Discrete Mathematics
Lecture Notes
Incomplete Preliminary Version

Instructor: László Babai

Last revision: June 22, 2003
Last update: October 12, 2004

Contents

List of Figures

Chapter 1

Logic

TO BE WRITTEN.

1.1 Problems

1.2 Quantifier notation

Quantifier notation: $\forall$ - "universal quantifier," $\exists$ - "existential quatifier."

$(\forall x)$ is read as "for all x"

$(\exists x)$ is read as "there exists x **such that**"

$(\forall x, \text{statement}(x))$ is read as "for all x such that statement(x) holds,..."

Example. $(\forall x \neq 0)(\exists y)(xy = 1)$ says that every x other than zero has a multiplicative inverse. The validity of this statement depends on the universe over which the variables range. The statement holds (is true) over $\mathbb{R}$ (real numbers) and $\mathbb{Q}$ (rational numbers) but does not hold over $\mathbb{Z}$ (integers) or $\mathbb{N}$ (nonnegative integers). It holds over $\mathbb{Z}_m$ (the set of residue classes modulo m) if m is prime but not if m is composite. (Why?)

1.3 Problems

Several of the problems below will refer to the *divisibility* relation between integers.

Definition 1.3.1. Let a, b be integers. We say that $a \mid b$ ("a divides b") if $(\exists x)(ax = b)$. (The universe of the quantifiers is $\mathbb{Z}$, the set of integers (positive, negative, zero).)

From this definition we see that $7 \mid 21$ (because $x = 3$ satisfies $7x = 21$); $5 \mid -5$ (because $x = -1$ satisfies $5x = -5$); $0 \mid 0$ (because $x = 17$ (or any other x) satisfies $0x = 0$).

Does our conclusion $0 \mid 0$ violate the prohibition agains division by zero? By no means; division by zero continues to be a no-no. But read the definition of divisibility: it involves *multiplication*, not division. Nothing can stop us from *multiplying* a number by zero.

Remark. Most (but not all) Discrete Mathematics texts deliberately misstate the definition of divisibility to exclude $0 \mid 0$ from the definition. This abomination stems from many textbook authors' contempt for their readers' intellect; the result is a multitude of unnecessary case distinctions, destroying a fundamental element of mathematical aesthetics. (To these authors, for instance, $x \mid x$ does not hold for all x; there is an exception: $x = 0$. And then, to them, $x - y$ does not always divide $x^2 - y^2$; to them, the cases when $x = y$ are exceptions.) We do not follow this deplorable textbook trend; to us (as well as to any mathematician), $(\forall x)(x \mid x)$ and $(\forall x)(\forall y)(x - y \mid x^2 - y^2)$.

Exercise 1.3.2. Restate the following statements in plain English and prove them. The universe is $\mathbb{Z}$.

(a) $(\forall x)(x \mid x)$. In particular, $0 \mid 0$.

(b) $(\forall x)(\forall y)(x - y \mid x^2 - y^2)$.

(c) $(\forall x)(1 \mid x)$.

(d) $(\forall x)(x \mid 0)$.

(e) $(\forall x)(\text{ if } (\forall y)(x \mid y) \text{ then } x = \pm 1)$.

(f) $(\forall x)(\text{ if } (\forall y)(y \mid x) \text{ then } x = 0)$.

Definition 1.3.3. (Congruence) Let a, b, m be integers. We say that $a \equiv b \pmod{m}$ ("a is congruent to b modulo m") if $m \mid a - b$.

Examples: $11 \equiv -10 \pmod{-7}$ because $-7 \mid 11 - (-10) = 21$. Two integers are congruent modulo 2 exactly if they have the same parity (both are even or both are odd).

Exercise 1.3.4. Prove the following statements. The universe is $\mathbb{Z}$.

(a) $(\forall x)((\forall y)(\forall z)(y \equiv z \pmod{x}) \iff x = \pm 1)$.

(b) $(\forall x)(\forall y)(x \equiv y \pmod{0} \iff x = y)$.

(c) $(\forall x \neq \pm 1)(\forall y)(\exists z)(y \not\equiv z \pmod{x})$.

Exercise 1.3.5. Decide whether each of the following statements is true or false. *State* and *prove* your answers. In these statements, the universe for the variables x, y, k is $\mathbb{Z}$, the set of *integers*. Warning: in interpreting the formulas, *the order of the quantifiers matters!* $(\forall x)(\forall y)(P(x,y))$ is the same as $(\forall y)(\forall x)(P(x,y))$; $(\exists x)(\exists y)(P(x,y))$ is the same as $(\exists y)(\exists x)(P(x,y))$; but $(\forall x)(\exists y)(P(x,y))$ is NOT the same as $(\exists y)(\forall x)(P(x,y))$!

(a) $(\forall x)(\forall y)(x+y \mid x^2-y^2)$.

(b) $(\forall x)(\forall y)(x+y \mid x^2+y^2)$.

(c) $(\exists x)(\forall y)(x+y \mid x^2+y^2)$.

(d) $(\forall x)(\exists y)(x^2+y^2 \equiv 1 \pmod{x+y})$.

(e) $(\forall x)(\forall y)(\forall k)$ (if $k \geq 1$ then $x^k \equiv y^k \pmod{x-y}$).

(f) $(\forall x)(\exists y)(x \neq y$ and $x \mid y$ and $x \equiv y \pmod 7)$.

(g) $(\exists y)(\forall x)(x \neq y$ and $x \mid y$ and $x \equiv y \pmod 7)$.

(h) $(\forall x)(\forall y)($ if $x \mid y$ and $x \neq y$ then $x < y)$.

Exercise 1.3.6. True or false (prove your answer):

$$(\forall x)(\exists y)(\forall z)((x-5y)z \not\equiv 1 \pmod{17}).$$

(The universe of the variables is the set of integers.)

Negation of quantified formulas. If A is a statement then $\neg A$ denotes its negation; so $\neg A$ is true if and only if A is false. $\Leftrightarrow$ denotes logical equivalence ("if and only if").

Exercise 1.3.7. Let $P(x)$ be a statement in variable x.

(a) Prove: $\neg(\forall x)(P(x)) \Leftrightarrow (\exists x)(\neg P(x))$.

(b) Prove: $\neg(\exists x)(P(x)) \Leftrightarrow (\forall x)(\neg P(x))$.

(c) Let $Q(x,y)$ be a statement in two variables. Prove: $\neg(\forall x)(\exists y)(Q(x,y)) \Leftrightarrow (\exists x)(\forall y)(\neg Q(x,y))$.

Exercise 1.3.8. Let $P(x,y)$ be a statement about the variables x and y. Consider the following two statements: $A := (\forall x)(\exists y)(P(x,y))$ and $B := (\exists y)(\forall x)(P(x,y))$. The universe is the set of integers.

(a) Prove: $(\forall P)(B \Rightarrow A)$ ("B always implies A," i. e., for all P, if B is true then A is true).

(b) Prove: $\neg(\forall P)(A \Rightarrow B)$ (i. e., A does not necessarily imply B). In other words, $(\exists P)(A \not\Rightarrow B)$. To prove this, you need to construct a counterexample, i. e., a statement $P(x,y)$ such that the corresponding statement A is true but B is false. Make $P(x,y)$ as simple as possible. *Hint.* Three symbols suffice. These include x and y.

Quantifier alternation and games.

Exercise 1.3.9. Digest and generalize the following. Consider a chess-puzzle which says "white moves and wins in 2 moves." Let $W(x)$ denote the statement that the move x is available to White; and $B(x, y)$ that the move y is available to Black after White's move x; and $W(x, y, z)$ the statement that move z is avaliable to White after White moved x and Black moved y. Let $C(x, y, z)$ denote the statement that after moves x, y, z, Black is checkmated. Now the puzzle's claim can be formalized in the following quantified formula:

$$(\exists x, W(x))(\forall y, B(x, y))(\exists z, W(x, y, z))(C(x, y, z)).$$

1.4 Notation. Floor, ceiling

Notation: $\exp(x) = e^x$.

In combinatorial contexts, the symbol $[n]$ will be used to denote $\{1, 2, \ldots, n\}$. This is not be confused with the **floor** and **ceiling** notations: if x is a real number, the *floor* of x, denoted by $\lfloor x \rfloor$, is the greatest integer $\leq x$. For instance, $\lfloor 5.7 \rfloor = 5$, $\lfloor -5.7 \rfloor = -6$, $\lfloor 5 \rfloor = 5$. The *ceiling* of x, denoted by $\lceil x \rceil$, is the smallest integer $\geq x$. So $\lceil 5.7 \rceil = 6$, $\lceil -5.7 \rceil = -5$, $\lceil 5 \rceil = 5$.

Exercise 1.4.1. $|\lceil x \rceil - \lfloor x \rfloor| \leq 1$.

Exercise 1.4.2. Prove: $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x + y \rfloor$. When is the left-hand side strctly less than the right-hand side?

Exercise 1.4.3. (a) Prove: if k is a positive integer then

$$\left\lfloor \frac{x}{k} \right\rfloor = \left\lfloor \frac{\lfloor x \rfloor}{k} \right\rfloor.$$

(b) Show that this statement becomes false if k is not an integer: for every $k > 0$ that is not an integer, find x such that the two sides are not equal.

Exercise 1.4.4. Let p be a prime and p^s be the largest power of p which divides $n!$. Prove:

$$s = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor < \frac{n}{p-1}.$$

1.5 Limit of sequence

Definition 1.5.1 (finite limit of a sequence). Let $\{a_n\}$ be a sequence of real or complex numbers. We write $\lim_{n\to\infty} a_n = c$ (or simply $a_n \to c$) if

$$(\forall \epsilon > 0)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(|a_n - c| \leq \epsilon).$$

We say that a sequence *converges* if it has a finite limit.

Definition 1.5.2 (infinite limit of a sequence). Let a_n be a sequence of real or complex numbers. We write $\lim_{n\to\infty} a_n = \infty$ (or simply $a_n \to \infty$) if

$$(\forall L)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(a_n \geq L).$$

Exercise 1.5.3. Prove: $(\forall x \in \mathbb{R})\,(\lim_{n\to\infty}(1 + x/n)^n = e^x)$.

Exercise 1.5.4. (a) Consider the sequence $\{a_n\}$ defined by the recurrence $a_{n+1} = \sqrt{2}^{a_n}$ with the initial condition $a_0 = 1$. Prove that $\lim_{n\to\infty} a_n$ exists; find the limit.

(b) Prove that the previous statement becomes false if we replace $\sqrt{2}$ by 1.5. What is the largest number (in place of $\sqrt{2}$) for which the sequence converges?

1.6 Asymptotic Equality and Inequality

Often, we are interested in comparing the rate of growth of two functions, as inputs increase in length. Asymptotic equality is one formalization of the idea of two functions having the "same rate of growth."

Definition 1.6.1. We say a_n is *asymptotically equal* to b_n (denoted $a_n \sim b_n$) if $\lim_{n\to\infty} a_n/b_n = 1$. For the purposes of this definition only, we set $0/0 = 1$.

Observation. If $c \neq 0$ is a constant then the statement $a_n \sim c$ (where c means the sequence $c, c, \dots$) is equivalent to $a_n \to c$ (where c means the number c).

Exercise 1.6.2. Prove: $a_n \sim 0$ if and only if $(\exists n_0)(\forall n \geq n_0)(a_n = 0)$, i. e., $a_n = 0$ for all sufficiently large n.

Exercise 1.6.3. (a) Let $a_n \to \infty$. Prove: $a_n \sim \lfloor a_n \rfloor$.

(b) Let $a_n \to c$ be a finite limit. True or false:

(i) If c is not an integer then $a_n \not\sim \lfloor a_n \rfloor$.

(ii) If $c = 0$ then $a_n \not\sim \lfloor a_n \rfloor$.

(iii) If c is a positive integer then $a_n \sim \lfloor a_n \rfloor$.

Exercise 1.6.4. Let $\mathcal{S}$ denote the set of sequences of real or complex numbers. Prove that $\sim$ is an *equivalence relation* on $\mathcal{S}$, i. e., the relation "$\sim$" is

(a) *reflexive:* $a_n \sim a_n$;

(b) *symmetric:* if $a_n \sim b_n$ then $b_n \sim a_n$; and

(c) *transitive:* if $a_n \sim b_n$ and $b_n \sim c_n$ then $a_n \sim c_n$.

Exercise 1.6.5. Prove: if $a_n \sim b_n$ and $c_n \sim d_n$ then $a_n c_n \sim b_n d_n$. If, moreover, $c_n d_n \neq 0$ for all sufficiently large n then $a_n/c_n \sim b_n/d_n$. (Note that a finite number of undefined terms do not invalidate a limit relation.)

Exercise 1.6.6. Consider the following statement.

$$\text{If } a_n \sim b_n \text{ and } c_n \sim d_n \text{ then } a_n + c_n \sim b_n + d_n. \tag{1.1}$$

1. Prove that (1.1) is false.

2. Prove: if $a_n c_n > 0$ then (1.1) is true. *Hint.* Prove: if $a, b, c, d > 0$ and $a/b < c/d$ then $a/b < (a+c)/(b+d) < c/d$.

Exercise 1.6.7. 1. If $f(x)$ and $g(x)$ are polynomials with respective leading terms ax^n and bx^m then $f(n)/g(n) \sim (a/b)x^{n-m}$.

2. $\sin(1/n) \sim 1/n$.

3. $\ln(1 + 1/n) \sim 1/n$.

4. $\sqrt{n^2 + 1} - n \sim 1/2n$.

5. If f is a function, differentiable at zero, $f(0) = 0$, and $f'(0) \neq 0$, then $f(1/n) \sim f'(0)/n$. See that items 2–4 in this exercise follow from this.

Exercise 1.6.8. Find two sequences of positive real numbers, $\{a_n\}$ and $\{b_n\}$, such that $a_n \sim b_n$ but $a_n^n \not\sim b_n^n$.

Next we state some of the most important asymptotic relations in mathematics.

Theorem 1.6.9 (Stirling's Formula).

$$n! \sim \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}.$$

Exercise 1.6.10. Prove: $\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$.

Exercise 1.6.11. Give a very simple proof, without using Stirling's formula, that $\ln(n!) \sim n \ln n$.

Hint. It is obvious that $\ln(n!) \leq n \ln n$ (why?). According to the definition of limits, we need to prove that $(\forall \epsilon > 0)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(\ln(n!) > (1 - \epsilon) n \ln n$. For $\delta > 0$, let $N(n, \delta) = \prod\{k \,|\, n^{1-\delta} \leq k \leq n\}$. Observe that $n! \geq N(n, \delta) \geq n^{(1-\delta)s}$ where $s = n - \lceil n^{1-\delta} \rceil$.

Theorem 1.6.12 (The Prime Number Theorem). *Let $\pi(x)$ be the number of primes less than or equal to x.*

$$\pi(x) \sim \frac{x}{\ln x},$$

where ln *denotes the natural logarithm function.*

Exercise 1.6.13. Let p_n be the n-th prime number. Prove, using the Prime Number Theorem, that $p_n \sim n \ln n$.

Exercise 1.6.14. *Feasibility of generating random prime numbers.* Estimate, how many random ≤ 100-digit integers should we expect to pick before we encounter a prime number? (We generate our numbers by choosing the 100 digits independently at random (initial zeros are permitted), so each of the 10^{100} numbers has the same probability to be chosen.) Interpret this question as asking the reciprocal of the probability that a randomly chosen integer is prime.

Definition 1.6.15. A *partition* of a positive integer n is a representation of n as a sum of positive integers: $n = x_1 + \cdots + x_k$ where $x_1 \leq \cdots \leq x_k$. Let $p(n)$ denote the number of partitions of n.

Examples: $p(1) = 1$, $p(2) = 2$, $p(3) = 3$, $p(4) = 5$. The 5 representations of 4 are $4 = 4$; $4 = 1 + 3$; $4 = 2 + 2$; $4 = 1 + 1 + 2$; $4 = 1 + 1 + 1 + 1$. One of the most amazing asymptotic formulas in discrete mathematics gives the growth of $p(n)$.

Theorem 1.6.16 (Hardy-Ramanujan Formula).

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\frac{2\pi}{\sqrt{6}}\sqrt{n}\right). \tag{1.2}$$

Definition 1.6.17. Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. We say that a_n is *greater than or asymptotically equal to* b_n, denoted as $a_n \gtrsim b_n$ if $a_n \sim \max\{a_n, b_n\}$.

Exercise 1.6.18. Prove: $a_n \gtrsim b_n$ if and only if $b_n \sim \min\{a_n, b_n\}$.

Exercise 1.6.19. Prove: if $a_n \sim b_n$ then $a_n \gtrsim b_n$.

Exercise 1.6.20. Prove: if $a_n \gtrsim b_n$ and $b_n \gtrsim a_n$ then $a_n \sim b_n$.

Exercise 1.6.21. Prove: if $a_n \gtrsim b_n$ and $b_n \gtrsim c_n$ then $a_n \gtrsim c_n$.

Exercise 1.6.22. Conclude from the preceding exercises that the "$\gtrsim$" relation is a partial order on the set of asymptotic equivalence classes of sequences of real numbers.

Exercise 1.6.23. Prove: $a_n \gtrsim 0$ if and only if $(\exists n_0)(\forall n \geq n_0)(a_n \geq 0)$, i. e., $a_n \geq 0$ for all sufficiently large n.

Exercise 1.6.24. Prove: if $a_n \gtrsim b_n \geq 0$ and $c_n \gtrsim d_n \geq 0$ then $a_n + c_n \gtrsim b_n + d_n$.

Exercise 1.6.25. (a) Let $a_n, b_n \geq 0$. Prove that $a_n \gtrsim b_n$ if and only if $(\forall \epsilon > 0)(\exists n_0)(\forall n > n_0)(a_n \geq b_n(1 - \epsilon))$.

(b) Show that the same formula does not define the relation "$a_n \gtrsim b_n$" if we omit the condition $a_n, b_n \geq 0$.

Exercise 1.6.26. Assume $b_n \to \infty$ and $a_n \geq b_n^2 \ln b_n$. Prove: $b_n \lesssim c\sqrt{a_n / \ln a_n}$, where c is a constant. Determine the smallest value of c for which this statement follows from the assumptions.

1.7 Little-oh notation

Definition 1.7.1. We say that $a_n = o(b_n)$ ("a_n is little oh of b_n") if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

For the purposes of this definition only, we set $0/0 = 0$, so if $a_n = b_n = 0$ then $a_n = o(b_n)$ as well as $a_n \sim b_n$.

Exercise 1.7.2. Prove: if $a_n = o(b_n)$ and $a_n \sim b_n$ then for all sufficiently large n we must have $a_n = b_n = 0$.

Observation. $a_n = o(1)$ means $\lim_{n\to\infty} a_n = 0$.

Exercise 1.7.3. Prove: $a_n = o(b_n)$ if and only if there exists a sequence c_n such that $|c_n| \to \infty$ and $a_n = b_n/c_n$.

Exercise 1.7.4. Show: if $a_n = o(c_n)$ and $b_n = o(c_n)$ then $a_n \pm b_n = o(c_n)$.

