mathcal{F}$ is 2-*colorable* if it is possible to color the points of the underlying set in red and blue, so that no member of $\mathcal{F}$ is monochromatic. A family is k-*uniform* if all its members have size k.

In Chap. 2 (see Theorem 3.4) we proved that if the family $\mathcal{F}$ is relatively small then it is 2-colorable: Every k-uniform family of fewer than 2^{k-1} sets is 2-colorable.

Let us recall the argument. Suppose $\mathcal{F}$ is a k-uniform family with at most $2^{k-1} - 1$ sets. Consider a random coloring, each element independently colored red or blue with probability 1/2. Any one member of $\mathcal{F}$ will then be monochromatic with probability $2 \cdot 2^{-k} = 2^{1-k}$, and so the probability that *some* member will be monochromatic, does not exceed $|\mathcal{F}| \cdot 2^{1-k}$, which is strictly smaller than 1. Therefore, at least one coloring must leave no member of $\mathcal{F}$ monochromatic.

Now suppose that $\mathcal{F}$ has more than 2^k members. Then the above random coloring will be doomed since the chances of it to be a proper 2-coloring will tend to zero. Fortunately, we do not require a high probability of success, just a positive probability of success. For example, if $\mathcal{F}$ is a family of m mutually disjoint k-element subsets of some set, then the events A_i="the i-th member of $\mathcal{F}$ is monochromatic" are mutually independent, and so the probability that none of them holds is exactly $\left(1 - 2^{-(k-1)}\right)^m$, which is positive no matter how large m is. Therefore, $\mathcal{F}$ is 2-colorable.

Of course for general families $\mathcal{F}$, the events $A_1, \ldots, A_m$ are not independent as some pairs of members may intersect. In such situations the Lovász sieve shows its surprising power.

Theorem 19.5 (Erdős–Lovász 1975). *If every member of a k-uniform family intersects at most 2^{k-3} other members, then the family is 2-colorable.*

Proof. Suppose $\mathcal{F} = \{S_1, \ldots, S_m\}$ is a family of k-element subsets of some set X. Consider a random coloring of X, each point independently colored red or blue with probability 1/2. Let A_i denote the event that S_i is monochromatic. Then $\Pr[A_i] = p$ where $p = 2(1/2)^{|S_i|} = 2^{1-k}$. Our goal is to show that $\Pr[\overline{A}_1 \cdots \overline{A}_m] > 0$. Define a dependency graph by joining A_i and A_j if and only if $S_i \cap S_j \neq \emptyset$. By the assumption, this graph has degree at most $d = 2^{k-3}$. Since $4dp = d2^{3-k} \leq 1$, Lemma 19.1 yields the result. □

In the general (not necessarily uniform) case we have the following.

Theorem 19.6 (Beck 1980). *Let $\mathcal{F}$ be a family of sets, each of which has at least k ($k \geq 2$) points. Also suppose that for each point v,*

$$\sum_{S \in \mathcal{F} : v \in S} (1 - 1/k)^{-|S|} 2^{-|S|+1} \leq \frac{1}{k}.$$

Then $\mathcal{F}$ is 2-colorable.

Proof. Let $\mathcal{F} = \{S_1, \ldots, S_m\}$ and (again) color the points with red and blue at random, independently of each other and with probability 1/2. Let A_i denote the event that S_i is monochromatic; hence $\Pr[A_i] = 2^{-|S_i|+1}$. Consider the same dependency graph $G = (V, E)$ as above: $\{i, j\} \in E$ if and only if $S_i \cap S_j \neq \emptyset$. We shall prove that the condition of Lemma 19.2 is satisfied with

$$x_i := (1 - 1/k)^{-|S_i|}\, 2^{-|S_i|+1}.$$

Indeed, by the definition of the graph G, for every $i = 1, \ldots, m$ we have

$$\begin{aligned} x_i \prod_{\{i,j\} \in E} (1 - x_j) &\geq x_i \prod_{v \in S_i} \prod_{j : v \in S_j} (1 - x_j) \\ &\geq x_i \prod_{v \in S_i} \left[1 - \sum_{j : v \in S_j} x_j\right] \geq x_i (1 - 1/k)^{|S_i|}, \end{aligned}$$

since, by the condition of the theorem, $\sum_{j : v \in S_j} x_j \leq 1/k$. Thus,

$$x_i \prod_{\{i,j\} \in E} (1 - x_j) \geq x_i\, (1 - 1/k)^{|S_i|} = 2^{-|S_i|+1} = \Pr[A_i].$$

By the application of Lemma 19.2 we obtain $\Pr[\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n] > 0$, i.e., there is a 2-coloring in which no set of $\mathcal{F}$ is monochromatic. □

Later, Beck (1991) was even able to design an efficient randomized algorithm finding a desired coloring. This was the first time when an algorithmic version of the Lovász Local Lemma was found.

Let us now consider yet another coloring problem. Let $\mathcal{F}$ be a family of k-element sets and suppose that no point appears in more than l of its members. By induction on k, it can be shown (see Exercise 19.5) that then it is possible to color the points in $r = l(k-1)+1$ colors so that no member of $\mathcal{F}$ contains two points of the same color. On the other hand, if we have only $r < k$ colors, then every member of $\mathcal{F}$ will always have at least k/r points of the same color. Is it possible, also in this case (when $r < k$) to find a coloring such that no member has much more than k/r points of one color? The following result says that, if $k = l$ and if we have about $k/\log k$ colors, then such a coloring exists.

Theorem 19.7 (Füredi–Kahn 1986). *Let k be sufficiently large. Let $\mathcal{F}$ be a k-uniform family of sets and suppose that no point belongs to more than k sets of $\mathcal{F}$. Then it is possible to color the points in $r = \lfloor k/\log k \rfloor$ colors so that every member of $\mathcal{F}$ has at most $v = \lceil 2\mathrm{e}\log k \rceil$ points of the same color.*

In fact, Füredi and Kahn proved a stronger result, where $v = \lfloor 4.5 \log k \rfloor$ and the members of $\mathcal{F}$ have size at most k. The argument then is the same but requires more precise computations.

Proof. Color the points of X by r colors, each point getting a particular color randomly and independently with probability $1/r$. Let $A(S,i)$ denote the event that more than v points of S get color i. We are going to apply Lemma 19.1 to these events. Events $A(S,i)$ and $A(S',i')$ can be dependent only if $S \cap S' \neq \emptyset$. So, we consider the following dependency graph G for these events: the vertex set consists of the pairs (S,i) where $S \in \mathcal{F}$ and $1 \leq i \leq r$, and two vertices (S,i) and (S',i') are joined by an edge if and only if $S \cap S' \neq \emptyset$.

Let d be the maximum degree of G. By the condition on our family $\mathcal{F}$, every member can intersect at most $k(k-1)$ other members, implying that $d \leq (1 + k(k-1))r \leq k^3$. By Lemma 19.1, it remains to show that each of the events $A(S,i)$ can happen with probability at most $1/(4k^3)$.

Since $|S| = k$, the probability that only the points of a subset $I \subseteq S$ get color i, is $(1/r)^{|I|}(1-1/r)^{k-|I|}$. Summing over all subsets I of S, then the event $A(S,i)$ happens with probability at most

$$\sum_{t>v} \binom{k}{t} \left(\frac{1}{r}\right)^t \left(1 - \frac{1}{r}\right)^{k-t} \leq \binom{k}{v} \left(\frac{1}{r}\right)^v < \left(\frac{\mathrm{e}k}{vr}\right)^v \leq 2^{-v} < k^{-4}.$$

By Lemma 19.1, with positive probability, none of the events $A(S,i)$ will happen, and the desired coloring exists. □

19.4 The k-SAT problem

Let $x_1, \ldots, x_n$ be boolean variables. A literal is a boolean variable x_i or its negation $\overline{x_i}$. A *k-CNF formula* (conjunctive normal form) is an And of clauses, each being an Or of k literals. Such a CNF formula φ is *satisfiable* if these exists a truth assignment $a \in \{0,1\}^n$ for which $\varphi(a) = 1$.

The k-SAT problem is, given a k-CNF, to decide whether it is satisfiable or not; here $k \geq 3$ is assumed to be a fixed constant. Of course, this question can always be solved in 2^n trials: just test all 2^n possible assignments one by one. This dummy strategy will, however, take a huge amount of time on formulas with, say, $n = 100$ variables. Are there any quicker algorithms working, say in time n^c for some constant c? To show that no such algorithm exists is one of the central problems (if not the central) of the whole of computer science, and is known under the name "P vs. NP problem."

On the other hand, the Lovász Local Lemma gives us a tool to quickly recognize some satisfiable CNFs by just looking at their structure!

We say that two clauses *overlap* if they have a common variable x_i, regardless of whether the variable is negated or not in the clauses. In this case we also say that the clauses *share* the variable x_i.

Lemma 19.8. *Let φ be a k-CNF formula. If each of its clauses overlaps with at most 2^{k-2} clauses, then φ is satisfiable.*

Note that the total number n of variables is irrelevant here!

Proof. Consider a random experiment where the variables in φ are assigned truth values by independent tosses of a fair coin. Let A_i be the event that the i-th clause of φ is *not* satisfied. For this event to happen, all k literals in C_i must receive a "wrong" value. Hence, $p = \Pr[A_i] = 2^{-k}$. Further, each A_i is mutually independent of the set of all A_j such that the i-th clause C_i and the j-th clause C_j of φ do not overlap. Hence, the dependency graph of the events A_i has degree $d \leq 2^k/4$. Since $4dp \leq 42^{k-2}2^{-k} = 1$, Lemma 19.1 applies and φ will be satisfied with non-zero probability. □

This lemma is "non-constructive:" it gives no clue on how to find a satisfying assignment in a reasonable (polynomial) time.

Beck (1991) achieved a breakthrough by proving that a polynomial-time algorithm exists which finds a satisfying assignment to every k-CNF formula in which each clause has a neighbourhood of at most $2^{k/48}$ other clauses. His approach was deterministic and used the nonconstructive version of the Lovász Local Lemma as a key ingredient, basically proving that even after truncating clauses to a 48th of their size (a step used to simplify the formula and make it fall apart into small components), a solution remains guaranteed and can then be looked for by exhaustive enumeration. Alon (1991) simplified Beck's algorithm and analysis by introducing randomness and presented an algorithm that works up to neighbourhoods of $2^{k/8}$ in size. Czumaj and Scheideler (2000) later demonstrated that a variant of the method can be made to work for the non-uniform case where clause sizes vary. Srinivasan (2008) improved the bound of what was polynomial-time feasible to essentially $2^{k/4}$ by a more accurate analysis. Finally, Moser (2009) published a polynomial-time algorithm that can cope with neighbourhood size up to 2^{k-5} neighbours, which is asymptotically optimal with a constant gap. Recently, Moser and Tardos (2010) gave a randomized algorithm for the general (non-symmetric) version of the Lovász Local Lemma.

Theorem 19.9 (Moser 2009)**.** *There is a constant c such that, given a k-CNF formula φ with m clauses, none of which overlaps with more than $r = 2^{k-c}$ other clauses, one can find a satisfying assignment for φ in expected time polynomial in m.*

We are not going to prove this theorem in full detail. We rather give a coarse, intuitive and convincing argument that this "must" hold.

Moser's algorithm Solve(φ) is a randomized algorithm consisting of recursive calls of (also recursive) procedures Fix(C) for clauses C of φ: Pick a random assignment $a \in \{0, 1\}^n$; while there is an unsatisfied clause C, call Fix(C). The procedure Fix(C) itself is the following recursive procedure:

Step 1: Replace the variables of C with new random values.
Step 2: While there is a clause D that shares a variable with C that is not satisfied, call Fix(D).

First, observe that, *if* Fix(C) *terminates*, then every clause A that was satisfied before Fix(C) is called will remain satisfied after Fix(C) is called. This holds because each flipping of the variables in a clause C can only affect the values of clauses that share a common variable with C. So, if the value of A is turned from true to false at some moment of the execution of Fix(C), then Fix(A) is called.

By this observation, Solve makes at most m calls to Fix, *if* Fix(C) *always terminates.* So we need to show all the Fix(C) terminate. Suppose the algorithm makes at least s Fix calls including all the recursive ones. We will show that s is bounded by $O(m \log m)$, and thus the algorithm terminates in almost linear expected time.

This can be proved (at least at an intuitive level) by Kolmogorov complexity arguments. An extensive account concerning these arguments can be found in the book by Li and Vitányi (2008). Here we just describe this argument on an informal level.

The Kolmogorov complexity, $K(x)$, of a string x is the length of the string's shortest description in some fixed universal description language. Such a description language can be based on any programming language. If P is a program which outputs a string x, then P is a description of x. The length of the description is just the length of P as a character string. In determining the length of P, the lengths of any subroutines used in P must be accounted for. The length of any integer constant n which occurs in the program P is the number of bits required to represent n, that is (roughly) $\log_2 n$. For example, a huge string $x = 010101\cdots01$ with 01 repeated 2^{100} times can be described as a program "repeat 01 2^{100} times", whose length (after binary encoding) is only about 100 bits. In general, any string containing some repeating patterns has small Kolmogorov complexity. On the other hand, random strings are "resistant" against compression, and hence, have large Kolmogorov complexity.

It is straightforward to compute upper bounds for $K(x)$: simply compress the string x with some method, implement the corresponding decompresser in the chosen language, concatenate the decompresser to the compressed string, and measure the resulting string's length.

A string x is compressible by a number c if it has a description whose length does not exceed $|x| - c$, that is, if $K(x) \leq |x| - c$, where $|x|$ is the length of (number of symbols in) x. Otherwise x is incompressible by c. A string incompressible by 1 is said to be simply incompressible or *Kolmogorov random*; by the pigeonhole principle, incompressible strings must exist, since there are 2^n bit strings of length n but only $2^n - 1$ shorter strings, that is strings of length $n - 1$ or less.

For the same reason, "most" strings are complex in the sense that they cannot be significantly compressed: $K(x)$ is not much smaller than $|x|$, the length of x in bits. To make this precise, fix a value of n. There are 2^n binary strings of length n. Let x be a string chosen uniformly at random with probability 2^{-n}. It is easy to show that the probability that x *is* compressible by c is negligible: it is $2^{-c+1} - 2^{-n}$. To see this, it is enough to observe that the number of descriptions of length not exceeding $n - c$ is given by the geometric series: $1 + 2 + 2^2 + \cdots + 2^{n-c} = 2^{n-c+1} - 1$, and there remain at least $2^n - 2^{n-c+1} + 1$ binary strings of length n that are incompressible by c.

Now, the so-called *incompressibility argument* works as follows: In order to show that some condition holds, assume it does not hold and use this assumption to show that then some Kolmogorov random string x would have a description much shorter than $K(x)$.

After this short excursion into Kolmogorov complexity, let us return to Moser's algorithm. Fix a Kolmogorov random string x of length $n+sk$ (where n is the total number of variables) and assume the algorithm uses the first n bits as the initial assignment a, and k bits each to replace the variables in each Fix call. (If we choose the string x randomly then it will be Kolmogorov random with high probability.) The random string x is used in Step 1 of Fix(C) to replace the values of variables in C by "fresh" random values, and each time next k bits of x are used.

If we know which clause is being fixed, we know the clause is violated so we know all the bits of this clause and thus we learn k bits of x (recall that assignments used by an algorithm are from the string x). We then replace those bits with another part of x.

So we can describe x by the list of clauses we fix plus the remaining n bits of the final assignment. We can describe each clause C such that $\mathrm{Fix}(C)$ is called by Solve using $O(m\log m)$ bits. The remaining fixed clauses can be described by $\log_2 r+\alpha$ bits (for a constant α) because either it is one of r clauses that intersects the previous clause or we indicate the end of a recursive call (keeping track of the recursion stack). This is exactly the place where the compression comes: Since the clause was not satisfied, we reveal k bits of information about the string x, but since $r\le 2^{k-c}$, we can describe this information using only $k-c$ bits. Since the string x was Kolmogorov random, we must have

$$O(m\log m)+s(\log r+\alpha)+n\ge n+sk$$

or $s(k-\log r-\alpha)\le O(m\log m)$. Now, if $r\le 2^{k-c}$ for $c>\alpha$, then $k-\log r-\alpha$ is a positive constant (not exceeding k, which is also a constant), implying that $s=O(m\log m)$.

Exercises

19.1. Let $A_1,\ldots,A_n$ be events, and suppose that each of them is mutually independent of all the other events except for at most d of them. Let $0<\epsilon<1$, and assume that

$$\Pr[A_i]\le\frac{\epsilon}{n}\left(1-\frac{\epsilon}{n}\right)^d$$

for all $i=1,\ldots,n$ Prove that then $\Pr[\cap_i\overline{A_i}]\ge 1-\epsilon$.

19.2. Let $\mathcal{F}$ be a k-uniform k-regular family, i.e., each set has k points and each point belongs to k sets. Let $k\ge 10$. Show that then at least one 2-coloring of points leaves no set of $\mathcal{F}$ monochromatic.

19.3. The van der Waerden number $W(2,k)$ is the least number n such that any coloring of $\{1,2,\ldots,n\}$ in two colors gives a monochromatic arithmetic progression with k terms. Prove that $W(2,k)>2^k/(2\mathrm{e}k)$. *Hint*: Assume that $n\le 2^k/(2\mathrm{e}k)$ and observe that one progression with k terms intersects at most nk others.

19.4. (Erdős–Lovász 1975). Consider the colorings of real numbers in r colors. Say that a set of numbers is *multicolored* if it contains elements of all r colors. Fix a finite set X of real numbers, and let m be such that

$$4rm(m-1)\left(1-\frac{1}{r}\right)^m<1.$$

Using Lemma 19.1 prove that then, for any set S of m numbers there is an r-coloring under which every translate $x+S:=\{x+y:\ y\in S\}$, with $x\in X$, is multicolored. *Hint*: Take a random r-coloring $Y = \bigcup_{x \in X}(x + S)$ which, for every point $y \in Y$, takes a particular color randomly and independently with probability $1/r$. Consider events A_x saying that $x + S$ is not multicolored, and show that $\Pr[A_x] \leq r\left(1 - \frac{1}{r}\right)^m$. Also observe that for each point x there are at most $m(m-1)$ other points x' for which $(x + S) \cap (x' + S) \neq \emptyset$.

19.5. (Füredi–Kahn 1986). Let $\mathcal{F}$ be a family of *rank* a, i.e., each member has at most a points, and suppose that no point belongs to more than b members of $\mathcal{F}$. Prove that then it is possible to color the points in $r = (a-1)b + 1$ colors so that every member of $\mathcal{F}$ is differently colored, i.e., no member of $\mathcal{F}$ has two points of the same color. *Hint*: By induction on a. The case $a = 2$ is Exercise 4.27. For the induction step, select a sequence of points $V = \{x_1, x_2, \ldots, x_m\}$ by the following rule: at the i-th step take a set $F \in \mathcal{F}$ disjoint from $\{x_1, \ldots, x_{i-1}\}$, and let x_i be an arbitrary point in this set. If we delete the points V from all members of $\mathcal{F}$, we obtain a family $\mathcal{F}'$ of rank at most $a - 1$. By the induction hypothesis, $\mathcal{F}'$ can be differently colored using only $(a-2)b + 1$ colors. So, it remains to color the deleted points. For this, consider the graph $G = (V, E)$ where two points are joined by an edge iff both these points belong to the same member of $\mathcal{F}$. Show that $\Delta(G) \leq b - 1$ and apply Exercise 4.27.

20. The Deletion Method

As described in previous sections, the basic probabilistic method works as follows: trying to prove that an object with certain properties exists, one defines an appropriate probability space of objects and then shows that the desired properties hold in this space with positive probability. In this section, we consider situations where the "random" object does not have all the desired properties but may have a few "blemishes." With a small alteration, we remove the blemishes, giving the desired structure.

20.1 Edge clique covering

An edge clique covering of a graph G is a family of complete subgraphs of G such that every edge of G is an edge of at least one member of the family. The minimum cardinality of such a family is the *edge clique covering number* of G, denoted by $\mathrm{cc}(G)$.

The following general upper bound on this number was proved by Alon (1986b). A *non-neighbor* of a vertex v in a given graph is a vertex $u \neq v$ which is non-adjacent to v.

Theorem 20.1 (Alon 1986b)**.** *Let G be a graph on n vertices such that every vertex has at least one neighbor and at most d non-neighbors. Then*

$$\mathrm{cc}(G) = O(d^2 \ln n)\,.$$

As often happpens in probabilistic proofs, the following estimates turn out to be very useful (see inequalities (1.4) and (1.4) in Chap. 1.5): For every $t > 0$, $1 + t < \mathrm{e}^t$ and $1 - t > \mathrm{e}^{-t-t^2/2}$.

Proof. Consider the following procedure for choosing a complete subgraph of $G = (V, E)$. In the first phase, pick every vertex $v \in V$ independently, with probability $p = 1/(d + 1)$, to get a set W. In the second phase remove from

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science. An EATCS Series, DOI 10.1007/978-3-642-17364-6_20,

W all vertices having at least one non-neighbor in W. Clearly, the resulting set is the set of vertices of a complete subgraph of G.

Now apply the above procedure, independently, t times to get t complete subgraphs $K_1, \dots, K_t$ of G; here, t is a parameter to be specified soon. Let us estimate the expected value of the number of edges of G that are not covered by the union of the K_i's. Let uw be an edge of G, and fix an i, $1 \leq i \leq t$. Note that the edge uw is covered by K_i, if both its endpoints u and w and none of their $\leq 2d$ non-neighbors were chosen in the first phase. Hence

$$\Pr[K_i \text{ covers } uw] \geq p^2(1-p)^{2d} = \frac{1}{(d+1)^2}\left(1 - \frac{1}{d+1}\right)^{2d} \geq \frac{1}{\mathrm{e}^2(d+1)^2}\,.$$

By the union bound,

$$\Pr[\text{none of the } K_i\text{'s covers } uw] \leq \left(1 - \frac{1}{\mathrm{e}^2(d+1)^2}\right)^t \leq \exp(-t/\mathrm{e}^2(d+1)^2)\,.$$

By the linearity of expectation, the expected number of non-covered edges is at most

$$\binom{n}{2} \cdot \exp(-t/\mathrm{e}^2(d+1)^2)\,.$$

By taking $t := \lfloor cd^2 \log_2 n \rfloor$ for a sufficiently large (absolute) constant c, this expectation can be made < 1. Hence, by the pigeonhole principle of expectation, there is at least one choice of t complete subgraphs of G that form a clique covering of G and $\mathrm{cc}(G) \leq t$, as needed. □

20.2 Independent sets

Here is a short argument that gives roughly half of Turán's celebrated theorem. A set of vertices in a graph is *independent* if no two vertices from this set are joined by an edge; the *independence number* $\alpha(G)$ is the maximum number of vertices in G with no edges between them. Turán's theorem states: if G has n vertices and $nk/2$ edges, then $\alpha(G) \geq n/(k+1)$. We can get "halfway" to this result with the deletion method.

Theorem 20.2. *If a graph $G = (V, E)$ has n vertices and $nk/2$ edges, then $\alpha(G) \geq n/2k$.*

Proof (Spencer 1987). Form a random subset $\boldsymbol{S} \subseteq V$ of vertices by including each vertex independently with probability p. That is, we take a coin which comes up heads with probability p and flip it for each $x \in V$ to see if x is "chosen" to be in $\boldsymbol{S}$. Let X denote the number of vertices in $\boldsymbol{S}$, and let Y denote the number of edges of G, both ends of which lie in $\boldsymbol{S}$. For each edge $e \in E$, let Y_e be the indicator random variable for the event $e \subseteq \boldsymbol{S}$. Then for each edge $e \in E$, $\mathrm{E}[Y_e] = p^2$ as two vertices must be chosen for e to be inside $\boldsymbol{S}$. So, by linearity of expectation,

$$\mathrm{E}[Y] = \sum_{e \in E} \mathrm{E}[Y_e] = \frac{nk}{2} p^2.$$

Clearly, $\mathrm{E}[X] = np$, so, again by linearity of expectation,

$$\mathrm{E}[X - Y] = np - \frac{nk}{2} p^2.$$

We choose $p = 1/k$ to maximize this quantity, giving

$$\mathrm{E}[X - Y] = \frac{n}{k} - \frac{n}{2k} = \frac{n}{2k}.$$

Thus, there exists at least one point in the probability space for which the difference $X - Y$ is at least $n/2k$. That is, there is a set S which has at least $n/2k$ more vertices than edges. Delete one vertex from each edge from S leaving a set S'. This set S' is independent and has at least $n/2k$ vertices. □

20.3 Coloring large-girth graphs

A *cycle* of length k in a graph $G = (V, E)$ is a sequence of vertices $v_1, v_2, \ldots, v_k$ such that: $v_1 = v_k$, $v_i \neq v_j$ for all $1 < i < j \leq k$, and all the edges $\{v_i, v_{i+1}\}$ belong to E. The *girth* $g(G)$ of a graph is the length of the shortest cycle in G. Recall also that the *chromatic number* $\chi(G)$ is the minimal number of colors which we need to color the vertices of G so that no two vertices of the same color are joined by the edge.

A striking example of the deletion method is the proof that for fixed k and l, there are graphs with girth at least l and chromatic number at least k. This result was proved by Erdős, and is highly unintuitive. If the girth is large there is no simple reason why the graph could not be colored with a few colors: locally it is easy to color such a graph with three colors. Thus, we can force the chromatic number only by some global considerations and the deletion method helps in doing this. The proof we present here is a simplification of the Erdős proof, due to Alon and Spencer (1992).

Theorem 20.3 (Erdős 1959). *For all k, l there exists a finite graph G with $\chi(G) \geq k$ and $g(G) \geq l$.*

Since every color class must be an independent set, we have an obvious relation: $\chi(G) \geq n/\alpha(G)$. So, instead of showing that $\chi(G) \geq k$ it is sufficient to show that $\alpha(G) \leq n/k$. We will use this simple trick in the proof below.

Proof (Alon–Spencer 1992). Fix $\theta < 1/l$. Let n be large enough and let $\boldsymbol{G}$ be a random graph on n vertices, where each pair of nodes is joined by an edge with independent probability $p = n^{\theta-1}$. Let X be the number of cycles in $\boldsymbol{G}$ of length at most l. How many cycles $v_1, v_2, \ldots, v_i, v_1$ of length i can our graph have?

There are $(n)_i = n(n-1)\cdots(n-i+1)$ sequences $v_1, v_2, \ldots, v_i$ of distinct vertices, and each cycle is identified by $2i$ of those sequences: there are two possibilities to choose the "direction" and i possibilities to choose the first vertex of the cycle. Thus, for $3 \leq i \leq t$ there are $(n)_i/2i \leq n^i/2i$ potential cycles of length i, each of which is in $\boldsymbol{G}$ with probability p^i. By the linearity of the expectation

$$\mathrm{E}\left[X\right] = \sum_{i=3}^{l} \frac{(n)_i}{2i} p^i \leq \sum_{i=3}^{l} \frac{n^{\theta i}}{2i} = o(n)$$

as $\theta l < 1$. By Markov's inequality,

$$\Pr\left[X \geq n/2\right] \leq \frac{2\mathrm{E}\left[X\right]}{n} = o(1).$$

Set $x := \lceil \frac{3}{p} \ln n \rceil$, so that

$$\Pr\left[\alpha(\boldsymbol{G}) \geq x\right] \leq \binom{n}{x} (1-p)^{\binom{x}{2}} < \left[n\mathrm{e}^{-p(x-1)/2}\right]^x = o(1).$$

Let n be sufficiently large so that both these events have probability less than $1/2$. Then there is a specific G with less than $n/2$ "short" cycles, i.e., cycles of length less than l, and with $\alpha(G) < x \leq 3n^{1-\theta} \ln n$. Remove from G a vertex from each short cycle. This gives a graph G' which has no short cycles and still has at least $n/2$ vertices. Hence, G' has girth greater than l and $\alpha(G') \leq \alpha(G)$ (since $\alpha(G)$ cannot grow when we delete vertices). Thus

$$\chi(G') \geq \frac{|G'|}{\alpha(G')} \geq \frac{n/2}{3n^{1-\theta} \ln n} = \frac{n^\theta}{6 \ln n}.$$

To complete the proof, it remains to take n sufficiently large so that this is greater than k. □

20.4 Point sets without obtuse triangles

Around 1950, Erdős conjectured that every set of more than 2^n points in $\mathbb{R}^n$ determines at least one *obtuse angle*, that is, an angle that is strictly greater than $\pi/2$. In other words, any set of points in $\mathbb{R}^n$ which only has acute angles (including right angles) has size at most 2^n.

In 1962, Danzer and Grünbaum proved this conjecture. They also constructed configurations of $2n-1$ points in $\mathbb{R}^n$ with only acute angles, and conjectured that this may be best possible. But 21 years later, Erdős and Füredi – using a probabilistic argument – disproved this conjecture. It turns out that, if the dimension n is high, the bound $2n-1$ is not even near to the truth.

Theorem 20.4 (Erdős–Füredi 1983)**.** *For every $n \geq 1$ there is a set of at least $m = \lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^n \rfloor$ points in the n-dimensional Euclidean space $\mathbb{R}^n$, such that all angles determined by three points from the set are strictly less than $\pi/2$.*

The theorem is an easy consequence of the following lemma.

Lemma 20.5. *For every $n \geq 1$ there is a family $\mathcal{F}$ of $m = \lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^n \rfloor$ subsets of $\{1, \ldots, n\}$, such that there are no three distinct members A, B, C of $\mathcal{F}$ satisfying*

$$A \cap B \subseteq C \subseteq A \cup B. \tag{20.1}$$

Proof of Lemma 20.5. Let $\boldsymbol{A}$ be a random subset of $\{1, \ldots, n\}$, where each element appears randomly and independently with probability 1/2. Let $\mathcal{A}$ be a family of $2m$ independent copies of $\boldsymbol{A}$. For a triple $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ of sets in $\mathcal{A}$, what is the probability that they satisfy the condition (20.1)? This condition just means that for each $i = 1, \ldots, n$, neither $i \in \boldsymbol{A} \cap \boldsymbol{B}, i \notin \boldsymbol{C}$ nor $i \notin \boldsymbol{A} \cup \boldsymbol{B}, i \in \boldsymbol{C}$ hold. For each i, each of these two events happens with probability $(1/2)^3 = 1/8$. Therefore, the sets $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ satisfy the condition (20.1) with probability $(1-2/8)^n = (3/4)^n$. Since there are $3\binom{2m}{3}$ possible triples $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ (there are 3 possibilities to choose $\boldsymbol{C}$ in a triple), the expected number of triples that satisfy (20.1) is

$$3\binom{2m}{3}(3/4)^n = m(2m-1)(2m-2)(3/4)^n < m(2m)^2(3/4)^n \leq m,$$

where the last inequality follows from the choice of m.

Thus, there is a choice of a family $\mathcal{A}$ of $2m$ subsets of $\{1, \ldots, n\}$ in which the number of triples A, B, C satisfying (20.1) is at most m. By deleting one set from each such triple we obtain a family $\mathcal{F}$ of at least $2m - m = m$ subsets satisfying the assertion of the lemma. Notice that the members of $\mathcal{F}$ are all distinct since (20.1) is trivially satisfied if $A = C$. This completes the proof of the lemma. □

Proof of Theorem 20.4. We select the points of a set X in $\mathbb{R}^n$ from the points of the n-dimensional cube $\{0, 1\}^n$. We view the vertices of the cube, which are 0-1 vectors of length n, as the incidence vectors of subsets of an n-element set.

It is easy to verify that the three points a, b and c of the n-cube, corresponding to the sets A, B and C, respectively, determine a right angle at c if and only if (20.1) holds.

Indeed, the angle θ at c is the angle between the vectors $u = a - c$ and $v = b - c$. This angle can be computed from

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\| \cdot \|v\|},$$

where $\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$ is the scalar product and $\|u\| = \left(\sum_{i=1}^{n} u_i^2\right)^{1/2}$ is the norm of u. Since a, b and c are 0-1 vectors, the angle θ can be right if and only if $\langle u, v\rangle = 0$. This can happen if and only if $(a_i - c_i)(b_i - c_i) = 0$ for all $i = 1, \ldots, n$, which in its turn can happen if and only if for each i neither $a_i = b_i = 0, c_i = 1$ nor $a_i = b_i = 1, c_i = 0$ hold. This is precisely the condition (20.1), and the result follows immediately from Lemma 20.5. □

20.5 Affine cubes of integers

A collection C of integers is called an *affine d-cube* if there exist $d+1$ positive integers $x_0, x_1, \ldots, x_d$ so that

$$C = \left\{x_0 + \sum_{i \in I} x_i \;:\; I \subseteq \{1, 2, \ldots, d\}\right\}.$$

Such a cube is *replete* if all the sums are distinct, i.e., if $|C| = 2^d$. If an affine cube is generated by $x_0, x_1, \ldots, x_d$ then we write $C = C(x_0, x_1, \ldots, x_d)$. For example, $C(1, 1, 1) = \{1, 2, 3\}$ is not replete, while $C(1, 3, 9) = \{1, 4, 10, 13\}$ is a replete affine 2-cube. Note also that $C(x_0, x_1, \ldots, x_d)$ may be different from, say, $C(x_1, x_0, \ldots, x_d)$.

Typical extremal problems related to affine cubes are the following:

1. *Partition problem*: Given r and d, what is the smallest integer $n = H(r, d)$ such that, for any partition of $\{1, \ldots, n\}$ into r classes, at least one class contains an affine d-cube?
2. *Density problem*: Given a set of integers $A \subseteq \{1, 2, \ldots\}$, how large must A be to contain an affine d-cube?

Concerning the first (partition) problem, the existence of $H(r, d)$ for all r and d was first proved by Hilbert (1892). This result was then strengthened by van der Waerden (1927), whose celebrated theorem says that, for every r and d there is an integer $n = W(r, d)$ such that, for any partition of $\{1, \ldots, n\}$ into r classes, at least one class contains an arithmetic progression with d terms. We will prove this theorem in Chap. 26. Note that it implies Hilbert's result because any such progression $a, a + b, a + 2b, \ldots, a + (d - 1)b$ is also an affine $(d - 1)$-cube $C(a, b, b, \ldots, b)$.

At present not much is known about the rate of growth of van der Waerden's function $W(r, d)$ except that it is primitive recursive (Shelah 1988). For small values of r and d we know better bounds. Using the counting sieve we have proved that $W(2,d) > 2^{d/2}$ (see Theorem 25.3). Behrend (1949) proved a lower bound for the density of sets without arithmetic progressions of length three: there are subsets A of $[n]$ with $|A| \geq n \cdot 2^{-O(\sqrt{\log n})}$ and no length-three arithmetic progressions (see Theorem 25.6).

If we consider affine cubes instead of arithmetic progressions then the situation is better. In particular, the following bounds for Hilbert's function $H(r,d)$ are known. Brown et al. (1985) have shown that there exist constants $\epsilon > 0$ and $c > 0$ such that $r^{\epsilon d} \leq H(r,d) \leq r^{c^d}$, where $c \sim 2.6$ follows from Hilbert's original proof. Quite recently, these bounds were improved by Gunderson and Rödl (1998):

Theorem 20.6. *For any integers $d \geq 3$ and $r \geq 2$,*

$$r^{(1-\epsilon)(2^d-1)/d} \leq H(r,d) \leq (2r)^{2^{d-1}}, \tag{20.2}$$

where $\epsilon \to 0$ as $r \to \infty$.

The proof of the upper bound in (20.2) is based on the following density result, known as *Szemerédi's cube lemma* whose strengthened version, found by Graham (1981) and Graham, Rothschild and Spencer (1990), is as follows (see Sect. 25.2 for the proof).

Lemma 20.7. *Let $d \geq 2$ be given. Then, for every sufficiently large n, every subset A of $\{1,\ldots,n\}$ of size $|A| \geq (4n)^{1-1/2^{d-1}}$ contains an affine d-cube.*

The proof of the lower bound in (20.2) is based on the following lemma, whose proof gives one more illustration of the deletion method at work.

Lemma 20.8 (Gunderson–Rödl 1998)**.** *For each $d \geq 2$ and every set X of positive integers, there exists an $A \subseteq X$ with*

$$|A| \geq \tfrac{1}{8}|X|^{1-d/(2^d-1)},$$

which does not contain any replete affine d-cubes.

Proof. Fix $d \geq 3$, a set X, and let

$$p := |X|^{-d/(2^d-1)}.$$

Without loss of generality, we can assume that X is large enough so that $\frac{1}{8}p|X| > 2^d - 1$, because if not, then any set A of at most $2^d - 1$ elements would satisfy the lemma.

Let $\boldsymbol{Y}$ be a random subset of X whose elements are chosen independently with probability p.

Since any replete affine d-cube $C = C(x_0, x_1, \dots, x_d)$ is uniquely determined by $d+1$ distinct integers $x_0, x_0 + x_1, \dots, x_0 + x_d$ in X, the expected number of replete affine d-cubes in $\boldsymbol{Y}$ is bounded above by

$$\binom{|X|}{d+1} \cdot \Pr[C \subseteq \boldsymbol{Y}] = \binom{|X|}{d+1} \cdot p^{2^d}.$$

Therefore (by Markov's inequality), with probability at least 1/2, the number of replete affine d-cubes in $\boldsymbol{Y}$ does not exceed

$$2\binom{|X|}{d+1} p^{2^d} < \tfrac{1}{3}|X|p, \tag{20.3}$$

because

$$\begin{aligned} 2 \cdot \binom{|X|}{d+1} \cdot p^{2^d} &< 2\left(\frac{\mathrm{e}|X|}{d+1}\right)^{d+1} \cdot p^{2^d} \\ &= 2\left(\frac{\mathrm{e}}{d+1}\right)^{d+1} |X|^{d+1} \cdot |X|^{-d2^d/(2^d-1)} \\ &\le \tfrac{1}{3}|X|^{1-d/(2^d-1)} = \tfrac{1}{3}|X|p. \end{aligned}$$

On the other hand, the number $|\boldsymbol{Y}|$ of elements in a random subset $\boldsymbol{Y}$ of X is a binomially distributed random variable with expectation $|X|p$. It can be shown (see Exercise 20.2) that then

$$\Pr[|\boldsymbol{Y}| \le \tfrac{1}{2}|X|p] < 2\left(\frac{2}{e}\right)^{|X|p/2}. \tag{20.4}$$

Since $p|X| > 8(2^d - 1)$ and $d \ge 2$, this implies that

$$\Pr[|\boldsymbol{Y}| \ge \lfloor \tfrac{1}{2}|X|p \rfloor] > \frac{1}{2}. \tag{20.5}$$

Hence, there must exist an instance $Y \subseteq X$ of $\boldsymbol{Y}$ satisfying both above events; fix such Y. Due to (20.3) and (20.5), this set has at least $\lfloor \frac{1}{2}|X|p \rfloor$ elements and contains fewer that $\frac{1}{3}|X|p$ replete affine d-cubes. Deleting an element from each of these cubes, we get a set $A \subseteq Y$ with no replete affine d-cubes such that

$$|A| > |Y| - \tfrac{1}{3}|X|p \ge \lfloor \tfrac{1}{2}|X|p \rfloor - \tfrac{1}{3}|X|p > \tfrac{1}{3}|X|p,$$

where the last inequality follows from our assumption that $|X|p$ is larger than $8(2^d - 1) \ge 24$. □

Exercises

20.1. Prove the following analogon of Theorem 20.1 for bipartite graphs: If a bipartite $n \times n$ graph has minimum degree $n-d$, then its edges can be covered by at most $O(d \log n)$ complete bipartite subgraphs. *Hint*: Pick every vertex on the left side independently, with probability $p = 1/(d+1)$ to get a set S, and let T be the set of all vertices on the right side joined to *all* vertices in S. Show that $O(d \log n)$ such bipartite complete subgraphs $S \times T$ are enough to cover all edges.