Exercise 1.7.5. Consider the following statement:

$$\text{If } a_n = o(b_n) \text{ and } c_n = o(d_n) \text{ then } a_n + c_n = o(b_n + d_n). \tag{1.3}$$

1. Show that statement (1.3) is false.
2. Prove that statement (1.3) becomes true if we assume $b_n, d_n > 0$.

Exercise 1.7.6. Show that $a_n \sim b_n \iff a_n = b_n(1 + o(1))$.

Exercise 1.7.7. Use the preceding exercise to give a second proof of (1.1) when $a_n, b_n, c_n, d_n > 0$.

Exercise 1.7.8. Construct sequences $a_n, b_n > 1$ such that $a_n = o(b_n)$ and $\ln a_n \sim \ln b_n$.

Exercise 1.7.9. Let $a_n, b_n > 1$. (a) Prove that the relation $a_n = o(b_n)$ does NOT follow from the relation $\ln a_n = o(\ln b_n)$. (b) If we additionally assume that $b_n \to \infty$ then $a_n = o(b_n)$ DOES follow from $\ln a_n = o(\ln b_n)$.

1.8 Big-Oh, Omega, Theta notation (O, Ω, Θ)

Definition 1.8.1. We say that

1. $a_n = O(b_n)$ (a_n is "big oh" of b_n) if $|a_n/b_n|$ is bounded (0/0 counts as "bounded"), i. e.,
$$(\exists C > 0, n_0 \in \mathbb{N})(\forall n > n_0)(|a_n| \le C|b_n|).$$
2. $a_n = \Omega(b_n)$ if $b_n = O(a_n)$, i. e., if $|b_n/a_n|$ is bounded $(\exists c > 0, n_0 \in \mathbb{N})(\forall n > n_0)(|a_n| \ge c|b_n|)$
3. $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$, i. e.,
$$(\exists C, c > 0, n_0 \in \mathbb{N})(\forall n > n_0)(c|b_n| \le |a_n| \le C|b_n|).$$

Exercise 1.8.2. Suppose the finite or infinite limit $\lim_{n\to\infty} |a_n/b_n| = L$ exists. Then

(a) $b_n = o(a_n)$ if and only if $L = \infty$;

(b) $a_n = o(b_n)$ if and only if $L = 0$; and

(c) $a_n = \Theta(b_n)$ if and only if $0 < L < \infty$.

Exercise 1.8.3. Construct sequences $a_n, b_n > 0$ such that $a_n = \Theta(b_n)$ but the limit $\lim_{n\to\infty} a_n/b_n$ does not exist.

Exercise 1.8.4. Let $a_n, b_n > 0$. Show: $a_n = \Theta(b_n) \iff \ln a_n = \ln b_n + O(1)$.

Exercise 1.8.5. Show: if $a_n = O(c_n)$ and $b_n = O(c_n)$ then $a_n + b_n = O(c_n)$.

Exercise 1.8.6. Consider the statement "if $a_n = \Omega(c_n)$ and $b_n = \Omega(c_n)$ then $a_n + b_n = \Omega(c_n)$. (a) Show that this statement is false. (b) Show that if we additionally assume $a_n b_n > 0$ then the statement becomes true.

Exercise 1.8.7. Let $a_n, b_n > 1$. Suppose $a_n = \Theta(b_n)$. Does it follow that $\ln a_n \sim \ln b_n$?

1. Show that even $\ln a_n = \Omega(\ln b_n)$ does not follow.
2. Show that if $a_n \to \infty$ then $\ln a_n \sim \ln b_n$ follows.

Exercise 1.8.8. Let $a_n, b_n > 1$. Suppose $a_n = \Omega(b_n)$. Does it follow that $\ln a_n \gtrsim \ln b_n$?

1. Show that even $\ln a_n = \Omega(\ln b_n)$ does not follow.
2. Show that if $a_n \to \infty$ then $\ln a_n \gtrsim \ln b_n$ follows.

Exercise 1.8.9. Let $a_n, b_n > 0$. Consider the relations

(A) $a_n = O(2^{b_n})$ and (B) $a_n = 2^{O(b_n)}$.

(a) Prove: the relation (B) does NOT follow from (A).

(b) Prove: if $a_n > 0.01$ and $b_n > 0.01$ then (B) DOES follow from (A).

Note. $a_n = 2^{O(b_n)}$ means that $a_n = 2^{c_n}$ where $c_n = O(b_n)$.

Exercise 1.8.10. Prove: if $a_n = \Omega(b_n)$ and $a_n = \Omega(c_n)$ then $a_n = \Omega(b_n + c_n)$.

Note. We say that the "statement A *implies* statement B" if B follows from A.

Exercise 1.8.11. (a) Prove that the relations $a_n = O(b_n)$ and $a_n = O(c_n)$ do NOT imply $a_n = O(b_n + c_n)$.

(b) Prove that if $a_n, b_n > 0$ then the relations $a_n = O(b_n)$ and $a_n = O(c_n)$ DO imply $a_n = O(b_n + c_n)$.

Exercise 1.8.12. Prove: $\sum_{i=1}^{n} 1/i = \ln n + O(1)$.

1.9 Prime Numbers

Exercise$^+$ 1.9.1. Let $P(x)$ denote the product of all prime numbers $\le x$. Consider the following statement: $\ln P(x) \sim x$. Prove that this statement is equivalent to the Prime Number Theorem.

Exercise$^+$ 1.9.2. Prove, without using the Prime Number Theorem, that

$$\ln P(x) = \Theta(x).$$

Hint. For the easy upper bound, observe that the binomial coefficient $\binom{2n}{n}$ is divisible by the integer $P(2n)/P(n)$. This observation yields $P(x) \le 4^x$. For the lower bound, prove that if a prime power p^t divides the binomial coefficient $\binom{n}{k}$ then $p^t \le n$. From this it follows that $\binom{2n}{n}$ divides the product $P(2n)P((2n)^{1/2})P((2n)^{1/3})P((2n)^{1/4})\ldots$. Use the upper bound to estimate all but the first term in this product.

1.10 Partitions

Exercise 1.10.1. Let $p(n,k)$ denote the number of those partitions of n which have at most k terms. Let $q(n,k)$ denote the number of those partitions in which every term is $\le k$. Observe that $p(n,1) = q(n,1) = 1$ and $p(n,n) = q(n,n) = p(n)$. (Do!) Let $\widetilde{p}(n) = \sum_{i=0}^{n} p(i)$ and let $\widetilde{p}(n,k) = \sum_{i=0}^{n} p(i,k)$.

1. Prove: $p(n,k) = q(n,k)$.
2. Compute $p(n,2)$. Give a very simple formula.
3. Compute $p(n,3)$. Give a simple formula.
4. Prove: $\widetilde{p}(n) \le \widetilde{p}(n,k)^2$, where $k = \lfloor \sqrt{n} \rfloor$. *Hint.* Use part 1 of this exercise.

Exercise 1.10.2. Using the notation proved in Exercise 1.10.1, prove the following.

(a) $\widetilde{p}(n,k) < \binom{n+k}{k}$

(b) $\log p(n) = O(\sqrt{n} \log n)$. *Hint.* Use (a) and part 4 of Exercise 1.10.1.

Exercise$^+$ 1.10.3. Prove, without using the Hardy–Ramanujan formula, that

$$\ln p(n) = \Theta(\sqrt{n}).$$

Hint. $\ln p(n) = \Omega(\sqrt{n})$ is easy (2 lines). The upper bound is harder. Use the preceding exercise, especially item 4. When estimating $p(n, \sqrt{n})$, split the terms of your partition into sets $\{x_i \le \sqrt{n}\}$, $\{\sqrt{n} < x_i \le 2\sqrt{n}\}$, $\{2\sqrt{n} < x_i \le 4\sqrt{n}\}$, $\{4\sqrt{n} < x_i \le 8\sqrt{n}\}$, etc.

Exercise$^+$ 1.10.4. Let $p'(n)$ denote the number of partitions of n such that all terms are primes or 1. Example: $16 = 1+1+1+3+3+7$. Prove:

$$\ln p'(n) = \Theta\left(\sqrt{\frac{n}{\ln n}}\right).$$

Exercise 1.10.5. Let $r(n)$ denote the number of different integers of the form $\prod x_i!$ where $x_i \geq 1$ and $\sum x_i = n$. (The x_i are integers.) Prove:

$$p'(n) \leq r(n) \leq p(n).$$

OPEN QUESTIONS. Is $\log r(n) = \Theta(\sqrt{n})$? Or perhaps, $\log r(n) = \Theta(\sqrt{n/\log n})$? Or maybe $\log r(n)$ lies somewhere between these bounds?

1.11 Problems

Exercise 1.11.1. 1. (1 point) Describe in words what it means for a sequence a_n that $a_n = O(1)$ (big-Oh of 1).

2. (2 points) Suppose $a_n = O(1)$. Does it follow that the sequence a_n has a limit? (Prove your answer.)

3. (2 points) Suppose the sequence a_n has a finite limit. Does it follow that $a_n = O(1)$? Prove your answer.

Exercise 1.11.2. Let $a_n, b_n > 1$. True or false: if $a_n \sim b_n$ then $a_n^n = \Theta(b_n^n)$. Prove your answer.

Exercise 1.11.3. Prove: if $a_n, b_n, c_n, d_n > 0$ and $a_n = O(b_n)$ and $c_n = O(d_n)$ then $a_n + c_n = O(b_n + d_n)$. State the constant implicit in the conclusion as a function of the constants implicit in the conditions.

Exercise 1.11.4. Using the fact that $\ln x = o(x)$, prove that $(\ln y)^{100} = o(\sqrt{y})$. $(x, y \to \infty.)$ Do not use calculus.

Exercise 1.11.5. True or false (prove your answer):

$$2^{\binom{n}{2}} \sim 2^{n^2/2}.$$

Exercise 1.11.6. Construct two sequences, $\{a_n\}$ and $\{b_n\}$ such that $a_n > 1$, $b_n > 1$, $a_n \sim b_n$, and $a_n^n = o(b_n^n)$.

Exercise 1.11.7. Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers. Prove: if $a_n \to \infty$ and $a_n = \Theta(b_n)$ then $\ln(a_n) \sim \ln(b_n)$.

Last update: October 12, 2004

Exercise 1.11.8. Recall that a sequence $\{a_n\}$ is *polynomially bounded* if $(\exists C)(a_n = O(n^C))$. Decide whether or not each of the following sequences is polynomialy bounded. Prove your answers.

1. $n^3 \ln(n^2 + 5)$
2. $5^{\ln n}$
3. $\lfloor \ln n \rfloor !$

Exercise 1.11.9. Construct two sequences, $\{a_n\}$ and $\{b_n\}$ such that $a_n > 1$, $b_n > 1$, $a_n \sim b_n$, and $a_n^n = o(b_n^n)$.

Exercise 1.11.10. Let $f_n = (1 + 1/\sqrt{n})^n$ and $g_n = \mathrm{e}^{\sqrt{n}}$. Prove: $f_n = \Theta(g_n)$ but $f_n \not\sim g_n$. Show that in fact $\lim_{n\to\infty} f_n/g_n = 1/\sqrt{\mathrm{e}}$.

Exercise 1.11.11. Consider the statement

$\lim x^y = 1$ is "almost always true" as $x, y \to 0^+$.

Give a definition of "almost always" in this context, then prove the statement.

Exercise 1.11.12. Let $\{a_n\}$ be a sequence of positive integers, and assume $a_n \to \infty$. Let $b_n = \binom{a_n}{3}$. Prove that $a_n \sim c \cdot b_n^d$ for some constants c, d. Determine the values of c and d.

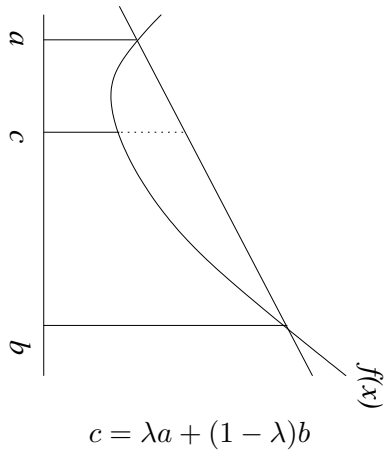


Figure 1.1: Definition of convexity

Definition 1.1.1. Let $f(x)$ be a real function defined over a finite or infinite interval. We say that $f(x)$ is a *convex* function if for all a, b in its domain and all real numbers λ in the interval $0 \le \lambda \le 1$, the inequality

$$f(\lambda a + (1-\lambda)b) \le \lambda f(a) + (1-\lambda)f(b)$$

holds. The function $g(x)$ is *concave* if $-g(x)$ is convex. See Figure 1.11.

Exercise 1.1.2. Prove the following sufficient condition of convexity: If $f(x)$ is twice differentiable and its second derivative is always ≥ 0 then $f(x)$ is convex.

Exercise 1.1.3. Prove the following sufficient condition of convexity: If $f(x)$ is continuous and the inequality $f\left(\dfrac{a+b}{2}\right) \le \dfrac{f(a)+f(b)}{2}$ holds for all a, b in its domain then $f(x)$ is convex.

Exercise 1.1.4. (a) The functions x^2, $\binom{x}{2}$, e^x are convex. (b) The functions $\sqrt{x}$, $\ln x$ are concave. (c) The function $\sin x$ is concave over the interval $[0, \pi]$ and convex over the interval $[\pi, 2\pi]$.

Exercise 1.1.5. (a) A continuous convex function is *unimodal:* it decreases to its minimum and then it increases. (b) If a continuous convex function is invertible then it is monotone (increasing or decreasing). (c) The inverse of a monotone increasing continuous convex function is concave. (d) The inverse of a monotone decreasing convex function is convex.

Theorem 1.1.6 (Jensen's Inequality). *If $f(x)$ is a convex function then for any choice of real numbers $x_1, \ldots, x_k$ from the domain of f,*

$$f\left(\frac{\sum_{i=1}^k x_i}{k}\right) \le \frac{\sum_{i=1}^k f(x_i)}{k}.$$

Exercise 1.1.7. Prove Jensen's Ineqality. *Hint.* Induction on k.

Exercise 1.1.8. Prove the inequality between the **arithmetic and quadratic means:** for all real $x_1, \ldots, x_k$,

$$\frac{x_1 + \cdots + x_k}{k} \leq \sqrt{\frac{x_1^2 + \cdots + x_k^2}{k}}.$$

Hint 1. Use the convexity of $f(x) = x^2$ and Jensen's Inequality.
Hint 2. Give a 1-line proof using the Cauchy–Schwarz Inequality.
Hint 3. Give a simple direct proof (do not use either Jensen's Inequality or Cauchy–Schwarz).

Exercise 1.1.9. In the proof of the Kővári–Sós–Turán theorem (Exercise 3.1.27), we applied Jensen's Inequality to $f(x) = \binom{x}{2} = x(x-1)/2$. Modify the proof so that Jensen's Inequality is avoided and the inequality between the arithmetic and quadratic means is used instead.

Exercise 1.1.10. Prove the inequality between the **arithmetic and geometric means:** if $x_1, \ldots, x_k > 0$ then

$$\frac{x_1 + \cdots + x_k}{k} \geq (x_1 x_2 \ldots x_k)^{1/k}.$$

Hint. Use the concavity of the natural logarithm function, ln.

1.2 Gcd, congruences

Exercise 1.2.1. Prove that the product of n consecutive integers is always divisible by $n!$. *Hint.* One-line proof.

Exercise 1.2.2. (The Divisor Game) Select an integer $n \geq 2$. Two players alternate naming positive divisors of n subject to the following rule: no divisor of any previously named integer can be named. The first player forced to name "n" loses. Example: if $n = 30$ then the following is a possible sequence of moves: 10, 3, 6, 15, at which point it is the first player's move; he is forced to say "30" and loses.

1. Find a winning strategy for the first player when n is a prime power; or of the form pq^k; p^kq^k; pqr; or $pqrs$, where p, q, r, s are prime and k is a positive integer.

2. Prove: $\forall n \geq 2$, the first player has a winning strategy. (*Hint:* prove, in two or three lines, the *existence* of a winning strategy.)

Notation 1.2.3. Let $\mathrm{Div}\,(n)$ denote the set of positive divisors of n.

Exercise 1.2.4. Prove, for all $a, b \in \mathbb{Z}$,

$$(\mathrm{Div}\,(a) \subseteq \mathrm{Div}\,(b)) \quad \Longleftrightarrow \quad a \,|\, b.$$

Exercise$^+$ 1.2.5. Prove: $(\forall a, b)(\exists d)(\mathrm{Div}\,(a) \cap \mathrm{Div}\,(b) = \mathrm{Div}\,(d))$. A nonnegative d satisfying this statement is called the g.c.d. of a and b. Note that $\text{g.c.d.}\,(a, b) = 0 \Longleftrightarrow a = b = 0$. Define l.c.m. analogously. When is $\text{l.c.m.}\,(a, b) = 0$?

Exercise 1.2.6. Prove: $\text{g.c.d.}\,(a^k - 1, a^\ell - 1) = a^d - 1$, where $d = \text{g.c.d.}\,(k, \ell)$.

Definition 1.2.7. The Fibonacci numbers are defined by the recurrence $F_n = F_{n-1} + F_{n-2}$, $F_0 = 0$, $F_1 = 1$.

Exercise$^+$ 1.2.8. Prove: $\text{g.c.d.}\,(F_k, F_\ell) = F_d$, where $d = \text{g.c.d.}\,(k, \ell)$.

Exercise 1.2.9. Prove: if $a \equiv b \pmod{m}$ then $\text{g.c.d.}\,(a, m) = \text{g.c.d.}\,(b, m)$.

Exercise 1.2.10. Prove: if $a, b \geq 0$ then $\text{g.c.d.}\,(a, b) \cdot \text{l.c.m.}\,(a, b) = ab$.

Exercise 1.2.11. Prove: congruence modulo m is an equivalence relation on $\mathbb{Z}$. The equivalence classes are called the *residue classes* mod m. There are m residue classes modulo m. Under the natural operations they form the ring $\mathbb{Z}/m\mathbb{Z}$. The additive group of this ring is cyclic.

Exercise 1.2.12. Prove that the sequence of Fibonacci numbers mod m is periodic. The length of the period is $\leq m^2 - 1$.

Exercise 1.2.13. An *integer-preserving polynomial* is a polynomial $f(x)$ such that $(\forall a \in \mathbb{Z})(f(a) \in \mathbb{Z})$. Prove that $f(x)$ is integer-preserving if and only if it can be written as

$$f(x) = \sum_{i=0}^{n} a_i \binom{x}{i} \tag{1.4}$$

with suitable integer coefficients a_i. Here

$$\binom{x}{i} = \frac{x(x-1)\dots(x-i+1)}{i!}; \qquad \binom{x}{0} = 1.$$

Exercise 1.2.14. A *congruence-preserving polynomial* is an integer-preserving polynomial such that $(\forall a, b, m \in \mathbb{Z})(a \equiv b \pmod m \Rightarrow f(a) \equiv f(b) \pmod m)$. Prove that $f(x)$ is congruence-preserving if and only if $(\forall i)(e_i \mid a_i)$ in the expression (1.4), where $e_i = \text{l.c.m.}\,(1, 2, \dots, i)$.

Exercise 1.2.15. A *multiplicative inverse* of a modulo m is an integer x such that $ax \equiv 1 \pmod m$; notation: $x = a^{-1} \bmod m$. Prove: $\exists a^{-1} \bmod m \iff \text{g.c.d.}\,(a, m) = 1$.

Exercise 1.2.16. (Wilson's theorem) Prove: $(p-1)! \equiv -1 \pmod p$. *Hint:* match each number with its multiplicative inverse in the product $(p-1)!$

Exercise 1.2.17. Prove: if $\text{g.c.d.}\,(a, p) = 1$ then $\prod_{j=1}^{p-1} j \equiv \prod_{i=1}^{p-1}(ai)$. *Hint.* Match terms on the right hand side with terms on the left hand side so that corresponding terms satisfy $j \equiv ai \pmod p$.

Theorem 1.2.18 (Fermat's little Theorem). If $\text{g.c.d.}\,(a, p) = 1$ then $a^{p-1} \equiv 1 \pmod p$.

Exercise 1.2.19. Infer Fermat's little Theorem from Exercise 1.2.17.

Exercise 1.2.20. Use the same idea to prove the **Euler–Fermat theorem:** if $\text{g.c.d.}\,(a, m) = 1$ then $a^{\varphi(m)} \equiv 1 \pmod m$. ($\varphi$ is Euler's φ function, see Definition 1.3.1).

Exercise 1.2.21. Prove: if p is a prime and f is a polynomial with integer coefficients then $f(x)^p \equiv f(x^p) \pmod p$. Here the congruence of two polynomials means coefficientwise congruence.

The multiplicative group $(\mathbb{Z}/m\mathbb{Z})^\times$ consists of the mod m residue classes relatively prime to m. Its order is $\varphi(m)$. For a review of related concepts in abstract algebra, see Chapter 6 (cf. especially Exercise 6.2.6).

Exercise$^+$ 1.2.22. Prove: if p is a prime then $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic (see Definition 6.1.11). A generator of this group is called a *primitive root mod p*.

Exercise$^+$ 1.2.23. Prove: if p is an odd prime then $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is cyclic.

Exercise$^+$ 1.2.24. If $k \geq 2$ then the group $(\mathbb{Z}/2^k\mathbb{Z})^\times$ is not cyclic but the direct sum of a cyclic group of order 2 and a cyclic group of order 2^{k-2}.

1.3 Arithmetic Functions

Definition 1.3.1 (Euler's Phi Function).

$$\begin{aligned}\varphi(n) &= \Big|\big\{k \in [n] \,:\, \text{g.c.d.}\,(k,n) = 1\big\}\Big| \\ &= \text{number of positive integers not greater than } n \text{ which are relatively prime to } n\end{aligned}$$

Exercise 1.3.2. Show that the number of complex primitive n-th roots of unity is $\varphi(n)$. Show that if $d|n$ then the number of elements of order d in a cyclic group of order n is $\varphi(d)$.

Exercise 1.3.3. Show

$$\sum_{d\,|\,n} \varphi(d) = n.$$

Exercise$^+$ 1.3.4. Let $D_n = (d_{ij})$ denote the $n \times n$ matrix with $d_{ij} = \text{g.c.d.}\,(i,j)$. Prove:

$$\det D_n = \varphi(1)\varphi(2)\cdots\varphi(n).$$

(*Hint.* Let $Z = (z_{ij})$ be the matrix with $z_{ij} = 1$ if $i \,|\, j$ and $z_{ij} = 0$ otherwise. Consider the matrix $Z^T F Z$ where F is the diagonal matrix with entries $\varphi(1), \ldots, \varphi(n)$ and Z^T is "Z-transpose" (reflection in the main diagonal).)

Definition 1.3.5 (Number of [positive] divisors).

$$d(n) = \Big|\big\{d \in \mathbb{N} : d \,|\, n\big\}\Big|$$

Exercise 1.3.6. Prove: $d(n) < 2\sqrt{n}$.

Exercise$^+$ 1.3.7. Prove: $(\forall \epsilon > 0)(\exists n_0)(\forall n > n_0)(d(n) < n^\epsilon)$. (*Hint.* Use a consequence of the Prime Number Theorem (Theorem 1.4.6 in the next section).) Prove that $d(n) < n^{c/\ln\ln n}$ for some constant c. The best asymptotic constant is $c = \ln 2 + o(1)$.

Exercise$^+$ 1.3.8. Prove that for infinitely many values of n the reverse inequality $d(n) > n^{c/\ln\ln n}$ holds (with another constant $c > 0$). (Again, use the PNT.)

Exercise$^+$ 1.3.9. Let $D(n) = (1/n)\sum_{i=1}^{n} d(i)$ (the average number of divisors). Prove: $D(n) \sim \ln(n)$. (*Comment.* If we pick an integer t at random between 1 and n then $D(n)$ will be the *expected number* of divisors of t. – Make your proof very simple (3 lines). Do not use the PNT.)

Exercise$^+$ 1.3.10. Prove: $(1/n)\sum_{i=1}^{n} d(i)^2 = \Theta((\ln n)^3)$.

Definition 1.3.11 (Sum of [positive] divisors).

$$\sigma(n) = \sum_{d\,|\,n} d$$

Definition 1.3.12 (Number of [distinct] prime divisors). Let $n = p_1^{k_1} \cdots p_r^{k_r}$ where the p_i are distinct primes and $k_i > 0$. Set $\nu(n) = r$ (number of distinct prime divisors; so $\nu(1) = 0$). Set $\nu^*(n) = k_1 + \cdots + k_r$ (total number of prime divisors; so $\nu^*(1) = 0$).

Exercise$^+$ 1.3.13. Prove that the expected number of distinct prime divisors of a random integer $i \in [n]$ is asymptotically $\ln \ln n$:

$$\frac{1}{n} \sum_{i=1}^{n} \nu(i) \sim \ln \ln n.$$

How much larger is ν^*? On average, not much. Prove that the average value of ν^* is also asymptotic to $\ln \ln n$.

Definition 1.3.14. n is **square-free** if $(\forall p \text{ prime })(p^2 \nmid n)$.

Definition 1.3.15 (Möbius Function).

$$\mu(n) = \begin{cases} 1 & n = 1 \\ (-1)^k & n = p_1 \cdots p_k \text{ where the } p_i \text{ are distinct } (n \text{ is square-free}) \\ 0 & \text{if } (\exists p)(p^2 \mid n) \end{cases}$$

Exercise 1.3.16. Let $\delta(n) = \sum_{d|n} \mu(d)$. Evaluate $\delta(n)$.

Definition 1.3.17 (Riemann zeta function). For $s > 1$ define the *zeta function* $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$.

Exercise 1.3.18. Prove Euler's identity:

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1 - \frac{1}{p^s}}.$$

Exercise 1.3.19. Prove:

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}.$$

Exercise 1.3.20. Prove:

$$(\zeta(s))^2 = \sum_{n=1}^{\infty} \frac{d(n)}{n^s}.$$

Exercise 1.3.21. Prove:

$$\zeta(s)(\zeta(s) - 1) = \sum_{n=1}^{\infty} \frac{\sigma(n)}{n^s}.$$

Exercise* 1.3.22. (Euler) Prove: $\zeta(2) = \pi^2/6$.

Exercise 1.3.23. Give a natural definition which will make following statement sensible and true: "the probability that a random positive integer n satisfies $n \equiv 3 \pmod 7$ is $1/7$." Our choice of a "random positive integer" should be "uniform" (obviously impossible). (*Hint.* Consider the integers up to x; then take the limit as $x \to \infty$.)

Exercise 1.3.24. Make sense out of the question "What is the probability that two random positive integers are relatively prime?" Prove that the answer is $6/\pi^2$. *Hint.* To prove that the required limit exists may be somewhat tedious. If you want to see the fun part, assume the existence of the limit, and prove in just two lines that the limit must be $1/\zeta(2)$.

Definition 1.3.25. Let F be a field. $f\colon \mathbb{N} \to F$ is called **multiplicative** if

$$(\forall a, b)(\text{g.c.d.}\,(a, b) = 1 \Rightarrow f(ab) = f(a)f(b)).$$

f is called **completely multiplicative** if

$$(\forall a, b)(f(ab) = f(a)f(b)).$$

f is called **additive** if

$$(\forall a, b)(\text{g.c.d.}\,(a, b) = 1 \Rightarrow f(ab) = f(a) + f(b)).$$

Exercise 1.3.26. Show that

1. φ, σ, d, and μ are multiplicative but not completely multiplicative
2. ν is additive and ν^* is completely additive. Log is completely additive.

Exercise 1.3.27. Show

1. $\varphi\left(p^k\right) = p^k - p^{k-1} = (p-1)p^{k-1}$
2. $d\left(p^k\right) = k + 1$
3. $\sigma\left(p^k\right) = \dfrac{p^{k+1} - 1}{p - 1}$

Exercise 1.3.28. Show

1. $\varphi\left(\prod_{i=1}^{r} p_i^{k_i}\right) = \prod_{i=1}^{r} (p_i - 1)p_i^{k_i - 1}$
2. $d\left(\prod_{i=1}^{r} p_i^{k_i}\right) = \prod_{i=1}^{r} (k_i + 1)$
3. $\sigma\left(\prod_{i=1}^{r} p_i^{k_i}\right) = \prod_{i=1}^{r} \dfrac{p_i^{k_i+1} - 1}{p_i - 1}$

Exercise 1.3.29. Show

$$\varphi(n) = n \prod_{\substack{p \mid N \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right)$$

Let F be a field and $f\colon \mathbb{N} \to F$. Define

$$g(n) = \sum_{d \mid n} f(d).$$

Exercise 1.3.30 (Möbius Inversion Formula). Show

$$f(n) = \sum_{d \mid N} g(d)\mu\left(\frac{n}{d}\right).$$

Exercise 1.3.31. Use the Möbius Inversion Formula together with Exercise 1.3.3 for a second proof of Exercise 1.3.29.

Exercise 1.3.32. Prove that the sum of the complex primitive n-th roots of unity is $\mu(n)$.

Definition 1.3.33. The **n-th cyclotomic polynomial** $\Phi_n(x)$ is defined as

$$\Phi_n(x) = \prod_{\omega}(x - \omega)$$

where the product ranges over all complex primitive n-th roots of unity. Note that the degree of $\Phi_n(x)$ is $\varphi(n)$. Also note that $\Phi_1(x) = x-1$, $\Phi_2(x) = x+1$, $\Phi_3(x) = x^2+x+1$, $\Phi_4(x) = x^2+1$, $\Phi_5(x) = x^4 + x^3 + x^2 + x + 1$, $\Phi_6(x) = x^2 - x + 1$.

Exercise 1.3.34. Prove that $\Phi_n(x)$ has integer coefficients. What is the coefficient of $x^{\varphi(n)-1}$?

Exercise 1.3.35. Prove: if p is a prime then $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$.

Exercise 1.3.36. Prove:

$$\Phi_n(x) = \prod_{d|n}(x^d - 1)^{\mu(n/d)}.$$

Exercise$^+$ 1.3.37. (Bateman) Let A_n denote the sum of the absolute values of the coefficients of $\Phi_n(x)$. Prove that $A_n < n^{d(n)/2}$. Infer from this that $A_n < \exp(n^{c/\ln\ln n})$ for some constant c. *Hint:* We say that the power series $\sum_{n=0}^{\infty} a_n x^n$ *dominates* the power series $\sum_{n=0}^{\infty} b_n x^n$ if $(\forall n)(|b_n| \le a_n)$. Prove that the power series

$$\prod_{d|n} \frac{1}{1 - x^d}$$

dominates $\Phi_n(x)$.

Note: Erdős proved that this bound is tight, apart from the value of the constant: for infinitely many values of n, $A_n > \exp(n^{c/\ln\ln n})$ for another constant $c > 0$.

Exercise$^+$ 1.3.38. (Kronecker) Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a monic polynomial of degree n (i. e., $a_n = 1$) with integer coefficients. Suppose all roots of f have unit absolute value. Prove that all roots of f are roots of unity. (In other words, if all algebraic conjugates of a complex algebraic number z have unit absolute value then z is a root of unity.)

1.4 Prime Numbers

Exercise 1.4.1. Prove:

$$\sum_{i=1}^{n} \frac{1}{i} = \ln n + O(1).$$

Exercise 1.4.2. Prove:

$$\prod_{p \le x} \frac{1}{1 - 1/p} = \sum{}' \frac{1}{i},$$

where the product is over all primes $\le x$ and the summation extends over all positive integers i with no prime divisors greater than x. In particular, the sum on the right-hand side converges. It also follows that the left-hand side is greater than $\ln x$.

Exercise 1.4.3. Prove: $\sum 1/p = \infty$. (*Hint.* Use the preceding exercise. Take natural logarithms; use the power series expansion of $\ln(1 - z)$. Conclude that $\sum_{p \le x} 1/p > \ln\ln x + O(1)$. (In other words, $\sum_{p \le x} 1/p - \ln\ln x$ is bounded from below.))

Exercise$^+$ 1.4.4. Prove: $\sum_{p \le x} 1/p = \ln\ln x + O(1)$. (In other words, $|\sum_{p \le x} 1/p - \ln\ln x|$ is bounded.)

Exercise$^+$ 1.4.5. Prove $\varphi(n) = \Omega\left(\frac{n}{\ln\ln n}\right)$ and find the largest implicit asymptotic constant.

Let $\pi(x)$ the number of primes less than or equal to x.

Theorem 1.4.6 (Prime Number Theorem)(Hadamard and de la Vallée Poussin, 1896).

$$\pi(x) \sim \frac{x}{\ln x}$$

Exercise 1.4.7. Use the PNT to show that $\lim_{n\to\infty} \frac{p_{n+1}}{p_n} = 1$, where p_n is the n-th prime.

Exercise 1.4.8. Use the PNT to prove $p_n \sim n \cdot \ln n$.

Exercise 1.4.9. Prove $\prod_{\substack{p \le x \\ p \text{ prime}}} p = \exp\big(x(1 + o(1))\big)$. Prove that this result is in fact equivalent to the PNT.

Exercise 1.4.10. Let $e_n = \text{l.c.m.}\,(1, 2, \ldots, n)$. Prove: $e_n = \exp\big(n(1 + o(1))\big)$. Prove that this result is in fact equivalent to the PNT.

Exercise 1.4.11. Prove: $\sum_{p \le x} p \sim x^2/(2 \ln x)$. (Use the PNT.)

Definition 1.4.12. A *permutation* is a bijection of a set to itself. The permutations of a set form a group under composition. The *symmetric group of degree* n is the group of all permutations of a set of n elements; it has order $n!$. The *exponent* of a group is the l.c.m. of the orders of all elements of the group.

Exercise 1.4.13. Prove: the exponent of S_n is e_n.

Exercise$^+$ 1.4.14. Let $m(n)$ denote the maximum of the orders of the elements in S_n. Prove: $m(n) = \exp(\sqrt{n \ln n}(1 + o(1)))$.

Exercise* 1.4.15. Let $a(n)$ denote the "typical" order of elements in S_n. Prove that $\ln a(n) = O((\ln n)^2)$. ("Typical" order means that 99% of the elements has order falling in the stated range. Here "99" is arbitrarily close to 100.) *Hint.* Prove that a typical permutation has $O(\ln n)$ cycles.

Erdős and Turán proved in 1965 that in fact $\ln a(n) \sim (\ln n)^2/2$.

Exercise 1.4.16. Prove from first principles: $\prod_{\substack{p<x \\ p \text{ prime}}} p < 4^x$. (*Hint:* if $n < p \le 2n$ then $p \mid \binom{2n}{n}$.)

Exercise 1.4.17. Prove: if $p > \sqrt{2n}$ then $p^2 \nmid \binom{2n}{n}$.

Exercise 1.4.18. Prove: if q is a prime power dividing $\binom{2n}{n}$ then $q \le n$. (*Hint.* Give a formula for the highest exponent of a prime p which divides $\binom{2n}{n}$. First, find a formula for the exponent of p in $n!$.)

Exercise 1.4.19. Prove from first principles: $\prod_{\substack{p<x \\ p \text{ prime}}} p > (2 + o(1))^x$. (*Hint.* Consider the prime-power decomposition of $\binom{x}{x/2}$. Show that the contribution of the powers of primes $\le \sqrt{x}$ is negligible.)

Exercise 1.4.20. Paul Erdős was an undergraduate when he found a simple proof of Chebyshev's theorem based on the prime factors of $\binom{2n}{n}$. Chebyshev's theorem is a precursor of the PNT; it says that

$$\pi(x) = \Theta\left(\frac{x}{\ln x}\right).$$

Following Erdős, prove Chebyshev's Theorem from first principles. The proof should be only a few lines, based on Exercises 1.4.16 and 1.4.19.

Exercise 1.4.21. Prove: for all integers x, either $x^2 \equiv 0 \pmod 4$ or $x^2 \equiv 1 \pmod 4$. (*Hint.* Distinguish two cases according to the parity of x [parity: even or odd].)

Exercise 1.4.22. $a^2 + b^2 \not\equiv -1 \pmod 4$.

Exercise 1.4.23. (a) Make a table of all primes ≤ 100. Next to each prime p write its expression as the sum of two squares if p can be so represented; otherwise write "NONE" next to p.

(b) Discover and state a very simple pattern as to which primes can and which primes cannot be represented as the sum of two squares. Your statement should go like this: "It seems from the table that a prime p can be represented as the sum of two squares if and only if either $p = 2$ or ***" where "***" stands for a very simple rule (less than half a line).

(c) Give a simple proof that the primes you believe cannot be represented as a sum of two squares indeed cannot. *Hint.* Use the previous exercise.

Exercise 1.4.24. Prove: if p is a prime number and $p \geq 5$ then $p \equiv \pm 1 \pmod 6$. *Hint.* There are only 6 cases to consider. (What are they?)

1.5 Quadratic Residues

Definition 1.5.1. a is a **quadratic residue** mod p if $(p \nmid a)$ and $(\exists b)(a \equiv b^2 \bmod p)$.

Exercise 1.5.2. Prove: a is a quadratic residue mod $p \iff a^{(p-1)/2} \equiv 1 \pmod p$.

Definition 1.5.3. a is a **quadratic non-residue** mod p if $(\forall b)(a \not\equiv b^2 \bmod p)$.

Exercise 1.5.4. Prove: a is a quadratic non-residue mod $p \iff a^{(p-1)/2} \equiv -1 \pmod p$.

Definition 1.5.5 (Legendre Symbol).

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue mod } p \\ -1 & \text{if } a \text{ is a quadratic non-residue mod } p \\ 0 & \text{if } p \mid a \end{cases}$$

Let $\mathbb{F}_q$ be a finite field of odd prime power order q.

Definition 1.5.6. $a \in F_q$ is a **quadratic residue** if $a \neq 0$ and $(\exists b)(a = b^2)$.

Exercise 1.5.7. Prove: a is a quadratic residue in $\mathbb{F}_q \iff a^{(q-1)/2} = 1$.

Definition 1.5.8. $a \in \mathbb{F}_q$ is a **quadratic non-residue** if $(\forall b)(a \neq b^2)$.

Exercise 1.5.9. Prove: a is a quadratic non-residue in $\mathbb{F}_q \iff a^{(q-1)/2} = -1$.

Exercise 1.5.10. Prove: in $\mathbb{F}_q$, the number of quadratic residues equals the number of quadratic non-residues; so there are $(q-1)/2$ of each. (As before, q is an odd prime power.)