20.2. Let $S_n = X_1 + \cdots + X_n$ where X_i are independent random variables with $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1-p$. Show that

$$\Pr[S_n \leq pn/2] < 2(2/e)^{pn/2}.$$

Hint: Recall that $\Pr[S_n \leq pn/2] = \sum_{k \leq pn/2} \binom{n}{k} p^k (1-p)^{n-k}$; show that each term in this sum is more than two times larger than the previous one, and hence, $\Pr[S_n = pn/2 - k] < 2^{-k} \Pr[S_n = pn/2]$, for $k < pn/2$. Sum up and use the estimate for $\binom{n}{pn}$ obtained in Exercise 1.16.

20.3. Show that every graph on n vertices has an independent set of size at least $n/(2d)$, where d is the average degree of its vertices. *Hint*: Combine Theorem 20.2 with Euler's theorem from Chap. 1.

20.4. (Gunderson–Rödl 1998). Prove that if a finite collection X of distinct positive integers contains an affine d-cube then either X contains a replete affine d-cube or an arithmetic progression of length 3 (or both). *Hint*: First, show that if $C = C(x_0, x_1, \ldots, x_d) \subseteq X$ is an affine d-cube, but is not replete (i.e., $|C| < 2^d$) then we can find two subsets $I, J \subset [n]$ such that $I \neq J$ but $\sum_{i \in I \setminus J} x_i = \sum_{j \in J \setminus I} x_j \neq 0$, and consider the triple of integers

$$x_0 + \sum_{i \in I \cap J} x_i,\ x_0 + \sum_{i \in I} x_i,\ x_0 + \sum_{i \in I \cup J} x_i.$$

20.5. (*Zarankiewicz's problem*; Erdős–Spencer 1974). Let $k_a(n)$ be the minimal k such that all $n \times n$ 0-1 matrices containing more than k ones contain an $a \times a$ submatrix consisting entirely of ones (the "all-ones" submatrix). Prove that for every constant $a \geq 2$ there is an $\epsilon > 0$ such that $k_a(n) \geq \epsilon n^{2-2/a}$. *Hint*: Argue as in the proof of Theorem 20.2. Take a random $n \times n$ 0-1 matrix $\boldsymbol{A}$, each entry of which takes value 1 independently and with probability $p = n^{-2/a}$. Associate with each $a \times a$ submatrix e of $\boldsymbol{A}$ the indicator random variable Y_e for the event "e is an all-ones submatrix." Switch one entry of each such submatrix to 0 and argue that in the resulting matrix we still can expect at least $n^2 p - \binom{n}{a}^2 p^{a^2}$ ones. (A more accurate choice of p leads to a somewhat better bound $k_a(n) \geq \epsilon n^{2-2/(a+1)}$.)

20.6. (Reiman 1958). Improve the above bound on $k_a(n)$ when $a = 2$: show that $k_2(n) \geq (1 - o(1)) n^{3/2}$. *Hint*: Let $n = p^2 + p + 1$ for a prime p, and consider the incidence matrix of lines in a projective plane of order p (see Sect. 12.4).

20.7. (Spencer 1990). Let $\mathcal{F}$ be an r-uniform family of subsets on an n-element set. Supose that, in average, each element belongs to d members of $\mathcal{F}$. Prove that there exists a set S of elements such that S is independent (i.e., contains no member of $\mathcal{F}$) and has size $|S| \geq (1 - 1/r) \cdot n \cdot d^{-1/(r-1)}$.
Hint: Argue as in the proof of Theorem 20.2 with $p = d^{-1/(r-1)}$.

21. The Second Moment Method

The pigeonhole property of expectation says that a random variable X cannot always be smaller (or always greater) than its expectation $\mathrm{E}[X]$. The second moment property tells us more: if the variance of X is much smaller than $\mathrm{E}[X]^2$ then X is almost always near to $\mathrm{E}[X]$, that is, the values of X are concentrated around its expectation.

21.1 The method

Let X be a random variable and $\mathrm{Var}[X]$ be its variance,

$$\mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 .$$

The *second moment method* uses the fact that, if $\mathrm{Var}[X] = o(\mathrm{E}[X]^2)$, then $X \sim \mathrm{E}[X]$, i.e., X is almost always almost equal to its expectation $\mathrm{E}[X]$. This follows from Chebyshev's inequality

$$\Pr[|X - \mathrm{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2} \tag{21.1}$$

which itself is an easy consequence of Markov's inequality (see Exercise 21.1). In particular, setting $t = \mathrm{E}[X]$ we arrive at

$$\Pr[X = 0] \leq \Pr[X \geq 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2} . \tag{21.2}$$

The following refined form of this inequality is often useful:

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X^2]} . \tag{21.3}$$

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_21,

To prove this inequality, let I_X be the indicator random variable for the event $X \neq 0$. Using the Cauchy–Schwarz inequality (see Proposition 13.4) we obtain

$$\mathrm{E}\left[X\right]^2 = \mathrm{E}\left[I_X \cdot X\right]^2 \leq \mathrm{E}\left[I_X\right] \mathrm{E}\left[X^2\right] = \Pr\left[X \neq 0\right] \cdot \mathrm{E}\left[X^2\right] ,$$

or equivalently,

$$\Pr\left[X = 0\right] = 1 - \Pr\left[X \neq 0\right] \leq 1 - \frac{\mathrm{E}\left[X\right]^2}{\mathrm{E}\left[X^2\right]} = \frac{\mathrm{Var}\left[X\right]}{\mathrm{E}\left[X^2\right]} .$$

If $X = X_1 + \cdots + X_n$ is a sum of random variables, then the variance can be computed by the formula

$$\mathrm{Var}\left[X\right] = \sum_{i,j=1}^{n} \mathrm{Cov}\left(X_i, X_j\right) = \sum_{i=1}^{n} \mathrm{Var}\left[X_i\right] + \sum_{i \neq j} \mathrm{Cov}\left(X_i, X_j\right) ,$$

where $\mathrm{Cov}\left(X_i, X_j\right)$ is the *covariance* and is defined as

$$\mathrm{Cov}\left(X_i, X_j\right) := \mathrm{E}\left[X_i X_j\right] - \mathrm{E}\left[X_i\right] \cdot \mathrm{E}\left[X_j\right] .$$

In general, if X_i and X_j are independent, then $\mathrm{E}\left[X_i X_j\right] = \mathrm{E}\left[X_i\right] \cdot \mathrm{E}\left[X_j\right]$, and hence, $\mathrm{Cov}\left(X_i, X_j\right) = 0$. This often considerably simplifies the variance calculations.

Let us mention that there are several forms of Chebyshev's inequality – the usual form (stated above), and the following, less standard form (see, for example, Hardy, Littlewood and Pólya (1952), Theorem 43):

Proposition 21.1. *Let $a_1, \ldots, a_n$ be a non-decreasing sequence and $b_1, \ldots, b_n$ be a non-increasing sequence of non-negative numbers. Then,*

$$\sum_{i=1}^{n} a_i b_i \leq \frac{1}{n} \Big(\sum_{i=1}^{n} a_i \Big) \Big(\sum_{i=1}^{n} b_i \Big).$$

21.2 Distinct sums

Let $f(n)$ denote the maximal m such that there exists a set $x_1, \ldots, x_m$ of m distinct numbers in $[n] = \{1, \ldots, n\}$ all of whose sums are distinct. An example of such a set consists of all numbers of the form 2^i with $i \leq \log_2 n$; this shows that $f(n) \geq 1 + \lfloor \log_2 n \rfloor$. We now use the second moment method to show that this lower bound is almost optimal: $f(n) \leq (1 + o(1)) \log_2 n$.

Theorem 21.2. *$f(n) \leq \log_2 n + \log_2 \log_2 n + O(1)$.*

Proof. Let $x_1, \dots, x_m$ be a subset of $m = f(n)$ integers in $[n]$ all of whose sums are distinct. Let $I_1, \dots, I_m$ be independent random variables, each taking values 0 and 1 with equal probability 1/2. Consider the random variable $X = I_1 x_1 + \cdots + I_m x_m$. Then

$$\mathrm{E}\,[X] = \frac{x_1 + \cdots + x_m}{2} \quad \text{and} \quad \mathrm{Var}\,[X] = \frac{x_1^2 + \cdots + x_m^2}{4} \leq \frac{n^2 m}{4}\,.$$

Setting $Y := X - \mathrm{E}\,[X]$ and using Chebyshev's inequality with $t := 2\sqrt{\mathrm{Var}\,[X]} \leq n\sqrt{m}$, after reversing the inequality we obtain

$$\Pr\,[|Y| \leq t] \geq 1 - \frac{1}{4} = 0.75\,.$$

On the other hand, due to the assumption that all sums of $x_1, \dots, x_m$ are distinct, the probability that X takes a particular value is either 0 or 2^{-m}. In particular, $\Pr\,[Y = s] \leq 2^{-m}$ for every integer s in the interval $[-t, t]$. Since there are only $2t + 1$ such integers, the union bound implies that

$$\Pr\,[|Y| \leq t] \leq 2^{-m}(2t + 1)\,.$$

Comparing the above inequalities and remembering that $t \leq n\sqrt{m}$ leads to $0.75 \cdot 2^m \leq 2t + 1 \leq 2n\sqrt{m} + 1$, it follows that $2^m/\sqrt{m} \leq Cn$ for a constant C, and the desired upper bound on $m = f(n)$ follows. □

21.3 Prime factors

Number theory has its foundation in the Fundamental Theorem of Arithmetic, which states that every integer $x > 1$ can be written uniquely in the form

$$x = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\,,$$

where the p_i's are primes and the k_i's are positive integers. Given x, we are interested in the number r of *prime factors* of x, that is, in the number of distinct primes p_i in such a representation of x. This number of primes dividing x is usually denoted by $\nu(x)$.

An important result in number theory, due to Hardy and Ramanujan (1917) states that almost every integer number between 1 and n has about $\ln\ln n$ prime factors. "Almost all" here means all but $o(n)$ numbers.

Theorem 21.3. *Let $\alpha = \alpha(n)$ be an arbitrarily slowly growing function. Then almost all integers x in $[n]$ satisfy $|\nu(x) - \ln\ln n| \leq \alpha\sqrt{\ln\ln n}$.*

Proof (due to Turán 1934). Throughout this proof, let p, q denote prime numbers. We need two well known results from number theory, namely,

$$\sum_{p \le x} \frac{1}{p} \le \ln\ln x + O(1)\,, \tag{21.4}$$

$$\pi(x) = (1+o(1))\frac{x}{\ln x}\,, \tag{21.5}$$

where $\pi(x)$ denotes the number of primes smaller than x.

We now choose x randomly from the set $\{1,\ldots,n\}$. For prime p, let X_p be the indicator random variable for the event that p divides x, and let $X = \sum_{p \le x} X_p$; hence, $X = \nu(x)$.

Since x can be chosen in n different ways, and in $\lfloor n/p \rfloor$ cases it will be divisible by p, we have that

$$\mathrm{E}\left[X_p\right] = \frac{\lfloor n/p \rfloor}{n} \le \frac{1}{p}\,,$$

and by (21.4) we also have

$$\mathrm{E}\left[X\right] \le \sum_{p \le x} \frac{1}{p} \le \ln\ln n + O(1)\,.$$

Now we bound the variance

$$\mathrm{Var}\left[X\right] = \sum_{p \le x} \mathrm{Var}\left[X_p\right] + \sum_{p \ne q \le n} \mathrm{Cov}\left(X_p X_q\right) \le \mathrm{E}\left[X\right] + \sum_{p \ne q \le n} \mathrm{Cov}\left(X_p X_q\right)\,,$$

since $\mathrm{Var}\left[X_p\right] \le \mathrm{E}\left[X_p\right]$. Observe that $X_p X_q = 1$ if and only if both p and q divide x, which further implies that pq divides x. In view of this we have

$$\begin{aligned}\mathrm{Cov}\left(X_p X_q\right) = \mathrm{E}\left[X_p X_q\right] - \mathrm{E}\left[X_p\right]\mathrm{E}\left[X_q\right] &= \frac{\lfloor n/(pq) \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \cdot \frac{\lfloor n/q \rfloor}{n} \\ &\le \frac{1}{pq} - \Big(\frac{1}{p} - \frac{1}{n}\Big)\Big(\frac{1}{q} - \frac{1}{n}\Big) \\ &\le \frac{1}{n}\Big(\frac{1}{p} + \frac{1}{q}\Big)\,.\end{aligned}$$

Then by (21.5)

$$\sum_{p \ne q \le n} \mathrm{Cov}\left(X_p X_q\right) \le \frac{2\pi(n)}{n} \sum_{p \le n} \frac{1}{p} = O\Big(\frac{\ln\ln n}{\ln n}\Big) \to 0\,.$$

Applying Chebyshev's inequality with $t = \alpha\sqrt{\ln\ln n}$ yields the desired result. □

21.4 Separators

In complexity theory we often face the following problem. We have a set of players, each of whom can see some small portion of the input bits, and we want to split the players into two groups so that for each group there is a large set of "forbidden" bits which are seen by *no* member of that group. To avoid trivial situations, we also assume that every bit is seen by at least one player. This question can be formalized as follows.

Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a family of subsets of some set X. By a *separator* for $\mathcal{F}$ we will mean a pair (S, T) of disjoint subsets of X such that each member of $\mathcal{F}$ is disjoint from either S or from T; the *size* of such a separator is the minimum of $|S|$ and $|T|$.

To approach the question raised at the beginning, interpret X as the set of bits and let F_i be the set of bits seen by the i-th player. The problem is, given $\mathcal{F}$, to make the sets of "forbidden bits" S and T as large as possible. Intuitively, if no bit is seen by too many players then these sets should be large. Using the averaging principle for partitions (a prototype of Markov's inequality), we have shown in Exercise 2.8 that this is true if no bit is seen by too many players. Using the second moment method, Beame, Saks, and Thathachar (1998) have shown that we may relax this condition, and only require that on *average*, no bit is seen by too many players.

The *degree* d_x of a point x in $\mathcal{F}$ is the number of members of $\mathcal{F}$ that contain x. The *average degree* of $\mathcal{F}$ is

$$d = \frac{1}{|X|} \sum_{x \in X} d_x \, .$$

Theorem 21.4 (Beame–Saks–Thathachar 1998)**.** *Let $\mathcal{F}$ be a family of non-empty sets of an n-element set, each containing at most r points. Let d be the average degree of $\mathcal{F}$. Then, $\mathcal{F}$ has a separator of size at least $(1-\delta)2^{-d}n$, where*

$$\delta = \sqrt{\frac{dr2^{d+1}}{n}}.$$

In particular, if $4rd2^{d+1} \le n$, then $\mathcal{F}$ contains a separator of size at least $n/2^{d+1}$.

Proof. Let $X = \bigcup_{F \in \mathcal{F}} F$ and $n = |X|$. Color each set $F \in \mathcal{F}$ red or blue uniformly and independently with probability 1/2. Define $\boldsymbol{S}$ (respectively, $\boldsymbol{T}$) to be the set of points x such that *every* set that contains x is colored red (respectively, blue). Since every element of X occurs in at least one set, it follows that $\boldsymbol{S}$ and $\boldsymbol{T}$ are disjoint. Moreover, for each $F \in \mathcal{F}$, either $F \cap \boldsymbol{S}$ or $F \cap \boldsymbol{T}$ is empty. To complete the proof, we show that with positive probability both $\boldsymbol{S}$ and $\boldsymbol{T}$ have at least $(1-\delta)2^{-d}n$ elements.

Let Z_x be the indicator random variable for the event "$x \in \boldsymbol{S}$." By the definition of $\boldsymbol{S}$, this event occurs with probability 2^{-d_x}, implying that

$$\mathrm{E}\left[Z_x\right] = \Pr\left[Z_x = 1\right] = 2^{-d_x}.$$

Let $Z = \sum_x Z_x$ and observe that $Z = |\boldsymbol{S}|$. Using the arithmetic-geometric mean inequality (1.16), we obtain

$$\mathrm{E}\left[Z\right] = \sum_x \mathrm{E}\left[Z_x\right] = \sum_x 2^{-d_x} \geq n \cdot 2^{-\sum_x d_x/n} = n2^{-d}. \tag{21.6}$$

Using the second moment argument, we show below that Z is close to its expected value with high probability. By Chebyshev's inequality (21.1) we need only to upper-bound the variance

$$\mathrm{Var}\left[Z\right] = \sum_x \mathrm{Var}\left[Z_x\right] + \sum_{x \neq y} \mathrm{Cov}\left(Z_x, Z_y\right), \tag{21.7}$$

where $\mathrm{Cov}\left(Z_x, Z_y\right) = \mathrm{E}\left[Z_x Z_y\right] - \mathrm{E}\left[Z_x\right]\mathrm{E}\left[Z_y\right]$ is the covariance of Z_x and Z_y. Consider the first term in the right-hand side of (21.7). For any x, Z_x is a Bernoulli random variable, so $\mathrm{Var}\left[Z_x\right] = \mathrm{E}\left[Z_x\right] - \mathrm{E}\left[Z_x\right]^2 \leq \mathrm{E}\left[Z_x\right]$, implying that

$$\sum_x \mathrm{Var}\left[Z_x\right] \leq \mathrm{E}\left[Z\right]. \tag{21.8}$$

To bound the second term in the right-hand side of (21.7), observe that if no member of $\mathcal{F}$ contains both x and y, then Z_x and Z_y are independent, implying that $\mathrm{Cov}\left(Z_x, Z_y\right) = 0$. Thus, we are only interested in those pairs (x, y) such that *some* member of $\mathcal{F}$ contains both x and y. For any fixed x, the number of such pairs (x, y) is at most $(r-1)d_x$. For each such pair,

$$\mathrm{Cov}\left(Z_x, Z_y\right) \leq \mathrm{E}\left[Z_x Z_y\right] \leq \mathrm{E}\left[Z_x\right] \leq 2^{-d_x}.$$

Therefore,

$$\sum_{x \neq y} \mathrm{Cov}\left(Z_x, Z_y\right) \leq (r-1)\sum_x d_x 2^{-d_x}.$$

The last term above can be bounded as follows. Order the x's so that the sequence $\{d_x\}$ is non-decreasing. Now the second Chebyshev inequality (Proposition 21.1) can be applied to the sequences $\{d_x\}$ and $\{2^{-d_x}\}$. We obtain

$$\sum_{x \neq y} \mathrm{Cov}\left(Z_x, Z_y\right) \leq \frac{r-1}{n}\left(\sum_x 2^{-d_x}\right)\left(\sum_x d_x\right) = d(r-1)\mathrm{E}\left[Z\right] \tag{21.9}$$

because $(\sum_x d_x)/n = d$, and $\sum_x 2^{-d_x} = \mathrm{E}\left[Z\right]$. Substitute the bounds (21.8) and (21.9) into (21.7). We obtain

$$\mathrm{Var}\left[Z\right] \leq (d(r-1)+1)\mathrm{E}\left[Z\right] \leq dr\mathrm{E}\left[Z\right],$$

where the last inequality holds because each $x \in X$ occurs in at least one set, implying that $d = \sum_x d_x/n \geq 1$. Using Chebyshev's inequality (21.1), we have

$$\Pr[Z < (1-\delta)\cdot \mathrm{E}[Z]] < \frac{\mathrm{Var}[Z]}{\delta^2 \cdot \mathrm{E}[Z]^2} \le \frac{dr}{\delta^2 \mathrm{E}[Z]}.$$

Substituting for δ its value as given by the statement of the theorem, and using the inequality (21.6), we obtain

$$\Pr[Z < (1-\delta)\cdot \mathrm{E}[Z]] < \frac{drn}{dr2^{d+1}\mathrm{E}[Z]} = \frac{n}{2^{d+1}\mathrm{E}[Z]} \le \frac{n}{2^{d+1}2^{-d}n} = \frac{1}{2}.$$

In a similar fashion, we obtain $\Pr[|\boldsymbol{T}| < (1-\delta)\cdot \mathrm{E}[Z]] < 1/2$. Thus, with positive probability, both $\boldsymbol{S}$ and $\boldsymbol{T}$ have size at least

$$(1-\delta)\cdot \mathrm{E}[Z] \ge (1-\delta)2^{-d}n.$$

We conclude that there is a coloring of the sets in $\mathcal{F}$ such that the induced S and T satisfy the theorem. □

21.5 Threshold for cliques

The second moment method is a useful tool for determining the *threshold function* of an event, i.e., a threshold such that below it, the probability of the event tends to 0, and above it, the probability tends to 1.

A *k-clique* is a complete graph on k vertices. We write $\omega(G) \ge k$ if the graph G contains a k-clique. A *random graph* $\boldsymbol{G}(n,p)$ is a graph on n vertices where each edge appears independently with probability p.

How large does p have to be before a random graph G is very likely to contain a 4-clique? The answer is remarkable: there is a sharply defined *threshold value* of p such that, if p is above this value then G is almost certain to contain a 4-clique, and if p is below it then G is almost certain *not* to contain such a clique.

Theorem 21.5. *The threshold for a random graph $\boldsymbol{G}(n,p)$ to contain a* 4*-clique is $p = n^{-2/3}$.*

Proof. For a subset of four vertices S, let A_S be the event that S induces a clique in $\boldsymbol{G}(n,p)$, and let X_S be the indicator variable of A_S. Clearly, $\Pr[X_S = 1] = p^6$ for every S (every 4-clique has six edges). We define $X = \sum X_S$. Our goal is to show that $\Pr[X \ge 1]$ tends to 0 if $p \ll n^{-2/3}$, and to 1 if $p \gg n^{-2/3}$.

The first claim follows from Markov's inequality:

$$\Pr[X \ge 1] \le \mathrm{E}[X] = \binom{n}{4}\cdot p^6 \sim \frac{n^4p^6}{24} \to 0, \quad \text{if } p \ll n^{-2/3}.$$

To prove the second claim, suppose $p \gg n^{-2/3}$. We must show that $\Pr[X = 0]$ tends to 0. By the second moment method,

$$\Pr[X = 0] \le \frac{\operatorname{Var}[X]}{\operatorname{E}[X]^2},$$

so all what we need is to estimate the variance. As we have already mentioned above, the variance can be written in the form

$$\operatorname{Var}[X] = \sum_S \operatorname{Var}[X_S] + \sum_{S \ne T} \operatorname{Cov}(X_S, X_T). \tag{21.10}$$

Since X_S is an indicator random variable, its variance $\operatorname{Var}[X_S] = \operatorname{E}[X_S] - \operatorname{E}[X_S]^2 \le \operatorname{E}[X_S] = p^6$ as six different edges must lie in $\boldsymbol{G}(n,p)$. Since there are $\binom{n}{4} = O(n^4)$ sets S, the total contribution of the first sum in (21.10) is $O(n^4p^6)$. Let us now estimate the contribution of pairs $S \ne T$.

If the events A_S and A_T are independent, then $\operatorname{Cov}(X_S, X_T) = 0$, and these pairs contribute nothing. Now, since $S \ne T$, the events A_S and A_T can be dependent if and only if the cliques S and T have common edges – that is, if and only if $|S \cap T| = 2$ or $|S \cap T| = 3$.

There are $O(n^6)$ pairs S, T with $|S \cap T| = 2$ and for each of these $\operatorname{Cov}(X_S, X_T) \le \operatorname{E}[X_S X_T] = p^{11}$ as $S \cup T$ has 11 different edges. So, the total contribution of these pairs is $O(n^6p^{11})$.

Similarly, there are $O(n^5)$ pairs S, T with $|S \cap T| = 3$ and for each of these $\operatorname{E}[X_S X_T] = p^9$, since in this case $S \cup T$ has 9 different edges; and the total contribution of these pairs is $O(n^5p^9)$.

Putting all this together in (21.10) gives

$$\operatorname{Var}[X] = O\left(n^4p^6 + n^6p^{11} + n^5p^9\right) = o(n^8p^{12}) = o(\operatorname{E}[X]^2),$$

since $p \gg n^{-2/3}$. Therefore, $\Pr[X = 0] = o(1)$ and hence $\boldsymbol{G}(n,p)$ is almost certain to contain a 4-clique. □

Let us consider a more general question: given a graph G, what is the threshold function for the property that a random graph $\boldsymbol{G}(n,p)$ contains a copy of G as an induced subgraph? (Recall that induced subgraphs are obtained by deleting vertices together with all the edges incident to them.) In the previous section we have solved this question for the case when G is a complete graph on 4 vertices. What about other graphs? Using the second moment method, a surprisingly general answer to this question was found.

The *density* of a graph $G = (V, E)$ is the fraction $d(G) := |E|/|V|$. The *subgraph density* $m(G)$ of a graph is the maximum density $d(H)$ of its subgraph $H = (V', E')$. A graph G is *balanced* if $d(H) \le d(G)$ for all subgraphs H of G.

It turns out that $n^{-1/m(G)}$ is the right threshold for an arbitrary (!) balanced graph G. This fundamental result was proved by Erdős and Rényi

(1960). Bollobás (1981) and Ruciński and Vince (1986) extended it by removing this restriction on G (of being balanced).

Theorem 21.6 (Erdős–Rényi 1960)**.** *Let G be a balanced graph on k vertices. The threshold for the property that a random graph $\boldsymbol{G}(n,p)$ contains a subgraph isomorphic to G is $n^{-1/m(G)}$.*

This theorem can be proved in a similar way to Theorem 21.5. But the computations are more involved, and we omit the proof – the interested reader can find it in the book of Alon and Spencer (1992) or in the survey paper of Karoński (1995).

Exercises

21.1 (Chebyshev's Inequality)**.** If X is a non-negative random variable with mean $\mu = \mathrm{E}[X]$ and a a real number, then Markov's inequality gives an upper bound $\Pr[X \geq a] \leq \mu/a$ (see Theorem 18.21). Use this inequality to prove the following *Chebyshev's Inequality.* Let X be a random variable with mean $\mu = \mathrm{E}[X]$ and variance $\sigma^2 = \mathrm{Var}[X]$. Then for any real $a > 0$,

$$\Pr[|X - \mu| \geq a\sigma] \leq \frac{1}{a^2}\,.$$

Hint: Apply Markov's inequality to a random variable $Y := (X - \mu)^2$.

21.2. Show that Chebyshev's inequality cannot be improved. *Hint*: Consider a random variable X taking its values in $\{-1, 0, +1\}$ with probabilities $\Pr[X = -1] = 1/2a^2$, $\Pr[X = 0] = 1 - 1/a^2$ and $\Pr[X = +1] = 1/2a^2$.

21.3 (Cantelli's Inequality)**.** Prove the following one-sided Chebyshev's inequality:

$$\Pr[X \geq \mu + a\sigma] \leq \frac{1}{1 + a^2}\,.$$

Hint: Let $t \geq -\mu$ be a parameter, consider the random variable $Y := (X + t)^2$, compute its expectation, apply Markov's inequality and show that the probability $\Pr[X \geq \mu + a\sigma]$ is minimized when $t = \sigma^2/a - \mu$.

21.4. Let $A_1, \ldots, A_n$ be arbitrary events. Define $a = \sum_{i=1}^{n} \Pr[A_i]$ and $b = \sum_{i<j} \Pr[A_i \cap A_j]$. Prove that

$$\Pr[\overline{A}_1 \cdots \overline{A}_n] \leq \frac{a + 2b}{a^2} - 1$$

and

$$\Pr[A_1 \cup \cdots \cup A_n] \geq \frac{a^2}{a + b}\,.$$

Hint: Let X be the number of A_i's that occur. For the first inequality, use Chebyshev's inequality to show that $\Pr[X = 0] \le a^{-2}\mathrm{E}[(X-a)^2]$, and use the linearity of expectation to expand the righthand expression. For the second inequality, rewrite (21.3) as $\Pr[X > 0] \ge \mathrm{E}[X]^2 / \mathrm{E}[X]^2$.

21.5. Let $k > 0$ be an integer, and let $p = p(n)$ be a function of n such that $p \ge (6k \ln n)/n$ for large n. Prove that "almost surely" the random graph $\boldsymbol{G} = \boldsymbol{G}(n,p)$ has no large independent set of vertices. Namely, show that $\Pr\left[\alpha(\boldsymbol{G}) \ge \frac{n}{2k}\right] \to 0$ as $n \to \infty$.

21.6. Let $\epsilon > 0$ and $p = p(n) > 0$, and let $r \ge (1+\epsilon)(2\ln n)/p$ be an integer-valued function of n. Show that "almost surely" the random graph $\boldsymbol{G}(n,p)$ does *not* contain r independent vertices.

21.7. A *forest* is a graph without non-trivial cycles (i.e., cycles of length ≥ 3). Prove that, if $np \to 0$ as $n \to \infty$, then $\Pr[\boldsymbol{G}(n,p)$ is a forest$] \to 1$. *Hint*: Count the expected number of cycles and apply Markov's inequality.

21.8. Let p be constant. Show that with a probability approaching 1 the graph $\boldsymbol{G}(n,p)$ has the property that every pair of its vertices has a common neighbor, i.e., a vertex, adjacent to both of them. *Hint*: Consider indicator random variables X_{ij} for the event that i and j do not have a common neighbor. Argue that $\mathrm{E}[X_{ij}] = (1-p^2)^{n-2}$ and apply Markov's inequality.

21.9. Prove that $p = \ln n/n$ is a threshold probability for the disappearance of isolated vertices. *Hint*: Consider the random variable $X = X_1 + \ldots + X_n$ where X_i indicates whether vertex i is isolated in $\boldsymbol{G}(n,p)$. When estimating the variance of X, observe that $X_iX_j = 1$ iff both vertices are isolated. That requires forbidding $2(n-2)+1$ edges, so $\mathrm{E}[X_iX_j] = (1-p)^{2n-3}$.

22. The Entropy Function

Although the concept of entropy was originally a thermodynamic construct, it has been adapted to other fields, including information theory. In this theory, entropy is a measure of the uncertainty associated with a random variable. It measures the average information content one is missing when one does not know the value of the random variable. The concept was introduced by Claude E. Shannon in his 1948 paper "A Mathematical Theory of Communication."

A fair coin has an entropy of one bit. However, if the coin is not fair, then the uncertainty is lower (if asked to bet on the next outcome, we would bet preferentially on the most frequent result), and thus the Shannon entropy is lower.

In this chapter we will consider some applications of this concept—the Shannon entropy—in combinatorics.

22.1 Quantifying information

The *information size* $h_0(A)$ of a finite set A is the number of bits that is necessary to encode each element of A separately, that is,

$$h_0(A) = \log_2 |A| .$$

If we have two sets A and B, then $h_0(A \times B) = h_0(A) + h_0(B)$. This justifies the logarithm. The next notion, the *information gain*, is also intuitive.

Suppose we need to uncover a certain English word. We manage to obtain one letter, say, an "e." But this letter is common in English, so this provides little information. If, on the other hand, the letter that we discover is "z" (the least common in English), the search has been narrowed and we obtain more information. The notion of "information gain" quantifies this situation.

Let (A, p) be a discrete probability space. That is, A is a finite set, and each element $a \in A$ has probability $p_a = \Pr[a]$. As before, the probability of

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_22,

an event $B \subseteq A$ is $p(B) = \sum_{a \in B} p_a$. The *information gain*

$$H(B|A) := \log_2 \frac{1}{p(B)} = -\log_2 p(B)$$

measures the gain obtained by the knowledge that the outcome belongs to the set B. The information gain has the following additivity property. Let $B \subset C \subset A$. The gain for knowing that the outcome is in C is $H(C|A) = -\log_2 p(C)$. The gain for knowing that it is in B, after knowing that it is in C, is

$$H(B|C) = -\log_2 \Pr[B \mid C] = -\log_2 \frac{p(B)}{p(C)}\,.$$

It follows that $H(B|A) = H(C|A) + H(B|C)$, as it should be.

22.2 Limits to data compression

Let $A = \{a_1, \ldots, a_n\}$ be a finite set of letters (or symbols). A (binary) *code* is an assignment $a_i \mapsto c(a_i)$ of binary strings (codewords) $c(a_i)$ to symbols. A code is a *prefix code* if no codeword occurs as the beginning of another codeword. Such codes are uniquely decodable codes where a sequence of binary numbers can be decoded sequentially. That is, one reads the binary numbers from the left, one by one, until one recognizes the code of a letter. One extracts the letter, and reads the next binary numbers until the next letter is identified.

The folowing lemma tells us that we cannot have short codewords for all letters: if some letters have short codewords, there must be letters with long ones.

Lemma 22.1 (Kraft Inequality). *For every prefix code with codeword lengths $\ell_1 \leq \ell_2 \leq \ldots \leq \ell_n$, we have that*

$$\sum_{i=1}^{n} 2^{-\ell_i} \leq 1\,. \tag{22.1}$$

Conversely, for a given set of natural numbers $\ell_1 \leq \ell_2 \leq \ldots \leq \ell_n$, satisfying the above inequality, there exists a prefix code with those codeword lengths.

By using Jensen's inequality $f(\sum_i \lambda_i x_i) \leq \sum_i \lambda_i f(x_i)$ with $\lambda_i = 1/n$ and $f(x) = 2^{-x}$, (22.1) implies that $\sum_{i=1}^{n} \ell_i \geq n \log_2 n$. That is, the average length of a codeword is always at least $\log_2 n$.

Proof. Any given prefix code can be represented by a binary tree of depth ℓ_n where the i-th codeword is represented by a path to a leaf at depth ℓ_i. This guarantees that no codeword is a prefix of another. For each leaf in such a code tree, consider the set of descendants $A_i \subseteq \{0,1\}^{\ell_n}$ that each would have at depth ℓ_n in a full tree. Then $A_i \cap A_j = \emptyset$ for all $i \neq j$, and $|A_i| = 2^{\ell_n - \ell_i}$. Thus, given that the total number of nodes at depth ℓ_n is 2^{ℓ_n},

$$2^{\ell_n} \sum_{i=1}^{n} 2^{-\ell_i} = \sum_{i=1}^{n} |A_i| = \left| \bigcup_{i=1}^{n} A_i \right| \leq 2^{\ell_n}$$

from which (22.1) follows.

Conversely, given any ordered sequence of n natural numbers, $\ell_1 \leq \cdots \leq \ell_n$ satisfying the Kraft inequality, one can construct a prefix code with codeword lengths equal to ℓ_i by pruning subtrees from a full binary tree of depth ℓ_n. First choose any node v_1 from the full tree at depth ℓ_1 and remove all of its descendents. This removes a $2^{-\ell_1}$ fraction of the nodes from the full tree from being considered for the rest of the remaining codewords. In the next iteration choose any node v_2 from the remaining tree at depth ℓ_2, not on the path to v_1, and remove all of its descendents. This removes a $2^{-\ell_2}$ fraction of the full tree for a total of $2^{-\ell_1} + 2^{-\ell_2}$. After m iterations, a $\sum_{i=1}^{m} 2^{-\ell_i}$ fraction of the full tree nodes are removed from consideration for any remaining codewords. But, by the assumption, this sum is less than 1 for all $m < n$, thus a prefix code with lengths ℓ_i can be constructed for all n letters in our alphabet. □

The following inequality turns out to be very useful when dealing with sums involving logarithms.

Lemma 22.2 (Gibbs' Inequality)**.** *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be positive real numbers such that $\sum_i y_i \leq \sum_i x_i$. Then*

$$\sum_{i=1}^{n} x_i \log_2 x_i \geq \sum_{i=1}^{n} x_i \log_2 y_i \,.$$

Proof. Multiplying both sides by $\ln 2$ we may assume that all logarithms are natural. Using the inequality $\ln x \leq x - 1$ (see 1.4), we get

$$\begin{aligned}\sum x_i \ln y_i - \sum x_i \ln x_i = \sum x_i \ln\left(\frac{y_i}{x_i}\right) &\leq \sum x_i \left(\frac{y_i}{x_i} - 1\right) \\ &= \sum y_i - \sum x_i \leq 0. \quad \square\end{aligned}$$

An important special case of *Shannon's noiseless coding theorem* gives a limit for data compression for *any* probability distribution $p_1, \ldots, p_n$ on $A = \{a_1, \ldots, a_n\}$.

Given a prefix code c, which encodes each letter a_i by a string $c(a_i) \in \{0,1\}^{\ell_i}$, the average length of a codeword is

$$L(A, c) := \sum_{i=1}^{n} p_i \ell_i \,.$$

If $p_1 = \ldots = p_n = 1/n$ then the Kraft inequality implies that $L(A,c) \geq \log_2 n$. Shannon's theorem extends this to arbitrary distributions $p_1, \ldots, p_n$ on $A = \{a_1, \ldots, a_n\}$. To do this, define the *entropy* of the probability distribution on A by:

$$\mathrm{H}(A) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}\,. \tag{22.2}$$

That is, $\mathrm{H}(A)$ is just the expectation of the random variable $a_i \mapsto \log_2(1/p_i)$. Note that $\mathrm{H}(A) = \log_2 n$ if all $p_i = 1/n$.

Theorem 22.3 (Limits to compression)**.** *For any alphabet A, and any probability distribution p on A, the average length $L(A,c)$ of an optimal prefix code c satisfies*

$$\mathrm{H}(A) \leq L(A,c) \leq \mathrm{H}(A) + 1\,.$$

Thus, the expected length is minimized and is equal to $\mathrm{H}(A)$ only if the code-lengths are equal to the Shannon information contents: $\ell_i = \log_2(1/p_i)$. That is, *frequent letters should be coded with small length.* This clearly comes at the expense of other letters, that will need longer strings in order for the code to be decodable.