Definition 1.5.11. Let q be an odd prime power. We define the **quadratic character** $\chi\colon \mathbb{F}_q \to \{0,1,-1\} \subset \mathbb{C}$ by

$$\chi(a) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue} \\ -1 & \text{if } a \text{ is a non-residue} \\ 0 & \text{if } a = 0 \end{cases}$$

Note that if $q = p$ (i.e. prime and not prime power) then $\chi(a) = \left(\frac{a}{p}\right)$.

Exercise 1.5.12. Prove χ is multiplicative.

Exercise 1.5.13. The Legendre Symbol is completely multiplicative in the numerator.

Exercise 1.5.14. Prove that -1 is a quadratic residue in $\mathbb{F}_q$ if and only if $q \equiv 1 \pmod 4$. *Hint.* Exercise 1.5.7.

Exercise 1.5.15. Prove that $\sum_{a\in\mathbb{F}_q} \chi(a(a-1)) = -1$. *Hint.* Divide by a^2.

Exercise 1.5.16. Prove that each of the four pairs $(\pm 1, \pm 1)$ occur a roughly equal number of times ($\approx q/4$) as $(\chi(a), \chi(a-1))$ $(a \in \mathbb{F}_q)$. "Roughly equal" means the difference is bounded by a small constant. Moral: for a random element $a \in \mathbb{F}_q$, the values of $\chi(a)$ and $\chi(a-1)$ are nearly independent.

Exercise 1.5.17. Let $f(x) = ax^2 + bx + c$ be a quadratic polynomial over $\mathbb{F}_q$ $(a, b, c \in \mathbb{F}_q$, $a \neq 0)$. Prove: if $b^2 - 4ac \neq 0$ then $|\sum_{a\in\mathbb{F}_q} \chi(f(a))| \leq 2$. What happens if $b^2 - 4ac = 0$?

1.6 Lattices and diophantine approximation

Definition 1.6.1. An n-dimensional **lattice** (grid) is the set L of all *integer* linear combinations $\sum_{i=1}^n a_i \mathbf{b}_i$ of a basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of $\mathbb{R}^n$ $(a_i \in \mathbb{Z})$. The set of *real* linear combinations with $0 \leq a_i \leq 1$ $(a_i \in \mathbb{R})$ form a **fundamental parallelepiped**.

Exercise 1.6.2. The volume of the fundamental parallelepiped of the lattice L is $\det(L) := |\det(\mathbf{b}_1, \ldots, \mathbf{b}_n)|$.

Exercise* 1.6.3. (Minkowski's Theorem) Let L be an n-dimensional lattice and let V be the volume of its fundamental parallelepiped. Let $A \subset \mathbb{R}^n$ be an n-dimensional convex set, symmetrical about the origin (i. e., $-A = A$), with volume greater than $2^n V$. Then $A \cap L \neq \{0\}$, i. e., A contains a lattice point other than the origin.
Hint. Linear transformations don't change the proportion of volumes, and preserve convexity and central symmetry. So WLOG $L = \mathbb{Z}^n$ with $\{\mathbf{b}_i\}$ the standard basis. The fundamental parallelepiped is now the unit cube C. Consider the lattice $2L = (2\mathbb{Z})^n$. Then the quotient space $\mathbb{R}^n/(2\mathbb{Z})^n$ can be identified with the cube $2C$ which has volume 2^n. Since A has volume $> 2^n$, there exist two points $u, v \in A$ which are mapped to the same point in $2C$, i. e., all coordinates of $u - v$ are even integers. Show that $(u - v)/2 \in A \cap L$.

Exercise 1.6.4. Finding "short" vectors in a lattice is of particular importance. Prove the following corollary to Minkowski's Theorem:

$$(\exists v \in L)\left(0 < \|v\|_\infty \leq (\det L)^{1/n}\right).$$

Definition 1.6.5. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. A *simultaneous ϵ-approximation* of the α_i is a sequence of fractions p_i/q with a common denominator $q > 0$ such that $(\forall i)(|q\alpha_i - p_i| \leq \epsilon)$.

Exercise$^+$ 1.6.6. (Dirichlet) $(\forall \alpha_1, \ldots, \alpha_n \in \mathbb{R})(\forall \epsilon > 0)(\exists$ an ϵ-approximation with the denominator satisfying $0 < q \leq \epsilon^{-n})$.
Hint. Apply the preceding exercise to the $(n+1)$-dimensional lattice L with basis $\mathbf{e}_1, \ldots, \mathbf{e}_n, \mathbf{f}$ where $\mathbf{f} = \sum_{i=1}^n \alpha_i \mathbf{e}_i + \epsilon^{n+1}\mathbf{e}_{n+1}$ and $\{\mathbf{e}_1, \ldots, \mathbf{e}_{n+1}\}$ is the standard basis.

The following remarkable result was first stated by Albert Girard (1540–1632) who may have found it on an empirical basis; there is no evidence that he could prove it. The first person to claim to have a proof was Pierre de Fermat (1601–1665). Fermat, however, never published anything mathematical and, while he claimed many discoveries in his correspondence or on the margins of his copy of Diophantus' *Arithmetic* (those marginal notes were later found and published by his son Samuel), there is no trace of proofs, except for one, in his entire extensive surviving correspondence. A century later Leonhard Euler (1707–1783) took great pride in providing proofs of Fermat's theorems, including this one. We give a more recent, devilishly clever proof, based on Minkowski's Theorem and found by Paul Turán (1910–1976).

Exercise* 1.6.7 (Girard-Fermat-Euler). Prove: a prime p can be written as the sum of two squares if and only if $p = 2$ or $p \equiv 1 \pmod 4$.
Hint. Necessity was established in Exercise 1.4.23. For sufficiency, assume $p \equiv 1 \pmod 4$. Then $\left(\frac{-1}{p}\right) = 1$ by Exercise 1.5.14 and therefore $(\exists a)(p \,|\, a^2 + 1)$. Consider the lattice (plane grid) $L \subset \mathbb{Z}^2$ consisting of all integral linear combinations of the vectors $(a, 1)$ and $(p, 0)$. Observe that if $(x, y) \in L$ then $p \,|\, x^2 + y^2$. Moreover, the area of the fundamental parallelogram of the lattice is p. Apply Minkowski's Theorem to this lattice to obtain a nonzero lattice point (x, y) satisfying $x^2 + y^2 < 2p$.

1.7 Introductory Problems: g.c.d., congruences, multiplicative inverse, Chinese Remainder Theorem, Fermat's Little Theorem

Notation: Unless otherwise stated, all variables in this chapter are *integers.* For $n \geq 0$, $[n] = \{1, 2, \ldots, n\}$. The formula $d \,|\, n$ denotes the relation "d divides n," i. e., $(\exists k)(n = dk)$. We also say "d is a divisor of n" or "n is a multiple of d." Note that $(\forall a)(a \,|\, a)$, including $0 \,|\, 0$ (even though we do not allow division by zero!). In fact $0 \,|\, n \iff n = 0$. Note also that $(\forall k\ (n \,|\, k)) \iff n = \pm 1$.

Notation 1.7.1. Let $\operatorname{div}(n)$ denote the set of divisors of n.

Examples. $\operatorname{div}(6) = \operatorname{div}(-6) = \{\pm 1, \pm 2, \pm 3, \pm 6\}$; $\operatorname{div}(1) = \{\pm 1\}$; $\operatorname{div}(0) = \mathbb{Z}$.

Exercise 1.7.2. Prove: $a \mid b \iff \operatorname{div}(a) \subseteq \operatorname{div}(b)$.

Exercise 1.7.3. Prove: $\operatorname{div}(a) = \operatorname{div}(b) \iff b = \pm a$.

Congruence notation. We write $a \equiv b \pmod{m}$ if $m \mid (a - b)$ ("a is congruent to b modulo m").

For instance, $100 \equiv 2 \pmod{7}$ (because $7 \mid 100 - 2 = 98 = 7 \cdot 14$); therefore, if today is Monday then 100 days from now it will be Wednesday (Monday +2). This example expains why *modular arithemtic* (calculations modulo m) are also referred to as "calendar arithmetic."

Division Theorem. $(\forall a)(\forall b \geq 1)(\exists q)(\exists r)(0 \leq r \leq b - 1 \text{ and } a = bq + r)$.

q is called the "integer quotient" and r the "remainder."

Exercise 1.7.4. Prove: $r \equiv a \pmod{b}$.

Remainder notation. The remainder r is denoted by the expression $(a \mod b)$. (Exercise 1.7.4 explains this notation; the congruence *relation* and the mod *function* should not be confused.) Examples: $(100 \mod 7) = 2$; $(-100 \mod 7) = 5$; $(98 \mod 7) = 0$; $(0 \mod 7) = 0$; $(a \mod 0)$ is undefined.

Common Divisor. The integer f is a common divisor of the integers a and b if $f \mid a$ and $f \mid b$.

Exercise 1.7.5. Prove: f is a common divisor of a and $b \iff \operatorname{div}(f) \subseteq \operatorname{div}(a) \cap \operatorname{div}(b)$.

Greatest Common Divisor. The integer d is a greatest common divisor of the integers a and b if

- d is a common divisor of a and b;
- every common divisor of a and b divides d.

Exercise 1.7.6. Prove: d is a greatest common divisor of a and $b \iff \operatorname{div}(d) = \operatorname{div}(a) \cap \operatorname{div}(b)$.

The existence of a greatest common divisor is not evident at all; it is an important basic theorem. Often we need the additional fact that the greatest common divisor can be written as a linear combination with integer coefficients: $d = au + bv$.

Exercise$^+$ 1.7.7. $(\forall a)(\forall b)(\exists u)(\exists v)(au + bv$ is a greatest common divisor of a and $b)$.

Exercise 1.7.8. Prove: if d is a greatest common divisor of a and b then $-d$ is also a greatest common divisor of a and b and there are no other greatest common divisors.

G.c.d. notation. g.c.d.(a, b) will denote the (unique) nonnegative greatest common divisor of the integers a and b.

Exercise 1.7.9. Prove: g.c.d.$(0, 0) = 0$.

Exercise 1.7.10. What are the common divisors of 0 and 0? Is 0 the "greatest"?

Exercise 1.7.11. (a) Prove: $(\forall a)$(g.c.d.$(a, a) = |a|$).

(b) Prove: $(\forall a)$(g.c.d.$(a, 0) = |a|$).

Note that each of these statements includes the fact that g.c.d.$(0, 0) = 0$.

The **Euclidean algorithm**, described in Euclid's *Elements* around 350 B.C.E., is an efficient method to calculate the g.c.d. of two positive integers. We describe the algorithm in *pseudocode.*

Euclidean Algorithm

INPUT: integers a, b.

OUTPUT: g.c.d.(a, b).

0 Initialize: $A := |a|$, $B := |b|$
1 **while** $B \geq 1$ **do**
2 division: $R := (A \mod B)$
3 $A := B$, $B := R$
4 **end**(**while**)
5 **return** A

The **correctness** of the algorithm follows from the following *loop invariant:*

$$\text{g.c.d.}(A, B) = \text{g.c.d.}(a, b).$$

Exercise 1.7.12. Prove that the statement above is indeed a *loop invariant*, i. e., prove that if the statement "g.c.d.$(A, B) =$ g.c.d.(a, b)" is true before an iteration of the **while** loop then it remains true after the execution of the **while** loop.

In addition, at the end we use the fact that g.c.d.$(A, 0) = A$.

Exercise 1.7.13. The **efficiency** of the Euclidean the algorithm follows from the observation that after every two rounds, the value of B is reduced to less than half. Prove this statement.

This implies that the number of rounds is $\leq 2n$ where n is the number of binary digits of b. Therefore the total number of bit-operations is $O(n^3)$, so this is a *polynomial-time algorithm.* (Good job, Euclid!)

Last update: October 12, 2004

Exercise 1.7.14. Use Euclid's algorithm to determine the g.c.d. of the following pairs of integers:

(a) (105; 480)

(b) (72,806; 13,587,574).

Exercise 1.7.15. Let n be a *positive* integer and let $d(n)$ denote the number of positive divisors of n. For instance, $d(1) = 1$, $d(2) = d(3) = d(5) = 2$, $d(4) = 3$, $d(6) = 4$. Prove your answers to the following questions.

(a) For what values of n is $d(n) = 2$?

(b) For what values of n is $d(n) = 3$?

(c) Prove: $(\forall n)(d(n) < 2\sqrt{n})$.

Exercise 1.7.16. (a) Let $a, b > 0$ and let us perform Euclid's algorithm to find the g.c.d. of a and b. Let $r_1, r_2, \ldots$ denote the successive remainders; let us use the notation $r_{-1} = a$ and $r_0 = b$. Prove: $(\forall i \geq -1)(r_{i+2} \leq r_i/2)$.

(b) Prove: if a has n bits (digits in binary) then the algorithm will terminate in $\leq 2n$ rounds (one round being a division to find the next remainder). *Hint:* use part (a).

Exercise 1.7.17. Recall that the *multiplicative inverse* of b modulo m, denoted by $x = (b^{-1} \pmod m)$, is an integer x such that $bx \equiv 1 \pmod m$. Find each of the following multiplicative inverses, or prove that the multiplicative inverse does not exist. Among the infinitely many values of the multiplicative inverse, find the smallest positive integer.

(a) $5^{-1} \pmod{17}$

(b) $39^{-1} \pmod{403}$

(c) $2^{-1} \pmod{2k+1}$ (where k is a given integer).

(d) $k^{-1} \pmod{2k+1}$. Find the inverse in the range $\{0, 1, \ldots, 2k\}$.

(e) $k^{-1} \pmod{3k+1}$. Find the inverse in the range $\{0, 1, \ldots, 3k\}$.

Exercise 1.7.18. Solve the following system of congruences:

$$\begin{aligned} x &\equiv 7 \pmod{16} \\ x &\equiv 3 \pmod{15} \\ x &\equiv 1 \pmod{11} \end{aligned}$$

Exercise 1.7.19. Decide whether or not the following system of congruences is solvable. If your answer is YES, find a solution. If your answer is NO, prove your answer.

$$\begin{aligned} x &\equiv 7 \pmod{13} \\ x &\equiv 3 \pmod{25} \\ x &\equiv 20 \pmod{39} \end{aligned}$$

Exercise 1.7.20. Prove whether or not the following system of congruences is solvable.

$$\begin{aligned} x &\equiv 7 \pmod{18} \\ x &\equiv 7 \pmod{12} \\ x &\equiv 1 \pmod{6} \end{aligned}$$

Exercise 1.7.21. Consider the statement "if $a \equiv 1 \pmod 5$ and $b \equiv 1 \pmod 5$ then g.c.d.$(a, b) \equiv 1 \pmod 5$." Find infinitely many counterexamples.

Exercise 1.7.22. The **Fibonacci numbers** are defined as follows: $F_0 = 0, F_1 = 1$, and for $n \geq 2$, $F_n = F_{n-1} + F_{n-2}$. So $F_2 = 1, F_3 = 2, F_4 = 3, F_5 = 5, F_6 = 8$, $F_7 = 13$, $F_8 = 21$, etc. Prove: for all $n \geq 1$,

(a) g.c.d. $(F_{n-1}, F_n) = 1$.

(b) $|F_n^2 - F_{n-1}F_{n+1}| = 1$.

(c) If g.c.d. $(m, n) = d$ then g.c.d. $(F_m, F_n) = F_d$.

(d) If g.c.d. $(m, n) = d$ then g.c.d. $(a^m - 1, a^n - 1) = a^d - 1$.

Hint: For parts (a) and (b), use mathematical induction.

Exercise 1.7.23. Calculate $(a \bmod m)$ where $a = 3^{114,555}$ and $m = 173$. Recall that the expression $(a \bmod m)$ denotes the smallest nonnegative remainder of the division of a by m. *Hint.* Fermat's little Theorem (Theorem 1.2.18).

Exercise 1.7.24. (a) Prove: if m is a prime and $x^2 \equiv 1 \pmod m$ then $x \equiv \pm 1 \pmod m$ (i. e., either $x \equiv 1 \pmod m$, or $x \equiv -1 \pmod m$).

(b) Prove that (a) becomes false if we omit the condition that m is a prime. (Give a counterexample.)

(c) Prove that (a) is false for every m of the form $m = pq$ where p, q are distinct odd primes. In other words, show that $(\forall p, q)(\exists x)($ if p, q are distinct odd primes then $x^2 \equiv 1 \pmod{pq}$ but $x \not\equiv \pm 1 \pmod{pq})$. *Hint.* Observe that $a \equiv b \pmod{pq} \Leftrightarrow a \equiv b \pmod p$ and $a \equiv b \pmod q$. Work separately modulo each prime; combine your results using the Chinese Remainder Theorem.

Exercise 1.7.25. Prove: $\forall x(x^2 \not\equiv -1 \pmod{419})$.
Hint. Proof by contradiction. Use Fermat's little Theorem (Theorem 1.2.18). (419 is a prime.)

Exercise 1.7.26. (a) Prove: if g.c.d.$(a, 85) = 1$ then $a^{33} \equiv a \pmod{85}$). *Hint.* $85 = 5 \cdot 17$, so two numbers are congruent modulo 85 if and only if they are congruent modulo 5 as well as modulo 17. Prove the stated congruence modulo 5 and modulo 17.

(b) True or false (prove your answer): if 85 does not divide a then $a^{32} \equiv 1 \pmod{85}$).

Exercise 1.7.27. True or False. If False, give a counterexample.

1. If g.c.d. $(a, b) = 0$ then $a = b = 0$.
2. If l.c.m. $(a, b) = 0$ then $a = b = 0$.
3. If $a \equiv b \pmod{24}$ then $a \equiv b \pmod{6}$ and $a \equiv b \pmod{4}$.
4. If $a \equiv b \pmod{6}$ and $a \equiv b \pmod{4}$ then $a \equiv b \pmod{24}$.

Exercise 1.7.28. Consider the following statement:

Statement. a^{15} is a multiplicative inverse of a modulo 17.

1. Define what it means that "x is a multiplicative inverse of a modulo m."
2. Give infinitely many counterexamples to the statement above.
3. State a very simple necessary and sufficient condition for the statement to be true. Prove your answer.

Exercise 1.7.29. Prove: $(\forall a)(a^{37} \equiv a \pmod{247})$. *Hint.* $247 = 13 \cdot 19$.

Exercise 1.7.30. Prove: if a is an odd integer then

$$a^{67} \equiv a \pmod{12,328}.$$

Hint. $12,328 = 8 \cdot 23 \cdot 67$.

Exercise 1.7.31. Prove: the congruence $x^2 \equiv -1 \pmod{103}$ has no solution. (103 is a prime number.) *Hint.* FℓT.

Exercise 1.7.32. Let $1 \leq a_1 < \cdots < a_{n+1} \leq 2n$ be $n + 1$ distinct integers between 1 and $2n$. Prove:

(a) $(\exists i, j)(i \neq j$ and g.c.d.$(a_i, a_j) = 1)$.

(b) $(\exists i, j)(i \neq j$ and $a_i \mid a_j)$. *Hint.* Pigeon-hole principle.

Exercise 1.7.33. Let p be a prime number. Find all solutions to the following congruence. Prove your answer.

$$x^p \equiv x^{3p} \pmod{p}.$$

Exercise 1.7.34. In this problem, the universe of the variable x is the set of integers. Prove:

$$(\forall x)(x^{21} \equiv x \pmod{55}).$$

Chapter 2

Counting

To Be WRITTEN

2.1 Problems

2.2 Binomial coefficients

Exercise 2.2.1. For $n \geq 5$, let $S_n = \binom{5}{5} + \binom{6}{5} + \cdots + \binom{n}{5}$. Prove that

$$S_n = \binom{n+1}{6}.$$

Hint: mathematical induction. Make your proof very simple. You should not need any calculations, just use what we learned in class about binomial coefficients.

Exercise 2.2.2. Prove: if p is a prime number and $1 \leq k \leq p-1$ then p divides the binomial coefficient $\binom{p}{k}$.

Exercise 2.2.3. Give closed form expressions (no product symbols or dot-dot-dots) of the binomial coefficients below, using "old" binomial coefficients:

(a) $\binom{-1}{k}$

(b) $\binom{-1/2}{k}$

where k is a positive integer.

Exercise 2.2.4. Let O_n denote the number of odd subsets of an n-set and E_n the number of even subsets of an n-set. For $n \geq 1$, prove that $O_n = E_n$. Give

(a) a bijective (combinatorial) proof;

(b) an algebraic proof. (Use the Binomial Theorem for the algebraic proof.)

Exercise 2.2.5. Give a closed form expression for

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \binom{n}{6} + \dots .$$

Exercise$^+$ 2.2.6. Give a closed form expression for

$$\binom{n}{0} + \binom{n}{4} + \binom{n}{8} + \binom{n}{12} + \dots .$$

Hint. Apply the Binomial Theorem to $(1+x)^n$; substitute $1, i, -1, -i$ for x (where $i = \sqrt{-1}$).

Exercise 2.2.7. Prove: $\binom{2n}{n} < 4^n$. Do NOT use Stirling's formula. Your proof should be just one line.

Exercise 2.2.8. Let $n \geq 7$. Count those strings of length n over the alphabet $\{A, B\}$ which contain at least $n-3$ consecutive A's.
Hint. Inclusion–exclusion.

Exercise 2.2.9. Prove: if $1 \leq k \leq n$ then

$$\binom{n}{k} \geq \left(\frac{n}{k}\right)^k .$$

Your proof should be no more than a couple of lines.

Exercise 2.2.10. Prove: if $1 \leq k \leq n$ then

$$\binom{n}{k} < \left(\frac{\mathrm{e}n}{k}\right)^k .$$

Hint. Use the Binomial Theorem and the fact that $(\forall x \neq 0)(\mathrm{e}^x > 1 + x)$. (Note that Stirling's formula is of no use; it would only prove things for "large enough n.")

Exercise$^+$ 2.2.11. Prove: if $1 \leq k \leq n$ then

$$\sum_{j=0}^{k} \binom{n}{j} < \left(\frac{\mathrm{e}n}{k}\right)^k .$$

Hint. As in the previous exercise.

Exercise 2.2.12. (a) Evaluate the sum $S_n = \sum_{i=0}^{\infty} \binom{n}{i} 2^i$. Your answer should be a very simple closed-form expression (no summation symbols or dot-dot-dots).

(b) Let b_n be the largest term in the sum S_n. Prove: $b_n = \Theta(S_n/\sqrt{n})$.

Exercise 2.2.13. An airline wishes to operate m routes between a given set of n cities. Count the number of possibilities. (A "route" is a pair of cities between which the airline will operate a direct flight. The cities are given, the routes need to be selected. There are no "repeated routes.") Your answer should be a very simple formula.

Exercise 2.2.14. Evaluate the following sums. In each case, your answer should be a simple closed-form expression.

1. $\sum_{i=1}^{n} 4^{n-i}$
2. $\sum_{i=1}^{n} \binom{n}{i} 4^{n-i}$

Exercise 2.2.15. Out of n candidates, an association elects a president, two vice presidents, and a treasurer. Count the number of possible outcomes of the election. (Give a simple expression. State, do not prove.)

Exercise 2.2.16. State your answers as very simple expressions.

1. Count the strings of length 3 (3-letter "words") over an alphabet of n characters.
2. What is the answer to the previous question if no repeated letters are allowed?

Exercise 2.2.17. Evaluate the expression $\binom{0.4}{2}$. Give your answer as a decimal.

Exercise 2.2.18. Pascal's Identity states that $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$. Give a combinatorial proof.

Exercise 2.2.19. We have 5 red beads and 11 blue beads. Count the necklaces that can be made out of these 16 beads. A "necklace" is an arrangement of the beads in a circle. The necklace obtained by rotating the circle does not count as a different necklace. Give a simple expression; do not evaluate.

Exercise 2.2.20. Use the idea of the preceding problem to prove that if a and b are relatively prime then $a + b \mid \binom{a+b}{a}$.

Exercise 2.2.21. Let $a_1, \ldots, a_k$ be positive integers. Prove: the least common multiple $L = \text{l.c.m.}\,(a_1, \ldots, a_k)$ can be expressed through g.c.d's of subsets of the a_i as follows:

$$L = \prod_{I \subseteq [k]} \left(\text{g.c.d.}\,(a_i \,:\, i \in I)\right)^{(-1)^{|I|+1}}.$$

Before attempting to solve this problem for all k, write down the expressions you get for $k = 2$ and $k = 3$ (without the product sign).

2.3 Recurrences, generating functions

Exercise 2.3.1. Let F_n denote the n-th Fibonacci number. ($F_0 = 0, F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$.) Prove: $F_0 + F_1 + \cdots + F_n = F_{n+2} - 1$.

Exercise 2.3.2. Let $a_0 = 3$, $a_1 = 1$, and $a_n = a_{n-1} + a_{n-2}$ ($n \geq 2$) (Fibonacci recurrence with different initial values).

(a) Give a closed-form expression for the generating function $f(x) = \sum_{n=0}^{\infty} a_n x^n$.

(b) Using the generating function, find a closed-form expression for a_n. Show all your work.

Exercise 2.3.3. Let $b_0 = 1$ and $b_n = 3b_{n-1} - 1$ ($n \geq 1$).

(a) (4 points) Give a closed-form expression for the generating function $g(x) = \sum_{n=0}^{\infty} b_n x^n$.

(b) (4 points) Using the generating function, find a closed-form expression for b_n. Show all your work.

Exercise 2.3.4. What is the generating function of each of the following sequences? Give a closed-form expression. Prove your answers.

(a) $a_n = n$.

(b) $b_n = \binom{n}{2}$.

(c) $c_n = n^2$.

(d) $d_n = 1/n!$.

(e) $e_n = 1/n$.

Exercise 2.3.5. If the generating function of the sequence $\{a_n\}$ is $f(x)$, what is the generating function of the sequence $b_n = na_n$? Your answer should be a very simple expression involving $f(x)$ (less than half a line).

Last update: October 12, 2004

Exercise 2.3.6. Let $m_0 = 1$, $m_1 = 2$, and $m_n = m_{n-1} + m_{n-2} + 1$. Express m_n through the Fibonacci numbers. Your expression should be very simple, less than half a line. Do not use generating functions. *Hint.* Tabulate the sequence. Compare with the Fibonacci numbers. Observe the pattern, prove by induction. Watch the subscripts.

Exercise 2.3.7. The sequence $\{a_n\}$ satisfies the recurrence $a_n = 5a_{n-1} - 6a_{n-2}$. Suppose the limit $L = \lim_{n\to\infty} a_n/a_{n-1}$ exists. Determine L.

Exercise 2.3.8. Let the sequence $\{b_n\}$ be defined by the recurrence $b_n = (b_{n-1} + 1)/n$ with initial value $b_0 = 0$. Let $f(x) = \sum_{n=0}^{\infty} b_n x^n$ be the generating function of the sequence. Write a differential equation for f: express $f'(x)$ in terms of $f(x)$ and x. Your expression should be very simple and closed-form.

Exercise 2.3.9. Let r_n be the number of strings of length n over the alphabet $\{A, B\}$ without consecutive A's (so $r_0 = 1$, $r_1 = 2$, $r_2 = 3$). Prove: $r_n \sim c\gamma^n$ where $\gamma = (1 + \sqrt{5})/2$ is the golden ratio. Determine the constant c. Prove your answers.

Chapter 3

Graphs and Digraphs

3.1 Graph Theory Terminology

The graph theoretic terminology we use in class differs from that of many texts. Here we make comparisons with Rosen's text (used in past years) and Anderson's (the current text). Please remember the differences listed below and use the terminology of these lecture notes when it differs from the text's. All concepts refer to a (simple) graph $G = (V, E)$.

Exercises. The unmarked exercises are routine, the exercises marked with a "plus" (+) are creative, those marked with an asterisk (*) are challenging; those marked with two asterisks are gems of mathematical ingenuity.

A **graph** (in Rosen's text: *simple graph*) is a pair $G = (V, E)$ where V is the set of **vertices** and E is the set of **edges**. An **edge** is an unordered pair of vertices. Two vertices joined by an edge are said to be **adjacent**. Two vertices are **neighbors** if they are adjacent. The **degree** $\deg(v)$ of vertex v is the number of its neighbors. A graph is **regular** of degree r if all vertices have degree r. The *complement* $\overline{G}$ of the graph G is the graph $\overline{G} = (V, \overline{E})$ where $\overline{E}$ is the complement of E with respect to the set $\binom{V}{2}$ of all pairs of vertices. So $\overline{G}$ has the same set of vertices as G; two distinct vertices are adjacent in $\overline{G}$ if and only if they are not adjacent in G.

An *isomorphism* between the graphs $G = (V, E)$ and $H = (W, F)$ is a bijection $f : V \to W$ from V to W which preserves adjacency, i. e., $(\forall x, y \in V)(x$ is adjacent o y in $G \Leftrightarrow f(x)$ and $f(y)$ are adjacent in H. Two graphs are *isomorphic* if there *exists* an isomorphism between them.

Exercise 3.1.1. Draw two non-isomorphic regular graphs of the same degree on 6 vertices. Prove that your graphs are not isomorphic.

Exercise 3.1.2. Prove: $\sum_{v \in V} \deg(v) = 2|E|$.

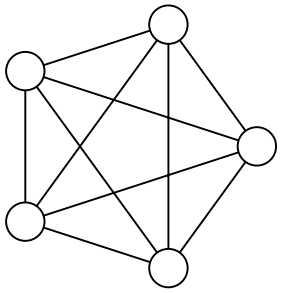


Figure 3.1: The complete graph K_5.

The number of vertices will usually be denoted by n.

Exercise 3.1.3. Observe: $|E| \leq \binom{n}{2}$.

Exercise 3.1.4. Observe: $|E(G)| + |E(\overline{G})| = \binom{n}{2}$.

Exercise 3.1.5. A graph is *self-complementary* if it is isomorphic to its complement. (a) Construct a self-complementary graph with 4 vertices. (b) Construct a self-complementary graph with 5 vertices. (c) Prove: if a graph with n vertices is self-complementary then $n \equiv 0$ or 1 (mod 4).

Exercise 3.1.6. (a) Prove: if $b_n = 2^{\binom{n}{2}}$ and $a_n = b_n/n!$ then $\log_2 a_n \sim \log_2 b_n$.

(b) Let $G(n)$ denote the number of non-isomorphic graphs on n vertices. Prove: $a_n \leq G(n) \leq b_n$.

(c)* Prove: $G(n) \sim a_n$. *Hint.* Reduce this question to the following: The expected number of automorphisms of a random graph in $1 + o(1)$. (Automorphism = self-isomorphism, i. e., an adjacency preserving permutation of the set of vertices.)

Complete graphs, complete bipartite graphs, subgraphs

In a **complete graph**, all pairs of vertices are adjacent. The complete graph on n vertices is denoted by K_n. It has $\binom{n}{2}$ edges. See Figure 3.1.

The vertices of a **complete bipartite graph** are split into two subsets $V = V_1 \dot\cup V_2$; and $E = \{\{x, y\} : x \in V_1, y \in V_2\}$ (each vertex in V_1 is adjacent to every vertex in V_2). If $k = |V_1|$ and $\ell = |V_2|$ then we obtain the graph $K_{k,\ell}$. This graph has $n = k + \ell$ vertices and $|E| = k\ell$ edges. See Figure 3.2.

The graph $H = (W, F)$ is a **subgraph** of $G = (V, E)$ if $W \subseteq V$ and $F \subseteq E$.

$H = (W, F)$ is a **spanning subgraph** of G if H is a subgraph and $V = W$.

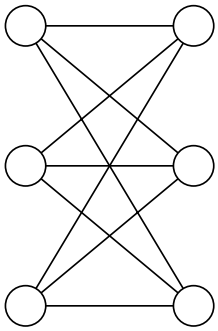


Figure 3.2: The complete bipartite graph $K_{3,3}$.

H is an **induced subgraph** of G if H is a subgraph of G and $(\forall x, y \in W)(x$ and y are adjacent in $H \Leftrightarrow x$ and y are adjacent in G. (So to obtain an induced subgraph, we may delete some vertices and the edges incident with the deleted vertices but no more edges.)

Exercise 3.1.7. Observe: (a) Every graph on n vertices is a spanning subgraph of K_n. (b) All induced subgraphs of a complete graph are complete.

Exercise 3.1.8. Let G be a graph with n vertices and m edges. Count the (a) induced subgraphs of G; (b) the spanning subgraphs of G. Both answers should be very simple expressions.

Exercise 3.1.9. Count those spanning subgraphs of K_n which have exactly m edges.

Exercise 3.1.10. Prove: if G is a bipartite graph with n vertices and m edges then $m \leq \lfloor n^2/4 \rfloor$.

Exercise$^+$ 3.1.11. (Mandel–Turán) Prove: if G is triangle-free ($K_3 \not\subseteq G$) then $|E| \leq \lfloor n^2/4 \rfloor$. Show that this bound is tight for every n. *Hint.* Use induction from in increments of 2; delete both vertices of an edge for the inductive step.

Walks, paths, cycles, trees

This is the area where our terminology most differs from the texts. Following common usage in the graph theory literature, we use the two simplest terms, *path* and *cycle*, to denote the two most central concepts (see below). For reasons that boggle the mind, textbook authors tend to use up these two terms for concepts of lesser importance, and use compound terms such as "simple path" and "simple cycle," or, astoundingly, have no term at all, to designate the two central concepts.

- **walk** (in both texts: *path*) of length k: a sequence of $k+1$ vertices $v_0, \ldots, v_k$ such that v_{i-1} and v_i are adjacent for all i.
- **trail** (in Rosen's text: *simple path*): a walk without repeated edges.

- **path:** a walk without repeated vertices. P_{k+1} denotes a path of length k (it has $k+1$ vertices). See Figure 3.3. (This all-important concept has no name in Rosen's text; it is called "simple path" in Anderson's. Note that the term "path" in both texts and even "simple path" in Rosen's text allow vertices to be repeated.) IMPORTANT: for the purposes of counting, two paths involving the same set of k vertices and the same set of $k-1$ connecting edges count as the same path. In other words, the paths $(v_0, \ldots, v_k)$ and $(v_k, \ldots, v_0)$ (the same path travelled backward) count as the same path. So each path of length ≥ 1 is defined by two walks (forward, backward).
- **closed walk** of length k a walk $v_0, \ldots, v_k$ where $v_k = v_0$. (Rosen's text confounds readers by using two terms, *circuit* and *cycle* for this concept; this important concept has no name in Anderson's text.)
- **closed trail:** a closed walk without repeated edges (this relatively unimportant concept gets the undeserved name "cycle" in Anderson's text);
- **cycle of length** k or k**-cycle**: a closed walk of length $k \geq 3$ with no repeated vertices except that $v_0 = v_k$. Notation: C_k. See Figure 3.4. (This all-important concept has no name at all in Rosen's text. Anderson uses the term "simple cycle" but then, to confuse the reader, he won't say "simple k-cycle;" in his terminology, as in ours, a k-cycle is automatically "simple" (has no repreated vertices) so in his inconsistent terminology, a "cycle of length k" is not necessarily a "k-cycle;" only "simple cycles" of length k are k-cycles.) IMPORTANT: for the purposes of counting, two cycles involving the same set of k vertices and the same set of k connecting edges count as the same cycle. In other words, the cycle $(v_0, v_1, \ldots, v_{k-1}, v_0)$ counts as the same as its cyclic shifts, $(v_1, v_2, \ldots, v_{k-1}, v_0, v_1)$, $(v_2, v_3, \ldots, v_{k-1}, v_0, v_1, v_2)$, etc., and also as its reverse, $(v_0, v_{k-1}, v_{k-2}, \ldots, v_1, v_0)$ and its cyclic shifts. So each cycle of length k is defined by $2k$ distinct closed walks of length k.
- a graph G is **connected** if there is a path between each pair of vertices.
- a **tree** is a connected graph without cycles. See Figure 3.5.
- H is a **spanning tree** of G if H is a tree and it is a spanning subgraph of G.

Exercise 3.1.12. Let G be a regular graph of degree d. Count the walks of length k starting at a given vertex v. (Your answer should be a very simple formula. Note that a walk is permitted to backtrack a step just made.)

Exercise 3.1.13. Count the paths of length k in the complete graph K_n. Your answer should be a simple formula. (Self-check: make sure your answer is consistent with the observations that (a) the number of paths of length 1 in a graph is the number of edges; and (b) the number of paths of length 2 in K_3 is 3.)

Exercise 3.1.14. Count the cycles of length k in the complete graph K_n. Your answer should be a simple formula. (Self-check: make sure your answer is consistent with the observation that the number of cycles of length 3 in K_n is $\binom{n}{3}$.)

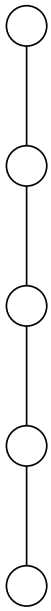


Figure 3.3: P_5, the path of length 4.

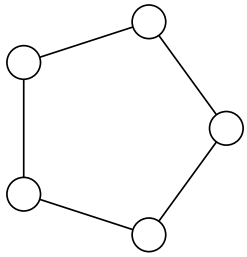


Figure 3.4: C_5, the cycle of length 5.

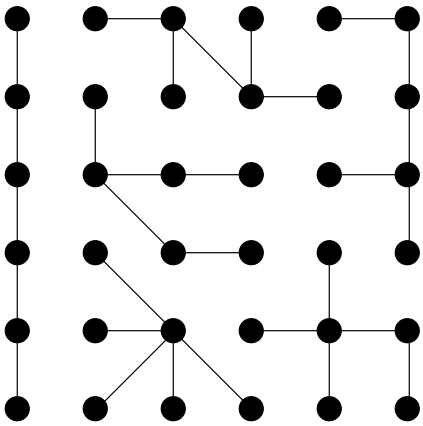


Figure 3.5: The trees on 6 vertices (complete list).

Exercise 3.1.15. Prove: if a vertex v has odd degree in the graph G then there exists another vertex w, also of odd degree, such that v and w are connected by a path.

Exercise 3.1.16. Prove that every tree with $n \geq 2$ vertices has at least two vertices of degree 1. *Hint.* Prove that the endpoints of a longest path in a tree have degree 1.

Exercise 3.1.17. Prove that every tree has $n-1$ edges. *Hint.* Induction. Use the preceding exercise.

Exercise 3.1.18. Prove: if G is a connected graph with n vertices then G has at least $n-1$ edges. If G is connedcted and has exactly $n-1$ edges then G is a tree.

Exercise 3.1.19. Verify that Figure 3.5 shows each 6-vertex tree exactly once. *Hint.* Without looking at Figure 3.5, produce your own systematic list of all trees on 6 vertices. Once done, compare your list with Figure 3.5.