Proof. For the lower bound, consider a prefix code c with lengths $\ell_i = |c(a_i)|$. Define $z := \sum_{j=1}^{n} 2^{-\ell_j}$ and $q_i := 2^{-\ell_i}/z$. By the Kraft inequality, we have that $z \leq 1$, and hence, $\log_2 z \leq 0$. Together with the Gibbs inequality, we obtain

$$L(A,c) = \sum_{i=1}^{n} p_i \ell_i = -\sum_{i=1}^{n} p_i \log_2 q_i - \log_2 z \geq -\sum_{i=1}^{n} p_i \log_2 p_i = \mathrm{H}(A)\ .$$

For the upper bound, define $\ell_i = \lceil -\log_2 p_i \rceil$. Then

$$\sum_{i=1}^{n} 2^{-\ell_i} \leq \sum_{i=1}^{n} p_i = 1\,.$$

This shows that the Kraft inequality holds for the lengths ℓ_i, so there exists a prefix code c with these lengths. The expected length is easy to estimate:

$$L(A,c) = \sum_{i=1}^{n} p_i \lceil -\log_2 p_i \rceil \leq \sum_{i=1}^{n} p_i(-\log_2 p_i + 1) = \mathrm{H}(A) + 1\,. \qquad \square$$

A basic problem in information theory deals with encoding large quantities of information. We start with a finite set A, which could for instance be the 26 letters from the Latin alphabet, or the 128 ASCII symbols. We consider a message (a string) that contains n symbols from A with n large. How many bits are required so that the message can be encoded without loss of information? The answer is given by the information size: $h_0(A^n) = n \cdot h_0(A)$.

The question becomes more interesting, and the answer more surprising, if we allow an error probability $\delta > 0$. We now seek to encode only messages that only use symbols in a subset $B \subseteq A$ of "typical" characters, such that $p(B) = \Pr[B] \geq 1 - \delta$. If a message turns out to contain symbols in $A \setminus B$, then we lose the information. The information size is given by

$$h_\delta(A) = \inf_{p(B) \geq 1-\delta} \log_2 |B| \,.$$

Notice that $h_\delta(A) \to h_0(A)$ as $\delta \to 0$. We want to say something about $h_\delta(A^n)$ as $n \to \infty$. That is, a string of n characters is selected at random using some probability. One wants to encode this string, to send (or to store) it, and to decode it. We assume that no error is committed during these operations, except that the string may lie outside the set B^n of codable strings. This modeling is behind all compression algorithms.

So, given an alphabet A with a probability distribution $a \mapsto p(a)$, we want to say something about $h_\delta(A^n)$ as $n \to \infty$. Notice that the probability of each string $a = (a_1, \ldots, a_n)$ in A^n is $p(a) = \prod_{i=1}^{n} p(a_i)$, that is, we assume independence. Both $h_\delta(A)$ and

$$\mathrm{H}(A) = \sum_{a \in A} p(a) \log_2 \frac{1}{p(a)}$$

are at most the information size $h_0(A) = \log_2 |A|$ of the alphabet A. And we also have that $h_0(A^n) = n \cdot h_0(A)$, but $h_\delta(A^n)$ can be smaller than $n \cdot h_\delta(A)$ in general. It turns out that $h_\delta(A^n) \approx n \cdot \mathrm{H}(A)$ for any $\delta > 0$. That is, if we allow a tiny error, and if our message is long enough, the number of required bits is roughly $n \cdot \mathrm{H}(A)$.

Theorem 22.4 (Shannon's source coding theorem). *For any $\delta > 0$,*

$$\lim_{n \to \infty} \frac{1}{n} h_\delta(A^n) = \mathrm{H}(A) \,.$$

Proof. This is a consequence of the *weak law of large numbers*: If X is a random variable with finite expectation $\mathrm{E}[X]$, and if $X_1, X_2, \ldots$ are independent copies of X, then

$$\frac{X_1 + \cdots + X_n}{n}$$

tends to $\mathrm{E}[X]$ with probability tending to 1, as $n \to \infty$. If we define a random variable $X : A \to \mathbb{R}$ by $X(a) := -\log_2 p(a)$, then

$$\mathrm{E}[X] = \sum_{a \in A} -p(a) \log_2 p(a) = \mathrm{H}(A) \,.$$

On the other hand, for a random string $(a_1, \ldots, a_n)$ in A^n we have that

$$-\log_2 p(a_1, \ldots, a_n) = -\log_2 \prod_{i=1}^{n} p(a_i) = \sum_{i=1}^{n} -\log_2 p(a_i) = \sum_{i=1}^{n} X_i \,.$$

Thus, for an arbitrary small constant $\epsilon > 0$, there exists a set $A_{n,\epsilon} \subseteq A^n$ such that $\lim_{n\to\infty} p(A_{n,\epsilon}) = 1$, and any $(a_1, \ldots, a_n) \in A_{n,\epsilon}$ satisfies

$$2^{-n(\mathrm{H}(A)+\epsilon)} \leq p(a_1, \ldots, a_n) \leq 2^{-n(\mathrm{H}(A)-\epsilon)} \,.$$

Since $1 \geq p(A_{n,\epsilon}) \geq |A_{n,\epsilon}| 2^{-n(\mathrm{H}(A)+\epsilon)}$, we have that $|A_{n,\epsilon}| \leq 2^{n(\mathrm{H}(A)+\epsilon)}$. For any $\delta > 0$, we can choose n large enough so that $p(A_{n,\epsilon}) > 1 - \delta$. Then

$$h_\delta\left(A^n\right) \leq \log_2 |A_{n,\epsilon}| \leq n(\mathrm{H}\left(A\right) + \epsilon) \,.$$

It follows that

$$\limsup_{n\to\infty} \frac{1}{n} h_\delta\left(A^n\right) = \mathrm{H}\left(A\right) \,.$$

For the lower bound, let $B_{n,\delta} \subseteq A^n$ be the minimizer for H_δ, that is, $p(B_{n,\delta}) \geq 1 - \delta$ and

$$h_\delta\left(A^n\right) = \log_2 |B_{n,\delta}| \geq \log_2 |B_{n,\delta} \cap A_{n,\epsilon}| \,.$$

Since $\lim_{n\to\infty} p(A_{n,\epsilon}) = 1$, we have that $p(\overline{A_{n,\epsilon}}) \leq \delta$ for n large enough. So,

$$p(B_{n,\delta} \cap A_{n,\epsilon}) = p(B_{n,\delta}) - p(B_{n,\delta} \cap \overline{A_{n,\epsilon}}) \geq 1 - 2\delta \,,$$

implying that

$$|B_{n,\delta} \cap A_{n,\epsilon}| \geq (1 - 2\delta) 2^{n(\mathrm{H}(A)-\epsilon)} \,.$$

We obtain

$$\frac{1}{n} h_\delta\left(A^n\right) \geq \frac{1}{n} \log_2(1 - 2\delta) + \mathrm{H}\left(A\right) - \epsilon \,.$$

This gives the desired lower bound for the lim sup, and Shannon's theorem follows. □

22.3 Shannon entropy

We now turn to a general notion of Shannon entropy and its properties.

Let X be a random variable taking values in some range B, and let p_b denote the probability that the value of X is b. The *binary entropy* of X, denoted by $\mathrm{H}\left(X\right)$, is just the expected information gain of X:

$$\mathrm{H}\left(X\right) := \sum_{b\in B} = p_b \log_2 \frac{1}{p_b} = -\sum_{b\in B} p_b \log_2 p_b$$

where $0 \log_2 0$ is interpreted as 0. For example, if X takes only two variables 0 and 1 with $\Pr\left[X = 1\right] = p$, then $\mathrm{H}\left(X\right) = -p \log_2 p - (1-p) \log_2(1-p)$. Since

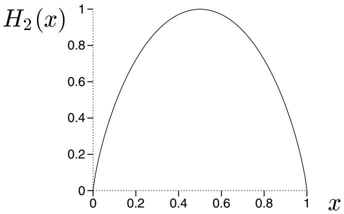


Fig. 22.1 The binary entropy function $H_2(x) = -x \log_2 x - (1-x)\log_2(1-x)$.

this quantity only depends on the probability distribution p, it is also often written as $H(p)$ or $H_2(p)$, and this is the binary entropy function, which we already introduced earlier.

Given an alphabet $A = \{a_1, \ldots, a_n\}$ together with a probability distribution $p_1, \ldots, p_n$, we can consider A as a random variable with $\Pr[A = i] = p_i$. Then $\mathrm{H}(A)$ is precisely the quantity (22.2) we considered above.

In general, entropy has the following basic properties:

(a) If $|B| = 2^t$ then $\mathrm{H}(X) \leq t$. Convexity arguments show that if the p_b's are smaller then the entropy must be larger. The extreme case is $p_b = 2^{-t}$ for all $b \in B$, then $\mathrm{H}(X) = t$. Moreover,

$$\sum_{b \in S} p_b \log_2 \frac{1}{p_b} \leq \log_2 |S| \tag{22.3}$$

for any subset $S \subseteq B$ (see Exercise 22.1).

(b) Entropy has the *concentration property.* If $\mathrm{H}(X) \leq t$ then there must be some b for which

$$\Pr[X = b] \geq 2^{-t}.$$

(c) Entropy is *subadditive*: if $X = (X_1, \ldots, X_n)$ is a random variable taking values in the set $B = B_1 \times \cdots \times B_n$, where each X_i is a random variable taking values in B_i, then

$$\mathrm{H}(X) \leq \sum_{i=1}^{n} \mathrm{H}(X_i) \; .$$

The first two properties follow directly from the definition. The last needs a proof, and we give it in Sect. 22.4.

If E is some event, it is natural to define the *conditional entropy* of X given E by

$$\mathrm{H}(X|E) := \sum_{b \in B} -\Pr[X = b \mid E] \cdot \log_2 \Pr[X = b \mid E] \; .$$

In the same way, if Y is any other random variable taking values in some range A, we define the conditional entropy of X given Y by

$$\mathrm{H}(X|Y) := \sum_{a\in A} \mathrm{H}(X|Y=a)\cdot \Pr[Y=a].$$

We think of $\mathrm{H}(X|Y)$ as the uncertainty of X given a particular value of Y, averaged over the range of values that Y can take. Fairly direct consequences of the definitions are:

(d) $\mathrm{H}(X|X)=0$;
(e) $\mathrm{H}(X|Y)=0$ if and only if $X=f(Y)$ for some function f;
(f) $\mathrm{H}(X|Y)=\mathrm{H}(X)$ if and only if X and Y are independent;
(g) $\mathrm{H}(X|Y,Z)\le \mathrm{H}(X|Y)$.

The main property of conditional entropy is the following:

$$\mathrm{H}(X,Y)=\mathrm{H}(Y)+\mathrm{H}(X|Y). \tag{22.4}$$

This equation also follows (though not so immediately) from definitions, and we leave the proof as an exercise. Using this equality we can derive the following analogue of the equality $\Pr[A\cap B]=\Pr[A]+\Pr[B]-\Pr[A\cup B]$:

$$\mathrm{H}(X,Y,Z)\le \mathrm{H}(X,Y)+\mathrm{H}(Y,Z)-\mathrm{H}(Y)\ . \tag{22.5}$$

Indeed

$$\begin{aligned}\mathrm{H}(X,Y,Z) &= \mathrm{H}(X,Y)+\mathrm{H}(Z|X,Y)\le \mathrm{H}(X,Y)+\mathrm{H}(Z|Y)\\ &= \mathrm{H}(X,Y)+\mathrm{H}(Y,Z)-\mathrm{H}(Y)\ .\end{aligned}$$

22.4 Subadditivity

Just like expectation has the additivity property for the *sums* $X=X_1+\cdots+X_n$, entropy has similar (though weaker) property for *strings* $X=(X_1,\ldots,X_n)$ of random variables.

Theorem 22.5. *If X and Y are two random variables taking only finitely many values, then $\mathrm{H}(X,Y)\le \mathrm{H}(X)+\mathrm{H}(Y)$, with equality only holding when X and Y are independent.*

Proof. Suppose X and Y take their values in A and B, respectively. Let $p_{a,b}$ denote the probability that $(X,Y)=(a,b)$, p_a denote the probability that $X=a$ and p_b denote the probability that $Y=b$. Since $\sum_{b\in B} p_{a,b}=p_a$ and $\sum_{a\in A} p_{a,b}=p_b$, we have

$$\begin{aligned} \mathrm{H}(X) + \mathrm{H}(Y) &= \sum_{a\in A} -p_a \log p_a + \sum_{b\in B} -p_b \log p_b \\ &= \sum_{a\in A}\sum_{b\in B} -p_{a,b} \log p_a + \sum_{b\in B}\sum_{a\in A} -p_{a,b} \log p_b \\ &= \sum_{(a,b)\in A\times B} -p_{a,b} \log(p_a \cdot p_b) \,. \end{aligned}$$

Since $\sum_{a,b} p_a \cdot p_b = \sum_a p_a \left(\sum_b p_b\right) = 1$, we can apply Lemma 22.2 to get

$$\mathrm{H}(X) + \mathrm{H}(Y) \geq \sum_{(a,b)\in A\times B} -p_{a,b} \log p_{a,b} = \mathrm{H}(X,Y) \ .$$

Equality holds only when $p_{a,b} = p_a p_b$ for all $a \in A$ and $b \in B$. But this is exactly the condition that X and Y are independent. □

Theorem 22.5 generalizes readily to more than two variables (we leave the proof as an exercise).

Theorem 22.6. *Let $X = (X_1, \ldots, X_n)$ be a random variable taking values in the set $B = B_1 \times \cdots \times B_n$, where each of the coordinates X_i of X is a random variable taking values in B_i. Then $\mathrm{H}(X) \leq \sum_{i=1}^n \mathrm{H}(X_i)$ with equality only holding when $X_1, \ldots, X_n$ are mutually independent.*

An interesting extension was proved by Chung, Frankl, Graham, and Shearer (1986). As in Theorem 22.6, let $X = (X_1, \ldots, X_n)$ be a random variable taking values in the set $B = B_1 \times \cdots \times B_n$, where each X_i is a random variable taking values in B_i. Also assume that all $B_i = \{0, 1\}$. For a subset of coordinates S, let X_S denote the random variable $(X_i)_{i\in S}$.

Theorem 22.7 (Generalized Subadditivity)**.** *Let $X = (X_1, \ldots, X_n)$ and B be as above and let $S_1, \ldots, S_m$ be subsets of $[n] = \{1, \ldots, n\}$ such that every $i \in [n]$ belongs to at least k of $S_1, \ldots, S_m$. Then*

$$\mathrm{H}(X) \leq \frac{1}{k} \sum_{i=1}^m \mathrm{H}(X_{S_i}) \ . \tag{22.6}$$

Proof. For $k = 1$ the assertion follows from Theorem 22.6. Now assume $k > 1$. Let ν denote the minimum number of S_i's whose union is $[n]$. We will prove (22.6) by induction on k and ν. If $\nu = 1$ then, say $S_1 = [n]$, and every point of $[n]$ belongs to at least $k-1$ of the sets $S_2, \ldots, S_m$. By induction (on k) we have in this case that

$$(k-1)\mathrm{H}(X) \leq \sum_{i=2}^m \mathrm{H}(X_{S_i})$$

and consequently, $k \cdot \mathrm{H}(X) \leq \sum_{i=1}^m \mathrm{H}(X_{S_i})$ since $X = X_{S_1}$.

Suppose $\nu > 1$. We may assume w.l.o.g. that $S_1 \cup S_2 \cup \cdots \cup S_\nu = [n]$. Let $S_1' := S_1 \cup S_2$ and $S_2' := S_1 \cap S_2$. Clearly, every element of $[n]$ is in at least k of the sets $S_1', S_2', S_3, \ldots, S_m$. Moreover, already $\nu - 1$ of these sets cover $[n]$ because $[n] = S_1' \cup S_3 \cup \cdots \cup S_\nu$. By induction (on ν),

$$k \cdot \mathrm{H}(X) \leq \sum_{i=3}^{m} \mathrm{H}(X_{S_i}) + \mathrm{H}\left(X_{S_1'}\right) + \mathrm{H}\left(X_{S_2'}\right) .$$

Since by (22.5) we have (we address this conclusion in the exercises) that

$$\mathrm{H}(X_{S_1 \cup S_2}) \leq \mathrm{H}(X_{S_1}) + \mathrm{H}(X_{S_2}) - \mathrm{H}(X_{S_1 \cap S_2}) , \tag{22.7}$$

the desired inequality (22.6) follows. □

22.5 Combinatorial applications

Theorem 22.6 was used by Kleitman, Shearer, and Sturtevant (1981) to derive several interesting applications in extremal set theory. Their basic idea can be illustrated by the following simple corollary of Theorem 22.6.

Corollary 22.8. *Let $\mathcal{F}$ be a family of subsets of $\{1, 2, \ldots, n\}$ and let p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then*

$$\log_2 |\mathcal{F}| \leq \sum_{i=1}^{n} H(p_i) ,$$

where $H(y) := -y \log_2 y - (1-y)\log_2(1-y)$.

Proof. Associate each set $F \in \mathcal{F}$ with its incidence vector v_F, which is a binary vector of length n. Let $X = (X_1, \ldots, X_n)$ be the random variable taking values in $\{0,1\}^n$, where $\Pr[X = v_F] = 1/|\mathcal{F}|$ for all $F \in \mathcal{F}$. Clearly,

$$\mathrm{H}(X) = |\mathcal{F}|\left(- \frac{1}{|\mathcal{F}|} \log_2 \frac{1}{|\mathcal{F}|}\right) = \log_2 |\mathcal{F}|,$$

and since here $\mathrm{H}(X_i) = H(p_i)$ for all $1 \leq i \leq n$, the result follows from Theorem 22.6. □

This corollary supplies a quick proof for the following well-known estimate for sums of binomial coefficients (cf. Exercise 1.17).

Corollary 22.9. *For every integer n and for every real $0 < p \leq 1/2$,*

$$\sum_{i \leq np} \binom{n}{i} \leq 2^{nH(p)}.$$

Proof (due to P. Frankl). Let $\mathcal{F}$ be the family of all subsets of cardinality at most pn of $\{1, 2, \ldots, n\}$. If p_i is the fraction of subsets of $\mathcal{F}$ that contain i then $p_1 = \ldots = p_n$. Counting in two ways we obtain that $\sum_{i=1}^{n} p_i \le pn$, and hence $p_i \le p$ for all i. Since the function $H(p)$ is increasing for $0 \le p \le 1/2$ this, together with Corollary 22.8, implies that

$$\sum_{i \le np} \binom{n}{i} = |\mathcal{F}| \le 2^{\sum_{i=1}^{n} H(p_i)} \le 2^{nH(p)},$$

as needed. □

The following easy consequence of Theorem 22.7 tells us that a family cannot have many members if its "projections" are small. For a family $\mathcal{F} \subseteq 2^{[n]}$ of subsets of $[n] = \{1, \ldots, n\}$, the *projection* of $\mathcal{F}$ onto a subset $S \subseteq [n]$ is the family of all intersections $S \cap A$ with $A \in \mathcal{F}$.

Theorem 22.10 (Product Theorem)**.** *Let $S_1, \ldots S_m$ be subsets of $[n]$ such that each element of $[n]$ is contained in at least k of them. Let $\mathcal{F} \subseteq 2^{[n]}$ and let $\mathcal{F}_i$ be the projection of $\mathcal{F}$ onto S_i. Then*

$$|\mathcal{F}|^k \le \prod_{i=1}^{m} |\mathcal{F}_i| \,.$$

Proof. Let $B_1 = \ldots = B_n = \{0, 1\}$, and let $X = (X_1, \ldots, X_n)$ be the random variable taking values in $B_1 \times \cdots \times B_n$, where for each $F \in \mathcal{F}$, X is equal to the incidence vector of F with probability $1/|\mathcal{F}|$. By Theorem 22.7, $k \cdot \mathrm{H}(X) \le \sum_{i=1}^{m} \mathrm{H}(X_{S_i})$. But $\mathrm{H}(X) = \log_2 |\mathcal{F}|$, whereas $\mathrm{H}(X_{S_i}) \le \log_2 |\mathcal{F}_i|$, implying the desired result. □

The generalized subadditivity of the entropy function can be used to prove some non-trivial "intersection theorems" (cf. Chap. 7). The following three results were obtained by Chung, Frankl, Graham, and Shearer (1986).

Recall that a family $\mathcal{F} \subseteq 2^{[n]}$ is *intersecting* if $A \cap B \ne \emptyset$ for all $A, B \in \mathcal{F}$. If $\mathcal{F}$ is intersecting then $|\mathcal{F}| \le 2^{n-1}$, since $\mathcal{F}$ cannot contain both a set and its complement. Moreover, this is optimal (just take the family of all subsets of $[n]$ containing one fixed point). To make the question more interesting we can require that the members of $\mathcal{F}$ not just intersect, but that these intersections contain at least one of the given configurations. We first consider one easy example.

Theorem 22.11. *Let $\mathcal{F} \subseteq 2^{[n]}$ a family, and suppose that the intersection of any two of its members contains a pair of consecutive numbers. Then*

$$|\mathcal{F}| \le 2^{n-2} \,.$$

Note that this upper bound is optimal: just let $\mathcal{F}$ be the family of all subsets containing the set $\{1, 2\}$

Proof. By our assumption, for every $A, B \in \mathcal{F}$ there must be an i such that $\{i, i+1\} \subseteq A \cap B$. Let S_0 and S_1 be the set of all even and odd numbers in $[n]$, respectively. Consider the projections $\mathcal{F}_a = \{A \cap S_a : A \in \mathcal{F}\}$ of our family $\mathcal{F}$ onto these two sets, $a = 0, 1$. Note that the intersection of any two members of $\mathcal{F}_a$ has the form $(A \cap B) \cap S_a$ for some $A, B \in \mathcal{F}$. Since $|\{i, i+1\} \cap S_a| = 1$ for every i and a, both of the families $\mathcal{F}_a$ are intersecting, and hence, $|\mathcal{F}_a| \leq 2^{|S_a|-1}$ for both $a = 0, 1$. Using Theorem 22.10 (with $k = 1$) we conclude that

$$|\mathcal{F}| \leq |\mathcal{F}_0| \cdot |\mathcal{F}_1| \leq 2^{|S_0|-1} \cdot 2^{|S_1|-1} = 2^{n-2}. \qquad \square$$

Theorem 22.10 also has not so trivial consequences. For example, one may take U to be the set of all $\binom{n}{2}$ edges of a complete graph K_n. We can look at the subsets $A \subseteq U$ as (labeled) subgraphs of K_n. (Dealing with "labeled" subgraphs means that we do not identify isomorphic ones.)

Theorem 22.12. *Suppose that $\mathcal{F}$ is a family of (labeled) subgraphs of K_n such that for all $A, B \in \mathcal{F}$, the graph $A \cap B$ does not contain any isolated vertices. Then*

$$|\mathcal{F}| \leq 2^{\binom{n}{2} - \frac{n}{2}}.$$

Note that this requirement is rather severe: each of the n vertices must be incident with at least one common edge in $A \cap B$. Is this upper bound optimal? (See Exercise 22.9.)

Proof. Choose S_i to be the star at the i-th vertex, i.e., S_i consists of all $n-1$ edges $\{i, j\}$, $j \neq i$. Clearly, every edge is in exactly two of $S_1, \ldots, S_n$. Consider the projections $\mathcal{F}_i := \{A \cap S_i : A \in \mathcal{F}\}$ of $\mathcal{F}$ onto these stars, $i = 1, 2, \ldots, n$. Note that the intersection of any two members of $\mathcal{F}_i$ has the form $(A \cap B) \cap S_i$ for some $A, B \in \mathcal{F}$, and hence, is non-empty because the vertex i cannot be isolated in the subgraph $A \cap B$. Thus, each of the families $\mathcal{F}_i$ is intersecting, and hence, $|\mathcal{F}_i| \leq 2^{|S_i|-1} = 2^{n-2}$ for all $i = 1, 2, \ldots, n$. Applying Theorem 22.10 (with $k = 2$) we conclude that

$$|\mathcal{F}| \leq \Big(\prod_{i=1}^{n} |\mathcal{F}_i|\Big)^{1/2} \leq 2^{n(n-2)/2} = 2^{\binom{n}{2} - \frac{n}{2}}. \qquad \square$$

Let us say that a family $\mathcal{F}$ of subgraphs of K_n is *triangle-intersecting* if $A \cap B$ contains a triangle for all $A, B \in \mathcal{F}$.

Theorem 22.13. *Let $n \geq 4$ be even and let $\mathcal{F}$ be a family of (labeled) subgraphs of K_n. If $\mathcal{F}$ is triangle-intersecting then*

$$|\mathcal{F}| \leq 2^{\binom{n}{2} - 2}.$$

It is not known if this bound is optimal, i.e., if $2^{\binom{n}{2}-2}$ can be replaced by $2^{\binom{n}{2}-3}$, the number of subgraphs of K_n containing a fixed triangle.

Proof. We choose S_i, $1 \le i \le m := \frac{1}{2}\binom{n}{n/2}$, to be all possible disjoint unions of two complete (labeled) subgraphs on $n/2$ vertices each. That is, each S_i has the form $K_U \cup K_{\overline{U}}$ for some subset of vertices $U \subseteq \{1, \ldots, n\}$ with $|U| = n/2$. Note that the intersection of any two members of $\mathcal{F}_i$ has the form $(A \cap B) \cap S_i$ for some $A, B \in \mathcal{F}$ and $S_i = K_U \cup K_{\overline{U}}$. By our assumption, $A \cap B$ must contain three edges e_1, e_2, e_3 forming a triangle. Since no triangle can lie entirely in a bipartite graph, at least one of these three edges must belong to S_i. So, each projection $\mathcal{F}_i$ is an intersecting family, implying that $|\mathcal{F}_i| \le 2^{|S_i|-1}$. Each of the graphs S_i has $s := 2\binom{n/2}{2}$ edges, and each edge of K_n is in $k := \binom{n-2}{n/2}$ of the S_i's. By Theorem 22.10,

$$|\mathcal{F}| \le \Big(\prod_{i=1}^{m} 2^{|S_i|-1}\Big)^{1/k} = 2^{(s-1)m/k}.$$

Substituting the values of s, m and k, we conclude that

$$\begin{aligned}\frac{(s-1)m}{k} &= \frac{1}{2}\left[2\binom{n/2}{2} - 1\right]\binom{n}{n/2}\binom{n-2}{n/2}^{-1} \\ &\le \binom{n}{2} - \frac{n(n-1)}{n(n/2-1)} \le \binom{n}{2} - 2\,. \quad \square\end{aligned}$$

Exercises

22.1. Prove the properties (a), (b) and (d)—(g) of the entropy. *Hint*: To (22.3): use Gibbs' inequality with $y_b = 1/|S|$.

22.2. Prove the equation (22.4).

22.3. Let X be a random variable taking its values in some set B, and let $Y = f(X)$ where f is some function on B. Prove that $\mathrm{H}(Y) \le \mathrm{H}(X)$. Show that equality holds if and only if f is one-to-one on the set of all $b \in B$ such that $\Pr[X = b] > 0$.

22.4. Show that for any random variables $\mathrm{H}(X, Y) \ge \mathrm{H}(X)$.

22.5. Show that, for any random variable X, $\mathrm{H}(X^2|X) = 0$, but give an example to show that $\mathrm{H}(X|X^2)$ is not always zero. *Hint*: Let X take the values $+1$ and -1 with equal probability.

22.6. The random variable Y takes the integer values $1, 2, \ldots, 2n$ with equal probability. The random variable X is defined by $X = 0$ if Y is even, and $X = 1$ if Y is odd. Show that $\mathrm{H}(Y|X) = \mathrm{H}(Y) - 1$, but that $\mathrm{H}(X|Y) = 0$.

22.7. Prove the inequality (22.7). *Hint*: Take $X = X_{S_1 \setminus S_2}$, $Y = X_{S_1 \cap S_2}$, $Z = X_{S_2 \setminus S_1}$ and apply (22.5).

22.8. Use Lemma 22.2 to prove that the entropy function is *concave* (or *convex down*) in the following sense. Assume that X and Y are random variables distributed over $\{1, \ldots, n\}$ (these two random variables may be dependent). Let $\alpha \in [0, 1]$ be any real number. Define a new random variable Z, also distributed over $\{1, \ldots, n\}$, with probability distribution

$$\Pr[Z = z] = \alpha \Pr[X = z] + (1 - \alpha) \Pr[Y = z] \; .$$

Prove that then

$$\mathrm{H}(Z) \geq \alpha \mathrm{H}(X) + (1 - \alpha) \mathrm{H}(Y) \; .$$

22.9. Show that for even n the bound of Theorem 22.12 is best possible. *Hint*: Consider the family of all subgraphs of K_n containing a fixed matching.

23. Random Walks

There are n rocks forming a circle in a river, and there is a frog on one of these rocks. The frog remains stationary with probability 1/3, jumps to the right rock with probability 1/3 and jumps to the left one with probability 1/3. The frog problem is: where will the frog be after time t?

This is, perhaps, the most popular illustration of what is known as a *random walk*, a concept which arises in many models of mathematics and physics. The reader can find such applications in any standard probability book containing a chapter about random walks or Markov chains. Besides these, random walks have found interesting applications in the theory of computing as well. In this chapter we present some of them.

23.1 The satisfiability problem

A *k-CNF* (conjunctive normal form) is an And of an arbitrary number of *clauses*, each being an Or of k literals; a *literal* is a variable x_i or its negation $\overline{x}_i$. Given such a CNF F, we seek an assignment of constants 0 and 1 to variables such that all the clauses are satisfied. For example, if $F = (x_1 \vee \overline{x}_2)(x_2 \vee x_3)(\overline{x}_1 \vee \overline{x}_3)$, then $(1, 1, 0)$ is a satisfying assignment for this CNF. A CNF is *satisfiable* if it has at least one satisfying assignment.

The k-SAT problem is: Given a k-CNF F, decide whether it is satisfiable. It is well known that for any (even fixed) $k \geq 3$, this problem is very hard—it is a so-called "NP-complete" problem. It is therefore not very likely that the problem can be solved in polynomial (in the number n of variables) time. In contrast, the case $k = 2$ is much easier.

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_23,

23.1.1 Papadimitriou's algorithm for 2-SAT

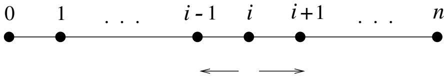
Let F be a 2-CNF, and suppose that we know that it is satisfiable. How quickly can we find a satisfying assignment? Papadimitriou (1991) proposed the following simple randomized procedure.

Suppose we start with an arbitrary assignment of values to the literals. As long as there is a clause that is unsatisfied, we modify the current assignment as follows: we choose an arbitrary unsatisfied clause and pick one of the (two) literals in it uniformly at random; the new assignment is obtained by complementing the value of the chosen literal. After each step we check whether there is an unsatisfied clause; if not, the algorithm terminates successfully with a satisfying assignment.

Theorem 23.1 (Papadimitriou 1991)**.** *Suppose that F is a satisfiable 2-CNF in n variables. Then, with probability at least $1/2$, the above algorithm will find a satisfying assignment in $2n^2$ steps.*

Proof. Fix an arbitrary satisfying assignment $a \in \{0,1\}^n$ for F, and refer to the values assigned by a to the literals as the "correct values."

The progress of the above algorithm can be represented by a particle moving between the integers $\{0, 1, \dots, n\}$ on the real line. The position of the particle indicates how many variables in the current solution have "incorrect values," i.e., values different from those in a. At each iteration, we complement the current value of one of the literals of some unsatisfied clause, so that the particle's position changes by 1 at each step. In particular, a particle currently in position i, for $0 < i < n$, can only move to positions $i - 1$ or $i + 1$:



Let $t(i)$ denote the expected number of steps which a particle, starting in position i, makes until it reaches position 0. Our goal is to show that $t(i) \le n^2$ for all i.

A particle at location n can only move to $n-1$, and the process terminates when the particle reaches position 0 (although it may terminate earlier at some other position with a satisfying assignment other than a). Hence, $t(n) \le t(n-1) + 1$ and $t(0) = 0$. In general, we have that

$$t(i) = p_{i,i-1} \cdot (1 + t(i-1)) + p_{i,i+1} \cdot (1 + t(i+1)),$$

where $p_{i,j}$ is the probability that the particle moves from position i to position $j \in \{i-1, i+1\}$.

The crucial observation is the following: in an unsatisfied clause at least one of the literals has an incorrect value. Thus, with probability at least $1/2$ we decrease the number of variables having false values. The motion of the particle thus resembles a random walk on the line where the particle moves from the i-th position ($0 < i < n$) to position $i-1$ with probability $p_{i,i-1} \geq 1/2$. This implies that

$$t(i) \leq \frac{t(i-1) + t(i+1)}{2} + 1.$$

Replace the obtained inequalities by equations

$$\begin{aligned} x(0) &= 0, \\ x(i) &= \frac{x(i-1) + x(i+1)}{2} + 1, \\ x(n) &= x(n-1) + 1. \end{aligned}$$

This resolves to $x(1) = 2n - 1$, $x(2) = 4n - 4$ and in general $x(i) = 2in - i^2$. Therefore, $t(i) \leq x(i) \leq x(n) = n^2$, as desired.

By Markov's inequality, a random variable can take a value 2 times larger than its expectation with probability at most 1/2. Thus, the probability that the particle will make more than $2 \cdot t(i)$ steps to reach position 0 from position i, is smaller than 1/2. Hence, with probability at least 1/2 the process will terminate in at most $2n^2$ steps, as claimed. □

23.1.2 Schöning's algorithm for 3-SAT

Can one design a similar algorithm also for 3-SAT? In the algorithm for 2-SAT above the randomness was only used to flip the bits—the initial assignment can be chosen arbitrarily: one could always start, say, with a fixed assignment $(1, 1, \ldots, 1)$. But what if we choose this initial assignment at random? If a formula is satisfiable, then we will "catch" a satisfying assignment with probability at least 2^{-n}. Interestingly, the success probability can be substantially increased to about $(3/4)^n$ via the following simple algorithm proposed by Schöning (1999):

1. Pick an initial assignment $a \in \{0, 1\}^n$ uniformly at random. The assignment a can be obtained as a result of n independent experiments, where at the i-th experiment we flip a coin to determine the i-th bit of a.
2. If a satisfies all clauses of F, then stop with the answer "F is satisfiable."
3. If F is not satisfied by a, then pick any of its unsatisfied clauses C, choose one of C's literals uniformly at random, flip its value, and go to step (2).
4. Repeat (3) n times.

For a satisfiable 3-CNF F, let $p(F)$ be the probability that Schöning's algorithm finds a satisfying assignment, and let $p(n) = \min p(F)$ where the minimum is over all satisfiable 3-CNFs in n variables. So, $p(n)$ lower bounds the success probability of the above algorithm.

It is clear that $p(n) \geq (1/2)^n$: any fixed satisfying assignment a^* will be "caught" in Step (1) with probability 2^{-n}. It turns out that $p(n)$ is much larger—it is at least about $p = (3/4)^n$. Thus, the probability that after, say, $t = 30(4/3)^n$ re-starts we will not have found a satisfying assignment is at most $(1-p)^t \leq \mathrm{e}^{-pt} = \mathrm{e}^{-30}$, an error probability with which everybody can live quite well.

Theorem 23.2 (Schöning 1999). *There is an absolute constant $c > 0$ such that*

$$p(n) \geq \frac{c}{n}\left(\frac{3}{4}\right)^n .$$

Proof. Let F be a satisfiable 3-CNF in n variables, and fix some (unknown for us) assignment a^* satisfying F. Let $\mathrm{dist}(a, a^*) = |\{i : \ a_i \neq a_i^*\}|$ be the Hamming distance between a and a^*. Since we choose our initial assignment a at random,

$$\Pr[\mathrm{dist}(a, a^*) = j] = \binom{n}{j} 2^{-n} \qquad \text{for each } j = 0, 1, \ldots, n.$$

Hence, if q_j is the probability that the algorithm finds a^* when started with an assignment a of Hamming distance j from a^*, then the probability q that the algorithm finds a^* is

$$q = \sum_{j=0}^{n} \binom{n}{j} 2^{-n} q_j .$$

To lower bound this sum, we concentrate on the value $j = n/3$. As in the case of 2-CNFs, the progress of the above algorithm can be represented by a particle moving between the integers $0, 1, \ldots, n$ on the real line. The position of the particle indicates how many variables in the current solution have "incorrect values," i.e., values different from those in a^*. If C is a clause not satisfied by a current assignment, then $C(a^*) = 1$ implies that in Step (3) a "right" variable of C (that is, one on which a differs from a^*) will be picked with probability at least $1/3$. That is, the particle will move from position i to position $i-1$ with probability at least $1/3$, and will move to position $i+1$ with probability at most $2/3$. We have to estimate the probability $q_{n/3}$ that the particle reaches position 0, if started in position $n/3$.

Let A be the event that, during n steps, the particle moves $n/3$ times to the right and $2n/3$ times to the left. Then

$$q_{n/3} \geq \Pr[A] = \binom{n}{n/3}\left(\frac{1}{3}\right)^{2n/3}\left(\frac{2}{3}\right)^{n/3} .$$

Now we use the estimate

$$\binom{n}{\alpha n} \geq \frac{1}{O(\sqrt{n})} 2^{n \cdot H(\alpha)} = \frac{1}{\Theta(\sqrt{n})}\left[\left(\frac{1}{\alpha}\right)^{\alpha}\left(\frac{1}{1-\alpha}\right)^{1-\alpha}\right]^n ,$$

where $H(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$ is the binary entropy function (see Exercise 1.16). Therefore, setting $\alpha = 1/3$,

$$\begin{aligned} q &\geq \binom{n}{n/3} q_{n/3} 2^{-n} \\ &\geq \binom{n}{n/3}^2 \left(\frac{1}{3}\right)^{2n/3} \left(\frac{2}{3}\right)^{n/3} 2^{-n} \\ &\geq \frac{1}{\Theta(n)} \left[3^{2/3} \left(\frac{3}{2}\right)^{4/3} \left(\frac{1}{3}\right)^{2/3} \left(\frac{2}{3}\right)^{1/3} 2^{-1} \right]^n \\ &= \frac{1}{\Theta(n)} \left(\frac{3}{4}\right)^n . \end{aligned} \qquad \square$$

23.2 Random walks in linear spaces

Let V be a linear space over $\mathbb{F}_2$ of dimension d, and let $\boldsymbol{v}$ be a random vector in V. Starting with $\boldsymbol{v}$, let us “walk” over V by adding independent copies of $\boldsymbol{v}$. (Being an independent copy of $\boldsymbol{v}$ does not mean being identical to $\boldsymbol{v}$, but rather having the same distribution.) What is the probability that we will reach a particular vector $v \in V$? More formally, define

$$\boldsymbol{v}^{(r)} = \boldsymbol{v}_1 \oplus \boldsymbol{v}_2 \oplus \cdots \oplus \boldsymbol{v}_r ,$$

where $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$ are independent copies of $\boldsymbol{v}$. What can be said about the distribution of $\boldsymbol{v}^{(r)}$ as $r \to \infty$? It turns out that, if $\Pr[\boldsymbol{v} = 0] > 0$ and $\boldsymbol{v}$ is not concentrated in some proper subspace of V, then the distribution of $\boldsymbol{v}^{(r)}$ converges to a uniform distribution, as $r \to \infty$. That is, we will reach each vector of V with almost the same probability!