Exercise 3.1.20. Draw a copy of each 7-vertex tree. Make sure you don't miss any, and you do not repeat, i. e., no pair of your drawings represent isomorphic trees. State the number of trees you found. List the trees in some systematic fashion; state your system.

Exercise 3.1.21. Prove: in a tree, all longest paths share a common vertex.

Exercise 3.1.22. Prove: in a tree on n vertices, the number of longest paths is at most $\binom{n-1}{2}$. *Hint.* Use the preceding exercise.

Exercise$^+$ 3.1.23. Let $d_1, \ldots, d_n$ be positive integers such that $\sum_{i=1}^n d_i = 2n-2$. Consider those spanning trees of K_n which have degree d_i at vertex i. Count these spanning trees; show that their number is

$$\frac{(n-2)!}{\prod_{i=1}^n (d_i-1)!}.$$

Exercise** 3.1.24. (Cayley) The number of spanning trees of K_n is n^{n-2}. *Hint.* This amazingly simple formula is in fact a simple consequence of the preceding exercise. Use the Multinomial Theorem.

Exercise 3.1.25. Let $t(n)$ denote the number of non-isomorphic trees on n vertices. Use Cayley's formula to prove that $t(n) > 2.7^n$ for sufficiently large n (i. e., $(\exists n_0)(\forall n > n_0)(t(n) > 2.7^n)$).

Exercise 3.1.26. Count the 4-cycles in the complete bipartite graph $K_{m,n}$. (You need to count those subgraphs which are isomorphic to C_4.) (*Comment.* Note that $K_{2,2}$ is isomorphic to C_4, the 4-cycle, therefore $K_{2,2}$ has exactly one 4-cycle. Check also that $K_{2,3}$ has three 4-cycles. Make sure that your answer to the general case conforms with this observation. Your answer should be a very simple formula.

Exercise* 3.1.27. (Kővári–Sós–Turán) Prove: if G has no 4-cycles ($C_4 \not\subseteq G$) then $|E| = O(n^{3/2})$. Show that this bound is tight (apart from the constant implied by the big-Oh notation). *Hint.* Let N denote the number of paths of length 2 in G. Observe that $N = \sum_{i=1}^n \binom{d_i}{2}$

where d_i is the degree of vertex i. On the other hand, observe that $N \leq \binom{n}{2}$. (Why? Use the assumption that there are no 4-cycles!) Compare these two expressions for N and apply Jensen's Inequality to the convex function $\binom{x}{2}$.

Cliques, distance, diameter, chromatic number

- A **k-clique** is a subgraph isomorphic to K_k (a set of k pairwise adjacent vertices). $\omega(G)$ denotes the size (number of vertices) of the largest clique in G.
- An **independent set** or **anti-clique** of size k is the complement of a k-clique: k vertices, no two of which are adjacent. $\alpha(G)$ denotes the size of the largest independent set in G.
- The **distance** $\text{dist}(x, y)$ between two vertices $x, y \in V$ is the length of a shortest path between them. If there is no path between x and y then their distance is said to be infinite: $\text{dist}(x, y) = \infty$.
- The **diameter** of a simple graph is the maximum distance between all pairs of vertices. So if a graph has diameter d then $(\forall x, y \in V)(\text{dist}(x, y) \leq d)$ and $(\exists x, y \in V)(\text{dist}(x, y) = d)$.
- The **girth** of a graph is the length of its shortest cycle. If a graph has no cycles then its girth is said to be infinite.

  Examples (verify!): trees have infinite girth; the $m \times n$ grid (Figure 3.6) has girth 4 if $m, n \geq 2$; $K_{m,n}$ has girth 4 if $m, n \geq 2$, K_n has girth 3; the Petersen graph (Figure 3.8) has girth 5.
- A **legal k-coloring** of a graph is a function $c : V \to [k] = \{1, \ldots, k\}$ such that adjacent vertices receive different colors, i. e., $\{u, v\} \in E \Rightarrow c(u) \neq c(v)$. A graph is **$k$-colorable** if there exists a legal k-coloring. The **chromatic number** $\chi(G)$ of a graph is the smallest k such that G is k-colorable.
- A graph is **bipartite** if it is 2-colorable.
- A **Hamilton cycle** is a cycle of length n, i. e., a cycle that passes through all vertices. G is **Hamiltonian** if it has a Hamilton cycle.
- A **Hamilton path** is a path of length $n-1$, i. e., a path that passes through all vertices.

Exercise 3.1.28. State the diameter of each of the following graphs: (a) P_n (the path of length $n-1$: this graph has n vertices and $n-1$ edges); (b) C_n (the n-cycle); (c) K_n (the complete graph on n vertices); (d) $K_{n,m}$ (complete bipartite graph);

Exercise 3.1.29. Disprove the following statement: "the diameter of a graph is the length of its longest path." Prove that the statement is true for trees.

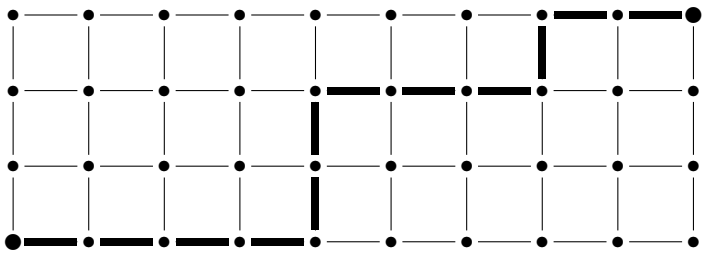


Figure 3.6: The 4×10 grid, with a shortest path between opposite corners highlighted.

Exercise 3.1.30. The $k \times \ell$ grid has $k\ell$ vertices (Figure 3.6). Count its edges.

Exercise 3.1.31. Verify that the diameter of the $k \times \ell$ grid is $k + \ell - 2$.

Exercise 3.1.32. Let u and v be two opposite corners of the $k \times \ell$ grid. Count the shortest paths between u and v. Your answer should be a very simple expression. *Hint.* Think of each shortest path as a sequence of North or East moves, represented as a string over the alphabet $\{N, E\}$.

Exercise 3.1.33. Prove: the bipartite graphs are exactly the subgraphs of the complete bipartite graphs.

Exercise 3.1.34. Prove: a graph is bipartite if and only if it has no odd cycles.

Exercise 3.1.35. We color the vertices of a bipartite graph G red and blue (legal coloring). Assume G has 30 red vertices (all other vertices are blue). Suppose each red vertex has degree 6 and each blue vertex has degree 5. What is the number of blue vertices? Prove your answer.

Exercise 3.1.36. Let us pick 3 distinct vertices at random in a bipartite graph G with n vertices. Prove that the probability that we picked an independent set is $\geq 1/4 - o(1)$ (as $n \to \infty$).

Exercise 3.1.37. For every $n \geq 1$, name a graph with n vertices, at least $(n^2 - 1)/4$ edges, and no cycles of length 5.

Exercise 3.1.38. Prove: if every vertex of a graph has degree $\leq d$ then the graph is $d + 1$-colorable (i. e., $\chi(G) \leq d + 1$).

Exercise 3.1.39. For every n, construct a 2-colorable graph with n vertices such that every vertex has degree $\geq (n - 1)/2$. (Moral: low degree is a sufficient but not a necessary condition of low chromatic number.)

Exercise 3.1.40. (Chromatic number vs. independence number) Prove: if G is a graph with n vertices then $\alpha(G)\chi(G) \geq n$.

Exercise 3.1.41. Give a formal definition of "3-colorable graphs." Watch your quantifiers.

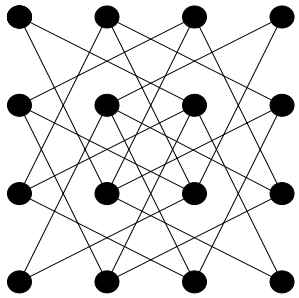


Figure 3.7: Graph of knight moves on a 4×4 chessboard

Exercise$^+$ 3.1.42. Construct a graph G on 11 vertices such that G is triangle-free ($K_3 \not\subseteq G$) and G is NOT 3-colorable. Prove that your graph has the stated properties. *Hint.* Draw your graph so that it has a rotational symmetry of order 5 (rotation by $2\pi/5$ should not change the picture).

Exercise* 3.1.43. Prove: $(\forall k)(\exists G)(\chi(G) \geq k$ and G is triangle-free.)

The following celebrated result is one of the early triumphs of the "Probabilistic Method." You can find the elegant proof in the book by Alon and Spencer.

Theorem 3.1.44. (Erdős, 1959) *Prove: $(\forall k, g)(\exists G)(\chi(G) \geq k$ and G has girth $\geq g$.)*

Exercise 3.1.45. Count the Hamilton cycles in the complete graph K_n.

Exercise 3.1.46. Count the Hamilton cycles in the complete bipartite graph $K_{r,s}$. (Make sure you count each cycle only once – note that $K_{2,2}$ has exactly one Hamilton cycle.)

Exercise 3.1.47. Prove that all grid graphs have a Hamilton path.

Exercise 3.1.48. Prove: the $k \times \ell$ grid is Hamiltonian if and only if $k, \ell \geq 2$ and $k\ell$ is even. (Your proofs should be very short, only one line for non-Hamiltonicity if $k\ell$ is odd.)

Exercise 3.1.49. Prove that the dodecahedron is Hamiltonian. (Lord Hamilton entertained his guests with this puzzle; hence the name.)

Exercise 3.1.50. (a) Prove: the graph of the knight's moves on a 4×4 chessboard (Figure 3.7) has no Hamilton path. Find an "Ah-ha!" proof: just "one line" after the following Lemma.

(b) Lemma. If a graph has a Hamilton path then after deleting k vertices, the remaining graph has $\leq k + 1$ connected components.

Exercise 3.1.51. We have a standard (8×8) chessboard and a set of 32 dominoes such that each domino can cover two neighboring cells of the chessboard. So the chessboard can be covered with the dominoes. Prove: if we remove the top left and the bottom right corner cells of the chessboard, the remaining 62 cells cannot be covered by 31 dominoes. Find an "Ah-ha!" proof (elegant, no case distinctions.)

Exercise 3.1.52. A mouse finds a $3 \times 3 \times 3$ chunk of cheese, cut into 27 blocks (cubes), and wishes to eat one block per day, always moving from a block to an adjacent block (a block that touches the previous block along a face). Moreover, the mouse wants to leave the center cube last. Prove that this is impossible. Find two "Ah-ha!" proofs; one along the lines of the solution of Exercise 3.1.50, the other inspired by the solution of Exercise 3.1.51.

Exercise 3.1.53. Prove that the Petersen graph (Figure 3.8) is not Hamiltonian; its longest cycle has 9 vertices. (No "Ah-ha!" proof of this statement is known.)

Exercise 3.1.54. Prove: if G is regular of degree r and G has girth ≥ 5 then $n \geq r^2 + 1$. (n is the number of vertices.) Show that $n = r^2 + 1$ is possible for $r = 1, 2, 3$.

Exercise 3.1.55. (a) Prove: if a graph G with n vertices is regular of degree r and has diameter 2 then $n \leq r^2 + 1$.

(b) Prove that if G is as in part (a) and $n = r^2 + 1$ then G has girth 5.

(c) Show that there exists a graph G satisfying the conditions of part (a) and the equation $n = r^2 + 1$ if $r = 2$ or $r = 3$ (what is the name of your graph?). *Remark.* $n = r^2 + 1$ is possible also if $r = 7$ (the "Hoffman–Singleton graph"). It is known **(Hoffmann–Singleton, 1960)** that the only values of r for which $n = r^2 + 1$ is conceivable are $2, 3, 7$, and 57. The proof is one of the gems of the applications of linear algebra (the Spectral Theorem) to graph theory. The question whether or not $r = 57$ can actually occur is open.

Exercise 3.1.56. An *automorphism* of the graph G is a $G \to G$ isomorphism. (a) Count the automorphisms of K_n, C_n, P_n, Q_n. (b)$^+$ Show that the dodecahedron has 120 automorphisms. (c)$^+$ Show that the Petersen graph (Figure 3.8) has 120 automorphisms.

Exercise$^+$ 3.1.57. (Requires some group theory.) The automorphisms of a graph form a *group* under composition. Prove: (a) The automorphism groups of the dodecahedron and of Petersen's graph are not isomorphic. *Hint:* the automorphism group of the dodecahedron has nontrivial center, i. e., it has an element, other than the identity, which commutes with all elements. (b) The automorphism group of the dodecahedron is isomorphic to $C_2 \times A_5$ where C_2 is the cyclic group of order 2 and A_5 is the alternating group of degree 5. (c) The automorphism group of the Petersen graph is isomorphic to S_5, the symmetric group of degree 5. (This statement has an "Ah-ha!" proof.)

Exercise 3.1.58. Decide whether or not Petersen's graph (Fig. 3.8) is isomorphic to the graph in Fig. 3.9.

Planarity

A *plane graph* is a graph drawn in the plane so that the lines (curves) representing the edges do not intersect (except at their end vertices). A graph is *planar* if it admits a plane drawing; such plane drawings are the *plane representations* of the graph. Of course a planar

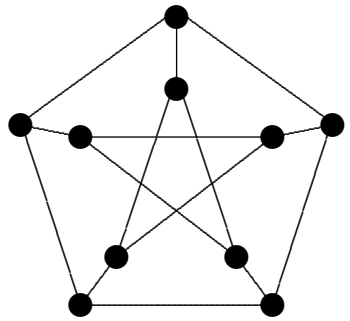


Figure 3.8: The Petersen graph.

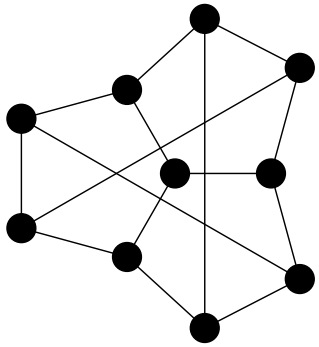


Figure 3.9: Is this graph isomorphic to Petersen's (Fig. 3.8)?

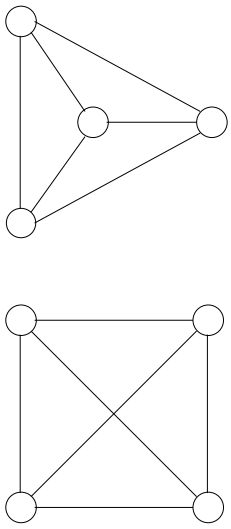


Figure 3.10: K_4 drawn two different ways. Only one is a plane graph.

graph may also have drawings that are not plane graphs (e. g., K_4 is a planar graph - a plane representation is a regular triangle with its center, with their conecting straight line segments; a drawing of K_4 which is not a plane graph is the square with all sides and diagonals–see Figure 3.10).

The *regions* of a plane graph are the regions into which the drawing divides the plane; so two points of the plane belong to the same region if they can be connected so that the connecting line does not intersect the drawing. Note that the infinite "outer region" counts as a region.

WARNING: it is incorrect to speak of regions of a *planar* graph; only a *plane* graph has regions. A planar graph may have many inequivalent plane representations; the sizes of the regions may depend on the representation.

Exercise 3.1.59. Prove: every plane representation of a tree has just one region. *Hint.* Induction (use the fact that the tree has a vertex of degree 1).

We need the following, highly nontrivial result.

Theorem 3.1.60. (Jordan's Curve Theorem) *Every plane representation of a cycle has two regions.*

Exercise 3.1.61. (Euler's formula) For a connected plane graph, let n, m, r denote the set of vertices, edges, and regions, respectively. Then $n - m + r = 2$. *Note* that this statement includes Jordan's Curve Theorem and the exercise before that. *Hint.* Induction on m. Unless the graph is a tree, delete an edge contained in a cycle; verify that this reduces the number of regions by 1. Trees are the base case.

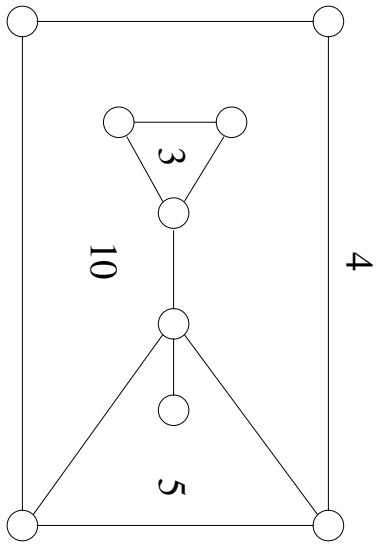


Figure 3.11: The numbers indicate the number of sides of each region of this plane graph.

Exercise 3.1.62. Verify that the Platonic solids satisfy Euler's formula.

Exercise 3.1.63. Let r_k denote the number of k-sided regions of a plane graph. (In a plane graph, an edge has two sides, and it is possible that both sides are incident with the same region. In such a case this edge contributes 2 to the number of sides of the region. See Figure 3.11.) Prove: $\sum_{k=3}^{n} r_k = 2m$.

Exercise 3.1.64. Prove: in a plane graph, $3r \leq 2m$.

Exercise 3.1.65. Prove: in a plane graph without triangles, $2r \leq m$.

Exercise 3.1.66. Prove: a planar graph with $n \geq 3$ vertices has $m \leq 3n - 6$ edges. *Hint.* Use Euler's formula and the inequality $3r \leq 2m$.

Exercise 3.1.67. Prove: a triangle-free planar graph with $n \geq 3$ vertices has $m \leq 2n - 4$ edges. *Hint.* Use Euler's formula and the inequality $2r \leq m$.

Exercise 3.1.68. Prove: the graphs K_5 and $K_{3,3}$ are not planar. *Hint.* Use the preceding two exercises.

A *subdivision* of a graph is obtained by subdividing some of its edges by new vertices. For instance, the cycle C_n is a subdivision of the triangle C_3; the path P_n is a subdivision of an edge. Two graphs are *homeomorphic* if both of them is a subdivision of the same graph. For instance, all cycles (including C_3) are homeomorphic. Homeomorphic planar graphs have identical plane drawings.

Kuratowski's celebrated theorem gives a *good characterization* of planarity.

Theorem 3.1.69. *A graph is planar if and only if it does not contain a subgraph homeomorphic to $K_{3,3}$ or K_5.*

The two minimal non-planar graphs, $K_{3,3}$ and K_5, are referred to as "Kuratowski graphs."

Exercise 3.1.70. Draw a BIPARTITE graph G which is NOT planar and does NOT contain a subdivision of $K_{3,3}$. Make a clean drawing; your graph should have no more than 20 edges. Prove that your graph has all the required properties.

Exercise 3.1.71. Prove: (a) if a connected graph G has n vertices and $n+2$ edges then G is planar. (b) Show that for every $n \geq 6$, statement (a) becomes false if we replace $n+2$ by $n+3$. (You must construct an infinite family of counterexamples, one graph for each $n \geq 6$.)

Exercise 3.1.72. Prove that every planar graph has a vertex of degree ≤ 5. *Hint.* $m \leq 3n-6$.

Exercise 3.1.73. Prove that every planar graph is 6-colorable. *Hint.* Induction, using the preceding exercise.

The famous **4-Color Theorem** of Appel and Haken asserts that every planar graph is 4-colorable. The proof considers hundreds of cases; no "elegant" proof is known.