Lemma 23.3 (Razborov 1988)**.** *Let V be a d-dimensional linear space over $\mathbb{F}_2$. Let $b_1, \ldots, b_d$ be a basis of V and*

$$p = \min\left\{\Pr[\boldsymbol{v} = 0], \Pr[\boldsymbol{v} = b_1], \ldots, \Pr[\boldsymbol{v} = b_d]\right\}.$$

Then, for every vector $u \in V$ and for all $r \geq 1$,

$$\left| \Pr\left[\boldsymbol{v}^{(r)} = u\right] - 2^{-d} \right| \leq \mathrm{e}^{-2pr}.$$

Proof. Let $\langle x, y \rangle = x_1 y_1 \oplus \cdots \oplus x_n y_n$ be the scalar product of vectors x, y over $\mathbb{F}_2$; hence $\langle x, y \rangle = 1$ if and only if the vectors x and y have an odd number of 1s in common. For a vector $w \in V$, let $p_w = \Pr[\boldsymbol{v} = w]$ and set

$$\Delta_v := \sum_{w \in V} p_w (-1)^{\langle w, v \rangle}. \tag{23.1}$$

Then, for every $u \in V$,

$$\begin{aligned}\sum_{v \in V} \mathit{\Delta}_v (-1)^{\langle u,v\rangle} &= \sum_{v \in V} \sum_{w \in V} p_w (-1)^{\langle u \oplus w, v\rangle} \\ &= \sum_{w \in V} p_w \sum_{v \in V} (-1)^{\langle u \oplus w, v\rangle} = 2^d p_u,\end{aligned}$$

since

$$\sum_{v \in V} (-1)^{\langle x,v\rangle} = \begin{cases} 0 & \text{if } x \neq 0; \\ 2^d & \text{if } x = 0. \end{cases} \tag{23.2}$$

Therefore,

$$p_u = 2^{-d} \sum_{v \in V} \mathit{\Delta}_v (-1)^{\langle u,v\rangle}. \tag{23.3}$$

Fix an arbitrary vector $u \in V$. We claim that

$$\Pr\left[\boldsymbol{v}^{(r)} = u\right] = 2^{-d} \sum_{v \in V} \mathit{\Delta}_v^r (-1)^{\langle u,v\rangle}. \tag{23.4}$$

We show this by induction on r. If $r = 1$ then equation (23.4) is just equation(23.3).

Suppose now that equation (23.4) holds for $\boldsymbol{v}^{(r-1)}$, and prove it for $\boldsymbol{v}^{(r)}$. We have

$$\begin{aligned}&\Pr\left[\boldsymbol{v}^{(r)} = u\right] = \Pr\left[\boldsymbol{v}^{(r-1)} \oplus \boldsymbol{v}_r = u\right] \\ &= \sum_{w \in V} \Pr\left[\boldsymbol{v}^{(r-1)} = w\right] \cdot \Pr\left[\boldsymbol{v}_r = u \oplus w\right] \\ &= \sum_{w} \left(2^{-d} \sum_{v} \mathit{\Delta}_v^{r-1} (-1)^{\langle w,v\rangle}\right)\left(2^{-d} \sum_{v'} \mathit{\Delta}_{v'} (-1)^{\langle u \oplus w, v'\rangle}\right) \\ &= 2^{-2d} \sum_{w} \sum_{v} \sum_{v'} \mathit{\Delta}_v^{r-1} \mathit{\Delta}_{v'} (-1)^{\langle w,v\rangle \oplus \langle u,v'\rangle \oplus \langle w,v'\rangle} \\ &= 2^{-2d} \sum_{v} \sum_{v'} \mathit{\Delta}_v^{r-1} \mathit{\Delta}_{v'} (-1)^{\langle u,v'\rangle} \sum_{w} (-1)^{\langle w, v \oplus v'\rangle} \\ &= 2^{-d} \sum_{v} \mathit{\Delta}_v^r (-1)^{\langle u,v\rangle}.\end{aligned}$$

Here, the last equality follows from (23.2), because the last sum is 2^d if $v' = v$, and is 0 otherwise.

By (23.1), $\mathit{\Delta}_0 = 1$. For each other vector $v \neq 0$, there exist vectors w_1, w_2 in $\{0, b_1, \ldots, b_d\}$ such that $\langle w_1, v\rangle \neq \langle w_2, v\rangle$, implying that

$$|\mathit{\Delta}_v| \leq 1 - 2p$$

(if $A + x = 1$ then $A - x = 1 - 2x$). This together with (23.4) yields the desired estimate:

$$\left|\Pr\left[\boldsymbol{v}^{(r)} = u\right] - 2^{-d}\right| \leq \max_{\substack{v \in V \\ v \neq 0}} |\Delta_v^r| \leq (1-2p)^r \leq \mathrm{e}^{-2pr}.$$

□

23.2.1 Small formulas for complicated functions

Razborov (1988) used Lemma 23.3 to establish the following (counterintuitive) phenomenon: some combinatorialy complicated graphs may be represented by small boolean circuits.

Boolean circuits compute boolean functions. So, our first goal is to associate graphs with boolean functions. For this, fix an arbitrary linear order $\preceq$ on $\{0,1\}^n$; for example, we can take $\preceq$ to be the lexicographic order: given two vectors $a \neq b$, look for the smallest index i on which these vectors differ, and set $a \prec b$ if and only if $a_i < b_i$. After we fix such an order on the n-cube, every boolean function $f(x_1, \ldots, x_n, \ldots, x_{2n})$ on $2n$ variables defines the following undirected graph $G(f)$ on $N = 2^n$ vertices:

- the vertices of $G(f)$ are vectors in $\{0,1\}^n$, and
- two vertices $a \neq b$ are joined by an edge if and only if $a \prec b$ and $f(a,b) = 1$.

Intuitively, if the graph $G(f)$ has a "complicated" combinatorial structure then the function f should be "hard" to compute (require large circuits or formulas). It turns out that this intuition is sometimes false!

Consider, for example, Ramsey graphs. Recall that a *clique* in a graph is a subset of vertices, each pair of which is joined by an edge. Similarly, an *independent set* is a set of vertices with no edge between them. Say that a graph is *K-Ramsey graph* if it contains no clique or independent set of size K.

Ramsey's theorem implies that no graph on $N = 2^n$ vertices is $n/2$-Ramsey. On the other hand, using the probabilistic argument we have already proved (see Theorem 4.17) that most graphs on $N = 2^n$ vertices are $2n$-Ramsey; let us call these graphs *strongly Ramsey.* However, no *explicit* construction of strongly Ramsey graphs is known. Explicit graphs that are $2^{\sqrt{n \log n}}$-Ramsey were constructed by Frankl and Wilson (1981) using a powerful linear algebra method (see Theorem 13.15 for the construction).

In the *bipartite* case the situation was even worse. Say that a bipartite graph on two sets of N vertices is a *K-Ramsey bipartite graph* if it has no $K \times K$ complete or empty bipartite subgraph. While Erdős' result on the abundance of $2n$-Ramsey graphs holds as is for bipartite graphs, until recently the best explicit construction of bipartite Ramsey graphs was $2^{n/2}$-Ramsey, using the Hadamard matrix (see Exercise 23.2). This was recently improved, first to $o(2^{n/2})$ by Pudlak and Rödl (2004), then to $2^{o(n)}$ by Barak et al. (2005), and to $2^{n^{o(1)}}$ by Barak et al. (2006).

This—the difficulty to explicitly construct strongly Ramsey graphs—serves as a serious indication that these graphs have a rather "complicated" combinatorial structure.

It is therefore surprising that boolean functions corresponding to these graphs can be computed by very small boolean formulas of depth-3 with And and Parity gates. These formulas have the form:

$$F = F_1 \oplus F_2 \oplus \cdots \oplus F_r,$$

where each F_i has the form

$$F_i = \bigwedge_{j=1}^{m} \bigoplus_{k=1}^{n} \lambda_{ijk} x_k \oplus \lambda_{ij}, \quad \text{with} \;\; \lambda_{ijk}, \lambda_{ij} \in \{0,1\}.$$

The *size* of such a formula F is the number rmn of literals in it. Let $L(f)$ denote the minimum size of such a formula computing f.

Theorem 23.4 (Razborov 1988)**.** *There exists a sequence f_n of boolean functions in $2n$ variables such that the graph $G(f_n)$ is strongly Ramsey and $L(f_n) = O(n^5 \log n)$.*

This result is even more surprising because Razborov (1987) had proved earlier that the model of constant depth formulae is rather weak: some seemingly "simple" boolean functions, like the majority function (which outputs 1 if and only if the input vector contains more 1s than 0s) require constant depth formulas of size exponential in n (cf. Sect. 18.7).

The proof of Theorem 23.4 is based on the following lemma about the distribution of a random depth-3 formula, which may be derived from Lemma 23.3.

Let $\boldsymbol{h}$ be a random boolean formula in n variables given by:

$$\boldsymbol{h} = \boldsymbol{\lambda}_0 \oplus \boldsymbol{\lambda}_1 x_1 \oplus \boldsymbol{\lambda}_2 x_2 \oplus \cdots \oplus \boldsymbol{\lambda}_n x_n,$$

where all $\boldsymbol{\lambda}_i$'s are independent random variables taking their values from $\{0,1\}$ with probability 1/2. Let

$$\boldsymbol{g} = \boldsymbol{h}_1 \wedge \boldsymbol{h}_2 \wedge \cdots \wedge \boldsymbol{h}_m,$$

where $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_m$ are independent copies of $\boldsymbol{h}$. Finally, let $\boldsymbol{f} = \boldsymbol{f}_{n,m,r}$ be a random boolean function given by

$$\boldsymbol{f} = \boldsymbol{g}_1 \oplus \boldsymbol{g}_2 \oplus \cdots \oplus \boldsymbol{g}_r,$$

where $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_r$ are independent copies of $\boldsymbol{g}$.

If $f : \{0,1\}^n \to \{0,1\}$ is a boolean function and $E \subseteq \{0,1\}^n$, then f_E denotes the function f restricted to E.

Lemma 23.5. *Let $E \subseteq \{0,1\}^n$ and $\phi : E \to \{0,1\}$ be a boolean function defined on E. If $|E| \leq 2^{m-1}$, then*

$$\left|\Pr\left[\boldsymbol{f}_E = \phi\right] - 2^{-|E|}\right| \leq \mathrm{e}^{-r/2^m}.$$

Proof. Recall that $\boldsymbol{f}$ is a sum modulo 2 of r independent copies of $\boldsymbol{g}$. We are going to apply Lemma 23.3 in the situation when $V = \{0,1\}^E$ is the linear space of all boolean functions defined on E, and $\boldsymbol{v} = \boldsymbol{g}_E$.

As the basis of our space V we take all $d = |E|$ boolean functions χ_a, $a \in E$, such that $\chi_a(b) = 1$ if and only if $a = b$.

It is clear that for every $a \in \{0,1\}^n$, $\Pr\left[\boldsymbol{h}(a) = 1\right] = 1/2$. Moreover, for $a \neq b$, $\boldsymbol{h}(a)$ and $\boldsymbol{h}(b)$ are two different linear forms $\neq 0$ of independent parameters $\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n$ that are uniformly distributed on $\{0,1\}$. Therefore, $\boldsymbol{h}(a)$ and $\boldsymbol{h}(b)$ are independent, implying that $\Pr\left[\boldsymbol{h}(a) = 1 \mid \boldsymbol{h}(b) = 1\right] = 1/2$, for any $a \neq b$. Since $\boldsymbol{g}$ is an And of m independent copies of $\boldsymbol{h}$, we obtain

$$\Pr\left[\boldsymbol{g}(a) = 1 \mid \boldsymbol{g}(b) = 1\right] = \Pr\left[\boldsymbol{g}(a) = 1\right] = 2^{-m}.$$

Using this and the condition $d \leq 2^{m-1}$ we obtain that

$$\Pr\left[\boldsymbol{g}_E \equiv 0\right] = 1 - \Pr\left[\exists a \in E : \boldsymbol{g}(a) = 1\right] \geq 1 - |E| \cdot 2^{-m} \geq 1/2$$

and

$$\begin{aligned}
\Pr\left[\boldsymbol{g}_E = \chi_a\right] &= \Pr\left[\boldsymbol{g}(a) = 1 \text{ and } \boldsymbol{g}(b) = 0 \text{ for all } b \in E, b \neq a\right] \\
&= \Pr\left[\boldsymbol{g}(a) = 1\right] \cdot \Pr\left[\forall b \neq a : \boldsymbol{g}(b) = 0 \mid \boldsymbol{g}(a) = 1\right] \\
&\geq 2^{-m}\left[1 - \sum_{b \in E, b \neq a} \Pr\left[\boldsymbol{g}(b) = 1 \mid \boldsymbol{g}(a) = 1\right]\right] \\
&\geq 2^{-m}\left(1 - d \cdot 2^{-m}\right) \geq 2^{-m-1}.
\end{aligned}$$

Thus, the minimum of $\Pr\left[\boldsymbol{g}_E \equiv 0\right]$ and $\Pr\left[\boldsymbol{g}_E = \chi_a\right]$ ($a \in E$) is at least 2^{-m-1}, and we can apply Lemma 23.3 with $p = 2^{-m-1}$. This completes the proof of Lemma 23.5. □

Proof of Theorem 23.4. Set $m := \lfloor 2\log_2 n + 3 \rfloor$, $r := \lfloor 40n^4 \rfloor$ and consider the random boolean function $\boldsymbol{f} = \boldsymbol{f}_{2n,m,r}$. Then the graph $G(\boldsymbol{f})$ is *not* strongly Ramsey if and only if there is a subset of vertices $S \subseteq \{0,1\}^n$ of size $|S| = 2n+1$ which either forms a clique or is an independent set in $G(\boldsymbol{f})$. This event happens only if $\boldsymbol{f}_E \equiv 1$ or $\boldsymbol{f}_E \equiv 0$, where E is the set of all pairs (a,b) such that $a, b \in S$ and $a \prec b$. As $|S| = 2n+1$, we have

$$|E| \leq 1 + 2 + \cdots + 2n = n(2n+1),$$

and hence, the condition $d \leq 2^{m-1}$ of Lemma 23.5 is satisfied (with $d = n(2n+1)$). Applying this lemma, we obtain that

$$\Pr\left[\boldsymbol{f}_E \equiv 1 \vee \boldsymbol{f}_E \equiv 0\right] = O(\mathrm{e}^{-r/2^m}) = O(2^{-n(2n+1)}),$$

and hence,

$$\begin{aligned}\Pr\left[G(\boldsymbol{f}) \text{ is not strongly Ramsey}\right] &\leq \binom{2^n}{2n+1} \cdot \Pr\left[\boldsymbol{f}_E \equiv 1 \vee \boldsymbol{f}_E \equiv 0\right] \\ &= O\left(\binom{2^n}{2n+1} \cdot 2^{-n(2n+1)}\right) = o(1)\,.\end{aligned}$$

This means that there exists a depth-3 formula $f = f_{2n,m,r}$ of size $2nmr = O(n^5 \log n)$ whose graph $G(f)$ is strogly Ramsey. □

23.3 Random walks and derandomization

Let $G = (V, E)$ be a d-regular graph on n vertices. A *random walk* starting in a vertex $v_0 \in V$ is a sequence $v_0, v_1, \ldots, v_t$ of vertices, where each v_{i+1} is a random neighbor of v_i chosen uniformly at random with probability $1/d$ among all its d neighbors. Note that one and the same vertex may appear in this sequence several times. In a *random t-walk* we first choose the start vertex v_0 uniformly at random, and then run a random walk of length t starting in v_0.

Given a subset of vertices $S \subseteq V$, what is the probability that a random t-walk $v_0, v_1, \ldots, v_t$ will never leave S? Small-degree graphs, where this probability is small, play an important role in many applications, one of them being "derandomization" of probabilistic algorithms, that is, reduction of the number of random bits used by algorithm.

We now give an upper bound on this probability in terms of the second eigenvalue of the graph G.

Theorem 23.6 (Hitting property of expander walks)**.** *Let $G = (V, E)$ be a d-regular n-vertex graph, and λ_2 the second-largest eigenvalue of its adjacency matrix A. Then for every subset $S \subseteq V$,*

$$\Pr\left[\textit{random } t\textit{-walk will stay inside } S\right] \leq \left(\frac{\lambda_2}{d} + \frac{|S|}{n}\right)^t .$$

Thus, if both λ_2/d and $|S|/n$ are less than $1/2$, then with exponentially large probability $1 - 2^{\Omega(-t)}$ a random t-walk will hit a vertex outside S.

Proof. For technical reasons it will be convenient to consider not the adjacency matrix of G itself, but its *normalized adjacency matrix* $A = (a_{ij})$ with $a_{ij} = 1/d$ if i and j are adjacent, and $a_{ij} = 0$ otherwise. The reason to consider this matrix (and not the adjacency matrix) is only technical: the matrix A is doubly-stochastic, that is, the entries in each row and each column sum up to 1. Note that $\lambda := \lambda_2/d$ is then the second-largest eigenvalue of A.

At the beginning, each of the vertices $1, 2, \ldots, n$ could be chosen as the start vector v_0 with the same probability $1/n$. So, the probability distribution for the start vertex v_0 is given by the vector

$$u = (1/n, 1/n \ldots, 1/n)\,.$$

Then Au is the probability distribution of the first reached vertex v_1, and $A^i u$ is the probability distribution for the i-th reached vertex v_i. Fix now a subset of vertices $S \subseteq V$. Let D be the “characteristic matrix” of S. This is a diagonal 0-1 matrix whose i-th diagonal entry is 1 iff $i \in S$.

Lemma 23.7. *The probability that the random t-walk does not leave S is precisely the sum of entries of the vector $(DA)^t Du = (DAD)^t u$.*

Proof. Induction on t. The base case is $t = 0$. The i-th entry of $(DA)^0 Du = Du$ is $1/n$ if $i \in S$, and is 0 if $i \notin S$. Hence, the sum $|S|/n$ of entries of Du is exactly the probability that the start vertex v_0 will belong to S.

Now assume the hypothesis holds up to some $t - 1$. Then the i-th entry of $(DA)^t Du$ is the probability that the random walk is at vertex i after t steps, and never leaves S until possibly the last step. Multiplying by D, we zero out all components for vertices not in S and leave the others unchanged. Thus, we obtain the probability that the random walk is at vertex i after t steps, and never leaves S. □

Lemma 23.8. *Let $\mu := |S|/n$, and let D be the characteristic matrix of S. Then for every vector w,*

$$\|DADw\| \leq (\lambda + \mu)\|w\|\,.$$

Proof. First note that there is no loss in assuming that w is supported by D, that is, $Dw = w$. Otherwise we may replace w by Dw. This leaves the left-hand side unchanged (since $D \cdot D = D$) and does not increase the right-hand side, since D is a contraction (that is, $\|Dw\| \leq \|w\|$). Similarly, we may assume that w is non-negative. Also, by linearity of both sides we may assume that $\sum_{i=1}^{n} w_i = 1$ and so w can be expressed as:

$$Dw = w = u + z \qquad \text{where } z \text{ is orthogonal to } u.$$

Since A is doubly-stochastic and all entries of u are the same (equal to $1/n$), it follows that

$$DADw = DAu + DAz = Du + DAz$$

and hence

$$\|DADw\| \leq \|Du\| + \|DAz\|\,.$$

We now prove that $\|Du\| \leq \mu \cdot \|w\|$ and $\|DAz\| \leq \lambda \cdot \|w\|$, which together imply the claim.

Since $\sum_i w_i = 1$, and w has at most $|S| = \mu n$ nonzero coordinates, the Cauchy–Schwarz inequality yields $1 = \sum_i w_i \leq \sqrt{\mu n} \cdot \|w\|$. Since $\|Du\| = \sqrt{\mu/n}$, we obtain $\|Du\| = \mu/\sqrt{\mu n} \leq \mu \cdot \|w\|$.

As for the second term, we have $\|Az\| \leq \lambda \cdot \|z\|$, since z is orthogonal to $\mathbf{1}$ and therefore is a linear combination of eigenvectors of A with eigenvalues of absolute values $\leq \lambda$. But $\|DAz\| \leq \|Az\|$, since D is a contraction. Also $w = u + z$ with $u \perp z$ implies that

$$\|z\|^2 = \sum_i z_i^2 \leq \sum_i (u_i^2 + z_i^2) = \sum_i (u_i + z_i)^2 = \|w\|^2 ,$$

where the second equality holds because $\sum_i u_i z_i = 0$. Hence,

$$\|DAz\| \leq \|Az\| \leq \lambda \cdot \|z\| \leq \lambda \cdot \|w\| ,$$

as needed. □

Now we turn to the actual proof of Theorem 23.6. Let p be the probability that the random t-walk will not leave the set S. By Lemma 23.7, we known that p is the sum of entries in the vector $x := (DA)^t Du$, where $u = (1/n, 1/n \ldots, 1/n)$. By the Cauchy–Schwarz inequality $\langle x, y\rangle \leq \|x\| \cdot \|y\|$, we have that $x_1 + \cdots + x_n \leq \sqrt{n} \cdot \|x\|$. Hence, since $\|u\| = 1/\sqrt{n}$, Lemma 23.8 implies that

$$\begin{aligned} p &\leq \sqrt{n} \cdot \|(DA)^t Du\| = \sqrt{n} \cdot \|(DAD)^t u\| \\ &\leq \sqrt{n} \cdot (\lambda + \mu)^t \|u\| = (\lambda + \mu)^t = \left(\frac{\lambda_2}{d} + \frac{|S|}{n}\right)^t , \end{aligned}$$

as claimed. □

To explain how Theorem 23.6 can be used to drastically reduce the number of random bits used by probabilistic algorithms, let us consider the following simplified scenario. We have a set $S \subseteq [n]$ of "good" elements in some large set $[n]$. We know nothing about S except that, say $|S| \geq n/2$. A probabilistic algorithm tries to find a good element with some large probability, at least $1 - 1/K$.

A trivial algorithm is to take t independent, uniformly distributed elements of $[n]$. Since $|S| \geq n/2$, the probability of missing S is at most 2^{-t}, so it is enough to take $t = \log_2 K$ random points. But to generate a random point we need $m = \log_2 n$ random 0-1 bits. Thus, our algorithm uses about $O(m \cdot \log K)$ random bits in total.

Using expanders we can reduce the number of random bits in our algorithm from $O(m \cdot \log K)$ to $O(m + \log K)$ as follows. Take an explicitly constructed expander G of *constant* degree d on $[n]$. Take a random t-walk, where t is chosen such that the probability of missing S is at most $1/K$. By Theorem 23.6, taking $t = O(\log K)$ suffices, provided $\lambda < (1 - \epsilon)d$ for some constant $\epsilon > 0$. We need about $m = \log n$ random bits to generate an initial vertex of a walk and a *constant* number $\log_2 d$ random bits in each step. Thus, the required number of random bits is at most $O(\log n + \log K)$.

Exercises

23.1. Let $P = (p_{i,j})$ be a *transition matrix* of a random walk on an undirected connected graph $G = (V, E)$ with n vertices $V = \{1, \ldots, n\}$. That is, $p_{i,j} = 1/d(i)$ if $\{i, j\} \in E$, and $p_{i,j} = 0$ otherwise. Let H_{ij} be the expected number of steps needed to visit state j for the first time, when starting from state i. Define the vector $\pi = (\pi_1, \ldots, \pi_n)$ by

$$\pi_i := \frac{d(i)}{2|E|} \qquad \text{for } i = 1, \ldots, n.$$

Show that $\sum_{i=1}^n \pi_i = 1$; $\pi \cdot P = \pi$, and $H_{ii} = 1/\pi_i$ for all $i = 1, \ldots, n$.

23.2. Let $N = 2^n$. The bipartite $N \times N$ Sylvester graph $G = (U \cup V, E)$ with $U = V = \mathbb{F}_2^n$ contains vectors in $\mathbb{F}_2^n$ as vertices, and two vectors $x \in U$ and $y \in V$ are adjacent iff their scalar product over $\mathbb{F}_2$ is equal to 1. Show that this graph is $2^{n/2}$-Ramsey. *Hint*: Use Lindsey's Lemma (Lemma 14.5).

23.3. (Razborov 1988). Let $V = \{1, \ldots, N\}$ be a set of players, $N = 2^n$. A *tournament* is an oriented graph $T = (V, E)$ such that $(i, i) \notin E$ for all $i \in V$, and for any two players $i \neq j$ exactly one of (i, j) and (j, i) belongs to E. A tournament is *transitive* if there exists a permutation σ of the players so that $(i, j) \in E$ if and only if $\sigma(i) < \sigma(j)$. Let $v(T)$ be the largest number of players in a transitive subtournament of T. It is known that: (a) $v(T) \geq n + 1$ for every tournament T, and (b) tournaments T with $v(T) \leq 2n + 1$ exist.

Every boolean function $f(x, y)$ on $2n$ variables defines a tournament $T(f)$ with $N = 2^n$ players in a natural way: players are vectors in $\{0, 1\}^n$, and player a beats player b iff either $a \prec b$ and $f(a, b) = 1$, or $b \prec a$ and $f(b, a) = 0$. Prove that there exists a sequence of boolean functions $f_n(x, y)$ such that $L(f_n) = O(n^5 \log n)$ and $v(T(f)) \leq 2n + 1$. *Hint*: Argue as in the proof of Theorem 23.4. Instead of sets V of size $2n + 1$ take sets of size $2n + 2$, and instead of the event "$\boldsymbol{f}_E = 1$ or $\boldsymbol{f}_E = 0$" consider the event "$\boldsymbol{f}_E$ induces a transitive subtournament of $T(\boldsymbol{f})$."

24. Derandomization

Probabilistic thinking turns out to be uncannily effective for proving the existence of combinatorial objects. Such proofs can often be converted into randomized algorithms. There exist efficient randomized algorithms for problems that are not known to have efficient deterministic solutions. Even when both deterministic as well as randomized algorithms are available for a problem, the randomized algorithm is usually simpler. This fact may be enough to favor the randomized algorithm in practice. Sometimes, the route to deterministic solution is via a randomized one: after a randomized algorithm has been discovered, we may be able to remove the use of randomness. We will illustrate such "derandomization" techniques.

24.1 The method of conditional probabilities

The aim of this method is to convert probabilistic proofs of *existence* of combinatorial structures into efficient deterministic algorithms for their actual *construction*. The idea is to perform a binary search of the sample space Ω for a good point. At each step, the current sample space is split into two equal halves and the *conditional probability* of obtaining a good point is computed for each half. The search is then restricted to the half where the conditional probability is higher. The search terminates when only one sample point (which must be good) remains. This method is applicable to large sample spaces Ω since it requires only $\log_2 |\Omega|$ steps. In situations where the corresponding conditional probabilities can be effectively computed (or at least approximated) this approach works pretty well. To explain the idea, let us consider the following problem.

Given a 3-CNF formula $F(x_1, \ldots, x_n)$, we want to find an assignment of values 0 or 1 to $x_1, \ldots, x_n$ satisfying as many clauses as possible. If we assign to each variable the value 0 or 1 at random independently and with equal probability, then we may expect that at least $7/8$ fraction of clauses will be

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_24,

satisfied, just because each clause is satisfied with probability $1 - 2^{-3} = 7/8$ (see Proposition 24.2).

But where is the assignment? The argument above guarantees only the *existence* of such an assignment and gives no idea about how to *find* it. An exhaustive search will always lead us to the desired assignment. But this dummy strategy will require exponential (in n) number of steps. Can we do better? It turns out that we can "derandomize" the probabilistic *proof* of existence so that it leads to a *deterministic* algorithm which is only polynomial in the length of the input formula.

Before we turn to a formal description of the method, let us first try to solve our special problem with the help of a chimpanzee (this beautiful explanation is due to Maurice Cochand).

We build a binary tree whose 2^n leaves correspond to the 2^n possible assignments. Leaves are close to the sky, as they should be. Going up to the left branch at level i corresponds to choosing the value 0 for x_i, going up to the right gives x_i the value 1.

In order to motivate the chimpanzee for this fascinating problem, we attach at every leaf of the tree a black box containing a number of bananas equal to the number of clauses satisfied by the assignment corresponding to that leaf. We do then invite the chimpanzee to go up in order to bring down one of the black boxes, making him most clearly the potential benefit of the operation.

We repeat this experiment many times, with many different trees corresponding to as many formulas F, having different number of variables and clauses. The chimpanzee *never* looked at the list of clauses (although he was allowed to do it), did not even care about the number of variables. He moved up quickly along the tree, and *always* brought back a box having a number of bananas at least equal to 7/8 times the number of clauses!

We asked him for his secret (he definitely had one, this was more than luck!). For a number of bananas we do not dare to mention here, he gave the following answer:

"Embarrassingly simple," he said. "At every junction I do the same: because of the weight, the branch supporting the subtree having the biggest number of bananas is not as steep as the other one, there I go!"

24.1.1 A general frame

Suppose we have a sample space Ω, and assume, for simplicity, that it is symmetric (i.e., each point has probability $1/|\Omega|$) and that $\Omega = \{0,1\}^n$. Let $A_1, \ldots, A_m$ be a collection of events, and consider the random variable $X = X_1 + \cdots + X_m$ where X_i is the indicator random variable for A_i. Hence, $\mathrm{E}[X] = \sum_{i=1}^{m} \Pr[A_i]$. Also suppose that we have a *proof* that $\mathrm{E}[X] \geq k$. So, there is a point $(a_1, \ldots, a_n)$ in the sample space in which at least k of the events hold. Our objective is to find such a point *deterministically.*

Introduce n random variables $Y_1, \ldots, Y_n$ where each Y_i takes value 0 or 1 independently with equal probability. We find the bits $a_1, a_2, \ldots$ sequentially as follows. Assume $a_1, \dots, a_j$ have already been fixed. Our goal is to choose a_{j+1}. We make this choice based on the value of "conditional" expectation

$$\mathrm{E}[X|a_1, \dots, a_j] := \sum_{i=1}^{m} \Pr[A_i \,|\, a_1, \dots, a_j],$$

where here and in what follows "$a_1, \dots, a_j$" stands for the event that $Y_1 = a_1, \dots, Y_j = a_j$. By Adam's Theorem (Exercise 3.1), for each choice of $a_1, \dots, a_j$ and for each event A_i, the conditional probability

$$\Pr[A_i \,|\, a_1, \dots, a_j]$$

of the event A_i given the values $Y_1 = a_1, \dots, Y_j = a_j$ is the average

$$\frac{\Pr[A_i \,|\, a_1, \dots, a_j, 0] + \Pr[A_i \,|\, a_1, \dots, a_j, 1]}{2}.$$

of the two conditional probabilities corresponding to the two possible choices for Y_{j+1}. Consequently,

$$\begin{aligned} \mathrm{E}[X|a_1, \dots, a_j] &= \frac{\mathrm{E}[X|a_1, \dots, a_j, 0] + \mathrm{E}[X|a_1, \dots, a_j, 1]}{2} \\ &\leq \max\left\{\mathrm{E}[X|a_1, \dots, a_j, 0],\, \mathrm{E}[X|a_1, \dots, a_j, 1]\right\}. \end{aligned}$$

Therefore, if the values a_{j+1} are chosen, each one in its turn, so as to maximize the value of $\mathrm{E}[X|a_1, \dots, a_{j+1}]$, then this value cannot decrease. Since this value is k at the beginning, it follows that it is at least k at the end. But at the end, each a_i is fixed, and hence the value of $\mathrm{E}[X|a_1, \dots, a_n]$ is precisely the number of events that hold at the point $(a_1, \dots, a_n)$, showing that our procedure works.

Note that the procedure above is efficient provided n is not too large (as is usually the case in combinatorial examples) and, more importantly, provided the conditional probabilities $\Pr[A_i \,|\, a_1, \dots, a_j]$ can be computed efficiently.

To see how the "chimpanzee algorithm" from the previous section fits in this general frame, just observe that the total weight of the bananas in the subtree reached by the chimpanzee after j moves $a_1, \dots, a_j$ is the number of clauses X in F, times the conditional expectation $\mathrm{E}[X|a_1, \dots, a_j]$.

24.1.2 Splitting graphs

A widely applied remark of Paul Erdős is that a graph with m edges always contains a bipartite subgraph of at least $m/2$ edges. This fact has a quick probabilistic proof.

Theorem 24.1 (Erdős 1965c)**.** *Every graph with m edges always contains a bipartite subgraph of at least $m/2$ edges.*

Proof. Let $G = (V, E)$ with the vertex set $V = \{1, \ldots, n\}$. Take a random subset $\boldsymbol{U} \subseteq V$ given by $\Pr[i \in \boldsymbol{U}] = 1/2$, these probabilities being mutually independent. Call an edge $e = \{i, j\}$ *crossing* if exactly one of i, j is in $\boldsymbol{U}$. Let X be the number of crossing edges. Then $X = \sum_{e \in E} X_e$, where X_e is the indicator random variable for the edge e being crossing. For a given edge e, $\mathrm{E}[X_e] = 1/2$ as two fair coin flips have probability $1/2$ of being different. By linearity of expectation,

$$\mathrm{E}[X] = \sum_{e \in E} \mathrm{E}[X_e] = \frac{|E|}{2}.$$

Thus, $X \geq |E|/2$ for some choice of $\boldsymbol{U}$, and the set of those (corresponding to this particular $\boldsymbol{U}$) edges forms the desired bipartite subgraph. □

The proof of this theorem gives us a *randomized* algorithm to find a bipartite subgraph whose expected number of edges is at least $|E|/2$. Moreover, Luby (1986) has shown how it can be converted to a linear time *deterministic* algorithm.

We use the conditional expectations to derandomize the algorithm. Introduce n random variables $Y_1, \ldots, Y_n$ where $Y_i = 1$ if $i \in \boldsymbol{U}$, and $Y_i = 0$, otherwise. Select an $a_1 \in \{0, 1\}$ such that $\mathrm{E}[X|Y_1 = a_1] \geq \mathrm{E}[X|Y_1 = a_1 \oplus 1]$, and set $Y_1 = a_1$. Repeat this process for all Y_i's. At the end we have an assignment for all Y_i's such that $\mathrm{E}[X|Y_1 = a_1, \ldots, Y_n = a_n] \geq |E|/2$. But X is no longer a random variable at this point (since $\boldsymbol{U}$ is completely defined), so $X \geq |E|/2$.

What is the running time? To determine each a_i, we need to count the number of edges between vertices in the current U and vertex i, the number of edges between vertices in the current $V \setminus U$ and vertex i. If the former is smaller than the latter, we set $a_i = 1$; otherwise, we set $a_i = 0$. This means that we only need to check the edges incident to vertex i to determine a_i. So the running time of this algorithm is $O(n + |E|)$.

24.1.3 Maximum satisfiability: the algorithmic aspect

Recall that a *k-CNF formula* (or conjunctive normal form) 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 k *literals*, each literal being either a variable x_i or a negated variable $\overline{x}_i$. An *assignment* is a mapping which assigns each variable one of the values 0 or 1. An assignment satisfies a clause if it satisfies at least one of its literals.

Proposition 24.2. *For any k-CNF formula there is an assignment that satisfies at least* $(1 - 2^{-k})$ *fraction of its clauses.*

Proof. Let F be a k-CNF formula on n variables $x_1, \ldots, x_n$ with m clauses. Assign each variable x_i the value 0 or 1 independently at random with probability $1/2$. Since each clause will be satisfied with probability $1 - 2^{-k}$, the expected number of satisfied clauses is $m(1-2^{-k})$. By the pigeonhole principle of expectation, there must exist an assignment satisfying this many of clauses. □

Using conditional expectations, this proof can be transformed to a deterministic algorithm. If $Z = Z(x_1, \ldots, x_n)$ is the number of satisfied by our random assignment clauses, then (as before)

$$\begin{aligned} \mathrm{E}[Z|(x_1,\ldots,x_i)=(a_1,\ldots,a_i)] &= \frac{1}{2}\mathrm{E}[Z|(x_1,\ldots,x_{i+1})=(a_1,\ldots,a_i,0)] \\ &\quad + \frac{1}{2}\mathrm{E}[Z|(x_1,\ldots,x_{i+1})=(a_1,\ldots,a_i,1)], \end{aligned}$$

for $i=0,\ldots,n-1$. We iteratively choose $a_1, a_2, \ldots, a_n$ to maximize the conditional expectation. The above equality implies that in doing so the conditional expectation never goes below the starting point $m(1-2^{-k})$ guaranteed by Proposition 24.2.

24.2 The method of small sample spaces

The problem with randomized algorithms is that usually their sample spaces are of exponential size, so that an exhaustive search is impossible. It turns out however, that a (uniform) sample space Ω associated with a randomized algorithm *always* contains a polynomial-sized subspace $S \subseteq \Omega$ which still has a good point for each possible input, i.e., for every input x there is a point $r \in S$ such that $\mathcal{A}(x,r) = f(x)$.

Theorem 24.3 (Adleman 1978)**.** *There exists a set $S \subseteq \Omega$ of size $|S| \leq n$ such that for every input $x \in \{0,1\}^n$ there is at least one good point in S.*

Proof. Let $M = (m_{x,r})$ be a $2^n \times |\Omega|$ 0-1 matrix whose rows are labeled by inputs and columns by points in the sample space, such that $m_{x,r} = 1$ if r is a good point for x, and $m_{x,r} = 0$, otherwise.

The desired set $S \subseteq \Omega$ is constructed iteratively. Initially, S is empty. Since $\Pr[\mathcal{A}(x,\boldsymbol{r}) = f(x)] \geq 1/2$, each row of M has at least $|\Omega|/2$ ones. Thus, for at least one column r in M, at least half of its entries must be 1s, and hence, there exists a point $r \in \Omega$ that is good for at least half of the inputs x. We add this point r to S, delete all the rows in M corresponding to inputs for which r is good, and repeat the argument for the resulting submatrix. After at most $\log_2 2^n = n$ iterations there will be no rows, and the obtained set S of points has the desired properties. □

Unfortunately, the above result is highly non-constructive and it cannot be used to actually derandomize algorithms. This difficulty was overcome in certain cases by constructing a (different) polynomial-sized sample spaces.

24.2.1 Reducing the number of random bits

Let $f(x)$ be some function; for simplicity assume that its domain is the n-cube $\{0,1\}^n$. A randomized algorithm $\mathcal{A}$ for f works as follows. For a given input x, the algorithm first performs a sequence of coin flips to produce a random string $\boldsymbol{r} \in \{0,1\}^m$, and then computes the value $\mathcal{A}(x, \boldsymbol{r})$. Hence, for each input x, the output of the algorithm is a random value. The algorithm computes f with error probability $\epsilon \leq 1/2$ if for each input $x \in \{0,1\}^n$,

$$\mathrm{Pr}_{\boldsymbol{r}}\left[\mathcal{A}(x,\boldsymbol{r}) \neq f(x)\right] \leq \epsilon\,.$$

For various reasons it is important to keep the number m of random bits as small as possible. If $m \gg \log_2 n$, then the number of random bits can be reduced to about $\log_2 n$.

Theorem 24.4 (Newman 1991)**.** *For every $\delta > 0$ there is a randomized algorithm $\mathcal{B}$ of error probability $\epsilon + \delta$ and the same run-time for f which uses only about* $\log(n/\delta^2)$ *random bits.*

Proof. Let $Z(x, \boldsymbol{r})$ be the indicator random variable for the event that $\mathcal{A}(x,\boldsymbol{r}) \neq f(x)$. Because $\mathcal{A}$ computes f with ϵ error, we have $\mathrm{E}_{\boldsymbol{r}}\left[Z(x,\boldsymbol{r})\right] \leq \epsilon$, for all x. We will build a new algorithm $\mathcal{B}$, which uses fewer random bits, using the probabilistic method.

Let t be a parameter (to be fixed) and $r_1, \ldots, r_t$ be strings in $\{0,1\}^m$. For such strings, define an algorithm $\mathcal{B}_{r_1,\ldots,r_t}$ as follows: For each input x, choose $\boldsymbol{i} \in [t]$ uniformly at random (using $\log_2 t$ random bits) and then compute $\mathcal{A}(x, r_{\boldsymbol{i}})$. We now show that there exist strings $r_1, \ldots, r_t$ such that $\mathrm{E}_{\boldsymbol{i}}\left[Z(x, r_i)\right] \leq \epsilon + \delta$, for all inputs x.

To do this, we choose the t strings $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_t$ independently at random. Consider a particular input x and compute the probability that $\mathrm{E}_{\boldsymbol{i}}\left[Z(x, \boldsymbol{r}_{\boldsymbol{i}})\right] > \epsilon + \delta$, where $\boldsymbol{i}$ in this expectation is uniformly distributed in $[t]$. This probability is exactly the probability that the sum $X := \sum_{j=1}^{t} Z(x, \boldsymbol{r}_j)$ is larger than $(\epsilon+\delta)t$. Since $\mathrm{E}_{\boldsymbol{r}}\left[Z(x,\boldsymbol{r})\right] \leq \epsilon$, we have that $\mathrm{E}\left[X\right] \leq \epsilon t$. Chernoff's inequality (see Theorem 18.22) implies that

$$\Pr[X \geq \epsilon t + \delta t] \leq \mathrm{e}^{-\delta^2 t^2/2t} = \mathrm{e}^{-\delta^2 t/2}\,.$$

By choosing $t = \Theta(n/\delta^2)$, this is smaller than 2^{-n}. Thus, for a random choice of $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_t$ the probability that $\mathrm{E}_{\boldsymbol{i}}\left[Z(x, \boldsymbol{r}_{\boldsymbol{i}})\right] > \epsilon + \delta$ for *some* input x is smaller than $2^n \cdot 2^{-n} = 1$. This implies that there exists a choice of $r_1, \ldots, r_t$, where for *every* x the error of the algorithm $\mathcal{B}_{r_1,\ldots,r_t}$ is at most $\epsilon + \delta$. Finally note that the number of random bits used by this algorithm is $\log_2 t = O(\log(n/\delta^2))$. □

24.2.2 k-wise independence

Another way of constructing a small sample space is by showing that the probabilistic choices of the randomized algorithm are only required to be k-wise independent; then a sample space of size $O(n^k)$ suffices. Another way is to consider "small bias" probability spaces, i.e., to construct small probability spaces that "behave similarly" to larger probability spaces in certain senses.

Let $X_1, \ldots, X_n$ be random variables taking their values in a finite set S. These variables are *k-wise independent* if any k of them are mutually independent, i.e., if for every sequence $(s_{i_1}, \ldots, s_{i_k})$ of k values $s_{i_j} \in S$,

$$\Pr[X_{i_1} = s_{i_1}, \ldots, X_{i_k} = s_{i_k}] = \prod_{j=1}^{k} \Pr[X_{i_j} = s_{i_j}].$$

To illustrate how k-wise independence can help us to derandomize probabilistic proofs, let us look more carefully at the proof of Theorem 24.1. This theorem states that every graph $G = (V, E)$ contains a bipartite subgraph of at least $|E|/2$ edges. That is, there is a subset $U \subseteq V$ of vertices such that at least one half of the edges in E join vertices from U with those from $V \setminus U$. We used random variables to produce the desired (random) subset $\boldsymbol{U}$. Namely, for each vertex $i \in V$, we flip a coin to decide whether to include this vertex into the set $\boldsymbol{U}$ or not. This requires $n = |V|$ coin flips, and hence, the whole sample space is huge – it has 2^n points.

A closer look at this proof shows that the independence is used only to conclude that, for any two vertices $i \neq j$, the events $i \in \boldsymbol{U}$ and $j \in \boldsymbol{U}$ are independent. We need this independence to show that

$$\Pr[i \in \boldsymbol{U}, j \notin \boldsymbol{U}] = \Pr[i \in \boldsymbol{U}] \cdot \Pr[j \notin \boldsymbol{U}] = 1/4.$$

So, in this proof 2-wise independence of the indicator random variables X_i for the events "$i \in \boldsymbol{U}$" suffices. This observation allows us to substantially reduce the size of a sample space as follows (see Exercise 24.3 for more direct construction).

Look at our random set of vertices $\boldsymbol{U} \subseteq V$ as a sequence of random colorings $X_1, \ldots, X_n : V \to \{0, 1\}$, where $X_i = 1$ iff $i \in \boldsymbol{U}$. Our goal is to construct as small as possible sample space for these colorings, in which they are pairwise independent.

Suppose for simplicity that $n = 2^d$ for some d, and identify the vertices with the elements of the field $\mathbb{F}_n$. Choose two elements a and b of this field randomly and independently, and define for each element i the random variable $Z_i = a \cdot i + b$.

Claim 24.5. $Z_1, \ldots, Z_n$ are 2-wise independent.

Proof.

$$\begin{aligned}\Pr[Z_i = x, Z_j = y] &= \Pr[ai + b = x, aj + b = y]\\ &= \Pr\left[a = \frac{x-y}{i-j}, b = \frac{yi - xj}{i-j}\right] = \frac{1}{n^2}\\ &= \Pr[Z_i = x] \cdot \Pr[Z_j = y]. \qquad \square\end{aligned}$$

Encode the elements of $\mathbb{F}_n$ by binary strings of length d, and let X_i be the first bit of the code of i-th element. By the claim, the random variables X_i are also 2-wise independent and uniform on $\{0, 1\}$. Now color the vertex i by the color X_i. Each coloring so obtained is defined by the pair (a, b) of elements from our field $\mathbb{F}_n$. Thus, the whole sample space has size only n^2, and we can perform an exhaustive search of it to find the desired coloring.

Another example is Theorem 4.17 from Sect. 4.9 saying that the edges of K_n can be colored in two colors so that we get no monochromatic $K_{2\log n}$. In this proof, a variable X_i gives the color of the i-th edge. Their independence is used only to estimate the probability that all the edges of some fixed clique on $2 \log n$ vertices, receive the same color. Hence, once again, the k-wise independence with $k = \binom{2\log n}{2} = O((\log n)^2)$ is sufficient. The reader is encouraged to convince himself/herself that in most of the previous proofs k-wise independence with $k \ll n$ works.

Now suppose that for some probabilistic proof, k-wise independence is enough. One may expect that then a sample space Ω of a size much smaller than 2^n would suffice. How much?

In combinatorial terms, we are looking for a set $\Omega \subseteq \{0, 1\}^n$ with the following property: for every set of k coordinates, each vector from $\{0, 1\}^k$ is a projection (onto these k coordinates) *of one and the same* number of vectors in Ω. Thus, if we let $X = (X_1, \ldots, X_n)$ be a string chosen uniformly from Ω then, for any k indices $i_1 < i_2 < \cdots < i_k$ and any k-bit string $\alpha \in \{0, 1\}^k$,

$$\Pr[(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) = \alpha] = 2^{-k},$$

i.e., the coordinates X_i are k-wise independent.

Taking the duals of binary BCH codes, it is possible, for every fixed k, to construct a k-wise independent sample space Ω of size $|\Omega| = O(n^{\lfloor k/2 \rfloor})$. The construction can be found, for example, in the book of Alon and Spencer (1992); see also Exercise 17.1. The idea is to show that the dual code is not only (n, k)-universal (as we have proved in Sect. 17.4) but is such in a very strong sense: for every set of k coordinates, every 0-1 vector of length k is a projection of one and the same number of code words onto these coordinates.

It is natural to ask if this construction is optimal. It turns out that, indeed, the bound $n^{\lfloor k/2 \rfloor}$ cannot be improved, up to a constant factor (depending on k). Say that a random variable is *almost constant* if it takes a single value with probability 1. Let $m(n, k)$ denote the following sum of binomial coefficients:

$$m(n,k) := \sum_{i=0}^{k/2} \binom{n}{i} \qquad \text{if } k \text{ is even},$$

and

$$m(n,k) := \sum_{i=0}^{(k-1)/2} \binom{n}{i} + \binom{n-1}{(k+1)/2} \qquad \text{if } k \text{ is odd.}$$

Observe that for every fixed k, $m(n,k) = \Omega(n^{\lfloor k/2 \rfloor})$.

Theorem 24.6 (Alon–Babai–Itai 1986)**.** *Assume that the random variables $X_1, \ldots, X_n$ over a sample space Ω are k-wise independent and none of them is almost constant. Then $|\Omega| \geq m(n,k)$.*

Note that we assume neither that the variables X_i are $(0,1)$-variables nor that Ω is a symmetric space.

Proof. We can assume that the expected value of each X_i is 0 (since otherwise we can replace X_i by $X_i - \mathrm{E}[X_i]$). For a subset $S \subseteq \{1, \ldots, n\}$, define

$$\alpha_S := \prod_{i \in S} X_i.$$

Since no X_i is almost constant and since the variables are k-wise independent,

$$\mathrm{E}[\alpha_S \alpha_S] = \prod_{i \in S} \mathrm{E}[X_i^2] > 0 \tag{24.1}$$

for all S satisfying $|S| \leq k$. Similarly (and since $\mathrm{E}[X_i] = 0$), for all $S \neq T$ satisfying $|S \cup T| \leq k$, we have

$$\mathrm{E}[\alpha_S \cdot \alpha_T] = \prod_{i \in S \cap T} \mathrm{E}[X_i^2] \cdot \prod_{i \in (S \cup T) \setminus (S \cap T)} \mathrm{E}[X_i] = 0. \tag{24.2}$$

Now let $S_1, \ldots, S_m$ be all the subsets of $\{1, \ldots, n\}$ such that the union of each two is of size at most k. Then $m = m(n,k)$. (Take all sets of size at most $k/2$, and if k is odd add all the subsets of size $(k+1)/2$ containing 1.)

To complete the proof, we show that the functions $\alpha_{S_1}, \ldots, \alpha_{S_m}$ (considered as real vectors of length $|\Omega|$) are linearly independent. Since their number m cannot then exceed the dimension $|\Omega|$, this will imply the result.

To prove the linear independence, take a linear combination $\sum_{i=1}^m \lambda_i \alpha_{S_i}$. Then for every j, multiplying by α_{S_j} and computing expected values we obtain, by (24.2),

$$0 = \sum_{i=1}^m \lambda_i \mathrm{E}[\alpha_{S_i} \cdot \alpha_{S_j}] = \lambda_j \mathrm{E}[\alpha_{S_j} \cdot \alpha_{S_j}].$$

By (24.1), this implies that $\lambda_j = 0$ for all j, and the required linear independence follows. □

Finally, let us mention one recent result in computational complexity concerning the k-independence. Let $f : \{0,1\}^n \to \{0,1\}$ be boolean function. Let

$\boldsymbol{x}$ be a random vector in $\{0,1\}^n$ with some probability distribution. Say that $\boldsymbol{x}$ *ϵ-fools* the function f if $|\mathrm{E}_{\boldsymbol{x}}[f(\boldsymbol{x})] - \mathrm{E}_{\boldsymbol{y}}[f(\boldsymbol{y})]| < \epsilon$, where $\boldsymbol{y}$ is a random vector *uniformly* distributed in $\{0,1\}^n$. That is, the function f cannot distinguish $\boldsymbol{x}$ from a uniformly distributed random vector. Let $F(n,s)$ be the set of all boolean functions in n variables computable by constant-depth circuits using at most s Not and unbounded fanin And, Or gates.

Braverman (2009) proved that, if k is poly-logarithmic in s/ϵ, then *every* function in $F(n,s)$ is ϵ-fooled by a k-independent random vector $\boldsymbol{x}$. That is, shallow circuits of constant depth cannot distinguish k-independent random vectors from truly random ones.

24.3 Sum-free sets: the algorithmic aspect

In previous sections we considered two general approaches toward derandomizing of probabilistic proofs. In this section we will give one example to demonstrate that sometimes the desired polynomial-time algorithm is hidden in the existence proof *itself*.

A subset B of an additive group is called *sum-free* if $x+y \notin B$ for all $x,y \in B$. Erdős (1965a) and Alon and Kleitman (1990) have proved that every finite set A of integers has a sum-free subset B, with $|B| \geq |A|/3$. The proof is probabilistic (see Theorem 18.2) and the question was whether there exists a deterministic algorithm for the selection of such a subset B, which runs in time polynomial in the (binary) size of the problem, that is in $\sum_{a \in A} \log_2 |a_i|$.

Kolountzakis (1994) has shown that, with a slight modification, the proof of Theorem 18.2 can be transformed to such an algorithm.

For a prime p let (as before) $\mathbb{Z}_p = \{0,1,\ldots,p-1\}$ be the field of the integers mod p, and let $\mathbb{Z}_p^* = \{1,2,\ldots,p-1\}$ be the corresponding multiplicative group in $\mathbb{Z}_p$.

Lemma 24.7. *Let p be a prime number of the form $p = 3k+2$. Then, for every nonnegative weight function on $\mathbb{Z}_p^*$, there is a sum-free subset $E \subseteq \mathbb{Z}_p^*$ whose weight is at least one third of the total weight.*

Proof. Let $w(x)$ be a nonnegative weight function defined on $\mathbb{Z}_p^*$. Write $S = \{k+1, k+2, \ldots, k+(k+1)\}$, and observe that S is a sum-free subset in $\mathbb{Z}_p$ and $|S| = k+1 \geq (p-1)/3$. Let $W = \sum_{x \in \mathbb{Z}_p^*} w(x)$ be the total weight of all elements, and let the random variable $\boldsymbol{t}$ be uniformly distributed in $\mathbb{Z}_p^*$. Write $f(\boldsymbol{t}) := \sum w(x)$, where the sum is over all x for which $x \cdot \boldsymbol{t} \in S$, and the product $x \cdot \boldsymbol{t}$ is computed in $\mathbb{Z}_p$. If ξ_x denotes the indicator random variable for the event $x \cdot \boldsymbol{t} \in S$, then $\Pr[\xi_x = 1] = |S|/(p-1)$ and $f(\boldsymbol{t}) = \sum_{x \in \mathbb{Z}_p^*} w(x) \cdot \xi_x$. Hence,

$$\mathrm{E}[f(\boldsymbol{t})] = \sum_{x \in \mathbb{Z}_p^*} w(x)\mathrm{E}[\xi_x] = W \cdot (|S|/(p-1)) \geq W/3$$

By the pigeonhole property of the expectation, there is some $t \in \mathbb{Z}_p^*$ for which $f(t) \geq W/3$. Define $E := t^{-1}S$. This set is sum-free and, since $x \cdot t \in S$ iff $x \in t^{-1}S$, has weight

$$\sum_{x \in E} w(x) = \sum_{x : x \cdot \in S} w(x) = f(t) \geq W/3 . \qquad \square$$

We now turn this proof into an algorithm. Given a set A of integers of (binary) size $\ell := \sum_{a \in A} \log_2 |a|$, our goal is to find a sum-free subset B, with $|B| \geq |A|/3$, in time polynomial in ℓ. We assume that ℓ is large.

First, observe that the number of prime factors of an integer x is at most $\log_2 x$. This means that the number of prime factors which appear in the factorization of *any* element of A is at most ℓ. The Prime Number Theorem (see, for example Zagier (1997)) says that for every pair b, c of relatively prime positive integers, the number of primes $p \leq x$ such that p is of the form $p = bk + c$, asymptotically equals to $\chi(x) = x/(\varphi(b) \cdot \ln x)$, where $\varphi(b)$ is the Euler totient function (the number of numbers in $\mathbb{Z}_p$ relatively prime to b). In our case $b = 3$ and $c = 2$; hence, $\varphi(b) = 2$. Recall that the numbers in A have at most ℓ prime factors in total. So, if we take $x = c\ell \log_2 \ell$ for a large enough constant c (independent on ℓ), then $\chi(x)$ is strictly larger than ℓ, implying that there must be a prime $p \leq c\ell \log_2 \ell$ of the form $p = 3k + 2$ which does not divide any member of A.

Define now the weight $w(x)$ of a number $x \in \mathbb{Z}_p^*$ as the number of elements $a \in A$ such that $a \bmod p = x$. Since p does not divide any member of A, all residues $x = a \bmod p$ of numbers $a \in A$ belong to $\mathbb{Z}_p^*$ (none of them is equal to 0), implying that the sets $A_x = \{a \in A : a \bmod p = x\}$ with $x \in \mathbb{Z}_p^*$ form a partition of A. Thus, $W = \sum_{x \in \mathbb{Z}_p^*} w(x) = |A|$ and, using Lemma 24.7, we can find a sum-free subset $E \subseteq \mathbb{Z}_p^*$ for which the set $B = \{a \in A \colon a \bmod p \in E\}$ has at least $W/3 = |A|/3$ elements. This set B is sum-free since $x + y = z$ for some $x, y, z \in B$ would imply $x + y = z \bmod p$ and E would not be sum-free.

In summary, the steps of our algorithm are the following.

1. Compute all primes up to $c\ell \log_2 \ell$.
2. Find a prime $p = 3k + 2$ which divides no element of A.
3. Compute the values $w(x)$ for all $x \in \mathbb{Z}_p^*$.
4. Find by exhaustive search a $t \in \mathbb{Z}_p^*$ for which $f(t) > |A|/3$ (Lemma 24.7 guarantees that such t exists) and compute the set $E = t^{-1}S$.
5. Construct the set $B = \{a \in A : a \bmod p \in E\}$.

It is easy to verify (do this!) that all these steps can be carried out in time polynomial in ℓ.

Exercises

24.1. Use the method of conditional probabilities to derandomize the proof of Theorem 4.17 and Theorem 25.3.

24.2. Let $G = (V, E)$ be a graph with $n = 2m$ vertices. Improve the lower bound $|E|/2$ on the size of a cut in G (proved in Theorem 24.1) to $|E|$ times $m/(2m-1)$. *Hint*: Follow the argument of Theorem 24.1 with another probability space: choose $\boldsymbol{U} \subseteq V$ uniformly from among all m-element subsets of V. Observe that then any edge has probability $m/(2m-1)$ of being crossing.

24.3. Let $\boldsymbol{r}$ be a random vector uniformly distributed in $\mathbb{F}_2^d$. With each vector $a \in \mathbb{F}_2^d$ associate a random variable $X_a = \langle a, \boldsymbol{r}\rangle$ whose value is the scalar product over $\mathbb{F}_2$ of this vector with $\boldsymbol{r}$. Show that these random variables are 2-wise independent. *Hint*: Exercise 18.11.

24.4. Let $X_1, \ldots, X_n$ be pairwise independent random variables with the identical expectation, denoted μ, and identical variance, denoted σ^2. Let $Z = (\sum X_i)/n$. Use Chebyshev's inequality to prove that

$$\Pr\left[|Z - \mu| \geq \lambda\right] \leq \frac{\sigma^2}{\lambda^2 n}.$$

Hint: Consider the random variables $Y_i := X_i - \mathrm{E}\left[X_i\right]$. Note that the Y_i's are pairwise independent, and each has zero expectation. Apply Chebyshev's inequality to the random variable $Z = (\sum X_i)/n$, and use the linearity of expectation to show that

$$\Pr\left[|Z - \mu| \geq \lambda\right] \leq \mathrm{E}[(\sum Y_i)^2]/(\lambda^2 \cdot n^2)\,.$$

Then (again using the linearity of expectation) show that $\mathrm{E}[(\sum Y_i)^2] = n \cdot \sigma^2$.

24.5. Let $m > 4$, and let H be an $m \times n$ 0-1 matrix, the average density of (i.e., the average number of 1s in) each row of which does not exceed p, $0 \leq p < 1$. Show that then, for every constant $\delta > 0$, there is an $m \times t$ submatrix H' of H such that $t = O(\log(m/\delta^2))$ and each row of H' has average density at most $p + \delta$. *Hint*: Let ξ be a random variable uniformly distributed in $\{1, \ldots, n\}$ and let $\xi_1, \ldots, \xi_t$ be its independent copies, $t = \lceil 4p \log(m/\delta^2)\rceil$. First, observe that with probability strictly larger than 1/2 all the selected columns $\xi_1, \ldots, \xi_t$ are distinct. Next, fix a row $x_1, \ldots, x_n$ of H, and consider the 0-1 random variables $X_i = x_{\xi_i}$, for $i = 1, \ldots, t$. Observe that $\Pr[X_i = 1] \leq p$ and apply the Chernoff inequality (Theorem 18.22) to show that the average density $(\sum X_i)/t$ of 1s in the selected columns can exceed $p + \delta$ with probability at most $1/m^2$. Since we have only m rows, with probability at least $1 - 1/m > 1/2$, all the rows of the selected submatrix will have average density at most $p + \delta$.

24.6. Let $f(x)$ be a boolean function on n variables $x = (x_1, \ldots, x_n)$. Let $F(x, y)$ be a formula with an additional set of boolean variables $y = (y_1, \ldots, y_m)$. The *size* $|F|$ of the formula F is the number of leaves in it. Let

$\boldsymbol{y}$ be a random vector taking its values in $\{0,1\}^m$ independently and with equal probability 2^{-m}. Suppose that $F(x,y)$ computes f with (one-sided) *failure probability* p. That is, for every input $a \in \{0,1\}^n$, $\Pr[F(a,\boldsymbol{y}) \neq f(a)]$ is zero if $f(a)=0$, and is at most p if $f(a)=1$.

(a) Use Adleman's theorem to show that, if $p \leq 1/2$, then f can be computed by a usual (deterministic) formula of size $O(n \cdot |F|)$.

(b) The formula $F(a,\boldsymbol{y})$ can be written in the form

$$F(a,\boldsymbol{y}) = \sum F(x,b) \cdot X_b,$$

where the sum is over all $b \in \{0,1\}^m$ and X_b is the indicator random variable for the event "$\boldsymbol{y} = b$." This formula uses m random bits (to chose a particular formula $F(x,b)$). Use Exercise 24.5 to essentially reduce this number of random bits until $O\left(\log(m/\delta^2)\right)$ at the cost of a slight increase of failure probability by δ. Namely, prove that there is a subset $B=\{b_1,\ldots,b_t\}$ of $t=O(m/\delta^2)$ vectors such that the formula

$$F'(x,\boldsymbol{z}) = \sum_{i=1}^{t} F(x,b_i) \cdot Y_i$$

computes the same boolean function f with failure probability at most $p+1/4$; here $\boldsymbol{z}$ is a random variable taking its values in $\{1,\ldots,t\}$ independently and with equal probability, and Y_i is the indicator random variable for the event "$\boldsymbol{z} = i$."

Part V
Fragments of Ramsey Theory

25. Ramseyan Theorems for Numbers

In 1930 Frank Plumpton Ramsey wrote a paper *On a problem in formal logic* which initiated a part of discrete mathematics nowadays known as Ramsey Theory. At about the same time B.L. van der Waerden (1927) proved his famous Ramsey-type result on arithmetical progressions. A few years later Ramsey's theorem was rediscovered by P. Erdős and G. Szekeres (1935) while working on a problem in geometry. In 1963 A.W. Hales and R.I. Jewett revealed the combinatorial core of van der Waerden's theorem and proved a general result which turned this collection of separate ingenious results into Ramsey Theory.

In this chapter we discuss several Ramsey-type problems in additive number theory.

25.1 Arithmetic progressions

Let $W(r,k)$ be the least number n such that any coloring of $\{1,2,\ldots,n\}$ in r colors gives a monochromatic arithmetic progression with k terms, i.e., for any such coloring there exist integers a,b such that all the points

$$a, a+b, a+2b, \ldots, a+(k-1)b$$

get the same color. In other words, a sequence $a_1, a_2, \ldots, a_k$ of numbers is an arithmetic progression if and only if each its element a_i $(1 < i < k)$ is the arithmetic mean $a_i = (a_{i-1} + a_{i+1})/2$ of its two neighbors. The existence of $W(r,k)$ for any r and k is the celebrated theorem of van der Waerden (1927).

Theorem 25.1 (Van der Waerden 1927)**.** *For every choice of positive integers r and k, there exists a positive integer $n = W(r,k)$ such that for every coloring of the set of integers $\{1,\ldots,n\}$ in r colors at least one arithmetic progression with k terms will be monochromatic.*

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_25,

In Sect. 26.1 we will show how this theorem can be derived using a very powerful Ramsey-type result due to Hales and Jewett (1963).

Using much more involved techniques, Szemerédi (1975) obtained the following improvement of Van der Waerden's theorem.

Theorem 25.2 (Szemerédi 1975)**.** *For any $c > 0$ and $k \geq 3$, there is n_0 such that for any $n \geq n_0$ and any set $S \subseteq [n]$ of size $|S| \geq cn$, S contains an arithmetic progression of length k.*

It can be seen that this implies Van der Waerden's theorem, since we can set $c = 1/r$ and for any r-coloring of $[n]$, one color class contains at least cn elements. Szemerédi's theorem is a "density type" statement: *any* sufficiently large subset must contain a long arithmetic progression.

Using this fundamental result, Green and Tao (2008) were able to prove that, for every k, there exists a length-k arithmetic progression consisting entirely of prime numbers.

How fast does the number $W(r,k)$ in Van der Waerden's theorem grow? Easy probabilistic argument shows that the growth is exponential, even for $r = 2$.

Theorem 25.3. *$W(2,k) > 2^{k/2}$. That is, the set $\{1,\ldots,n\}$ may be two-colored so that no $2\log n$-term arithmetic progression is monochromatic.*

Proof. Color $\{1,\ldots,n\}$ randomly. That is, we flip a coin n times to determine a color of each point. For each arithmetic progression S with k terms, let A_S be the event that S is monochromatic. Then $\Pr[A_S] = 2 \cdot 2^{-|S|} = 2^{-k+1}$. There are no more than $\binom{n}{2}$ progressions (since each is uniquely determined by its first and second elements); so if $\binom{n}{2}2^{-k+1} < 1$, we have that $\Pr\left[\bigcup A_S\right] \leq \sum \Pr\left[A_S\right] < 1$, and the desired coloring exists. Therefore $W(2,k)$ must be larger than any n for which $\binom{n}{2}2^{-k+1} < 1$, e.g., for which $n^2 \leq 2^k$. □

Using the Lovász Local Lemma this lower bound can be improved to $W(2,k) > 2^k/(2\mathrm{e}k)$ (see Exercise 19.3). However, even this lower bound is very far from the best known upper bound due to Timothy Growers:

$$\log_2 \log_2 W(r,k) \leq r^{2^{2^{k+9}}} .$$

Ron Graham conjectures that $W(2,k) \leq 2^{k^2}$.

Let us ask a slightly different question: What is the smallest number $r = r_k(n)$ of colors needed to color $1,2,\ldots,n$ such that no length-k arithmetic progression is colored monochromatically? That is, $r = r_k(n)$ is the minimal number for which $W(r,k) > n$.

Determining the true rate of growth of $r_k(n)$ is a difficult problem. It is known that $r_k(n)$ is unbounded for fixed k, however, for $k \geq 4$ it can only be shown to grow extremely slowly.

We now relate $r_k(n)$ to another Ramsey-like function. Let $A_k(n)$ be the size $|S|$ of a largest subset $S \subseteq \{1,\ldots,n\}$ that contains no arithmetic progression of length k.

Theorem 25.4.
$$\frac{n}{A_k(n)} \leq r_k(n) \leq \frac{(4n+2)\ln n}{A_k(n)}\,.$$

Proof. Lower bound. Let $r = r_k(n)$. The set $\{1,\ldots,n\}$ can be r-colored in such a way that there are no length-k monochromatic arithmetic progressions. By the pigeonhole principle, some color must be used at least n/r times, and hence, some color class S has size $|S| \geq n/r$ and has no arithmetic progression of length k. Therefore, $A_k(n) \geq |S| \geq n/r$, implying that $r \geq n/A_k(n)$.

Upper bound. Assume that S is a subset of $[n] = \{1,\ldots,n\}$ that contains no arithmetic progressions of length k, and set $\ell := (4n+2)\ln n/|S|$. We demonstrate that the upper bound on $r_k(n)$ holds by proving that it is possible to color the set $[n]$ with $cn\ln n/|S|$ colors so that none of length-k monochromatic progressions are left monochromatic. Since the set S has no arithmetic progression of length k, it is enough to show that we can cover the set $\{1,\ldots,n\}$ by at most ℓ translates $S+t_1,\ldots,S+t_\ell$ of S; here, as before, $S+t = \{a+t \,:\, a \in S\}$. Having such a covering, we can define the coloring $\chi : [n] \to [\ell]$ by setting $\chi(x) = \min\{i \,:\, x \in S+t_i\}$. So, it is enough to prove the following

Claim 25.5. Let $S \subseteq [n]$. No more than $O(n\log n/|S|)$ translates of S are needed to cover $[n]$.

We use a probabilistic argument. Pick the number $\boldsymbol{t}$ in the interval $-n$ to n at random independently and uniformly; hence, each number t in this interval is picked with the same probability $p = 1/(2n+1)$. Note that a number $x \in [n]$ belongs to a translate $S+\boldsymbol{t}$ iff $\boldsymbol{t} = x-a$ for some $a \in S$. So, $\Pr[x \in S+\boldsymbol{t}] = p|S|$.

Now let $\boldsymbol{t}_1,\ldots,\boldsymbol{t}_\ell$ be ℓ independent copies of $\boldsymbol{t}$. The probability that a fixed number $x \in [n]$ is covered by none of the ℓ translates $S+\boldsymbol{t}_i$, $i=1,\ldots,\ell$ is $(1-p|S|)^\ell$. Since we have only n numbers x in $[n]$, the probability that some of the numbers remains uncovered is at most

$$n(1-p|S|)^\ell = n\left(1-\frac{|S|}{2n+1}\right)^\ell \leq \exp\left(\ln n - \frac{\ell|S|}{2n+1}\right) = \frac{1}{n} < 1\,. \qquad \square$$

In the case $k=3$ we have the following lower bound on $A_k(n)$. Recall that a length-3 arithmetic progression $a, a+d, a+2d$ is just a set of three distinct numbers $x=a$, $y=a+2d$ and $z=a+d$ such that $x+y=2z$.

Theorem 25.6 (Behrend 1949). *If n is a large enough, then $A_3(n) \geq ne^{-O(\sqrt{\ln n})}$.*

Proof. The proof relies on the geometrical observation that a straight line can intersect a sphere in $\mathbb{Z}^m$ in at most two points. In other words, the set $\{x \in \mathbb{Z}^m \,:\, \|x\| = r\}$ with $r > 0$ and $m \geq 1$ cannot contain an arithmetic progression of length three, that is, three vectors x, y, z in $\mathbb{Z}^m$ such that

$x+y=2z$. So, we only need to map this example back to $[n]=\{1,\ldots,n\}$. Let m, M be large integers which we shall determine later, and consider the spheres

$$S(r)=\{x\in[M]^m\ :\ x_1^2+\cdots+x_m^2=r^2\}\,.$$

Note that as r^2 ranges from m to mM^2, these sets cover the cube $[M]^m$, which is of cardinality M^m. By the pigeonhole principle, there must exist a radius $\sqrt{m}\le r_0\le\sqrt{m}M$ such that the sphere $S=S(r_0)$ in $[M]^m$ has cardinality

$$|S|\ge\frac{M^m}{mM^2-m}>\frac{M^{m-2}}{m}\,.$$

Now we map S to $\{1,\ldots,n\}$ via a mapping

$$P(x)=P(x_1,\ldots,x_m):=\frac{1}{2M}\sum_{i=1}^{m}x_i(2M)^i\,.$$

It is then not difficult to check that: (i) P is injective, (ii) $x+y=2z$ whenever $P(x)+P(y)=2P(z)$, and (iii) $\max_{x\in S}P(x)\le(2M)^m$. Therefore, if we set $M:=\lceil n^{1/m}/2\rceil$, it follows that the set $P(S)$ lies in $\{1,\ldots,n\}$ and contains no arithmetic progression of length three. Setting $m:=\sqrt{\log_2 n}$ we see that $P(S)$ has cardinality

$$|P(S)|=|S|\ge\frac{n^{1-2/m}}{m2^m}\ge n\exp(-c\sqrt{\ln n})\,. \qquad \square$$

Theorems 25.4 and 25.6 yield the following upper bound on $r_3(n)$.

Corollary 25.7. *At most* $\mathrm{e}^{O(\sqrt{\ln n})}$ *colors are enough to color* $\{1,2,\ldots,n\}$ *so that no length-*3 *arithmetic progression is colored monochromatically.*

25.2 Szemerédi's cube lemma

A collection C of integers is called an *affine d-cube* if there exist $d+1$ positive integers $x_0,x_1,\ldots,x_d$ so that

$$C=\left\{x_0+\sum_{i\in I}x_i\ :\ I\subseteq\{1,2,\ldots,d\}\right\}.$$

If an affine cube is generated by $x_0,x_1,\ldots,x_d$ then we write

$$C=C(x_0,x_1,\ldots,x_d).$$

For example, $C(1,1,1) = \{1,2,3\}$, $C(1,3,9) = \{1,4,10,13\}$ and $C(1,2,2,2) = \{1,3,5,7\}$. In particular, every arithmetic progression $a, a+b, a+2b, \ldots, a+db$ is an affine d-cube $C(a,b,b,\ldots,b)$.

The following result is a version of Van der Waerden's theorem for affine cubes.

Lemma 25.8. *For every $d, r \geq 1$ there exists an $n = S(d,r)$ with the following property. If we color the set $\{1,\ldots,n\}$ in r colors then all the elements of at least one affine d-cube lying in this set will receive the same color.*

Proof. We argue by induction on d. The case $d = 1$ is obvious, so assume that $n = S(r, d-1)$ exists and take $S(r,d) := r^n + n$.

Suppose $N \geq S(r,d)$ and $\{1,\ldots,N\}$ is colored in r colors. Consider the colors of the strings of n consecutive numbers

$$i, i+1, \ldots, i+n-1 \quad \text{for } 1 \leq i \leq r^n + 1.$$

We have $r^n + 1$ such strings of numbers but only r^n possible strings of their colors (we have only r colors in our disposal). By the pigeonhole principle, some two strings

$$\begin{array}{l} i,\ i+1,\ \ldots,\ i+n-1; \\ j,\ j+1,\ \ldots,\ j+n-1, \end{array}$$

with $i < j$, will receive the same sequence of colors. That is, for each x in $\{i, i+1, \ldots, i+n-1\}$, the numbers x and $x + (j-i)$ receive the same color.

By the choice of $n = S(r, d-1)$, the set $\{i, i+1, \ldots, i+n-1\}$ contains a monochromatic affine $(d-1)$-cube $C(x_0, x_1, \ldots, x_{d-1})$. But then all the numbers of the affine d-cube $C(x_0, x_1, \ldots, x_{d-1}, j-i)$ have the same color. Since $j - i \leq r^n$, this cube lies in $\{1,\ldots,N\}$, and we are done. □

The following important result of Szemerédi (1960) implies that in every coloring of $\{1,\ldots,n\}$ by $r \leq (4n)^{1/2^{d-1}}/4$ colors, at least one color class will contain an affine d-cube.

Lemma 25.9 (Szemerédi's Cube Lemma). *Let $d \geq 2$ be given. Then, for every sufficiently large n, every subset A of $\{1,\ldots,n\}$ of size*

$$|A| \geq (4n)^{1-1/2^{d-1}}$$

contains an affine d-cube.

Proof. We will iteratively use the following fact: if $B \subseteq \{1,\ldots,n\}$, $|B| \geq 2$, then there is a $i \geq 1$ such that the set $B_i := \{b \in B : b + i \in B\}$ has size

$$|B_i| > \frac{|B|^2}{4n}. \tag{25.1}$$

This immediately follows from the equality (see Exercise 25.9):

$$\sum_{i=1}^{n-1} |B_i| = \binom{|B|}{2}. \tag{25.2}$$

Applying this fact to the set A, we will find $i_1 \geq 1$ such that

$$|A_{i_1}| > \frac{|A|^2}{4n} \geq \frac{(4n)^{2-2/2^{d-1}}}{4n} = (4n)^{1-1/2^{d-2}}.$$

Similarly, applying the fact to the set A_{i_1}, we will find $i_2 \geq 1$ such that

$$|A_{i_1,i_2}| = |(A_{i_1})_{i_2}| > \frac{|A_{i_1}|^2}{4n} \geq \frac{(4n)^{2-2/2^{d-2}}}{4n} = (4n)^{1-1/2^{d-3}}.$$

Continuing this process, we will find $i_1, i_2, \ldots, i_{d-1}$ such that

$$|A_{i_1,i_2,\ldots,i_{d-1}}| > (4n)^{1-1/2^{d-d}} = 1.$$

Since this set still has at least 2 elements, we can apply the fact once more and conclude that the set $A_{i_1,i_2,\ldots,i_d}$ contains at least one element b_0. Observe that

$$\begin{aligned} A_{i_1} &= \{b : \ b \in A,\, b + i_1 \in A\}, \\ A_{i_1,i_2} &= \{b : \ b \in A,\, b + i_1 \in A,\, b + i_2 \in A,\, b + i_1 + i_2 \in A\} \end{aligned}$$

determines an affine 2-cube $C(b, i_1, i_1) \subseteq A$, and so on. Hence, the last set $A_{i_1,i_2,\ldots,i_d}$ determines an affine d-cube $C(b_0, i_1, \ldots, i_d)$, and this cube lies entirely in A. □

In the proof above no attempt was made to get the best constant. In particular, we were very generous when deriving (25.1) from (25.2). Using more precise estimate at this step, Gunderson and Rödl (1998) improved the bound to

$$|A| \geq 2^{1-1/2^{d-1}} (\sqrt{n} + 1)^{2-1/2^{d-1}}.$$

25.3 Sum-free sets

Recall that one of the earliest result in Ramsey theory—the Schur theorem (Theorem 4.14)—states that for any $r \geq 2$ there is $n > 3$ such that for any r-coloring of $\{1, 2, \ldots, n\}$, there are three integers of the same color and such that $x + y = z$.

In the wake of Schur's theorem, many people have studied so-called *sum-free sets*, i.e., subsets A of the positive integers such that $x, y \in A$ implies $x + y \notin A$. In particular, people have examined the question of how large a sum-free set can be.

This question has a natural generalization to arbitrary Abelian groups. An *Abelian group* is a nonempty set G together with an operation $(x, y) \mapsto x + y$, called addition, which is associative $(x+y)+z = x+(y+z)$ and commutative $x + y = y + x$. Moreover, there must be an element $0 \in G$ (called a zero) such that $x + 0 = x$ for all $x \in G$, and every $x \in G$ must have an inverse $-x$ such that $(-x) + x = 0$. Standard examples of Abelian groups are: the set $\mathbb{Z}$ of integers and the set $\mathbb{Z}_n$ of residues modulo n.