Exercise 3.1.74. Prove: if a planar graph G has n vertices then $\alpha(G) \geq n/6$. (Recall that $\alpha(G)$ denotes the maximum number of independent vertices in G.) *Hint.* Use the preceding exercise.

Prove that every triangle-free planar graph has a vertex of degree ≤ 3. *Hint.* $m \leq 2n-4$.

Exercise 3.1.75. Prove that every triangle-free planar graph is 4-colorable.

Ramsey Theory

The Erdős–Rado arrow notation $n \to (k, \ell)$ means that every graph on n vertices either has a clique of size $\geq k$ or an independent set of size $\geq \ell$. In other words, if we color the edges of K_n red and blue, there will either be an all-red K_k or an all-blue K_ℓ.

Exercise 3.1.76. Prove: (a) $6 \to (3,3)$; (b) $10 \to (4,3)$; (c) $n \to (n,2)$.

Exercise 3.1.77. (Erdős–Szekeres, 1933)

$$\binom{r+s}{r} \to (r+1, s+1).$$

Hint. Induction on $r+s$.

Exercise 3.1.78. Prove: $n \to (k,k)$ where $k = \lceil \log_2 n/2 \rceil$.

Exercise 3.1.79. Define and prove: $17 \to (3,3,3)$.

3.2 Digraph Terminology

A **directed graph** (digraph, for short), is a pair $G = (V, E)$, where V is the set of "vertices" and E is a set of ordered pairs of vertices called "edges:" $E \subseteq V \times V$.

Exercise 3.2.1. If G has n vertices and m edges then $m \leq n^2$.

"Graphs," also referred to as **undirected graphs**, can be represented as digraphs by introducing a pair of directed edges, (u, v) and (v, u), for every undirected edge $\{u, v\}$ of a graph. (So the digraph G corresponding to the graph G_0 has twice as many edges as G_0.)

Adjacency. We say that u is adjacent to v, denoted $u \to v$, if $(u, v) \in E$. *Self-adjacency* may occur; an edge $(u, u) \in E$ is called a **loop.**

We shall say that a digraph is **undirected** if the adjacency relation is symmetric ($u \to v$ implies $v \to u$). We say that a digraph is a "graph" if it is undirected and has no *loops*, i. e., no self-adjacencies ($v \not\to v$).

The **converse** of a digraph $G = (V, E)$ is the digraph $G^{op} = (V, E^{op})$ where E^{op} consists of all edges of G reversed: $E^{op} = \{(v, u) : (u, v) \in E\}$. Note that G is **undirected** if and only if $G = G^{op}$. – The superscript "tr" refers to "transpose," for a reason to be clarified below.

Orientations of a graph. Let $G_0 = (V, E_0)$ be a graph. We say that the digraph $G = (V, E)$ is an **orientation** of G_0 if for each edge $\{u, v\} \in E_0$, exactly one of (u, v) and (v, u) belongs to E.

Exercise 3.2.2. Suppose the graph G_0 has n vertices and m edges. Count the orientations of G_0.

Tournaments are orientations of complete graphs. So in a tournament $G = (V, E)$, for every pair of vertices $u, v \in V$, exactly one of the following holds: (a) $u = v$; (b) $u \to v$; (c) $v \to u$. We often think of the vertices of a tournament as players in a round-robin tournament without ties or rematches. Each player plays against every other player exactly once; $u \to v$ indicates that player u beat player v.

Exercise 3.2.3. Count the tournaments on a given set of n vertices. Is the similarity with the number of graphs a coincidence?

Neighbors. If $u \to v$ in a digraph then we say that v is an **out-neighbor** or **successor** of u; and u is an **in-neighbor** or **predecessor** of v.

Degrees. The **out-degree** $\deg^+(v)$ of vertex v is the number of its out-neighbors; the **in-degree** $\deg^-(v)$ of v is the number of its in-neighbors.

Exercise 3.2.4. Prove: if the digraph $G = (V, E)$ has n vertices and m edges then

$$\sum_{v \in V} \deg^+(v) = \sum_{v \in V} \deg^-(v) = m.$$

Exercise 3.2.5. Prove: if every vertex of a digraph G has the same out-degree d^+ and the same in-degree d^- then $d^+ = d^-$.

An **isomorphism** between the digraphs $G = (V, E)$ and $H = (W, F)$ is a bijection $f : V \rangle W$ from V to W which preserves adjacency, i. e., $(\forall x, y \in V)(x \to_G y \Leftrightarrow f(x) \to_H f(y)$. Two digraphs are **isomorphic** if there *exists* an isomorphism between them.

Let p be a prime. An integer z is a **quadratic residue** modulo p if $z \not\equiv 0 \pmod{p}$ and $(\exists x)(x^2 \equiv z \pmod{p})$.

Exercise 3.2.6. List the quadratic residues modulo 5 and modulo 7.

Exercise 3.2.7. Prove that if p is an odd prime then the number of non-congruent quadratic residues modulo p is $(p-1)/2$.

Exercise$^+$ 3.2.8. Prove: -1 is a quadratic residue mod p if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.

Paley graphs/tournaments. Let p be an odd prime. Let $V = \{0, 1, \ldots, p-1\}$. Let us set $u \to v$ if $u - v$ is a quadratic residue mod p. $(0 \le u, v \le p - 1.)$

Exercise 3.2.9. Prove: the preceding construction defines a tournament (the **Paley tournament**) if $p \equiv -1 \pmod{4}$; and it defines a graph (the **Paley graph**) if $p \equiv 1 \pmod{4}$.

A digraph is **self-converse** if it is isomorphic to its converse.

Exercise$^+$ 3.2.10. Prove: (a) The Paley tournaments are self-converse. (b) The Paley tournaments are self-complementary.

Exercise* 3.2.11. (Erdős.) We say that a tournament is **k-paradoxical** if to every k players there exists a player who beat all of them. Prove that if $n > 2k^2 2^k$ then there exists a k-paradoxical tournament on n vertices. *Hint.* Use the probabilistic method: prove that *almost all tournaments* are k-paradoxical.

Exercise** 3.2.12. (Graham – Spencer) If p is a prime, $p \equiv -1 \pmod{4}$ and $p > 2k^2 4^k$ then the Paley tournament on p vertices is k-paradoxical. *Hint.* The proof uses André Weil's character sum estimates.

Last update: October 12, 2004

Directed walks, paths, cycles

- **(directed) walk** (in text: *path*) of length k: a sequence of $k+1$ vertices $v_0, \ldots, v_k$ such that $(\forall i)(v_{i-1} \to v_i)$.
- **(directed) trail** (in text: *simple path*): a walk without repeated edges.
- **(directed) path:** (this all-important concept has no name in the text): a walk without repeated vertices. (Note that the terms "path" and even "simple path" in the text allow vertices to be repeated.) $\vec{P}_{k+1}$ denotes a directed path of length k (it has $k+1$ vertices)
- **closed (directed) walk** (in text: *circuit* or *cycle*) of length k: a (directed) walk $v_0, \ldots, v_k$ where $v_k = v_0$.
- **(directed) cycle of length** k or k**-cycle**: (this all-important concept has no name in the text): a closed walk of length k with no repeated vertices except that $v_0 = v_k$. Notation: $\vec{C}_k$.
- a vertex v is **accessible** from a vertex u if there exists a $u \to \cdots \to v$ directed path.

Exercise 3.2.13. Prove that the relation "u and v are mutually accessible from each other" is an **equivalence relation** on the set of vertices of the digraph G, i. e., this relation is *reflexive, symmetric, and transitive.*

- The **strong components** of G are the equivalence classes of this relation, i. e., the *maximal* subsets of the vertex set consisting of mutually accessible vertices. The vertex set of G is the disjoint union of the strong components. In other words, **each vertex belongs to exactly one strong component.** So the vertices u and v **belong to the same strong component** if they are mutually accessible from each other.
- a digraph G is **strongly connected** if there is a (directed) path between each pair of vertices, i. e., all vertices belong to the same strong component. (There is just one strong component.)
- an *undirected walk (path, cycle, etc.)* in a digraph is a walk (path, cycle, etc.) in the undirected graph obtained by ignoring orientation.
- a digraph is **weakly connected** if there is an undirected path between each pair of vertices.

Exercise 3.2.14. Prove that a weakly connected digraph has $\geq n-1$ edges; a strongly connected digraph has $\geq n$ edges.

Exercise$^+$ 3.2.15. Prove: if $(\forall v \in V)(\deg^+(v) = \deg^-(v))$ and G is weakly connected then G is strongly connected.

- A **Hamilton cycle** in a digraph is a (directed) cycle of length n, i. e., a cycle that passes through all vertices. G is **Hamiltonian** if it has a Hamilton cycle.
- A **Hamilton path** in a digraph is a (directed) path of length $n-1$, i. e., a path that passes through all vertices.

Exercise 3.2.16. Prove that every tournament has a Hamilton path.

Exercise$^+$ 3.2.17. Prove that every strongly connected tournament is Hamiltonian.

- A DAG **(directed acyclic graph)** is a digraph with no directed cycles.

Exercise 3.2.18. Prove that for every n, there exists exactly one tournament (up to isomorphism) which is a DAG.

- A **topological sort** of a digraph is an ordering of its vertices such that all edges go "forward:" if $u \to v$ then u precedes v in the ordering.

For example, if $V = \{1, 2, \ldots, n\}$ and $u \to v$ means $u \neq v$ and $u|v$ (u divides v) then the natural ordering of integers is a topological sort; but it is not the only possible topological sort of this digraph.)

Exercise 3.2.19. Prove that the "divisibility digraph" described in the preceding paragraph has at least $\lfloor n/2 \rfloor !$ topological sorts.

Exercise 3.2.20. Prove that a digraph G can be topologically sorted if and only if G is a DAG. – Note that this is a **good characterization**: the existence of an object (topological sort) is shown to be equivalent to the nonexistence of another (directed cycle).

The adjacency matrix

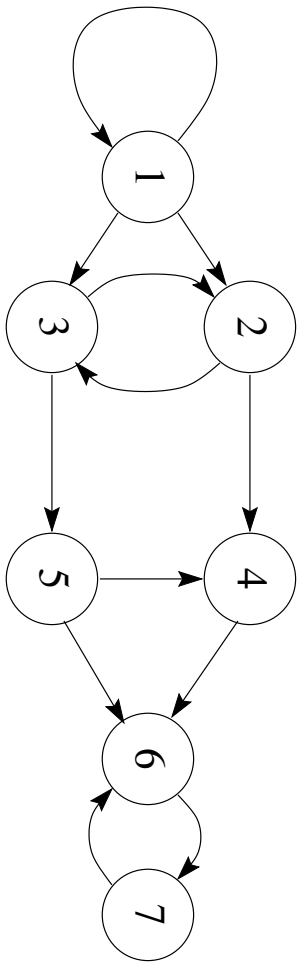
Let $G = (V, E)$ be a digraph; assume $V = [n] = \{1, 2, \ldots, n\}$. Consider the $n \times n$ matrix $A_G = (a_{ij})$ defined as follows: $a_{ij} = 1$ if $i \to j$; and $a_{ij} = 0$ otherwise. A_G is the **adjacency matrix** of G.

Exercise 3.2.21. Prove: The adjacency matrix of the G^{op} (the converse of G) is A_G^{tr} (the transpose of A_G. In particular, the digraph G is undirected if and only if A_G is a symmetric matrix.

Exercise$^+$ 3.2.22. (Counting walks) For $k \geq 0$, let a_{ijk} denote the **number of directed walks** of length k from vertex i to vertex j. Consider the matrix $A_G(k)$ which has a_{ijk} as its entry in row i, column j. Prove: $A_G(k) = A_G^k$. *Hint.* Induction on k.

Exercise 3.2.23. Let T be a tournament with n vertices. Prove: if all vertices have the same out-degree then n is odd.

Exercise 3.2.24. List the strong components of the digraph in the figure below. State the number of strong components. Recall that two vertices x and y belong to the same strong component if either $x = y$ or there exists $x \to y$ and $y \to x$ directed walks. The strong components are the equivalence classes of this equivalence relation, so each strong component is either a single vertex or a maximal strongly connected subgraph.



Exercise 3.2.25. Let $p_1, \ldots, p_k$ be distinct prime numbers and let $n = \prod_{i=1}^k p_i$. Let D denote the set of positive divisors of n.

1. Determine $|D|$ (the size of D). (Your answer should be a very simple formula.)
2. We define a digraph G with vertex set $V(G) := D$ by setting $i \to j$ if $j \mid i$ and i/j is a prime number ($i, j \in D$). Determine the number of directed paths from n to 1 in G. (Again, your answer should be a very simple formula.)
3. Prove that this digraph is self-converse (isomorphic to the digraph obtained by reversing all arrows). (You need to state a bijection $f : D \mapsto D$ which reverses all arrows. You should define f by a very simple formula.)

Definition 3.2.26. Let v be a vertex in a directed graph. The *period* of v is defined as the g.c.d. of the lengths of all closed walks containing v.

Exercise 3.2.27. Let G be a directed graph. Prove: if $v, w \in V$ are two vertices in the same strong component of G then their periods are equal.

Chapter 4

Finite Probability Spaces

4.1 Finite Probability Spaces and Events

Definition 4.1.1. A **finite probability space** is a finite set $\Omega \neq \emptyset$ together with a function $\Pr : \Omega \to \mathbf{R}^+$ such that

1. $\forall \omega \in \Omega,\ \Pr(\omega) > 0$
2. $\sum_{\omega \in \Omega} \Pr(\omega) = 1.$

The set Ω is the **sample space** and the function Pr is the **probability distribution**. The elements $\omega \in \Omega$ are called **atomic events** or **elementary events**. An **event** is a subset of Ω. For $A \subseteq \Omega$, we define the **probability** of A to be $\Pr(A) := \sum_{\omega \in A} \Pr(\omega)$. In particular, for atomic events we have $\Pr(\{\omega\}) = \Pr(\omega)$; and $\Pr(\emptyset) = 0$, $\Pr(\Omega) = 1$. The **trivial events** are those with probability 0 or 1, i. e. $\emptyset$ and Ω.

The **uniform distribution** over the sample space Ω is defined by setting $\Pr(\omega) = 1/|\Omega|$ for every $\omega \in \Omega$. With this distribution, we shall speak of the **uniform probability space** over Ω. In a uniform space, calculation of probabilities amounts to counting: $\Pr(A) = |A|/|\Omega|$.

Exercise 4.1.2. In the card game of bridge, a deck of 52 cards are evenly distributed among four players called North, East, South, and West. What sample space does each of the following questions refer to: (a) What is the probability that North holds all the aces? (b) What is the probability that each player holds one of the aces? – These questions refer to uniform probability spaces. Calculate the probabilities.

Exercise 4.1.3. We flip n coins. (a) What is the size of the sample space of this experiment? Show that the sample space has size 2^n. Is this a uniform probability space? (b) What is the probability that exactly k coins come up heads? (c) What is the probability that an even number of coins come up heads?

Exercise 4.1.4. We flip $2n$ coins. Let ξ denote the number of Heads among the first n coins and η the number of Heads among the last n coins. Let p_n denote the probability of the event that $\xi = \eta$. (a) Determine p_n. Give a very simple closed-form expression (no summation symbols or ellipses (dot-dot-dots). (b) Asymptotically evaluate p_n. Prove that there exists a constant c such that $p_n \sim c/\sqrt{n}$. Determine c.

Observation 4.1.5. $\Pr(A \cup B) + \Pr(A \cap B) = \Pr(A) + \Pr(B)$.

Definition 4.1.6. Events A and B are **disjoint** if $A \cap B = \emptyset$.

Consequence 4.1.7. $\Pr(A_1 \cup \cdots \cup A_k) \leq \sum_{i=1}^{k} \Pr(A_i)$, *and equality holds if and only if the* A_i *are pairwise disjoint.*

Definition 4.1.8. If A and B are events and $\Pr(B) > 0$ then the **conditional probability of** A **relative to** B, written $\Pr(A|B)$, is given by $\Pr(A|B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$.

We note that B can be viewed as a sample space with the probabilities being the conditional probabilities under condition B.

Note that $\Pr(A \cap B) = \Pr(A|B) \Pr(B)$.

Exercise 4.1.9. Prove: $\Pr(A \cap B \cap C) = \Pr(A|B \cap C) \Pr(B|C) \Pr(C)$.

Exercise 4.1.10. We roll three dice. What is the probability that the sum of the three numbers we obtain is 9? What is the probability that the first die shows 5? What is the conditional probability of this event assuming the sum of the numbers shown is 9? – What is the probability space in this problem? How large is the sample space?

A **partition** of Ω is a family of pairwise disjoint events $H_1, \ldots, H_m$ covering Ω:

$$\Omega = H_1 \cup \ldots \cup H_k, \qquad H_i \cap H_j = \emptyset. \tag{4.1}$$

The sets H_i are the *classes* of the partition. We assume that each class is nonempty.

Exercise 4.1.11. *"Theorem of Complete Probability."* Prove: given a partition $(H_1, \ldots, H_k)$ of Ω, we have

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A|H_i) \Pr(H_i). \tag{4.2}$$

for any event $A \subseteq \Omega$.

The significance of this formula is that the conditional probabilities are sometimes easier to calculate than the left hand side.

Definition 4.1.12. Events A and B are **independent** if $\Pr(A \cap B) = \Pr(A) \Pr(B)$.

Exercise 4.1.13. Pick a card at random from the standard deck of 52 cards. Are the following events independent: "the card picked is a King;" and "the card picked is of the 'Heart suit' "?

Exercise 4.1.14. We flip $2n+1$ coins. Are the following events independent: "the majority of the coins comes up Heads"; and "an even number of coins comes up heads."

Exercise 4.1.15. If $\Pr(B) > 0$ then: A and B are independent $\iff \Pr(A|B) = \Pr(A)$.

Exercise 4.1.16. Prove: if A and B are independent events then $\bar{A}$ and B are also independent, where $\bar{A} = \Omega \setminus A$.

Exercise 4.1.17. Let us consider a uniform probability space over a sample space whose cardinality is a prime number. Prove that no two non-trivial events can be independent.

Note that the trivial events are independent of any other events, i. e. if a trivial event is added to a collection of independent events, they remain independent.

The events A and B are said to be **positively correlated** if $\Pr(A \cap B) > \Pr(A)\Pr(B)$. They are **negatively correlated** if $\Pr(A \cap B) < \Pr(A)\Pr(B)$.

Exercise 4.1.18. Are the two events described in Exercise 4.1.10 positively, or negatively correlated, or independent?

Exercise 4.1.19. We roll a die. Consider the following events: A : "the number shown is odd"; B : "the number shown is prime"; (c) "the number shown is $\equiv 1 \pmod 4$. Decide whether the following pairs of events are independent; if they are not, determine whether they are positively or negaively correlated: (i) A and B; (ii) A and C; (iii) B and C.

Exercise 4.1.20. Prove: two events A, B are positively (negatively) correlated if and only if $\Pr(B|A) > \Pr(B)$ ($\Pr(B|A) < \Pr(B)$, resp.).

Definition 4.1.21. Events $A_1, \ldots, A_k$ are **independent** if for all subsets $S \subseteq \{1, \ldots, k\}$, we have

$$\Pr(\cap_{i \in S} A_i) = \prod_{i \in S} \Pr(A_i). \tag{4.3}$$

Note that if $k \geq 3$, then the statement that events $A_1, \ldots, A_k$ are independent is stronger than pairwise independence. For example, pairwise independence does not imply triplewise independence. For added emphasis, independent events are sometimes called *fully* independent, or *mutually* independent, or *collectionwise* independent.