For a subset S of an Abelian group G, let $\alpha(S)$ denote the cardinality of the largest sum-free subset of S. The following upper bound is immediate (see Exercise 25.3): for any finite Abelian group G,

$$\alpha(G) \leq |G|/2\,.$$

If we take $G = \mathbb{Z}_n$ for an even n, then the set

$$A = \{1, 3, \ldots, n-1\}$$

is clearly sum-free, and hence, in this case $\alpha(G) = |G|/2$.

In the case of odd n we may derive a better upper bound using a beautiful theorem of Kneser about sums of finite subsets of an Abelian group. Here we will prove a special case of this result which still has many applications in additive number theory and the proof of which is particularly simple.

25.3.1 Kneser's theorem

If $A, B \subseteq G$ are subsets of an Abelian group G, then by $A + B$ we denote the set of all its elements of the form $a + b$ with $a \in A$ and $b \in B$. A *subgroup* of G is a subset $H \subseteq G$ which itself forms a group; it is *proper* if $H \neq G$.

The following theorem has many applications in additive number theory.

Theorem 25.10 (Kneser 1955)**.** *Let G be an Abelian group, $G \neq \{0\}$, and let A, B be nonempty finite subsets of G. If $|A| + |B| \leq |G|$, then there exists a proper subgroup H of G such that $|A + B| \geq |A| + |B| - |H|$.*

Proof. We proceed by induction on $|B|$. If $|B| = 1$, then

$$|A + B| = |A| = |A| + |B| - 1 \geq |A| + |B| - |H|$$

for every subgroup H.

Let $|B| > 1$, and suppose that the theorem holds for all pairs A', B' of finite nonempty subsets of G such that $|B'| < |B|$. We distinguish two cases.
Case 1: $a + b - c \in A$ for *all* $a \in A$ and $b, c \in B$.

In this case $A + b - c = A$ for all $b, c \in B$. Let H be the subgroup of G generated by all elements of the form $b - c$, where $b, c \in B$. Then $|B| \leq |H|$ and $A + H = A \neq G$. Therefore, H is a proper subgroup of G, and

$$|A + B| \geq |A| \geq |A| + |B| - |H|.$$

Case 2: $a+b-c \notin A$ for *some* $a \in A$ and $b,c \in B$.

Let $e := a-c$ and define the subsets

$$A' = A \cup (B+e),\ B' = B \cap (A-e).$$

Note that $b \notin B'$, and hence, B' is a proper subset of B, because otherwise b would have a form $x-a+c$ for some $x \in A$, and hence,

$$a+b-c = a+(x-a+c)-c = x \in A,$$

a contradiction. Also, $c \in B'$ (because $0 \in A-a$), and hence, B' is nonempty. Therefore, we can apply the induction hypothesis to A' and B', and deduce that there exists a proper subgroup H of G such that

$$|A'+B'| \geq |A'|+|B'|-|H|. \tag{25.3}$$

It remains to observe that

$$\begin{aligned} A'+B' &= [A\cup(B+e)]+[B\cap(A-e)] \\ &\subseteq (A+B)\cup[(B+e)+(A-e)] = A+B \end{aligned}$$

and

$$\begin{aligned} |A'|+|B'| &= |A\cup(B+e)|+|B\cap(A-e)| \\ &= |A\cup(B+e)|+|(B+e)\cap A| \\ &= |A|+|B+e| = |A|+|B|. \end{aligned}$$

□

The following fact is a special version of Kneser's theorem (we leave the proof as Exercise 25.7; an alternative proof is given in Sect. 16.3.4):

Theorem 25.11 (Cauchy–Davenport)**.** *If p is a prime, and A, B are two non-empty subsets of $\mathbb{Z}_p$, then $|A+B| \geq \min\{p,\ |A|+|B|-1\}$.*

Kneser's theorem immediately yields the following upper bound on the size of sum-free subset in Abelian groups.

Corollary 25.12. *Let G be a finite Abelian group and let p be the smallest prime divisor of $|G|$. Then $\alpha(G) \leq (p+1)|G|/(3p)$.*

Proof. If $A \subseteq G$ is a sum-free set, then $|A+A| \leq |G|-|A|$ because $A+A$ and A are disjoint. Since $|A| \leq |G|/2$ (see Exercise 25.3), we can apply Theorem 25.10 and deduce

$$|G|-|A| \geq |A+A| \geq 2|A|-|H|$$

for some proper subgroup H of G. Since, by Lagrange's theorem, the order $|H|$ of any subgroup H of G divides the order $|G|$ of the group G, we have $|H| \leq |G|/p$. Therefore,

$$3|A| \leq |G| + |H| \leq (1 + 1/p)|G|,$$

and the desired result follows. □

What about the lower bounds for $\alpha(G)$?

We have already seen that for $G = \mathbb{Z}_n$ with even n, we have an equality $\alpha(G) = |G|/2$. If $G = \mathbb{Z}$ is the group of integers, then

$$\alpha(S) > |S|/3$$

for any finite subset $S \subseteq \mathbb{Z} \setminus \{0\}$ (we have proved this fact in Sect. 18.2 using the probabilistic argument). For other Abelian groups the situation is not so clear.

The following naive argument shows that in this case also

$$\alpha(G) \geq \sqrt{|G|} - 1.$$

To show this, let A be a maximal sum-free subset. If $a \not\in A$, then $A \cup \{a\}$ is not sum-free by the assumption, so we can write $a = s_1 + s_2$ for some $s_1, s_2 \in A$. Therefore $|G \setminus A| \leq |A|^2$, from which the desired inequality $|A| \geq \sqrt{|G|} - 1$ follows.

Better lower bounds can be derived using an improvement of Theorem 25.10, also due to Kneser, which states that with the same hypotheses, either $|A + B| \geq |A| + |B|$ or $|A + B| \geq |A| + |B| - |H|$ for some proper subgroup H such that $H + A + B = A + H$ (see, for example, Street (1972) for the proof). Here we only mention that the best know lower bound for an arbitrary finite Abelian group G is $\alpha(G) \geq 2|G|/7$. More information about the properties of sum-free sets can be found, for example, in Nathanson (1996).

25.4 Sum-product sets

For every $A \subset \mathbb{R}$ we let $A+A = \{a+b : \ a, b \in A\}$ and $A \cdot A = \{ab : \ a, b \in A\}$. An old conjecture of Erdős states that, for every $\epsilon > 0$, every sufficiently large finite set $A \subset \mathbb{R}$ satisfies

$$\max\{|A + A|, |A \cdot A|\} \geq |A|^{2-\epsilon} .$$

That is, this conjecture asserts that every set of numbers A must have either a large sum-set $A + A$ or a large product set $A \cdot A$. The conjecture is central to our understanding of the interplay between the additive and multiplicative properties of a set of numbers.

Erdős and Szemerédi (1983) were the first to prove that there exists $\delta > 0$ so that $\max\{|A+A|, |A \cdot A|\} \geq |A|^{1+\delta}$ for all sufficiently large sets A. This parameter δ has been steadily improved by a number of authors. One highlight in this sequence is a proof by Elekes (1997) that δ may be taken arbitrarily close to 1/4. His argument utilizes a clever application of the Szemerédi–Trotter theorem on point-line incidences (see Theorem 18.7).

Recall that the Szemerédi–Trotter theorem asserts the following: If $2 \leq k \leq \sqrt{N}$ and if we take any set of N points in the plane, then it is not possible to draw more than $O(N^2/k^3)$ lines so that each of them contains at least k of the points.

Theorem 25.13 (Elekes 1997)**.** *There is an absolute constant $\epsilon > 0$ such that, for every set A of non-negative real numbers,*

$$\max\{|A+A|, |A \cdot A|\} \geq \epsilon |A|^{5/4}\,.$$

Proof. Let $n = |A|$, and consider the following n^2 straight lines

$$f_{a,b}(x) := a(x-b) = ax - ab \qquad \text{for } a, b \in A.$$

Observe that, for every $a, b \in A$, the function maps at least n elements $b + c$ with $c \in A$ to some elements $f_{a,b}(b+c) = a \cdot c$ of $A \cdot A$. From a geometric point of view, this means that the graph of each of these $m = n^2$ lines $f_{a,b}(x)$ contains $k = n$ or more points of $P := (A + A) \times (A \cdot A)$. By applying the Szemerédi–Trotter theorem to P with $k = n$ and $N = |P|$, we get $n^2 = O(|P|^2/n^3)$, that is

$$|A+A| \cdot |A \cdot A| = |P| = \Omega(n^{5/2})\,. \qquad \square$$

In the case of *finite* fields we have the following result.

Theorem 25.14 (Garaev 2007)**.** *Let p be a prime number, and $A \subseteq \mathbb{F}_p \setminus \{0\}$. Then the number of elements in at least one of the sets $A + A$ or $A \cdot A$ is at least an absolute constant times*

$$\min\left\{\sqrt{p|A|},\ \frac{|A|^2}{\sqrt{p}}\right\}.$$

In particular, if $|A| \approx p^{2/3}$ then this minimum is about $|A|^{5/4}$.

Proof (due to Solymosi 2009). His idea is a very clever application of the expander mixing lemma (Lemma 15.5). As in Sect. 15.2.1, consider a graph G whose vertices are $n = p(p-1)$ pairs (a, b) of elements of a finite field $\mathbb{Z}_p$ with $a \neq 0$, and two vertices (a, b) and (c, d) are joined by an edge iff $ac = b + d$ (all operations modulo p). We already know that this graph is $(p-1)$-regular and that the second largest eigenvalue λ of its incidence matrix is smaller than $\sqrt{3p}$ (see Lemma 15.6). So, if we define $S, T \subseteq V$ by

$$S = (A \cdot A) \times (-A) \quad \text{and} \quad T = (A^{-1}) \times (A + A)$$

then the expander mixing lemma (Lemma 15.2) tells us that

$$\begin{aligned} e(S,T) &\le \frac{(p-1)|S||T|}{p(p-1)} + \lambda\sqrt{|S||T|} = \frac{|S||T|}{p} + \lambda\sqrt{|S||T|} \\ &< \frac{|A \cdot A||A + A||A|^2}{p} + \sqrt{3p|A \cdot A||A + A||A|^2}\,, \end{aligned}$$

where the second inequality used $\lambda < \sqrt{3p}$. But for every $a, b, c \in A$ there is an edge between vertices $(ab, -c) \in S$ and $(b^{-1}, a+c) \in T$, so that $e(S,T) \ge |A|^3$. Thus, if we set $N := |A + A||A \cdot A|$, then rearranging the resulting inequality

$$\begin{aligned} |A|^3 \le e(S,T) &\le \frac{N|A|^2}{p} + |A|\sqrt{3pN} \\ &= \sqrt{N}\left(\frac{\sqrt{N}|A|^2}{p} + |A|\sqrt{3p}\right) \end{aligned}$$

gives

$$\sqrt{N} > \left(\frac{\sqrt{N}}{p|A|} + \frac{\sqrt{3p}}{|A|^2}\right)^{-1}.$$

Now, since $(x+y)^{-1} \ge \frac{1}{2}\min\{x^{-1}, y^{-1}\}$ for positive x and y, we find that

$$\sqrt{N} \ge \epsilon \cdot \min\left\{\frac{p|A|}{\sqrt{N}},\ \frac{|A|^2}{p^{1/2}}\right\}$$

with $\epsilon = 1/2\sqrt{3}$, which in turn implies

$$\sqrt{N} \ge \epsilon \cdot \min\left\{\sqrt{p|A|},\ \frac{|A|^2}{\sqrt{p}}\right\}.$$

To finish the proof, we need only use the two-term arithmetic-geometric mean inequality:

$$\max\{|A + A|, |A \cdot A|\} \ge \frac{|A \cdot A| + |A + A|}{2} \ge \sqrt{|A \cdot A||A + A|} = \sqrt{N}\,.$$

□

Exercises

25.1. Prove a lower bound for the general van der Waerden's function $W(r,k)$. *Hint*: Modify the proof of Theorem 25.3 to the case of more than two colors.

25.2. Prove the following general version of Schur's theorem. For every r and $l \geq 2$, there exists a positive integer n such that for every partition $A_1, \ldots, A_r$ of the set $\{1, \ldots, n\}$ into r classes one of the classes contains l (not necessarily distinct) numbers $x_1, \ldots, x_l$ such that $x_1 + \ldots + x_{l-1} = x_l$. *Hint*: Take $n = R_r(2; l)$ and assign every pair $\{x, y\}$ the color i if $|x - y| \in A_i$.

25.3. Show that for any finite Abelian group G, $\alpha(G) \leq |G|/2$. *Hint*: If S is a sum-free set then $S + S$ and S are disjoint.

25.4. Give a detailed proof of the last two statements in the proof of Kneser's theorem that $A' + B' \subseteq A + B$ and $|A'| + |B'| = |A| + |B|$.

25.5. Let G be a finite Abelian group, and let A and B be subsets of G such that $|A| + |B| > |G|$. Show that then $A + B = G$. *Hint*: For every $x \in G$, the set $A \cap (x - B)$ has at least $|A| + |B| - |G| \geq 1$ elements.

25.6. Let A and B be finite subsets of an Abelian group G. For $x \in G$, let $r(x)$ be the number of representations of x as the sum of elements from A and B, that is, $r(x)$ is the number of ordered pairs $(a, b) \in A \times B$ such that $x = a + b$. Prove the following: if $|A| + |B| \geq |G| + t$ then $r(x) \geq t$. *Hint*: Take an $x \in G$ and show that $|A \cap (x - B)| \geq |A| + |B| - |G| \geq t$.

25.7. Show that Kneser's theorem (Theorem 25.10) implies the Cauchy–Davenport theorem. *Hint*: For a prime p, the only proper subgroup of $\mathbb{Z}_p$ is the trivial group $H = \{0\}$.

25.8. If A be a nonempty subset of an Abelian group G, then its *stabilizer* is the set $H(A) := \{x \in G : x + A = A\}$. Show that A is a subgroup if and only if $H(A) = A$.

25.9. Prove (25.2). *Hint*: Consider a complete graph whose vertices are elements of B; label the edge, joining two elements $a < b$, by their difference $b - a$, and observe that $|B_i|$ is precisely the number of edges labeled by i.

25.10. Let A be a finite set of non-zero numbers. Show that then both $|A+A|$ and $|A \cdot A|$ are at least $2|A| - 1$. Give examples of arbitrary large sets A matching these bounds. *Hint*: For the second part, consider arithmetic progressions $a, 2a, 3a, \ldots, na$ and geometric progressions $ar, ar^2, ar^3, \ldots, ar^n$.

25.11. Show that the number of sum-free subsets of $\{1, \ldots, n\}$ is at least $c2^{n/2}$ for some constant $c > 0$. *Hint*: We can take any set of odd numbers, or any set of numbers greater than $n/2$.

25.12. (Cameron–Erdős 1999). A sum-free subset S of $X = \{1, \ldots, n\}$ is *maximal* if none of the sets $S \cup \{x\}$ with $x \in X \setminus S$ is sum-free. Show that the number of maximal sum-free subsets of X is at least $2^{\lfloor n/4 \rfloor}$.

Hint: Let m be either n or $n-1$, whichever is even. Let S consist of m together with one of each pair of numbers $x, m-x$ for odd $x < m/2$. Show that every such set S is sum-free, and distinct sets S lie in distinct maximal sum-free sets.

26. The Hales–Jewett Theorem

In 1963 A. W. Hales and R. I. Jewett proved a very general result from which most of Ramsey-like theorems may be easily derived. Hales–Jewett theorem is presently one of the most useful techniques in Ramsey theory. Without this result, *Ramsey theory* would more properly be called *Ramseyan theorems.*

26.1 The theorem and its consequences

Let $[t] = \{1, \ldots, t\}$. Points in the cube $[t]^n$ are strings $x = (x_1, \ldots, x_n)$ with all $x_i \in [t]$. A subset $L \subseteq [t]^n$ is a *line* (or *combinatorial line*) if there exists a non-empty subset $I = \{i_1, \ldots, i_k\} \subset [n]$ and numbers a_i for $i \notin I$ (the *fixed positions* of L) such that

$$L = \{x \in [t]^n \,:\, x_i = a_i \text{ for } i \notin I \text{ and } x_{i_1} = x_{i_2} = \ldots = x_{i_k}\}\,.$$

If we introduce a new symbol $*$ to denote the "moving coordinates," then each line is defined by its *root* $\tau = (\tau_1, \ldots, \tau_n)$ with $\tau_i = a_i$ for $i \notin I$ and $\tau_i = *$ for $i \in I$. If we define $\tau(a)$ to be the string τ with all $*$-positions set to a, then

$$L = \{\tau(1), \tau(2), \ldots, \tau(t)\}\,.$$

Here is an example of a line rooted in $\tau = (1, 3, *, 2, *, 1)$ with active set $I = \{3, 5\}$ in $[5]^6$ with fixed positions $a_1 = 1$, $a_2 = 3$, $a_4 = 2$ and $a_6 = 1$:

$$L = \left\{\begin{array}{lll} 1\;3\;\mathbf{1}\;2\;\mathbf{1}\;1 & \tau(1) & \text{first point of } L \\ 1\;3\;\mathbf{2}\;2\;\mathbf{2}\;1 & \tau(2) & \\ 1\;3\;\mathbf{3}\;2\;\mathbf{3}\;1 & \tau(3) & \\ 1\;3\;\mathbf{4}\;2\;\mathbf{4}\;1 & \tau(4) & \\ 1\;3\;\mathbf{5}\;2\;\mathbf{5}\;1 & \tau(5) & \text{last point of } L \end{array}\right.$$

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6_26,

Note that every line in $[t]^n$ consists of exactly t points, and we have $(t+1)^n - t^n$ lines in $[t]^n$ (every root defines its own line).

We are using the alphabet $[t] = \{1, \ldots, t\}$ just for definiteness: one may take an arbitrary alphabet A with $|A| = t$ symbols. Say, if $A = \{0, 1\}$ then a combinatorial line in A^n is just a pair of two binary strings, one of which can be obtained from the other by changing some 0s to 1s. Viewing binary strings as characteristic vectors of subsets of $[n]$, each combinatorial line in $\{0, 1\}^n$ corresponds to a pair of subsets S, T such that $S \subset T$.

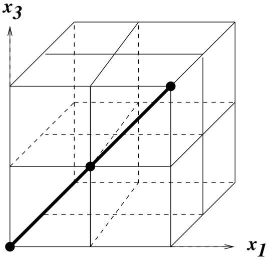


Fig. 26.1 A line $L = \{(0, 0, 0), (1, 0, 1), (2, 0, 2)\}$ rooted in $\tau = (*, 0, *)$

To match the classical parametric representation $x = a + \lambda b$ of a line in $\mathbb{R}^n$, observe that it corresponds to a combinatorial line rooted in τ, where $a_i = 0, b_i = 1$ if $\tau_i = *$, and $a_i = \tau_i, b_i = 0$ if $\tau_i \neq *$. Figure 26.1 shows a line $x = a + \lambda b$ with $a = (0, 0, 0)$ and $b = (1, 0, 1)$. Thus, the symbol $*$ indicates the moving coordinate. Note, however, that not every line in $\mathbb{R}^n$ is a combinatorial line.

Theorem 26.1 (Hales–Jewett 1963). *For every natural numbers t and r there exists a dimension $n = HJ(r, t)$ such that whenever $[t]^n$ is r-colored, there exists a monochromatic line.*

In fact, Hales and Jewett have proved this result for more general configurations, known as *combinatorial spaces*. Each combinatorial line has only one set of moving coordinates, determined by the occurrences of the special symbol $*$. A natural generalization is to allow several (disjoint) sets of moving coordinates.

A *combinatorial m-space $S_\tau \subseteq A^n$* is given by a word (a generalized root) $\tau \in (A \cup \{*_1, \ldots, *_m\})^n$, where $*_1, \ldots, *_m$ are distinct symbols not in the alphabet A. These symbols represent m mutually disjoint sets of moving coordinates. We require that each of these symbols occurs at least once in τ. Then S_τ is the set of all t^m words in A^n which can be obtained by simultaneously replacing each occurrence of these new symbols by symbols from A. Thus, combinatorial line is a just a combinatorial 1-space.

In the case of two-letter alphabet $A = \{0, 1\}$ there is a 1-1 correspondence between combinatorial spaces and subcubes of $\{0, 1\}^n$. For example, if $S_\tau \subseteq$

$\{0,1\}^6$ is a combinatorial 3-space given by $\tau = (1, *_1, 0, *_2, *_3, 1)$, then S_τ is exactly the set of all vectors in $\{0,1\}^6$ on which the monomial $x_1\overline{x}_3x_6$ takes value 1.

Theorem 26.2. *Let A be a finite alphabet of t symbols and let m, r be positive integers. Then there exists an integer $n = HJ(m, r, t)$ such that for every coloring of the cube A^n in r colors there exists a combinatorial m-space, which is monochromatic.*

This result can be derived from Theorem 26.1 (see Exercise 26.4). We will prove Theorem 26.1 itself in Sect. 26.2. Let us first show how Hales–Jewett theorem implies some classical results of Ramsey Theory.

26.1.1 Van der Waerden's theorem

Recall that an arithmetic progression of length t is a sequence of t natural numbers $a, a+d, a+2d, \ldots, a+(t-1)d$, each at the same distance $d \geq 1$ from the previous one. Thus, each arithmetic progression is a very regular configuration of numbers, just like cliques are in graphs.

In 1927 B.L. van der Waerden published a proof of the following Ramsey-type result for arithmetic progressions.

Theorem 26.3 (Van der Waerden 1927)**.** *For every choice of positive integers r and t, there exists a positive integer $N = W(r, t)$ such that for every coloring of the set of integers $\{1, \ldots, N\}$ in r colors at least one arithmetic progression with t terms will be monochromatic.*

Proof. Take $N := n(t-1)+1$ where $n = HJ(r, t)$ is from the Hales–Jewett theorem. Define a mapping $f : [t]^n \to \{1, \ldots, N\}$ which takes a word $x = (x_1, \ldots, x_n)$ to the sum $f(x) = x_1 + \ldots + x_n$ of its letters. The mapping f induces a coloring of $[t]^n$ in a natural manner: the color of a point $x \in [t]^n$ is the color of the number $f(x)$. It is not difficult to see that every combinatorial line $L_\tau = \{\tau(1), \tau(2), \ldots, \tau(t)\}$ is mapped by f to an arithmetic progression of length t: the difference between the integers corresponding to strings $\tau(i+1)$ and $\tau(i)$ is the same (and is equal to the number of $*$'s in τ). By the Hales–Jewett theorem, there is a monochromatic line that, in turn, translates back to a monochromatic arithmetical progression of length t, as desired. □

26.1.2 Gallai–Witt's Theorem

A multidimensional version of van der Waerden's theorem was proved independently by Gallai (= Grünwald), cf. Rado (1943), and Witt (1951).

A subset of vectors $U \subseteq \mathbb{Z}^m$ is a *homothetic copy* of a subset $V \subseteq \mathbb{Z}^m$ if there exists a vector $u \in \mathbb{Z}^m$ and a constant $\lambda \in \mathbb{Z}$, $\lambda > 0$ such that

$$U = u + \lambda V := \{u + \lambda v \ : \ v \in V\} .$$

Note that an arithmetic progression $a, a+b, a+2b, \ldots, a+kb$ in $\mathbb{Z}$ is a homothetic copy of $V = \{0, 1, \ldots, k\}$ with $u = a$ and $\lambda = b$.

Theorem 26.4 (Gallai–Witt). *Let the vectors of $\mathbb{Z}^m$ be finitely colored. Then every finite subset of $\mathbb{Z}^m$ has a homothetic copy which is monochromatic.*

Proof. Fix the number of colors r and a finite set of vectors $V = \{v_1, \ldots, v_t\}$ in $\mathbb{Z}^m$. We will consider these vectors as symbols of our alphabet $A = V$. Set $n = HJ(r,t)$ and consider the cube A^n; its elements are vectors $x = (x_1, \ldots, x_n)$, each of whose coordinates x_i is one of the vectors $v_1, \ldots, v_t$. As in the previous theorem, define a map $f : A^n \to \mathbb{Z}^m$ by $f(x) = x_1 + \ldots + x_n$. By the Hales–Jewett theorem, there is a monochromatic combinatorial line $L = \{\tau(v_1), \tau(v_2), \ldots, \tau(v_t)\} \subseteq A^n$. Let $I = \{i : \tau_i = *\}$ be the set of moving coordinates of L. Then $f(L)$ is the set of t vectors of the form $\lambda v_j + u$, $j = 1, \ldots, t$, where $\lambda = |I| > 0$ and $u = \sum_{i=1}^{t} a_i v_i$ with $a_i \in \mathbb{N}$ is the sum of fixed coordinates of the line L (the same vector v_i may appear several times in such coordinates). Hence, $f(L)$ is a homothetic copy of $V = \{v_1, \ldots, v_t\}$, as desired. □

26.2 Shelah's proof of HJT

Various proofs of Theorem 26.1 are known. The original proof of Hales and Jewett is relatively short but provides an upper bound for the function $HJ(r,t)$ which grows extremely fast. Shelah (1988) found a fundamentally new proof which yields a much smaller (in particular, a primitive recursive) upper bound for $HJ(r,t)$. Here we will follow the compact version of Shelah's proof from A. Nilli (1990) (c/o Noga Alon).

For each fixed number of colors r, we apply an induction on the number of symbols t in our alphabet A. For $t = 1$ the theorem is trivial. Assuming it holds for $t-1$ (and r) prove it for t. Set

$$n := HJ(r, t-1),$$

and define the following increasing sequence of dimensions $N_1, \ldots, N_n$ by

$$N_1 := r^{t^n} \quad \text{and} \quad N_i := r^{t^{n+\sum_{j=1}^{i-1} N_j}}.$$

Put $N := N_1 + \cdots + N_n$. We will prove that the dimension N has the desired property, i.e., that $HJ(r,t) \le N$. (This particular choice of the dimensions N_i will be important only in the proof of Claim 26.5 below).

To show this, let $A = \{0, 1, \ldots, t-1\}$ be an alphabet of t symbols, and let $\chi : A^N \to \{1, \ldots, r\}$ be a coloring of the N-cube A^N in r colors. Our goal is to prove that at least one combinatorial line will be monochromatic. The key to the whole proof is the following technical claim about the colors of neighboring words. We say that two words $a, b \in A^n$ are *neighbors* if they differ in exactly one coordinate, say, the i-th in which $a_i = 0$ and $b_i = 1$:

$$a = a_1 \dots a_{i-1}\, 0\, a_{i+1} \dots a_n$$
$$b = a_1 \dots a_{i-1}\, 1\, a_{i+1} \dots a_n$$

For a word $a = a_1 a_2 \dots a_n$ over A of length n and a sequence of n roots

$$\tau = \tau_1 \tau_2 \dots \tau_n,$$

the i-th of which has length N_i, let $\tau(a)$ denote the corresponding word of length N:

$$\tau(a) = \tau_1(a_1)\tau_2(a_2)\dots\tau_n(a_n).$$

That is, we replace each occurrence of $*$ in τ_1 by a_1, each occurrence of $*$ in τ_2 by a_2, and so on.

Claim 26.5. There exists a sequence of n roots $\tau = \tau_1\tau_2 \dots \tau_n$ as above, such that $\chi\big(\tau(a)\big) = \chi\big(\tau(b)\big)$ for any two neighbors $a, b \in A^n$.

Before proving the claim, let us look at how it implies the theorem. Using the coloring χ of the N-cube A^N, we define a coloring χ' of the n-cube $(A \setminus \{0\})^n$ by:

$$\chi'(a) := \chi\big(\tau(a)\big)\,,$$

where τ is from the claim. Since the alphabet $A \setminus \{0\}$ has only $t-1$ symbols and $n = HJ(r, t-1)$, we can apply the induction to this coloring χ'. By the induction hypothesis there exists a root

$$\nu = \nu_1\nu_2 \dots \nu_n \in \big((A \setminus \{0\}) \cup \{*\}\big)^n$$

such that the combinatorial line

$$L_\nu = \{\nu(1), \nu(2), \dots, \nu(t-1)\}$$

is monochromatic (with respect to χ'). Consider the string

$$\tau(\nu) = \tau_1(\nu_1)\tau_2(\nu_2)\dots\tau_n(\nu_n).$$

This string has length N and is a root since ν is a root (and hence, has at least one $*$). We claim that the corresponding line

$$L_{\tau(\nu)} = \big\{\tau\big(\nu(0)\big), \tau\big(\nu(1)\big), \dots, \tau\big(\nu(t-1)\big)\big\}$$

is monochromatic with respect to the original coloring χ. Indeed, the coloring χ' assigns the same color to all the words $\nu(1), \dots, \nu(t-1)$. Hence, by the definition of χ', the coloring χ assigns the same color to all the words $\tau\big(\nu(1)\big), \dots, \tau\big(\nu(t-1)\big)$. If ν contains only one $*$, then $\tau\big(\nu(0)\big)$ is a neighbor of $\tau\big(\nu(1)\big)$ and, by the claim, receives the same color (under χ). If ν has more $*$'s, then we can still reach the word $\tau\big(\nu(0)\big)$ from the word $\tau\big(\nu(1)\big)$ by passing through a sequence of neighbors

$$\begin{aligned}\tau\big(\nu(1)\big) &= \ldots 1 \ldots 1 \ldots 1 \ldots\\ &\ \ldots 0 \ldots 1 \ldots 1 \ldots\\ &\ \ldots 0 \ldots 0 \ldots 1 \ldots\\ \tau\big(\nu(0)\big) &= \ldots 0 \ldots 0 \ldots 0 \ldots\end{aligned}$$

and, by the claim, the word $\tau\big(\nu(0)\big)$ will receive the color of $\tau\big(\nu(1)\big)$. So, the whole line $L_{\tau(\nu)}$ is monochromatic, as desired.

It remains to prove Claim 26.5.

Recall that we want to find a sequence of n roots $\tau = \tau_1\tau_2\ldots\tau_n$, the i-th of which has length N_i, and such that $\chi(\tau(a)) = \chi(\tau(b))$ for any two neighbors $a, b \in A^n$.

We prove the existence of required roots τ_i by backward induction on i. Suppose we have already defined the roots $\tau_{i+1}, \ldots, \tau_n$. Our goal is to define the root τ_i.

Let $L_{i-1} := \sum_{j=1}^{i-1} N_j$ be the length of the initial segment $\tau_1\tau_2\ldots\tau_{i-1}$ of the sequence of roots we are looking for. The length of the i-th segment τ_i is N_i. For $k = 0, 1, \ldots, N_i$, let W_k denote the following word of length N_i:

$$W_k = \underbrace{0\ldots0}_{k}\,\underbrace{1\ldots1}_{N_i-k}.$$

For each $k = 0, 1, \ldots, N_i$, define the r-coloring χ_k of all words in $A^{L_{i-1}+n-i}$ as follows: let $\chi_k\left(x_1\,x_2\,\ldots\,x_{L_{i-1}}\,y_{i+1}\,\ldots\,y_n\right)$ be equal to

$$\chi\left(x_1\,x_2\,\ldots\,x_{L_{i-1}}\,W_k\,\tau_{i+1}(y_{i+1})\,\ldots\,\tau_n(y_n)\right).$$

We have $N_i + 1$ colorings $\chi_0, \chi_1, \ldots, \chi_{N_i}$, each being chosen from the set of at most

$$r^{\#\{\text{of words}\}} = r^{t^{L_{i-1}+n-i}} \le r^{t^{L_{i-1}+n}} = N_i$$

such colorings. By the pigeonhole principle, at least two of these colorings must coincide, i.e., $\chi_s = \chi_k$ for some $s < k$. Now define the desired root τ_i by

$$\tau_i := \underbrace{0\ldots0}_{s}\,\underbrace{*\ldots*}_{k-s}\,\underbrace{1\ldots1}_{N_i-k}.$$

One can easily check that the roots $\tau_1, \ldots, \tau_n$ defined by this procedure satisfy the assertion of the claim. Indeed, observe that $\tau_i(0) = W_k$ and $\tau_i(1) = W_s$. Hence, if we take any two neighbors in the i-th coordinate

$$\begin{aligned}a &= a_1\ldots a_{i-1}\,0\,a_{i+1}\ldots a_n\\ b &= a_1\ldots a_{i-1}\,1\,a_{i+1}\ldots a_n\end{aligned}$$

then

$$\tau(a) = \tau_1(a_1)\dots\tau_{i-1}(a_{i-1})\,\tau_i(0)\,\tau_{i+1}(a_{i+1})\dots\tau_n(a_n),$$
$$\tau(b) = \tau_1(a_1)\dots\tau_{i-1}(a_{i-1})\,\tau_i(1)\,\tau_{i+1}(a_{i+1})\dots\tau_n(a_n),$$

and since $\chi_s = \chi_k$,

$$\begin{aligned}\chi\big(\tau(a)\big) &= \chi\big(\tau_1(a_1)\,\dots\,\tau_{i-1}(a_{i-1})\,W_k\,\tau_{i+1}(a_{i+1})\,\dots\,\tau_n(a_n)\big)\\ &= \chi_k\big(\tau_1(a_1)\,\dots\,\tau_{i-1}(a_{i-1})\,a_{i+1}\,\dots\,a_n\big)\\ &= \chi_s\big(\tau_1(a_1)\,\dots\,\tau_{i-1}(a_{i-1})\,a_{i+1}\,\dots\,a_n\big)\\ &= \chi\big(\tau_1(a_1)\,\dots\,\tau_{i-1}(a_{i-1})\,W_s\,\tau_{i+1}(a_{i+1})\,\dots\,\tau_n(a_n)\big) = \chi\big(\tau(b)\big).\end{aligned}$$

This completes the proof of the claim, and thus, the proof of the Hales–Jewett theorem.

Exercises

26.1. Show that $HJ(r,2) \leq r$. That is, for every coloring of the cube $\{0,1\}^r$ in r colors at least one combinatorial line is monochromatic. *Hint*: Consider the words 0^i1^{r-i} for $i = 0,1,\dots,r$.

26.2. Let $N = W(2, t^2+1)$, where $W(r,t)$ is the van der Waerden's function, and let χ be a coloring of $\{1,\dots,N\}$ in two colors. Show that there exists a t-term arithmetic progression $\{a + i\cdot d \,:\, i = 0,1,\dots,t-1\}$ which together with its difference d is monochromatic, i.e., $\chi(d) = \chi(a + i\cdot d)$ for every $i < t$. *Hint*: Van der Waerden's theorem gives a monochromatic arithmetical progression $\{a + j\cdot d \,:\, j \leq t^2\}$ with t^2 terms. Then either some $j\cdot d$, with $1 \leq j \leq t$, gets the same color or all the numbers $d, 2d, \dots, td$ get the opposite color.

26.3. Say that a family $\mathcal{A} \subseteq 2^X$ of subsets of a finite set X is *r-regular* in X if for every coloring $\chi : X \to [r] = \{1,\dots,r\}$ of underlying elements in r colors, at least one member $A \in \mathcal{A}$ must be monochromatic, that is, $\chi(x) = c$ for some color $c \in [r]$ and all elements $x \in A$. For two families $\mathcal{A} \subseteq 2^X$ and $\mathcal{B} \subseteq 2^Y$, let $\mathcal{A}\otimes\mathcal{B}$ be the collection of all sets $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Prove the following:

> If $\mathcal{A}$ is r-regular in X, and $\mathcal{B}$ is $r^{|X|}$-regular in Y, then $\mathcal{A}\otimes\mathcal{B}$ is r-regular in $X \times Y$.

26.4. Use Theorem 26.1 to derive the Hales–Jewett theorem for combinatorial m-spaces (Theorem 26.2). *Hint*: Consider a new alphabet B of size $|B| = t^m$ whose symbols are all possible strings in A^m. If $n := HJ(r, t^m)$ then every r-coloring of B^n gives a combinatorial line L over the alphabet B. Argue that this line in B^n is a combinatorial m-space in A^{mn}.

27. Applications in Communication Complexity

Communication complexity is a basic part of the theory of computational complexity. We have k players who wish to collaboratively evaluate a given function $f(x_1, \ldots, x_n)$. The players have unlimited computational power but none of them has access to all inputs $x_1, \ldots, x_n$: each player can only see a part of them. The function f itself is known to all players. The players communicate by sending some bits of information about the inputs they can see. The communication complexity of f is then the minimal number of bits communicated on the worst-case input.

The case of two players ($k = 2$) is relatively well understood. In this case two players, Alice and Bob, wish to compute $f(x_1, x_2)$. Alice can only see x_1, and Bob only x_2. There is a rich literature concerning this two-party communication model—see the book by Kushilevitz and Nisan (1997) for an excellent survey. The case of three and more players is much less understood. The twist is that in this case the players share some inputs, and (at least potentially) can use this overlap to encode the information in some clever and non-trivial way. In this chapter we are mainly interested in this case of more than two players.

27.1 Multi-party communication

A general framework for multi-party communication complexity is as follows. Let $X = X_1 \times X_2 \times \cdots \times X_k$ be a Cartesian product of k n-element sets. There are k players $P_1, \ldots, P_k$ who wish to collaboratively evaluate a given function $f : X \to \mathbb{R}$ on every input $x \in X$. Each player has unlimited computational power and full knowledge of the function. However, each player has only partial information about the input $x = (x_1, \ldots, x_k)$: the i-th player P_i has access to all the x_j's *except* x_i. We can imagine the situation as k poker players sitting around the table, and each one is holding a number to his/her forehead for the others to see. Thus, all players know the function f but their access

An EATCS Series, DOI 10.1007/978-3-642-17364-6_27,

to the input vector is restricted: the first player sees the string $(*, x_2, \dots, x_k)$, the second sees $(x_1, *, x_3, \dots, x_k)$, ..., the k-th player sees $(x_1, \dots, x_{k-1}, *)$.

Players can communicate by writing bits 0 and 1 on a blackboard. The blackboard is seen by all players. The game starts with the blackboard empty. For each string on the blackboard, the protocol either gives the value of the output (in that case the protocol is over), or specifies which player writes the next bit and what that bit should be as a function of the inputs this player knows (and the string on the board). During the computation on one input the blackboard is never erased, players simply append their messages. The objective is to compute the function with as small an amount of communication as possible.

The *communication complexity* of a k-party game for f is the minimal number $C_k(f)$ such that on every input $x \in X$ the players can decide whether $f(x) = 1$ or not, by writing at most $C_k(f)$ bits on the blackboard. Put otherwise, $C_k(f)$ is the minimal number of bits written on the blackboard on the worst-case input.

It is clear that $C_k(f) \le \log_2 n + 1$ for any $f : X \to \{0, 1\}$: the first player writes the binary code of x_2, and the second player announces the result. But what about the lower bounds? The twist is that (for $k \ge 3$) the players share some inputs, and (at least potentially) can use this overlap to encode the information in some clever and non-trivial way (see Exercises 27.8, 27.9).

Still, we know that the access of each player is restricted: the i-th player cannot distinguish inputs differing only in the i-th coordinate. This leads to the following concept.

A *star* around a vector $x \in X$ is a set $S = \{x^1, \dots, x^k\}$ of k vectors in X, where x^i differs from x in exactly the i-th component. The vector x is a *center* of this star, and is not a part of the star!

Proposition 27.1. *If a k-party communication protocol gives the same answer on all k points of some star, then the protocol must give that same answer on its center.*

Proof. Take an arbitrary k-party communication protocol, and let $S = \{x^1, \dots, x^k\}$ be a star around some vector x. Assume that the protocol gives the same answer on all points of S. An important fact is that given the first l bits communicated by the players, the $(l+1)$-th bit communicated (transmitted, say, by the i-th player) must be defined by a function which does not depend on the i-th coordinate of the input: the i-th player cannot see it. Therefore, for every l, there is an i $(1 \le i \le k)$ such that the $(l+1)$-th communicated bit is the same for both inputs x and x^i. Since on all inputs $x^1, \dots, x^k$ the players behave in the same way (i.e., write the same string on the blackboard), it follows that they will also behave in the same way on the input x. □

Using Proposition 27.1 we can express the communication complexity in purely combinatorial terms.

A coloring $c : X \to \{1, \ldots, r\}$ of X is *legal* if it does not separate a star from its center: For every star S around some vector x, if all k points of S receive the same color, then x must also receive that color. In particular, a coloring is legal if it leaves no star monochromatic. A coloring *respects* a given function $f : X \to \mathbb{R}$, if $f(x) \neq f(y)$ implies $c(x) \neq c(y)$, that is, the function f must be constant in each color-class.

Define the chromatic number $\chi_k(f)$ of $f : X \to \mathbb{R}$ to be the minimum number of colors in a legal coloring of X respecting f.

Example 27.2. If we have only $k = 2$ players, then a function $f : X_1 \times X_2 \to \{0, 1\}$ can be viewed as a 0-1 matrix, and $\chi_2(f)$ in this case is exactly the smallest number of mutually disjoint monochromatic submatrices covering the whole matrix.

Proposition 27.3. *$C_k(f) \geq \log_2 \chi_k(f)$.*

Proof. Take an optimal protocol for the communication game for f. Color each vector $x \in X$ by the string which is written on the blackboard at the end of communication between the players on the input x. Since the protocol computes f, the coloring must respect f. We have $2^{C_k(f)}$ colors and, by Proposition 27.1, the coloring is legal. □

27.2 The hyperplane problem

To illustrate how the connection between communication complexity and colorings works in concrete situations, let us consider the *k-dimensional hyperplane problem.* Inputs to this problem are vectors $x = (x_1, \ldots, x_n)$ of integers $x_i \in [n]$. Given such an input, the players must decide whether $x_1 + \cdots + x_n = n$. That is, they must compute the function $h : [n]^k \to \{0, 1\}$ such that $h(x) = 1$ if and only if x belongs to the hyperplane

$$H = \{x \in [n]^k \ : \ x_1 + \cdots + x_k = n\} .$$

For this special function h the lower bound given by Proposition 27.3 is almost optimal (see Exercise 27.6 for a more general result).

Let r_H be the minimal number of colors needed to color the hyperplane H so that no star in H remains monochromatic.

Proposition 27.4. *$C_k(h) \leq k + \log_2 r_H$.*

Proof. Assume that all k players agree in advance on a coloring of H with r colors such that no star $S \subseteq H$ remains monochromatic. Given an input vector $x = (x_1, \ldots, x_k)$, x_i is the only component that the i-th player does not know. Let x^i denote the only vector in H that is consistent with the information available to the player i, that is, $x^i = (x_1, \ldots, x_{i-1}, x_i^*, x_{i+1}, \ldots, x_k)$ where $x_i^* := n - \sum_{j \neq i} x_j$. (If $x_i^* < 1$, then x^i is undefined.)

The protocol now works as follows. If, for some i, the i-th coordinate x_i^* does not belong to $[n]$, the i-th player immediately announces the result: $h(x) = 0$. Otherwise they proceed as follows.

Using $\log_2 r$ bits, the first player broadcasts the color of x^1. Then the i-th player ($i = 2, \ldots, k$) transmits a 1 if and only if the color of x^i matches the color of x^1. The process halts and accepts the input x if and only if all players agree, that is, broadcast 1s in the last phase.

If the actual input x is in H, then $x^i = x$ for all i, and all players will agree. Otherwise, the vectors $x^1, \ldots, x^k$ form a star (around $x \notin H$) lying entirely in the hyperplane H, and hence, cannot receive the same color. Thus, in this case the players correctly reject that input. □

We can upper bound r_H by a Ramsey-type function we have already considered in Sect. 25.1. Recall that $r_k(N)$ is the minimum number of colors needed to color $1, 2, \ldots, N$ so that no length-k arithmetic progression is colored monochromatically.

Proposition 27.5. *For $N = k^2 n$, we have $r_H \leq r_k(N)$.*

Proof. Define a mapping f from the hyperplane H to $\{1, \ldots, N\}$ by

$$f(x_1, x_2, \ldots, x_k) := x_1 + 2x_2 + \cdots + kx_k.$$

Color $1, \ldots, N$ with $r = r_k(N)$ colors, avoiding monochromatic length-k arithmetic progressions. Color each point $x \in H$ with the color of $f(x)$. We prove by contradiction that this coloring leaves no star in H monochromatic, implying that $r_H \leq r$, as desired.

Assume $x^1, \ldots, x^k$ is a monochromatic star in H around some vector y. By the definition, $x^i = y + \lambda_i e_i$ for some $\lambda_i \neq 0$ ($i = 1, \ldots, k$). Since each x^i is in the hyperplane H, it follows that $\lambda_1 = \lambda_2 = \ldots = \lambda_k = \lambda$. Consider the points $f(x^1), f(x^2), \ldots, f(x^k)$. The map f is linear, so $f(x^i) = f(y) + \lambda f(e_i)$. By the definition of f, $f(e_i) = i$; hence, the numbers $f(x^i) = f(y) + \lambda \cdot i$ ($i = 1, \ldots, k$) form a monochromatic arithmetic progression of length k, which is a contradiction. □

It is not difficult to show that the two-dimensional hyperplane problem cannot be solved by communicating fewer than $\Omega(\log n)$ bits (see Exercise 27.3). Interestingly, already three players can do the job much better!

We already know (see Corollary 25.7) that $r_3(N) \leq 2^{O(\sqrt{\ln N})}$. Together with Propositions 27.4 and 27.5, this gives the following surprising upper bound on the communication complexity of the three-dimensional hyperplane problem.

Theorem 27.6. *For any triple of numbers in $[n]$, three players can decide whether these numbers sum up to n by communicating only $O(\sqrt{\log n})$ bits.*

We now will use Ramsey-type arguments to show that the communication complexity of the k-dimensional hyperplane problem is non-constant, for any $k \geq 2$.

Theorem 27.7 (Chandra–Furst–Lipton 1983). *For any fixed $k \geq 2$, the communication complexity of the k-dimensional hyperplane problem goes to infinity as n goes to infinity.*

Proof. To get a contradiction, suppose there exists a constant r such that, for any n, there is a legal coloring $c : [n]^k \to [r]$ of the grid $[n]^k$ with r colors that respects the function h. Define the projection p from H to $[n]^{k-1}$ by

$$p(x_1, \ldots, x_k) = (x_1, \ldots, x_{k-1}) .$$

The mapping p is an injection. So c and p induce, in a natural way, an r-coloring of the points in $p(H)$. Let $m = \lfloor n/k \rfloor$, and consider the grid $[m]^{k-1}$. Since $(k-1)m \leq n$, this grid is a subset of $p(H)$ and so is also r-colored via p.

Take the set of k vectors $V = \{0, e_1, \ldots, e_{k-1}\}$, where e_i is the unit vector $e_i = (0, \ldots, 1, 0, \ldots, 0)$ with 1 in the i-th coordinate. If n (and hence, also m) is large enough, the Gallai–Witt theorem (Theorem 26.4) implies the existence of a homothetic copy $u + \lambda V = \{u, u + \lambda e_1, \ldots, x + \lambda e_{k-1}\}$ in $[m]^k$ with $\lambda > 0$, which is monochromatic. Consider the vector

$$y := (u_1, \ldots, u_{k-1}, n - s - \lambda),$$

where $s = u_1 + u_2 + \cdots + u_{k-1}$. Since $0 < s \leq (k-1)m$ and $0 < \lambda \leq m$, the vector y belongs to $[n]^k$. We now have a contradiction since the vectors

$$\begin{aligned} p^{-1}(u) &= (u_1, u_2, \ldots, u_{k-1}, n - s) \\ p^{-1}(u + \lambda e_1) &= (u_1 + \lambda, u_2, \ldots, u_{k-1}, n - s - \lambda) \\ &\vdots \\ p^{-1}(u + \lambda e_{k-1}) &= (u_1, u_2, \ldots, u_{k-1} + \lambda, n - s - \lambda) \end{aligned}$$

belong to the hyperplane H and form a monochromatic (under the original coloring c) star around the vector y. Since the coloring is legal, the center y of this star must also receive the same color. But the vector y does not belong to H because the sum of its components is $n - \lambda < n$. Thus, the coloring c does not respect the function h, a contradiction. □

27.3 The partition problem

The *partition function* is a boolean function $Part_{n,k}$ in nk variables arranged into an $n \times k$ matrix, and $Part_{n,k}(A) = 1$ iff each row of A contains exactly one 1. That is, if we think of the j-th column of A as representing a subset S_j of $[n]$, then $Part_{n,k}$ accepts a sequence $(S_1, \ldots, S_k)$ of subsets of $[n]$ iff these subsets form a partition of $[n]$.

We are interested in the k-party communication complexity of this function in the case when the j-th player can see the entire matrix A, except for its j-th column.

The case of two players ($k = 2$) is easy to analyze: in this case $\Omega(\log n)$ bits of communication are necessary (see Exercise 27.4).

Theorem 27.8 (Tesson 2003). *For every $k \geq 2$, the communication complexity of any k-party game for $Part_{n,k}$ is greater than any constant.*

Proof. Consider the input as a collection $(S_1, \ldots, S_k)$ of subsets of $[n]$. Every such input that is accepted by a communication protocol for $Part_{n,k}$ is such that, for every $i \in [n]$, the element i lies in exactly one of the S_i. We can therefore put these inputs into a one-to-one correspondence with n-tuples in $[k]^n$ by:

$$(S_1, \ldots, S_k) \mapsto (x_1, \ldots, x_n) \qquad \text{with } x_i = j \text{ iff } i \in S_j.$$

As an example for $k = 3$ and $n = 4$, an accepted input $(\{4\}, \{1, 3\}, \{2\})$ corresponds to the n-tuple $(2, 3, 2, 1)$.

Suppose that the k-party communication complexity of $Part_{n,k}$ is bounded by some constant c. To every input accepted by a protocol we assign one of 2^c colors corresponding to the communication history on this input. If n is large enough (it suffices to take $n = HJ(2^c, k)$), then the Hales–Jewett theorem (Theorem 26.1) implies that there must be a monochromatic combinatorial line $L = \{a^1, \ldots, a^k\}$ rooted in some string $\tau \in [k] \cup \{*\}$.

Let $T := \{i : \tau_i = *\}$; hence, $T \neq \emptyset$. Let S_j be the set of positions on which all points of L have value j. For example, if $L \subseteq [5]^6$ is a line rooted in $\tau = (1, 3, *, 2, *, 1)$,

$$L = \left\{ \begin{matrix} 1\,3\,\mathbf{1}\,2\,\mathbf{1}\,1 \\ 1\,3\,\mathbf{2}\,2\,\mathbf{2}\,1 \\ 1\,3\,\mathbf{3}\,2\,\mathbf{3}\,1 \\ 1\,3\,\mathbf{4}\,2\,\mathbf{4}\,1 \\ 1\,3\,\mathbf{5}\,2\,\mathbf{5}\,1 \end{matrix} \right\},$$

then $T = \{3, 5\}$, $S_1 = \{1, 6\}$, $S_2 = \{4\}$, $S_3 = \{2\}$ and $S_4 = S_5 = \emptyset$.

By definition of the above one-to-one correspondence, we have that the sets $T, S_1, \ldots, S_k$ form a partition of $[n]$, and all the inputs

$$(S_1 \cup T, S_2, \ldots, S_k), (S_1, S_2 \cup T, \ldots, S_k), \ldots, (S_1, S_2, \ldots, S_k \cup T)$$

induce the same communication history. But these k inputs form a (combinatorial) star around $(S_1, S_2, \ldots, S_k)$ and, by Proposition 27.1, must also be accepted by the protocol. However, $S_1 \cup S_2 \cup \cdots \cup S_k = [n] \setminus T \neq [n]$ does not cover the whole set $[n]$, so we get a contradiction: the protocol (wrongly) accepts an input which should be rejected. □

27.4 Lower bounds via discrepancy

As we have seen, Ramsey-type results can yield unexpectedly efficient communication protocols. However, the lower bounds on communication complexity obtained via these arguments, are rather weak. The highest known lower bounds were obtained using probabilistic arguments.

Recall that a coloring $c : X \to \{1, \ldots, r\}$ is legal if for every star S with center x either: (i) S is not monochromatic or (ii) all points of S receive the same color, but then the center x also receives that color.

The next proposition describes a combinatorial structure of color classes. Namely, each color class of a legal coloring must be a "cylinder intersection," a notion we already considered in Sect. 18.10. This notion generalizes that of a "submatrix" (see Example 18.16).

Recall that a subset $T_i \subseteq X$ is called a *cylinder* in the i-th dimension if membership in T_i does not depend on the i-th coordinate:

$$(x_1, \ldots, x_i, \ldots, x_k) \in T_i \text{ implies that } (x_1, \ldots, x_i', \ldots, x_k) \in T_i \text{ for all } x_i' \in X_i.$$

A subset $T \subseteq X$ is a *cylinder intersection* if it is an intersection $T = \bigcap_{i=1}^k T_i$, where T_i is a cylinder in the i-th dimension.

Proposition 27.9. *A set $T \subseteq X$ is a cylinder intersection if and only if, for every star $S \subseteq X$ around a vector $x \in X$, $S \subseteq T$ implies $x \in T$.*

Proof. The "only if" direction ($\Rightarrow$) is simple. Let $T = \bigcap_{i=1}^k T_i$ where T_i is a cylinder in the i-th dimension. If $S = \{x^1, \ldots, x^k\}$ is a star around some vector $x \in X$, and if $S \subseteq T$, then $x^i \in T \subseteq T_i$ and hence $x \in T_i$ for all $i = 1, \ldots, k$, implying that $x \in T$, as desired.

For the "if" direction ($\Leftarrow$), take an arbitrary subset $T \subseteq X$ and assume that T contains the center of every star it contains. For every $i = 1, \ldots, k$, let T_i be the set of all strings $x \in X$ such that x coincides with at least one string $x^i \in T$ in all but perhaps the i-th coordinate. By its definition, the set T_i is a cylinder in the i-th dimension. Hence, the set $T' = \bigcap_{i=1}^k T_i$ is a cylinder intersection. If a vector x belongs to T, then it also belongs to all the T_i, by their definition. This shows $T \subseteq T'$. To show that $T' \subseteq T$, take a vector $x \in T'$, and assume that $x \notin T$. But then $x \in T_i$ implies that there must be a vector $x^i \in T$ from which x differs in exactly the i-th coordinate. The vectors $x^1, \ldots, x^k$ form a star around x and are contained in T. Hence, vector x must belong to T as well. □

By Proposition 27.9, in every legal coloring of X, each color class must be a cylinder intersection. Together with Proposition 27.4, this implies that every

c-bit communication protocol for f gives us a partition of X into at most 2^c f-monochromatic cylinder intersections. The next question is: how can one show that any such partition must contain many cylinder intersections?

Recall (from Sect. 18.10) that the (normalized) *discrepancy* of a function $f : X \to \{-1, 1\}$ on a set T is defined by:

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

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

The following lemma allows one to prove *lower* bounds on multiparty communication complexity by proving *upper* bounds on the discrepancy.

Lemma 27.10. *For every $f : X \to \{-1, 1\}$, we have*

$$C_k(f) \geq \log_2 \frac{1}{\mathrm{disc}(f)} .$$

Proof. By Proposition 27.3, it is enough to show that, if a coloring $c : X \to \{1, \ldots, r\}$ is legal and respects the function f, then it must use $r \geq 1/\mathrm{disc}(f)$ colors. By Proposition 27.9, each color class $T = c^{-1}(i)$ must be a cylinder intersection. Since the coloring respects f, the function f must take the same value on all vectors in T, implying that $\mathrm{disc}_T(f) = |T|/|X|$. Thus, no color class T can have more than $|X|\cdot\mathrm{disc}_T(f) \leq |X|\cdot\mathrm{disc}(f)$ vectors. Since all $|X|$ vectors must be colored, at least $1/\mathrm{disc}(f)$ colors are necessary. The logarithm of this number gives the desired lower bound on $C_k(f)$. □

In general, $\mathrm{disc}(f)$ is very hard to estimate. Fortunately, using probabilistic arguments, we have already proved in Sect. 18.10 that the discrepancy can be bounded from above using the following more tractable measure.

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 multiset 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

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

be the expected value of the sign of a random cube D. We have already proved (see Theorem 18.17) that, for every function $f : X \to \{-1, 1\}$,

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

Together with Proposition 27.10, this gives the following general lower bound on the communication complexity.

Theorem 27.11. *For every $f : X \to \{-1, 1\}$,*

$$C_k(f) \geq \frac{1}{2^k} \log_2 \frac{1}{\mathcal{E}(f)} .$$

This is a very powerful result which allows us to show that, for any constant number k of players, some explicit functions require an almost maximal number of communicated bits. We demonstrate this by one function.

Say that a $(0, 1)$ matrix is *odd* if the number of its all-1 rows is odd. Note that, if the matrix has only two columns, then it is odd iff the scalar (or inner) product of these columns over $GF(2)$ is 1. By this reason, a boolean function, detecting whether a given matrix is odd, is called “generalized inner product” function. We will assume that input matrices have n rows and k columns.

That is, the *generalized inner product function* $\textsc{gip}(x)$ is a boolean function in kn variables, arranged in an $n \times k$ matrix $x = (x_{ij})$, and is defined by:

$$\textsc{gip}(x) = \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{k} x_{ij} .$$

We consider k-party communication gates for $\textsc{gip}(x)$, where the j-th player can see all but the j-th column of the input matrix x. Hence, the value of our function is determined by *rows*, whereas the players only have access to *columns*.

It is clear that $n + 1$ bits of communication is enough to determine the value of $\textsc{gip}(x)$: the first player announces the entire second column, and the second player announces the answer. In fact, $O(kn/2^k)$ bits of communication are already enough (see Exercise 27.9). We are now going to show that this last upper bound is almost optimal: $\Omega(n/4^k)$ bits are also necessary.

We have already shown in Sect. 18.10 (see Theorem 18.20) that the ±1 version $f(x) = (-1)^{\mathrm{GIP}(x)}$ has

$$\mathcal{E}(f) = \left(1 - \frac{1}{2^k}\right)^n \leq \mathrm{e}^{-n/2^k} .$$

Together with Theorem 27.11 this impies that the generalized inner product function has high multiparty communication complexity.

Theorem 27.12. *If the i-th player can see all but the i-th columns of a given $n \times k$ matrix, then k players cannot detect whether the matrix is odd by communicating fewer than $\Omega(n/4^k)$ bits.*

27.5 Making non-disjoint coverings disjoint

We have applied combinatorics (Ramsey theory and bounds on the discrepancy) to prove lower and upper bounds in communication complexity. In this section we give an example in the opposite direction: We show that communication games can be used to prove some purely combinatorial results.

Let X and Y be two finite sets. A *rectangle* is a subset $R \subseteq X \times Y$ of the form $R = R^0 \times R^1$ with $R^0 \subseteq X$ and $R^1 \subseteq Y$. That is, a subset R is a rectangle iff for every two points (x, y) and (x', y') of R, the combined points (x, y') and (x', y) belong to R as well. Note that R is a rectangle if and only if it is a cylinder intersection $R = T_0 \cap T_1$ with $T_0 = R^0 \times Y$ and $T_1 = X \times R^1$.

Let $f : X \times Y \to \{0, 1\}$ be a function. A rectangle $R \subseteq X \times Y$ is *f-monochromatic* if f takes the same value on all points of R.

Theorem 27.13. *If a rectangle can be covered by t not necessarily disjoint f-monochromatic rectangles, then it can be decomposed into at most $t^{O(\log t)}$ pairwise disjoint f-monochromatic rectangles.*

Proof. The function $f : X \times Y \to \{0, 1\}$ can be viewed as a 0-1 matrix, and $\chi_2(f)$ in this case is exactly the smallest number of mutually disjoint monochromatic matrices covering the whole matrix. Hence, our goal is to show that $\chi_2(f) \leq t^{\log_2 t}$. Since $\chi_2(f) \leq 2^{C_2(f)}$ (see Proposition 27.3), it is enough to design a 2-party communication protocol for f that uses

$$C_2(f) \leq (\log_2 t)^2$$

bits of communication. To design such a protocol, we first make one simple observation. Say that a rectangle $S = S^0 \times S^1$ intersects a rectangle $R = R^0 \times R^1$ in *rows*, if $S^0 \cap R^0 \neq \emptyset$, and intersects R in *columns*, if $S^1 \cap R^1 \neq \emptyset$.

Note that, $S \cap R \neq \emptyset$ if and only if S intersects R in rows *and* in columns. This immediately leads to the following observation about *disjoint* rectangles.

($*$) Let S be a rectangle and $\mathcal{R}$ a set of rectangles. If $S \cap R = \emptyset$ for all $R \in \mathcal{R}$, then either S intersects at most half of rectangles $R \in \mathcal{R}$ in rows or S intersects at most half of these rectangles in columns.

Now let $\mathcal{R}$ be a covering of $X \times Y$ by $|\mathcal{R}| = t$ f-monochromatic rectangles, and set $r := \lceil \log_2 t \rceil$. On all points in each rectangle $R \in \mathcal{R}$ the function f takes one and the same value 0 or 1; we call this value the *label* of the rectangle R. Say that two rectangles in $\mathcal{R}$ are *consistent* if they have the same label, and *inconsistent* otherwise. Note that rectangles with different labels must be disjoint. A rectangle $R = R^0 \times R^1$ *contains* a row x if $x \in R^0$, and *contains* a column y if $y \in R^1$.

The protocol consists of at most r rounds and in each round at most r bits are communicated. After each round the current set of rectangles is updated. Given an input (x, y), the goal is to decrease the number of rectangles in each round by at least one half. Let us call the two players Alice and Bob.

1. Alice checks whether all rectangles in $\mathcal{R}$ containing her row x are consistent. If yes, then the (unique) label i of all these rectangles is a correct answer, and she announces it.
2. Otherwise, Alice tries to find a rectangle $R \in \mathcal{R}$ containing x such that R intersects in *rows* at most half of the rectangles that are inconsistent with R. If such a rectangle R exists, then Alice sends its name (using r bits) to Bob and they both update $\mathcal{R}$ so that it only contains the rectangles that intersect with R in rows (the other rectangles cannot contain (x, y)).
3. If Alice is unable to find such a rectangle then she communicates this to Bob (using one bit).
4. Now is Bob's turn. Since Alice failed, our observation $(*)$ ensures that there must be a rectangle $R \in \mathcal{R}$ that contains y and intersects in *columns* at most half of the rectangles that are inconsistent with R. Bob takes any such rectangle R and sends its name (using r bits) to Alice and they both update $\mathcal{R}$ so that it only contains the rectangles that intersect with R in columns (the other rectangles cannot contain (x, y)). At this point the round is definitely over since they successfully eliminated at least half of the rectangles in $\mathcal{R}$, and we can proceed by induction.

In each round, the number of rectangles is decreased by at least one half. Hence, after at most $r = \log_2 t$ rounds the players will agree on a rectangle containing (x, y), and the label of this rectangle is the correct answer $f(x, y)$. □

Exercises

27.1. For a fixed vector $x \in X$, there are many (how many?) stars around it. How many colors do we need to leave none of them monochromatic?

27.2. Consider the following *equality function* $f : X_1 \times X_2 \to \{0, 1\}$ defined by: $f(x_1, x_2) = 1$ if and only if $x_1 = x_2$. Show that $\chi_2(f) = n$. *Hint*: If $\chi_2(f) < n$ then some color class contains two distinct vectors (x_1, x_1) and (x_2, x_2). What about the color of (x_1, x_2)?

27.3. Show that the two-dimensional hyperplane problem cannot be solved by communicating fewer than $\Omega(\log n)$ bits. *Hint*: Exercise 27.2.

27.4. Show that two players cannot solve the partition problem $Part_{n,2}$ by communicating fewer than $\Omega(\log n)$ bits. *Hint*: Exercise 27.2.

27.5. Three players want to compute the following boolean function $f(x, y, z)$ in $3n$ variables. Inputs x, y, z are vectors in $\{0,1\}^n$, and the function is defined by:

$$f(x,y,z) = \bigoplus_{i=1}^{n} \mathrm{Maj}(x_i, y_i, z_i)\,,$$

where $\mathrm{Maj}(x_i, y_i, z_i) = 1$ iff $x_i + y_i + z_i \geq 2$. Prove that $C_3(f) \leq 3$.

27.6. Prove the following extension of Proposition 27.4 to an arbitrary function $f : X \to \{0,1\}$. For a vector $x \in X$, its *i-th neighbor* is a vector $x^i = (x_1, \ldots, x'_i, \ldots, x_k)$, with $x'_i \in X_i$, such that $f(x^i) = 1$; if $f(x) = 1$, then x is a neighbor of itself. Let $c : X \to \{1, \ldots, r\}$ be a legal coloring (with respect to f). Let also $N(c) = \max\{N(c, x) : x \in X\}$, where $N(c, x)$ is the minimum, over all coordinates i, of the number of colors used by c to color the i-th neighbors of x. Then $C_k(f) \leq \log_2 r + k + N(c) \log_2 N(c)$. *Hint*: Given an input vector x, the i-th player can privately compute the set R_i of colors used by c to color the i-th neighbors of x. Show that $f(x) = 1$ if and only if $R_1 \cap R_2 \cap \cdots \cap R_k \neq \emptyset$.

27.7. Show that three players can compute the partition function $Part_{n,3}$ (introduced in Sect. 27.3) using $O(\sqrt{\log n})$ bits of communication. *Hint*: Theorem 27.6.

27.8. (Grolmusz 1994). Consider the following k-party communication game. Input is an $m \times k$ 0-1 matrix A, and the i-th player can see all A except its i-th column. Suppose that the players a priori know that some string $v = (0, \ldots, 0, 1, \ldots, 1)$ with the first 1 in position $t + 1$, does not appear among the rows of A. Show that then the players can decide if the number of all-1 rows is even or odd by communicating only t bits. *Hint*: Let y_i denote the number of rows of A of the form $(0, \ldots, 0, 1, \ldots, 1)$, where the first 1 occurs in position i. For every $i = 1, \ldots, t$, the i-th player announces the parity of the number of rows of the form $(0, \ldots, 0, *, 1, \ldots, 1)$, where the $*$ is at place i. Observe that this number is $y_i + y_{i+1}$. Subsequently, each player privately computes the mod 2 sum of all numbers announced. The result is $y_1 + y_{t+1}$ mod 2, where $y_{t+1} = 0$.

27.9. Use the previous protocol to show that (without any assumption) the players can decide if the number of all-1 rows is even or odd by communicating only $O(km/2^k)$ bits. *Hint*: Divide the matrix A into blocks with at most $2^{k-1} - 1$ rows in each. For each block there will be a string v' of length $k - 1$ such that neither $(0, v')$ nor $(1, v')$ occurs among the rows in that block. Using k bits the first player can make the string $(0, v')$ known to all players, and we are in the situation of the previous exercise.

27.10. (Due to Babai and Kimmel). Consider the following multiparty game with the *referee*. As before, we have an $m \times k$ 0-1 matrix A, and the i-th player can see all A except its i-th column. The restriction is that now the players do not communicate with each other but simultaneously write their messages on the blackboard. Using only this information (and without seeing the matrix A), an additional player (the referee) must compute the string $P(A) = (x_1, \ldots, x_m)$, where x_i is the sum modulo 2 of the number of $1'$s in the i-th row of A. Let N be the maximal number of bits which any player is allowed to write on any input matrix. Prove that $N \geq m/k$. *Hint*: For a matrix A, let $f(A)$ be the string $(p_1, \ldots, p_k)$, where $p_i \in \{0,1\}^N$ is the string written by the i-th player on input A. For each possible answer $x = (x_1, \ldots, x_m)$ of the referee, fix a matrix A_x for which $P(A_x) = x$. The correctness of the communication protocol ensures that $f(A_x) \neq f(A_y)$ for all $x \neq y$; hence, $2^{Nk} \geq 2^m$.

References

Adleman, L. (1978): Two theorems on random polynomial time, in: *Proc. of 19th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 75–83.

Ahlswede, R., El Gamal, A., and Pang, K.F. (1984): A two-family extremal problem in Hamming space, *Discrete Math.* **49**:1, 1–5.

Ajtai, M., Komlós, J., and Szemerédi, E. (1980): A note on Ramsey numbers, *J. Combin. Theory (A)* **29**, 354–360.

Ajtai, M., Chvátal, V., Newborn, M. M., and Szemerédi, E. (1982): Crossing-free subgraphs, *Ann. Discrete Math.* **12**, 9–12.

Alon, N. (1983): On the density of sets of vectors, *Discrete Math.* **46**, 199–202.

Alon, N. (1985): An extremal problem for sets with applications to graph theory, *J. Combin. Theory (A)* **40**, 82–89.

Alon, N. (1986a): Explicit constructions of exponential size families of k-independent sets, *Discrete Math.* **58**, 191–193.

Alon, N. (1986b): Covering graphs by the minimum number of equivalence relations, *Combinatorica* **6**(3), 201–206.

Alon, N. (1990a): On the rigidity of Hadamard matrices. Unpublished manuscript.

Alon, N. (1990b): The maximum number of Hamiltonian paths in tournaments, *Combinatorica* **10**, 319–324.

Alon, N. (1990c): Transversal numbers of uniform hypergraphs, *Graphs and Combinatorics* **6**, 1–4.

Alon, N. (1991): A parallel algorithmic version of the local lemma, *Random Structures & Algorithms* **2**:4, 367–378.

Alon, N. (1992): Choice numbers of graphs; a probabilistic approach, *Comb. Prob. Comput.* **1**, 107–114.

Alon, N. (1994): Probabilistic methods in extremal finite set theory, in: *Extremal Problems for Finite Sets*, Bolyai Society Mathematical Studies, Vol. 3, 39–58, János Bolyai Math. Society.

Alon, N. (1995): Tools from higher algebra, in: *Handbook of Combinatorics*, R. L. Graham, M. Grötschel and L. Lovász (eds.), Elsevier Science, Vol. 2, 1749–1784.

Alon, N. (1999): Combinatorial Nullstellensatz, *Comb. Prob. Comput.* **8**, 7–29.

Alon, N. and Boppana, R. (1987): The monotone circuit complexity of boolean functions, *Combinatorica* **7**:1, 1–22.

Alon, N., Babai, L., and Itai, A. (1986): A fast and simple randomized parallel algorithm for the maximal independent set problem, *J. of Algorithms* **7**, 567–583.

Alon, N., Bergmann, E.E., Coppersmith, D., and Odlyzko, A.M. (1988): Balancing sets of vectors, *IEEE Trans. Inf. Theory* **34**:1, 128–130.

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6,

Alon, N. and Dubiner, M. (1993): Zero-sum sets of prescribed size, in: *Combinatorics, Paul Erdős is Eighty*, Vol. 1, 33–50.

Alon, N. and Kleitman, D.J. (1990): Sum-free subsets, in: *A Tribute to Paul Erdős*, A. Baker, B. Bollobás, and A. Hajnál (eds.), Cambridge University Press, 13–26.

Alon, N. and Spencer, J. (1992): *The Probabilistic Method*, Wiley, 1992. Second edition: Wiley, 2000.

Alon, N., Friedland, S., and Kalai, G. (1984): Regular subgraphs of almost regular graphs, *J. Combin. Theory (B)* **37**(1), 79–91.

Alon, N., Bergmann, E.E., Coppersmith, D., and Odlyzko, A.M. (1988): Balancing sets of vectors, *IEEE Trans. Information Theory IT*-**34**, 128–130.

Alon, N., Babai, L., and Suzuki, H. (1991): Multilinear polynomials and Frankl–Ray-Chaudhuri–Wilson type intersection theorems, *J. Combin. Theory (A)* **58**:2, 165–180.

Alon N, McDiarmid C., and Molloy, M. (1996): Edge-disjoint cycles in regular directed graphs, *J. Graph Theory* **22**, 231–237.

Anderson, I. (1987): *Combinatorics of Finite Sets.* Oxford University Press.

Anderson, I. and Honkala, I. (1997): *A short course in combinatorial designs*, Internet Edition, Spring 1997.

Andreev, A.E. (1985): On a method for obtaining lower bounds for the complexity of individual monotone functions, *Soviet Math. Dokl.* **31**:3, 530–534.

Andreev, A.E. (1987): On one method of obtaining effective lower bounds on monotone complexity, *Algebra i Logica* **26**:1, 3–26. (Russian)

Aspnes, J., Beigel, R., Furst, M., and Rudich, S. (1994): The expressive power of voting polynomials, *Combinatorica*, **14**:2, 1–14.

Babai, L. (1988): A short proof of the nonuniform Ray-Chaudhuri–Wilson inequality, *Combinatorica* **8**:1, 133–135.

Babai, L. (1997): Communication complexity, in: *Proc. of 22nd Internat. Symp. on Mathematical Foundations of Computer Science*, Lect. Notes in Comput. Sci., Vol. 1295, 5–18, Springer.

Babai, L. and Frankl, P. (1992): *Linear Algebra Methods in Combinatorics*, Preliminary Version 2, University of Chicago.

Babai, L., Frankl, P., and Simon, J. (1986): The complexity classes in communication complexity, in: *Proc. of 27th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 337–347.

Babai, L., Fortnow, L., Levin, L.A., and Szegedy, M. (1991): Checking computations in polylogarithmic time, in: *Proc. of 23rd Ann. ACM Symp. on the Theory of Computing*, 21–31.

Babai, L., Nisan, N., and Szegedy, M. (1992): Multiparty protocols, pseudorandom generators for Logspace, and time-space trade-offs, *J. Comput. Syst. Sci.* **45**, 204–232.

Babai, L., Hayes, T., and Kimmel, P. (1998): The cost of the missing bit: communication complexity with help, in: *Proc. of 30th Ann. ACM Symp. on the Theory of Computing*, 673–682.

Barak, B., Kindler, G., Shaltiel, R., Sudakov, B., and Wigderson, A. (2005): Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors, in: *Proc. of 37th Ann. ACM Symp. on the Theory of Computing*, 1–10.

Barak, B., Rao, A., Shaltiel, R., and Wigderson, A. (2006): 2-source dispersers for sub-polynomial entropy and Ramsey graphs beating the Frankl-Wilson construction, in: *Proc. of 38th Ann. ACM Symp. on the Theory of Computing*, 671–680.

Barefoot, C., Casey, K., Fisher, D., Fraughnaugh, K., and Harary, F. (1995): Size in maximal triangle-free graphs and minimal graphs of diameter 2, *Discrete Math.* **138**, No. 1-3, 93–99.

Beame, P., Saks, M., and Thathachar, J. (1998): Time-space tradeoffs for branching programs, in: *Proc. of 39th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 254–

263.

Beck, J. (1978): On 3-chromatic hypergraphs, *Discrete Math.* **24**, 127–137.

Beck, J. (1980): A remark concerning arithmetic progressions, *J. Combin. Theory (A)* **29**, 376–379.

Beck, J. (1983): On the lattice property of the plane and some problems of Dirac, Motzkin, and Erdős in combinatorial geometry, *Combinatorica* **3**, 281–297.

Beck, J. (1991): An algorithmic approach to the Lovász Local Lemma, *Random Structures & Algorithms* **2**, 343–365.

Behrend, F.A. (1946): On sets of integers which contain no three elements in arithmetic progression, *Proc. Nat. Acad. Sci.* **23**, 331–332.

Berge, C. (1957): Two theorems in graph theory, *Proc. Nat. Acad. Sci.* **43**, 842–844.

Berge, C. (1989): *Hypergraphs*, Elsevier Science.

Birkhoff, G. (1946): Tres observaciones sobre el algebra lineal, *Rev. Fac. Ci. Exactas, Puras y Aplicadas Univ. Nac. Tucuman, Ser A* **5**, 147–151.

Blokhuis, A. (1981): A new upper bound for the cardinality of 2-distance sets in Euclidean space, Eindhoven Univ. Technology, Mem. 1981–04.

Blokhuis, A. (1987): More on maximal intersecting families of finite sets, *J. Combin.Theory (A)* **44**, 299–303.

Blokhuis, A. (1994): On the size of blocking sets in $PG(2,p)$, *Combinatorica* **14**:1, 111–114.

Blum, M. and Impagliazzo, R. (1987): Generic oracles and oracle classes, in: *Proc. of 28th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 118–126.

Bollobás, B. (1965): On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16**, 447–452.

Bollobás, B. (1978): *Extremal Graph Theory*, LMS Monographs, No. 11, Academic Press.

Bollobás, B. (1981): Threshold functions for small subgraphs, *Math. Proc. Cambridge Philos. Soc.* **90**, 197–206.

Bollobás, B. (1986): *Combinatorics: Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability*, Cambridge University Press.

Bollobás, B. and Thomason, A. (1981): Graphs which contain all small graphs, *European J. Combin.* **2**, 13–15.

Bondy, J.A. (1972): Induced subsets, *J. Combin. Theory (B)* **12**, 201–202.

Borodin, A., Dolev, D., Fich, F., and Paul, W. (1986): Bounds for width two branching programs, *SIAM J. Comput.* **15**:2, 549–560.

Borsuk, K. (1933): Drei Sätze über die n-dimensionale euklidische Sphäre, *Fund. Math.* **20**, 177–190.

Bose, R.C. (1949): A note on Fisher's inequality for balanced incomplete block designs, *Ann. Math. Stat.* **20**, 619–620.

Braverman M. (2009): Poly-logarithmic independence fools AC^0 circuits, in: *Proc. of 24th Ann. IEEE Conf. on Computational Complexity*, 3–8.

Brouwer, A.E. and Schrijver, A. (1978): The blocking number of an affine space, *J. Combin. Theory (A)* **24**, 251–253.

Brown, T.C., Chung, F.R.K., Erdős, P., and Graham R. L. (1985): Quantitative forms of a theorem of Hilbert, *J. Combin. Theory (A)* **38**, 210–216.

Bruen, A. (1970): Baer subplanes and blocking sets, *Bull. Amer. Math. Soc.* **76**, 342–344.

Buss, S. (1987): Polynomial size proofs of the propositional pigeonhole principle, *J. Symbolic Logic* **52**, 916–927.

Cameron, P.J. (1994): *Combinatorics: Topics, Techniques, Algorithms*, Cambridge University Press.

Cameron, P.J. and Erdős, P. (1999): Notes on sum-free and related sets, *Comb. Prob. Comput.* **8**, 95–107.

Chandra, A.K., Furst, M., and Lipton, R.J. (1983): Multi-party protocols, in: *Proc. of 15th Ann. ACM Symp. on the Theory of Computing*, 94–99.

Chandra, A.K., Kou, L., Markowsky, G., and Zaks, S. (1983): On sets of boolean n-vectors with all k-projections surjective, *Acta Informatica* **20**:2, 103-111.

Chernoff, H. (1952): A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23**, 493–509.

Chung, F.R.K. (1990): Quasi-random classes of hypergraphs, *Random Structures and Algorithms* **1**, 363–382.

Chung, F.R.K. and Tetali, P. (1993): Communication complexity and quasi randomness, *SIAM J. Discrete Math.* **6**, 110–123.

Chung, F.R.K., Frankl, P., Graham, R.L., and Shearer, J.B. (1986): Some intersection theorems for ordered sets and graphs, *J. Combin. Theory (A)* **43**, 23–37.

Chvátal, V. and Szemerédi, E. (1988): Many hard examples for resolution, *J. of the ACM* **35**:4, 759–768.

Cohen, G., Honkala, I., Litsyn, S., and Lobstein, A. (1997): *Covering Codes*, Math. Library, Vol. 54, North-Holland.

Conlon, D. (2008): A new upper bound for the bipartite Ramsey problem, *J. Graph Theory* **58**:4, 351–356.

Conlon, D. (2009): A new upper bound for diagonal Ramsey numbers, *Ann. of Math.* **170**:2, 941–960.

Corrádi, K. (1969): Problem at Schweitzer competition, *Mat. Lapok* **20**, 159–162.

Czumaj, A. and Scheideler, C. (2000): A new algorithm approach to the general Lovász Local Lemma with applications to scheduling and satisfiability problems, in: *Proc. of 32nd Ann. ACM Symp. on the Theory of Computing*, 38–47.

Danzer, L. and Grünbaum, B. (1962): Über zwei Probleme bezüglich konvexer Körper von P. Erdős und von V. L. Klee, *Math. Zeitschrift* **79**, 95-99.

De Bruijn, N.G. and Erdős, P. (1948): On a combinatorial problem, *Indagationes Mathematicae* **10**, 421–423.

De Bruijn, N.G., Tengbergen, C., and Kruyswijk, D. (1952): On the set of divisors of a number, *Nieuw Arch. Wiskunde* **23**, 191–193.

De Wolf, R. (2006): Lower bounds on matrix rigidity via a quantum argument, in: *Proc of 33rd Int. Colloq. on Automata, Languages and Programming*, Lect. Notes in Comput. Sci., Vol. 4051, 62–71.

Delsarte, Ph. and Piret, Ph. (1985): An extension of an inequality by Ahlswede, El Gamal and Pang for pairs of binary codes, *Discrete Math.* **55**, 313–315.

DeMillo, R. A. and Lipton, R.J. (1978): A probabilistic remark on algebraic program testing, *Inform. Process. Letters* , **7**(4), 193–195.

Deuber, W.A., Erdős, P., Gunderson, D.S., Kostochka, A.V., and Meyer A.G. (1997): Intersection statements for systems of sets, *J. Combin. Theory (A)* **79**:1, 118–132.

Deza, M. (1973): Une propriété extrémale des plans projectifs finis dans une classe des codes équidistantes, *Discrete Math.* **6**, 343–352.

Deza, M., Frankl, P., and Singhi, N.M. (1983): On functions of strength t, *Combinatorica* **3**, 331–339.

Dias da Silva J.A. and Hamidoune, Y.O. (1994): Cyclic spaces for Grassmann derivatives and additive theory, *Bull. London Math. Soc.* **26**, 140–146.

Dirichlet, P.G.L. (1879): *Vorlesungen über Zahlentheorie*, Vieweg, Braunschweig.

Dilworth, R.P. (1950): A decomposition theorem for partially ordered sets, *Ann. of Math.* **51**, 161–165.

Dvir, Z. (2009): On the size of Kakeya sets in finite fields, *J. AMS* **22** (4), 1093–1097.

Dyachkov, A.G. and Rykov, V.V. (1982): Bounds on the length of disjunctive codes, *Problemy Peredachi Informatsii* **18**:3, 7–13. (Russian)

Edmonds, J. (1971): Matroids and the greedy algorithm, *Math. Programming* **1**, 127–136.

Egerváry, E. (1931): On combinatorial properties of matrices, *Mat. Lapok* **38**, 16–28. (Hungarian with German summary)

Elekes, Gy. (1997): On the number of sums and products, *Acta Arithmetica*, **81**, 365–367.

Erdős, P. (1938): On sequences of integers no one of which divides the product of two others and on some related problems, *Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk* **2**, 74–82.

Erdős, P. (1945): On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51**, 898–902.

Erdős, P. (1947): Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53**, 292–294.

Erdős, P. (1959): Graph theory and probability, *Canadian J. Math.* **11**, 34–38.

Erdős, P. (1963a): On a problem of graph theory, *Math. Gaz.* **47**, 220–223.

Erdős, P. (1963b): On a combinatorial problem I, *Nordisk Tidskr. Informationsbehandlung (BIT)* **11**, 5–10.

Erdős, P. (1964a): On a combinatorial problem II, *Acta Math. Acad. Sci. Hungar.* **15**, 445–447.

Erdős, P. (1964b): On extremal problems of graphs and generalized graphs, *Israel J. Math.*, **2**, 183–190.

Erdős, P. (1965a): Extremal problems in number theory, *Proc. of the Symposium on Pure Mathematics*, **VIII**, AMS, 181–189.

Erdős, P. (1965b): On some extremal problems in graph theory, *Israel J. Math.*, **3** (1965), 113–116.

Erdős, P. (1967): On bipartite subgraphs of a graph, *Matematikai Lapok* **18**, 283–288. (Hungarian)

Erdős, P. and Füredi, Z. (1983): The greatest angle among n points in the d-dimensional Euclidean space, *Annals of Discrete Math.* **17**, 275–283.

Erdős, P. and Kleitman, D.J. (1968): On coloring graphs to maximize the proportion of multicolored k-edges, *J. Combin. Theory (A)* **5**, 164–169.

Erdős, P. and Lovász, L. (1975): Problems and results on 3-chromatic hypergraphs and some related questions, in: *Infinite and Finite Sets*, Coll. Math. Soc. J. Bolyai, 609–627.

Erdős, P. and Rado, R. (1960): Intersection theorems for systems of sets, *J. London Math. Soc.* **35**, 85–90.

Erdős, P. and Rényi, A. (1960): On the evolution of random graphs, *Publ. Math, Inst. Hung. Acad. Sci.* **5**, 17–61.

Erdős, P. and Spencer, J. (1974): *Probabilistic Methods in Combinatorics.* Academic Press, New York and London, and Akadémiai Kiadó, Budapest.

Erdős, P. and Szekeres, G. (1935): A combinatorial problem in geometry, *Composito Math.* **2**, 464–470.

Erdős, P. and Szemerédi. E. (1983): On sums and products of integers, In: Studies in Pure Mathematics, Akadámiai Kiadó - Birkhauser Verlag, Budapest-Basel-Boston, 213–218.

Erdős, P., Ginzburg, A., and Ziv, A. (1961): Theorem in the additive number theory, *Bull. Research Council Israel* **10F**, 41–43.

Erdős, P., Chao Ko and Rado, R. (1961): Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford (2)* **12**, 313–320.

Erdős, P., Hajnal, A., and Moon, J.W. (1964): A problem in graph theory, *Acta Math. Acad. Sci. Hungar.* **71**, 1107–1110.

Erdős, P., Frankl, P., and Füredi, Z. (1985): Families of finite sets in which no set is covered by the union of r others, *Israel J. Math.* **51**, 79–89.

Evans, T. (1960): Embedding incomplete Latin squares, *Amer. Math. Monthly* **67**, 958–961.

Frankl, P. (1977): A constructive lower bound for Ramsey numbers, *Ars Combin.* **3**, 297–302.

Frankl, P. (1982): An extremal problem for two families of sets, *European J. Combin.* **3**, 125–127.

Frankl, P. (1983): On the trace of finite sets, *J. Combin. Theory (A)* **34**, 41–45.

Frankl, P. (1988): Intersection and containment problems without size restrictions, in: *Algebraic, Extremal and Metric Combinatorics*, 1986, M. Deza and P. Frankl (eds.), Cambridge University Press.

Frankl, P. and Pach, J. (1983): On the number of sets in a null t-design, *European J. Combin.* **4**, 21–23.

Frankl, P. and Pach, J. (1984): On disjointly representable sets, *Combinatorica* **4**:1, 39–45.

Frankl, P. and Rödl, V. (1987): Forbidden intersections, *Trans. Amer. Math. Soc.* **300**, 259–286.

Frankl, P. and Wilson, R.M. (1981): Intersection theorems with geometric consequences, *Combinatorica* **1**, 357–368.

Fredman, M. and Komlós, J. (1984): On the size of separating systems and perfect hash functions, *SIAM J. Algebraic and Discrete Meth.* **5**, 61–68.

Freivalds, R. (1977): Probabilistic machines can use less running time, in: *Proc. of IFIP Congress 77*, 839–842, North-Holland, Amsterdam.

Friedman, J. (1984): Constructing $O(n \log n)$ size monotone formulae for the kth elementary symmetric polynomial of n boolean variables, in: *Proc. of 25th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 506–515.

Frobenius, G. (1917): Über zerlegbare Determinanten, *Sitzungsber. Königl. Preuss. Acad. Wiss.* **XVIII**, 274–277.

Füredi, Z. (1980): On maximal intersecting families of finite sets, *J. Combin. Theory (A)* **28**, 282–289.

Füredi, Z. (1984): Geometrical solution of an intersection problem for two hypergraphs, *European J. Combin.* **5**, 133–136.

Füredi, Z. (1995): Cross-intersecting families of finite sets, *J. Combin. Theory (A)* **72**, 332–339.

Füredi, Z. (1996): On r-cover-free families, *J. Combin. Theory (A)* **73**, 172–173.

Füredi, Z. and Kahn, J. (1989): On the dimension of ordered sets of bounded degree, *Order* **3**, 15–20.

Füredi, Z. and Tuza, Zs. (1985): Hypergraphs without large star, *Discrete Math.* **55**, 317–321.

Gál, A. (1998): A characterization of span program size and improved lower bounds for monotone span programs, in: *Proc. of 30th Ann. ACM Symp. on the Theory of Computing*, 429–437.

Galvin, F. (1994): A proof of Dilworth's chain decomposition theorem, *Amer. Math. Monthly* **101**:4, 352–353.

Galvin, F. (1995): The list chromatic index of a bipartite multigraph, *J. Combin. Theory (B)* **63**, 153–158.

Galvin, F. (1997): On a theorem of J. Ossowski, *J. Combin. Theory (A)* **78**, 151–153.

Garaev, M. Z. (2007): An explicit sum-product estimate in $\mathbb{F}_p$, *Int. Math. Res. Notices* **2007**, 11 pp.

Goodman, A.W. (1959): On sets of acquaintances and strangers at any party, *Amer. Math. Monthly* **66**, 778–783.

Graham, R.L. (1981): *Rudiments of Ramsey theory*, Regional conference series in mathematics, Vol. 45, American Math. Society (reprinted with corrections 1983).

Graham, R.L. and Kleitman, D.J. (1973): Intersecting paths in edge ordered graphs, *Period. Math. Hungar.* **3**, 141–148.

Graham, R. L. and Pollak, H. O. (1971): On the addressing problem for loop switching, *Bell Syst. Tech. J.* **50**, 2495–2519.

Graham, R.L. and Spencer, J. (1971): A constructive solution to a tournament problem, *Canad. Math. Bull.* **14**, 45–48.

Graham, R.L., Rothschild, B.L., and Spencer, J.H. (1990): *Ramsey Theory*, Second Edition, Wiley.

Graham, R., Grötschel, M., and Lovász, L. (eds.) (1995): *Handbook of Combinatorics*, Elsevier Science, North-Holland.

Green, B. and Tao, T. (2008): The primes contain arbitrarily long arithmetic progressions, *Annals of Math.* **167**, 481–547.

Grigni, M. and Sipser, M. (1995): Monotone separation of logarithmic space from logarithmic depth, *J. Comput. Syst. Sci.* **50**, 433–437.

Grolmusz, V. (1994): The BNS lower bound for multi-party protocols is nearly optimal, *Information and Computation* **112**:1, 51–54.

Grolmusz, V. (2000): Low rank co-diagonal matrices and Ramsey graphs, *Electr. J. Comb.* Nr. 7.

Gunderson, D.S. and Rödl, V. (1998): Extremal problems for affine cubes of integers, *Comb. Prob. Comput.* **7**:1, 65–79.

Gyárfás, A. (1987): Partition covers and blocking sets in hypergraphs, *MTA SZTAKI Tanulmáyok* **71**, Budapest. (Hungarian)

Gyárfás, A., Hubenko, A., and Solymosi, J. (2002): Large cliques in C_4-free graphs, *Combinatorica* **22**(2), 269–274.

Hagerup, T. and Rüb, Ch. (1989): A guided tour of Chernoff bounds, *Inform. Process. Letters* **33**, 305–308.

Hales, A.W. and Jewett, R.I. (1963): Regularity and positional games, *Trans. Amer. Math. Soc.* **106**, 222–229.

Hall, P. (1935): On representatives of subsets, *J. London Math. Soc.* **10**, 26–30.

Hardy, G. H. and Ramanujan, S. (1917): The normal number of prime factors of a number n, *Quart. J. Math.* **48**, 76–92.

Hardy, G.H., Littlewood, J.E., and Pólya, G. (1952): *Inequalities*, Cambridge University Press, 1952.

Hartmanis, J. and Hemachandra, L.A. (1991): One-way functions, robustness and non-isomorphism of NP-complete classes, *Theor. Comput. Sci.* **81**:1, 155–163.

Håstad, J. (1993): The shrinkage exponent is 2, in: *Proc. of 34th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 114–123.

Håstad, J., Jukna, S., and Pudlák, P. (1995): Top-down lower bounds for depth-three circuits, *Computational Complexity* **5**, 99–112.

Helly, E. (1923): Über Mengen konvexer Körper mit gemeinschaftlichen Punkten, *Jahresber. Deutsch. Math.-Verein* **32**, 175–176.

Hilbert, D. (1892): Über die Irreduzibilität ganzer rationalen Funktionen mit ganzzahligen Koeffizienten, *J. Reine Angew. Math.* **110**, 104–129.

Hirsch, E. A. (2000): SAT local search algorithms: worst-case study, *J. Autom. Reasoning* **24**(1/2), 127–143.

Hoory, S., Linial, N., and Wigderson A. (2006): Expander graphs and their applications, *Bull. of AMS* **43**(4), 439–561.

Hopcroft, J.E. and Karp, R.M. (1973): An $n^{5/2}$ algorithm for maximum matching in bipartite graphs, *SIAM J. Comput.* **4**, 225–231.

Hwang, F.K. and Sós, V.T. (1987): Non-adaptive hypergeometric group testing, *Studia Sci. Math. Hungar.* **22**, 257–263.

Ibarra, O.H. and Moran, S. (1983): Probabilistic algorithms for deciding equivalence of stright-line programs, *J. of the ACM* **30**:1, 217–228.

Jamison, R. (1977): Covering finite fields with cosets of subspaces, *J. Combin. Theory (A)* **22**, 253–266.

Janssen, J.C.M. (1992): The Dinitz problem solved for rectangles, *Bull. Amer. Math. Soc.* **29**, 243–249.

Johnson, D.S. (1974): Approximation algorithms for combinatorial problems, *J. Comput. Syst. Sci.* **9**, 256–278.

Johnson, S. (1986): A new proof of the Erdős–Szekeres convex k-gon result, *J. Combin. Theory (A)* **42**, 318–319.

Jordan, J. W. and Livné, R. (1997): Ramanujan local systems on graphs, *Topology* **36**(5), 1007–1024.

Kahn, J. and Kalai, G. (1993): A counterexample to Borsuk's conjecture, *Bull. Amer. Math. Soc.* **29**:1, 60–62.

Kalai, G. (1984): Intersection patterns of convex sets, *Israel J. Math.* **48**, 161–174.

Karchmer, M. and Wigderson, A. (1990): Monotone circuits for connectivity require super-logarithmic depth, *SIAM J. Discrete Math.* **3**, 255–265.

Karger, D.R. (1993): Global min-cuts in *RNC*, and other ramifications of a simple mincut algorithm, in: *Proc. 4th Ann. ACM-SIAM Symp. on Discrete Algorithms*, 21–30.

Karoński, M. (1995): Random graphs, in: *Handbook of Combinatorics*, R.L. Graham, M. Grötschel and L. Lovász (eds.), Elsevier-Science, Vol. 1, 351–380.

Kashin, B. and Razborov, A. (1998): Improved lower bounds on the rigidity of Hadamard matrices, *Matematicheskie Zametki* **63**(4), 535–540.

Katona, G.O. (1972): A simple proof of the Erdős–Chao Ko–Rado theorem, *J. Combin. Theory (B)* **13**, 183–184.

Katona,G.O. (1966): A theorem of finite sets, in: *Theory of Graphs* (Proc. of Conf. Tihany, Hungary 1966), Academic Press, 1968, 187–207.

Kautz, W.H. and Singleton, R.C. (1964): Nonrandom binary superimposed codes, *IEEE Trans. Inform. Theory* **10**, 363–377.

Kelly, L.M. (1947): Elementary problems and solutions. Isosceles n-points, *Amer. Math. Monthly* **54**, 227–229.

Khrapchenko, V.M. (1971): A method of determining lower bounds for the complexity of Π–schemes, *Math. Notes of the Acad. of Sci. of the USSR* **10**:1, 474–479.

Kim, J.H. (1995): The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures and Algorithms* **7**, 173–207.

Kleitman, D.J. (1966): Families of non-disjoint subsets, *J. Combin. Theory (A)* **1**, 153–155.

Kleitman, D.J. and Spencer, J. (1973): Families of k-independent sets, *Discrete Math.* **6**, 255–262.

Kleitman, D.J., Shearer, J. B., and Sturtevant, D. (1981): Intersection of k-element sets, *Combinatorica* **1**, 381–384.

Kneser, M. (1995): Ein Satz über abelsche Gruppen mit Anwendungen auf die Geometrie der Zahlen, *Math. Z.* **61**, 429-434.

Knuth, D.E. (1986): Efficient balanced codes, *IEEE Trans. Information Theory IT*-**32**, 51–53.

Kolountzakis, M.N. (1994): Selection of a large sum-free subset in polynomial time, *Inform. Process. Letters* **49**, 255–256.

König, D. (1931): Graphen und Matrizen, *Mat. és Fiz. Lapok* **38**, 116–119. (Hungarian)

Körner, J. (1986): Fredman–Komlós bounds and information theory, *SIAM J. Algebraic and Discrete Meth.* **7**, 560–570.

Körner, J. and Marton, K. (1988): New bounds for perfect hashing via information theory, *European J. Combin.* **9**, 523–530.

Kostochka, A.V. and Rödl, V. (1998): On large systems of sets with no large weak Δ-subsystem, *Combinatorica* **18**:2, 235–240.

Kostochka, A.V., Rödl, V., and Talysheva, L.A. (1999): On systems of small sets with no large Δ-system, *Comb. Prob. Comput.* **8**, 265–268.

Kővári, P., Sós, V.T., and Turán, P. (1954): On a problem of Zarankiewicz, *Colloq. Math.* **3**, 50–57.

Kruskal, J. (1963): The optimal number of simplices in a complex, in: *Math. Opt. Techniques*, Uviv. California Press, Berkeley, 1963, 251–278.

Kushilevitz, E. and Nisan, N. (1997): *Communication Complexity*, Cambridge University Press.

Kushilevitz, E., Linial, N., Rabinovitch, Y., and Saks, M. (1996): Witness sets for families of binary vectors, *J. Combin. Theory (A)* **73**:2, 376–380.

Kuzjurin, N. N. (2000): Explicit constructions of Rödl's asymptotically good packings and coverings, *Comb. Prob. Comput.* **9**(3), 265–276.

Larman, D.G. and Rogers, C.A. (1972): The realization of distances within sets in Euclidean space, *Mathematika* **12**, 1–24.

Larman, D.G., Rogers, C.A., and Seidel, J.J. (1977): On two-distance sets in Euclidean space, *Bull. London Math. Soc.* **9**, 261–267.

Leighton, F. T. (1984): New lower bound techniques for VLSI, *Math. Systems Theory* **17**(1), 47–70.

Li, M. and Vitányi, P. (2008): *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd Edition, Springer-Verlag.

Lieberher, K. and Specker, E. (1981): Complexity of partial satisfaction, *J. of the ACM* **28**:2, 411–422.

Lindstrom, B. (1993): Another theorem on families of sets, *Ars Combinatoria* **35**, 123–124.

Lipski, W. (1978): On strings containing all subsets as substrings, *Discrete Math.* **21**, 253–259.

Lovász, L. (1973): Coverings and colorings of hypergraphs, in: *Proc. 4th Southeastern Conf. on Comb.*, Utilitas Math., 3–12.

Lovász, L. (1975): On the ratio of optimal integral and fractional covers, *Discrete Math.* **13**, 383–390.

Lovász, L. (1977a): Certain duality principles in integer programming, *Ann. Discrete Math.* **1**, 363–374.

Lovász, L. (1977b): Flats in matroids and geometric graphs, in: *Combinatorial Surveys*, P.J. Cameron (ed.), Academic Press, 45–86.

Lovász, L. (1979): *Combinatorial Problems and Exercises.* Akadémiai Kiadó & North-Holland.

Lovász, L., Pelikán J., and Vesztergombi K. (1977): *Kombinatorika*, Tankönyvkiadó, Budapest, 1977. (German translation: Teubner, 1977)

Lovász, L., Shmoys, D.B., and Tardos, É. (1995): Combinatorics in computer science, in: *Handbook of Combinatorics*, R.L. Graham, M. Grötschel, and L. Lovász (eds.), Elsevier Science, Vol. 2, 2003–2038.

Lovász, L., Pyber, L., Welsh, D.J.A., and Ziegler, G.M. (1995): Combinatorics in pure mathematics, in: *Handbook of Combinatorics*, R.L. Graham, M. Grötschel, and L. Lovász (eds.), Elsevier Science, Vol. 2, 2039–2082.

Lubell, D. (1966): A short proof of Sperner's lemma, *J. Combin. Theory (A)* **1**:2, 402.

Lubotzky, A., Phillips, R. and Sarnak, P. (1988): Ramanujan graphs, *Combinatorica* **8**(3) 261–277.

Luby, M. (1986): A simple parallel algorithm for the maximal independent set, *SIAM J. Comput.* **15**, 1036–1053.

MacWilliams, F.J. and Sloane, N.J.A. (1977): *The theory of error correcting codes*, North-Holland, Amsterdam.

Majumdar, K.N. (1953): On some theorems in combinatorics relating to incomplete block designs, *Ann. Math. Stat.* **24**, 377–389.

Mantel, W. (1907): Problem 28, *Wiskundige Opgaven* **10**, 60–61.

Margulis, G. A. (1973): Explicit constructions of concentrators, *Probl. Peredachi Inf.* **9** (1973), 71–80 (in Russian). Translation: *Problems of Inf. Transm.* (1975), 323–332.

Matoušek, J. (2002): *Lectures on Discrete Geometry*, Graduate Texts in Mathematics, Vol. 212, Springer.

Menger, K. (1927): Zur allgemeinen Kurventheorie, *Fund. Math.* **10**, 95–115.

Meschalkin, L.D. (1963): A generalization of Sperner's theorem on the number of subsets of a finite set (in Russian), *Teor. Veroj. Primen.* **8**, 219–220. English translation in: *Theory of Probab. and its Appl.* **8** (1964), 204–205.

Mestre, J. (2006): Greedy in approximation algorithms, in: *Proc of. 14th Ann. European Symp. on Algorithms*, Lect. Notes in Comput. Sci., Vol. 4162, 528–539, Springer.

Moon, J.W. and Moser, L. (1962): On a problem of Turán, *Publ. Math. Inst. Hungar. Acd. Sci.* **7**, 283–286.

Morgenstern, M. (1994): Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q, *J. Combin. Theory (B)* **62**(1), 44–62.

Moser, R.A. (2009): A constructive proof of the Lovász local lemma, in *Proc. of 41st Ann. ACM Symp. on the Theory of Computing*, 343–350.

Moser, R.A. and Tardos, G. (2010): A constructive proof of the general Lovász Local Lemma, *J. of the ACM* **57**:2.

Moshkovitz, D. (2010): An alternative proof of the Schwartz-Zippel lemma, *Electronic Colloquium on Computational Complexity*, Report Nr. 96.

Motwani, R. and Raghavan, P. (1995): *Randomized Algorithms*, Cambridge University Press.

Motzkin, T.S. and Straus, E.G. (1965): Maxima for graphs and a new proof of a theorem of Turán, *Canadian J. Math.* **17**, 533–540.

Nagy, Z. (1972): A certain constructive estimate of the Ramsey number, *Matematikai Lapok* **23**, 301–302. (Hungarian)

Nathanson, M.B. (1996): *Additive Number Theory: Inverse Problems and Geometry of Sumsets*, Graduate Texts in Mathematics, Vol. 165, Springer.

Nechiporuk, E.I. (1966): On a boolean function, *Doklady Akademii Nauk SSSR* **169**:4, 765–766 (in Russian). English translation in: *Soviet Math. Dokl.* **7**:4, 999–1000.

Nesetril, J. and Rödl, V. (eds.) (1990): *Mathematics of Ramsey Theory*, Algorithms and Combinatorics, Vol. 5, Springer.

Newman, I. (1991): Private vs. common random bits in communication complexity, *Inform. Process. Letters* **39**, 67–71.

Nilli, A. (1990): Shelah's proof of the Hales–Jewett theorem, in: *Mathematics of Ramsey Theory*, J. Nesetril and V. Rödl, (eds.), 150–151, Springer.

Nilli, A. (1994): Perfect hashing and probability, *Comb. Prob. Comput.* **3**:3, 407–409.

Nisan, N. (1989): CREW PRAMS and decision trees, in: *Proc. of 21st Ann. ACM Symp. on the Theory of Computing*, 327–335.

Nisan, N. and Wigderson, A. (1995): On rank vs. communication complexity, *Combinatorica* **15**(4), 557–565.

Ossowski, J. (1993): On a problem of F. Galvin, *Congr. Numer.* **96**, 65–74.

Oxley, J. G. (1992): *Matroid Theory*, Oxford University Press.

Papadimitriou, C.H. (1991): On selecting a satisfying truth assignment, in: *Proc. of 32nd Ann. IEEE Symp. on Foundations of Comput. Sci.*, 163–169.

Paturi, R. and Zane, F. (1998): Dimension of projections in boolean functions, *SIAM J. Discrete Math.* **11**:4, 624–632.

Penfold Street, A. (1972): Sum-free sets, in: Lect. Notes in Math., Vol. 292, 123–272, Springer.

Petrenjuk, A.Ya. (1968): On Fisher's inequality for tactical configurations, *Mat. Zametki* **4**, 417–425. (Russian)

Plumstead, B.R., and Plumstead, J.B. (1985): Bounds for cube coloring, *SIAM J. Alg. Discr. Meth.* **6**:1, 73–78.

Pudlák, P. and Rödl, V. (2004): Pseudorandom sets and explicit constructions of Ramsey graphs, in: J. Krajiček (ed.) *Quaderni di Matematica*, Vol. 13, 327–346.

Pudlák, P., Razborov, A., and Savický, P. (1988): Observations on rigidity of Hadamard matrices. Unpublished manuscript.

Quine, W.V. (1988): Fermat's last theorem in combinatorial form, *Amer. Math. Monthly* **95**:7, 636.

Rado, R. (1942): A theorem on independence relations, *Quart. J. Math.* **13**, 83–89.

Rado, R. (1943): Note on combinatorial analysis, *Proc. London Math. Soc.* **48**, 122–160.

Ramsey, F.P. (1930): On a problem of formal logic, *Proc. of London Math. Soc.,* 2nd series, **30**, 264–286.

Ray-Chaudhuri, D.K., and Wilson, R.M. (1975): On t-designs, *Osaka J. Math.* **12**, 737–744.

Raz, R. (2000): The BNS–Chung criterion for multi-party communication complexity, *Computational Complexity* **9**, 113–122.

Razborov, A.A. (1985): Lower bounds for the monotone complexity of some boolean functions, *Soviet Math. Dokl.* **31**, 354-357.

Razborov, A.A. (1987): Lower bounds on the size of bounded-depth networks over a complete basis with logical addition, *Math. Notes of the Acad. of Sci. of the USSR* **41**:4, 333–338.

Razborov, A.A. (1988): Bounded-depth formulae over $\{\&, \oplus\}$ and some combinatorial problems, in: *Problems of Cybernetics, Complexity Theory and Applied Mathematical Logic*, S.I. Adian (ed.), VINITI, Moscow, 149–166. (Russian)

Razborov, A.A. (1990): Applications of matrix methods to the theory of lower bounds in computational complexity, *Combinatorica* **10**:1, 81–93.

Razborov, A.A. (1992): On submodular complexity measures, in: *Boolean Function Complexity*, M. Paterson (ed.), London Math. Society Lecture Note Series, Vol. 169, Cambridge University Press, 76–83.

Razborov, A.A. and Vereshchagin N.K. (1999): One property of cross-intersecting families, in: *Research Communications of the conference held in memory of Paul Erdős* in Budapest, Hungary, July 4–11, 1999, 218–220.

Rédei, L. (1934): Ein kombinatorischer Satz, *Acta Litt. Szeged* **7**, 39–43.

Reed, I. S. and Solomon, G. (1960): Polynomial codes over certain finite fields, *J. Soc. Indust. Appl. Math.* **8**, 300–304.

Reiman, I. (1958): Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* **9**, 269–279.

Rónyai, L., Babai, L., and Ganapathy,M.K. (2001): On the number of zero-patterns of a sequence of polynomials, *J. Amer. Math. Soc.*, 717–735.

Rosenbloom, A. (1997): Monotone circuits are more powerful than monotone boolean circuits, *Inform. Process. Letters* **61**:3, 161–164.

Ruciński, A. and Vince, A. (1986): Strongly balanced graphs and random graphs, *J. Graph Theory* **10**, 251–264.

Ruszinkó, M. (1994): On the upper bound of the size of the r-cover-free families, *J. Combin. Theory (A)* **66**, 302–310.

Rychkov K.L. (1985): A modification of Khrapchenko's method and its application to lower bounds for π-schemes of code functions, in: *Metody Diskretnogo Analiza*, 42 (Novosibirsk), 91–98. (Russian)

Ryser, H.J. (1951): A combinatorial theorem with an application to Latin rectangles, *Proc. Amer. Math. Soc.* **2**, 550–552.

Sauer, N. (1972): On the density of families of sets, *J. Combin. Theory (A)* **13**, 145–147.

Schmidt, W.M. (1976): Equations over finite fields, an elementary approach, Lect. Notes in Math., Vol. 536, Springer.

Schöning, U. (1999): A probabilistic algorithm for k-SAT and constraint satisfaction problems, in: *Proc. of 40th Ann. IEEE Symp. on Foundations of Comput. Sci.*, 410–414.

Schrijver,A. (2003): *Combinatorial Optimization*, Springer.

Schur, I. (1916): Über die Kongruenz $x^m + y^m \equiv z^m \pmod p$, *Jahresber. Deutsch. Math.-Verein* **25**, 114–116.

Schwartz, J.T. (1980): Fast probabilistic algorithms for verification of polynomial identities, *J. of the ACM* **27**:4, 701–717.

Seidenberg, A. (1959): A simple proof of a theorem of Erdős and Szekeres, *J. London Math. Soc.* **34**, 352.

Shelah, S. (1972): A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* **41**, 247–261.

Shelah, S. (1988): Primitive recursive bounds for van der Waerden's numbers, *J. Amer. Math. Soc.* **1**:3, 683–697.

Sipser, M. and Spielman, D.A. (1996): Expander codes *IEEE Trans. on Information Theory* **42**:6, 1710–1722.

Smetaniuk, B. (1981): A new construction on Latin squares I: A proof of the Evans conjecture, *Ars Combinatoria* **11**, 155–172.

Solymosi, J. (2005): On sum-sets and product-sets of complex numbers, *J. de Théorie des Nombres* **17**, 921–924.

Solymosi, J. (2009): Incidences and the spectra of graphs, in: Combinatorial Number Theory and Additive Group Theory, Advanced Courses in Mathematics — CRM Barcelona, Birkhäuser-Verlag, Basel, 299–314.

Spencer, J. (1977): Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20**, 69–76.

Spencer, J. (1987): *Ten lectures on the probabilistic method*, SIAM Regional Conference Series in Applied Mathematics, Vol. 52.

Spencer, J. (1990): Uncrowded hypergraphs, in: *Mathematics of Ramsey Theory*, Nesetril, J., and Rödl, V. (eds.), 253–262, Springer.

Spencer, J. (1995): Probabilistic methods, in: *Handbook of Combinatorics*, R.L. Graham, M. Grötschel and L. Lovász (eds.), Elsevier-Science, Vol. 2, 1786–1817.

Sperner, E. (1928): Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27**, 544–548.

Stein, S. K. (1974): Two combinatorial covering problems, *J. Combin. Theory (A)* **16**, 391–397.

Srinivasan, A. (2008): Improved algorithmic versions of the Lovász Local Lemma, in: *Proc. 19th Ann. ACM-SIAM Symp. on Discrete Algorithms*, 611—620.

Székely, L. A. (1997): Crossing numbers and hard Erdős problems in discrete geometry, *Comb. Prob. Comput.* **6**:3, 353–358.

Szele, T. (1943): Combinatorial investigations concerning complete directed graphs, *Matematikai Fizikai Lapok* **50**, 223–256 (Hungarian). German translation in: *Publ. Math. Debrecen* **13** (1966), 145–168.

Szeméredi, E. (1969): On sets of integers containing no four elements in aritmethic progression, *Acta. Math. Acad. Sci. Hungar.* **20**, 898–104.

Szeméredi, E. (1975): On sets of integers containing no k elements in arithmetic progression, *Acta Arithmetica* **27**, 199–245.

Szemerédi, E. and Trotter, W. T. (1983): Extremal problems in discrete geometry, *Combinatorica* **3**, 381–392.

Tardos, G. (1989): Query complexity, or why is it difficult to separate $NP^A \cap co - NP^A$ from P^A by a random oracle A?, *Combinatorica* **8**:4, 385–392.

Tesson, P. (2003): *Computational complexity questions related to finite monoids and semigroups*, PhD Thesis, McGill University, Montreal.

Trevisan, L. (2004): On local versus global satisfiability, *SIAM J. Discrete Math.* 17:4, 541–547.

Turán, P. (1934): On a theorem of Hardy and Ramanujan, *J. London. Math. Soc.***9**, 274–276.

Turán, P. (1941): On an extremal problem in graph theory, *Math. Fiz. Lapok* **48**, 436–452.

Tuza, Zs. (1984): Helly-type hypergraphs and Sperner families, *European J. Combin.* **5**, 185–187.

Tuza, Zs. (1985): Critical hypergraphs and intersecting set-pair systems, *J. Combin. Theory (B)* **39**, 134–145.

Tuza, Zs. (1989): The method of set-pairs for extremal problems in hypergraphs, in: *Finite and Infinite Sets*, A. Hajnal et al. (eds.), Colloq. Math. Soc. J. Bolyai, Vol. 37, 749–762, North-Holland, Amsterdam.

Tuza, Zs. (1994): Applications of the set-pair method in extremal hypergraphs, in: *Extremal Problems for Finite Sets*, P. Frankl et al. (eds.), Bolyai Society Mathematical Studies, Vol. 3, 479–514, János Bolyai Math. Society.

Tverberg, H. (1982): On the decomposition of K_n into complete bipartite graphs, *J. Graph Theory* **6**, 493–494.

Valiant, L. G. (1977): Graph-theoretic methods in low-level complexity, in *Proc. of 6th Conf. on Mathematical Foundations of Computer Science*, Lect. Notes in Comput. Sci., Vol. 53, 162–176, Springer.

Van der Waerden, B.L. (1927): Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15**, 212–216.

Vapnik, V.N. and Chervonenkis, A.Ya. (1971): On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and Appl.* **XVI**, 264–280.

Von Neumann, J. (1953): A certain zero-sum two-person game equivalent to the optimal assignment problem, in: *Contributions to the Theory of Games*, Vol. II, H.W. Kuhn (ed.), *Ann. of Math. Stud.* **28**, 5–12.

Ward, C. and Szabó, S. (1994): On swell colored complete graphs, *Acta Math. Univ. Comenianae* **LXIII**(2), 303–308.

Wei, V.K. (1981): A lower bound on the stability number of a simple graph. Technical Memorandum No. 81-11217-9, Bell Laboratories.

Weil, A. (1948): Sur les courbes algébraiques et les variétés qui s'en déduisent, *Actualités Sci. Ind.* **1041** (Herman, Paris).

Welsh, D.J.A. and Powell, M.B. (1967): An upper bound for the chromatic number of a graph and its applications to timetabling problems, *Computer Journal* **10**, 85–87.

Wigderson, A. (1993): The fusion method for lower bounds in circuit complexity, in: *Combinatorics, Paul Erdős is Eighty*, Vol. 1, 453–468.

Witt, E. (1951): Ein kombinatorischer Satz der Elementargeometrie, *Math. Nachr.* **6**, 261–262.

Wolff, T. (1999): Recent work connected with the Kakeya problem, *Prospects in Mathematics*, 129–162.

Yamamoto, K. (1954): Logarithmic order of free distributive lattices, *J. Math. Soc. Japan* **6**, 343–353.

Yannakakis, M. (1994): On the approximation of maximum satisfiability, J. of Algorithms **17**, 475–502.

Zagier, D. (1997): Newman's short proof of the prime number theorem, *Amer. Math. Monthly* **104**:8, 705–708.

Zarankiewicz, K. (1951): Problem P 101, *Colloq. Math.* **2**, 116–131.

Zippel, R.E. (1979): Probabilistic algorithms for sparse polynomials, in: *Proc. of EUROSAM 79*, Lect. Notes in Comput. Sci., Vol. 72, 216–226, Springer.

Index

S. Jukna, *Extremal Combinatorics*, Texts in Theoretical Computer Science.
An EATCS Series, DOI 10.1007/978-3-642-17364-6,