Exercise 4.1.22. Construct 3 events which are pairwise independent but not collectionwise independent. What is the smallest sample space for this problem?

(See the end of this section for more general problems of this type.)

Exercise 4.1.23. Prove: if the events A, B, C, D, E are independent then the events $A \setminus B$, $C \cup D$, and E are independent as well. Formulate a general statement, for n events grouped into blocks.

Exercise 4.1.24. We have n balls colored red, blue, and green (each ball has exactly one color and each color occurs at least once). We select k of the balls with replacement (independently, with uniform distribution). Let A denote the event that the k balls selected have the same color. Let p_r denote the conditional probaility that the first ball selected is red, assuming condition A. Define p_b and p_g analogously for blue and green outcomes. Assume $p_1 + p_2 = p_3$. Prove: $k \leq 2$. Show that $k = 2$ is actually possible.

Exercise 4.1.25. (Random graphs) Consider the uniform probability space over the set of all the $2^{\binom{n}{2}}$ graphs with a given set V of n vertices. (a) What is the probability that a particular pair of vertices is adjacent? Prove that these $\binom{n}{2}$ events are independent. (b) What is the probability that the degrees of vertex 1 and vertex 2 are equal? Give a closed-form expression. (c) If p_n denotes the probability calculated in part (b), prove that $p_n\sqrt{n}$ tends to a finite positive limit and determine its value. (c) How are the following two events correlated: A:"vertex 1 has degree 3"; B:"vertex 2 has degree 3"? Asymptotically evaluate the ratio $\Pr(A|B)/\Pr(A)$.

In exercises like the last one, one often has to estimate binomial coefficients. The following result comes in handy:

Stirling's formula.

$$n! \sim (n/e)^n\sqrt{2\pi n}. \tag{4.4}$$

Here the $\sim$ notation refers to *asymptotic equality:* for two sequences of numbers a_n and b_n we say that a_n and b_n are **asymptotically equal** and write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.

To "evaluate a sequence a_n asymptotically" means to find a simple expression describing a function $f(n)$ such that $a_n \sim f(n)$. Striling's formula is such an example. While such "asymptotic formulae" are excellent at predicting what happens for "large" n, they do not tell how large is large enough.

A stronger, non-asymptotic variant, giving useful results for specific values of n, is the following:

$$n! = (n/e)^n\sqrt{2\pi n}(1 + \theta_n/(12n)), \tag{4.5}$$

where $|\theta_n| \leq 1$.

Exercise 4.1.26. Evaluate asymptotically the binomial coefficient $\binom{2n}{n}$. Show that $\binom{2n}{n} \sim c \cdot 4^n/\sqrt{n}$ where c is a constant. Determine the value of c.

We mention some important asymptotic relations from number theory. Let $\pi(x)$ denote the number of all prime numbers $\leq x$, so $\pi(2) = 1$, $\pi(10) = 4$, etc. The **Prime Number Theorem** of Hadamard and de la Vallée-Poussin asserts that

$$\pi(x) \sim x/\ln x. \tag{4.6}$$

Another important relation estimates the sum of reciprocals of prime numbers. The summation below extends over all primes $p \leq x$.

$$\sum_{p \leq x} 1/p \sim \ln \ln x. \tag{4.7}$$

In fact a stronger result holds: there exists a number B such that

$$\lim_{x \to \infty} \left(\sum_{p \leq x} 1/p - \ln \ln x \right) = B. \tag{4.8}$$

(Deduce (4.7) from (4.8).)

Exercise 4.1.27. Assuming 100-digit integers are "large enough" for the Prime Number Theorem to give a good approximation, estimate the probability that a random integer with at most 100 decimal digits is prime. (The integer is drawn with uniform probability from all positive integers in the given range.)

Exercise 4.1.28. Construct a sample space Ω and events $A_1, \ldots, A_n$ ($\forall n \geq 2$) such that $\Pr(A_i) = 1/2$, every $n-1$ of the A_i are independent, but the n events are *not* independent.

Exercise 4.1.29. (*) Let $1 \leq k \leq n-1$. (a) Construct a sample space Ω and n events such that every k of these n events are independent; but no $k+1$ of these events are independent. (b) Solve part (a) under the additional constraint that each of the n events have probability $1/2$.

(Hint. Take a k-dimensional vector space W over a finite field of order $q \geq n$. Select n vectors from W so that any k are linearly independent. Let W be the sample space.)

Exercise 4.1.30. Suppose we have n independent nontrivial events. Prove: $|\Omega| \geq 2^n$.

Exercise 4.1.31. (Small sample space for pairwise independent events.) (a) For $n = 2^k - 1$, construct a probability space of size $n+1$ with n pairwise independent events each of probability $1/2$. (b)* Same for n a prime number of the form $n = 4k - 1$.

Exercise 4.1.32. (*) Prove: if there exist n pairwise independent nontrivial events in a probability space then $|\Omega| \geq n + 1$. (If this is too difficult, solve the special case when all events considered have probability $1/2$ and the space is uniform.)

4.2 Random Variables and Expected Value

Definition 4.2.1. A **random variable** is a function $\xi : \Omega \to \mathbf{R}$.

We say that ξ is **constant** if $\xi(\omega)$ takes the same value for all $\omega \in \Omega$.

Definition 4.2.2. The **expected value** of a random variable ξ is $\mathrm{E}(\xi) = \sum_{\omega\in\Omega} \xi(\omega)\Pr(\omega)$.

Proposition 4.2.3. *Let $\{u_1, \ldots, u_k\}$ be the set of (distinct) values taken by ξ. Let $p_i = \Pr(\xi = u_i)$, where the statement "$\xi = u_i$" refers to the event $\{\omega : \xi(\omega) = u_i\}$. Then $\mathrm{E}(\xi) = \sum_{i=1}^{k} u_i p_i$.*

Proof: Exercise.

Exercise 4.2.4.

$$\min \xi \le \mathrm{E}(\xi) \le \max \xi. \tag{4.9}$$

Throughout these notes, $\xi, \eta, \zeta, \vartheta$, and their subscripted versions refer to random variables.

Proposition 4.2.5. (Additivity of the Expected Value) *Let $\xi_1, \ldots, \xi_k$ be arbitrary random variables. Then*

$$\mathrm{E}(\xi_1 + \cdots + \xi_k) = \sum_{i=1}^{k} \mathrm{E}(\xi_i) \tag{4.10}$$

Proof: $\mathrm{E}(\sum_{i=1}^{k} \xi_i) = \sum_{\omega\in\Omega} (\xi_1(\omega) + \cdots + \xi_k(\omega)) \Pr(\omega) = \sum_{i=1}^{k} \sum_{\omega\in\Omega} \xi_i \Pr(\omega) = \sum_{i=1}^{k} \mathrm{E}(\xi_i).$ □

Exercise 4.2.6. (Linearity of the expected value.) If $c_1, \ldots, c_k$ are constants then

$$\mathrm{E}(\sum_{i=1}^{k} c_i \xi_i) = \sum_{i=1}^{k} c_i \mathrm{E}(\xi_i). \tag{4.11}$$

Definition 4.2.7. The **indicator variable** of an event $A \subseteq \Omega$ is the function $\vartheta_A : \Omega \to \{0, 1\}$ given by

$$\vartheta_A(\omega) = \begin{cases} 1 & \text{for } \omega \in A \\ 0 & \text{for } \omega \notin A \end{cases}$$

Exercise 4.2.8. The expected value of an indicator variable is $\mathrm{E}(\vartheta_A) = \Pr(A)$.

Indicator variables are particularly useful if we want to count events. Some of the exercises at the end of this section should serve as examples.

Exercise 4.2.9. (a) Every random variable ξ is a linear combination of indicator variables. (b) Given a random variable ξ there exist functions $f_1, \ldots, f_k$ such that the random variables $\xi_i := f_i(\xi)$ are indicator variables and ξ is a linear combination of the ξ_i.

We say that ξ is **nonnegative** if $\xi(\omega) \geq 0$ for all $\omega \in \Omega$.

Theorem 4.2.10 (Markov's Inequality). *If ξ is nonnegative then $\forall a > 0$,*

$$\Pr(\xi \geq a) \leq \frac{\mathrm{E}(\xi)}{a}.$$

Proof: Let $m = \mathrm{E}(\xi) > 0$. Then $m = \sum_i \mu_i \Pr(\xi = \mu_i) \geq$
$\geq \sum_{\mu_i \geq a} \mu_i \Pr(\xi = \mu_i)$ (we just omitted some terms; all terms are nonegative)
$\geq a \sum_{\mu_i \geq a} \Pr(\xi = \mu_i) = a \Pr(\xi \geq a)$ (sum of disjoint events).
So we have $m \geq a \Pr(\xi \geq a)$. □

Exercise 4.2.11. What is the expected number of runs of k heads in a string of n coin-flips? (A "run of k heads" means a string of k consecutive heads. Example: the string HHTHTTHHHT has 3 runs of 2 heads.) Prove your answer! *Hint.* Indicator variables.

Exercise 4.2.12. Suppose in a lottery you have to pick five different numbers from 1 to 90. Then five winning numbers are drawn. If you picked two of them, you win 20 dollars. For three, you win 150 dollars. For four, you win 5,000 dollars, and if all the five match, you win a million. (a) What is the probability that you picked exactly three of the winning numbers? (b) What is your expected win? (c) What does Markov's inequality predict about the probability that you'll win at least 20 dollars? (d) What is the actual probability that this happens?

Exercise 4.2.13. A club with 2000 members distributes membership cards numbered 1 through 2000 to its members at random; each of the 2000! permutations of the cards is equally likely. Members whose card number happens to coincide with their year of birth receive a prize. Determine the expected number of lucky members.

Exercise 4.2.14. What is the expected number of edges in a random graph? What is the expected number of triangles? (There are n vertices; each pair is adjacent with probability 1/2 independently.)

Exercise 4.2.15. Let n be a random integer, chosen uniformly between 1 and N. What is the expected number of distinct prime divisors of n? Show that the result is asymptotically equal to $\ln \ln N$ (as $N \to \infty$).

Exercise 4.2.16. The boss writes n different letters to n addressees whose addresses appear on n envelopes. The careless secretary puts the letters in the envelopes at random (one letter per envelope). Determine the expected number of those letters which get in the right envelope. Prove your answer. State the size of the sample space for this problem.

Exercise 4.2.17. For a permutation $\pi \in S_n$, let $c_k(\pi)$ denote the number of k-cycles in the cycle decomposition of π. (For instance, if $n = 7$ and $\pi = (13)(256)(47)$ then $c_2(\pi) = 2$, $c_3(\pi) = 1$, and $c_k(\pi) = 0$ for all $k \neq 2, 3$.) Pick π at random (from S_n). Calculate $E(c_k(\pi))$. Your answer should be a very simple expression (no factorials, no binomial coefficients, no summation). Prove your answer.

4.3 Standard deviation and Chebyshev's Inequality

Definition 4.3.1. The k^{th} **moment** of ξ is $\mathrm{E}(\xi^k)$. The k^{th} **central moment** of ξ is the k^{th} moment of $\xi - \mathrm{E}(\xi)$, i. e. $\mathrm{E}((\xi - \mathrm{E}(\xi))^k)$.

Definition 4.3.2. The **variance** of ξ is its second central moment, $\mathrm{Var}(\xi) := \mathrm{E}((\xi - \mathrm{E}(\xi))^2)$.

Note that the variance is always nonnegative. It is zero exactly if ξ is constant. (Why?)

Definition 4.3.3. The **standard deviation** of ξ is $\sigma(\xi) := \sqrt{\mathrm{Var}(\xi)}$.

Exercise 4.3.4. (Invariance under shifts.) Prove that if c is a constant then $\mathrm{Var}(\xi) = \mathrm{Var}(\xi + c)$; and consequently, $\sigma(\xi) = \sigma(\xi + c)$.

Exercise 4.3.5. Prove: if c is a constant then $\mathrm{Var}(c\xi) = c^2 \mathrm{Var}(\xi)$; and consequently, $\sigma(c\xi) = |c|\sigma(\xi)$.

Observation 4.3.6. $\mathrm{Var}(\xi) = \mathrm{E}(\xi^2) - (\mathrm{E}(\xi))^2$.

Corollary 4.3.7 (Cauchy-Schwarz inequality). $(\mathrm{E}(\xi))^2 \leq \mathrm{E}(\xi^2)$. □

Proof of Observation: Let $m = \mathrm{E}(\xi)$. Then $\mathrm{Var}(\xi) = \mathrm{E}((\xi - m)^2) = \mathrm{E}(\xi^2 - 2\xi m + m^2) = \mathrm{E}(\xi^2) - 2m\mathrm{E}(\xi) + \mathrm{E}(m^2) = \mathrm{E}(\xi^2) - 2mm + m^2 = \mathrm{E}(\xi^2) - m^2$. □

Chebyshev's inequality tells us that random variables don't like to stray away from their expected value by more than a small multiple of their standard deviation.

Theorem 4.3.8 (Chebyshev's Inequality). *Let $m = \mathrm{E}(\xi)$. Then for any number $a > 0$,*

$$\Pr(|\xi - m| \geq a) \leq \frac{\mathrm{Var}(\xi)}{a^2}. \tag{4.12}$$

Proof: Let $\eta = (\xi - m)^2$. Then, by definition, $\mathrm{E}(\eta) = \mathrm{Var}(\xi)$. We apply Markov's Inequality to the nonnegative random variable η: $\Pr(|\xi - m| \geq a) = \Pr(\eta \geq a^2) \leq \mathrm{E}(\eta)/a^2 = \mathrm{Var}(\xi)/a^2$. □

Exercise 4.3.9. A vertex z is a "common neighbor" of vertices x and y in a graph G if both x and y are adjacent to z in G. Let $N(x, y)$ denote the number of common neighbors of x and y. Prove that the following statement is true for *almost all* graphs $G = (V, E)$ with n vertices:

$$(\forall x \neq y \in V)(0.24n < N(x, y) < 0.26n).$$

In other words, if p_n denotes the probability of the event described by the displayed formula then $\lim_{n\to\infty} p_n = 1$.

Exercise 4.3.10. In its more common form the Cauchy-Schwarz inequality asserts that for any real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$ we have

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \geq \left(\sum_{i=1}^{n} x_i y_i\right)^2. \tag{4.13}$$

Deduce this inequality from Corollary 4.3.7.

Exercise 4.3.11. (Limit on negatively correlated events.) Suppose the events $A_1, \ldots, A_m$ each have probability $1/2$ and for each i, j, $\Pr(|A_i \cap A_j| \leq 1/5$. Prove: $m \leq 6$. Generalize the statement to events of probability p, with $p^2 - \epsilon$ in the place of $1/5$.

Exercise 4.3.12. Prove: if the k^{th} moment of ξ is zero for all odd integers $k > 0$ then $\Pr(\xi = u) = \Pr(\xi = -u)$ for all $u \in \mathbf{R}$.

4.4 Independence of random variables

Definition 4.4.1. $\xi_1, \ldots, \xi_k$ are **independent** if $\forall u_1, \ldots, u_k$,

$$\Pr(\xi_1 = u_1, \ldots, \xi_k = u_k) = \prod_{i=1}^{k} \Pr(\xi_i = u_i). \tag{4.14}$$

Exercise 4.4.2. Prove that the events $A_1, \ldots, A_k$ are independent if and only if their indicator variables are independent.

Exercise 4.4.3. Prove that the random variables $\xi_1, \ldots, \xi_k$ are independent if and only if for all choices of the numbers $u_1, \ldots, u_k$, the k events $\xi_1 = u_1, \ldots, \xi_k = u_k$ are independent. Show that this is also equivalent to the independence of all k-tuples of events of the form $\xi_1 < u_1, \ldots, \xi_k < u_k$.

Exercise 4.4.4. Prove: if $\xi_1, \ldots, \xi_k$ are independent then $f_1(\xi_1), \ldots, f_k(\xi_k)$ are also independent, where the f_i are arbitrary functions. For example, ξ_1^2, e^{ξ_2}, and $\cos(\xi_3)$ are independent.

Exercise 4.4.5. Prove: if ξ, η, ζ are independent random variables then $f(\xi, \eta)$ and ζ are also independent, where f is an arbitrary function. (For instance, $\xi + \eta$ and ζ, or $\xi\eta$ and ζ are independent.) Generalize this statement to several variables, grouped into blocks, and a function applied to each block.

Exercise 4.4.6. Let $\xi_1, \ldots, \xi_m$ be non-constant random variables over a sample space of size n. Suppose the ξ_i are 4-wise independent (every four of them are independent). Prove: $n \geq \binom{m}{2}$. *Hint.* Prove that the $\binom{m}{2}$ random variables $\xi_i \xi_j$ $(1 \leq i < j \leq m)$ are linearly independent over $\mathbb{R}$ (as members of the space of functions $\Omega \to \mathbb{R}$). To prove linear independence, first prove that w.l.o.g. we may assume $(\forall i)(E(\xi_i) = 0)$; then use the "inner product" argument, using the function $E(\zeta\eta)$ in the role of an "inner product" of the random variables ζ and η.

Theorem 4.4.7 (Multiplicativity of the expected value). *If $\xi_1, \ldots, \xi_m$ are independent, then*

$$\mathrm{E}(\prod_{i=1}^{m} \xi_i) = \prod_{i=1}^{m} \mathrm{E}(\xi_i). \tag{4.15}$$

Exercise 4.4.8. Prove this result for indicator variables.

Exercise 4.4.9. Prove: if ξ, η are independent, then one can write ξ as a sum $\xi = c_1\xi_1 + \ldots + c_k\xi_k$ and η as $\eta = d_1\eta_1 + \ldots + d_\ell\eta_\ell$ where the ξ_i and η_j are indicator variables and for every i, j, the variables ξ_i and η_j are independent.

Exercise 4.4.10. Combine the two preceding exercises to a proof of the Theorem for $m = 2$ variables.

Exercise 4.4.11. Deduce the general case from the preceding exercise by induction on m, using Exercise 4.4.5.

This sequence completes the proof of Theorem 4.4.7. □

While this result required the full force of independence of our random variables, in the next result, only pairwise independence is required.

Theorem 4.4.12 (Additivity of the variance). *Let $\eta = \xi_1 + \xi_2 + \cdots + \xi_k$. If $\xi_1, \ldots, \xi_k$ are* pairwise independent *then* $\mathrm{Var}(\eta) = \sum_{i=1}^{k} \mathrm{Var}(\xi_i)$.

Proof: By Exercise 4.3.4, we may assume that $\mathrm{E}(\xi_i) = 0$ (otherwise we replace each ξ_i by $\xi_i - \mathrm{E}(\xi_i)$; this will not change the variance, nor does it affect independence (why?)). Having made this assumption it follows that $\mathrm{E}(\eta) = 0$. Moreover, for $i \neq j$ we have $\mathrm{E}(\xi_i\xi_j) = \mathrm{E}(\xi_i)\mathrm{E}(\xi_j) = 0$ by pairwise independence.

It follows that $\mathrm{Var}(\xi_i) = \mathrm{E}(\xi_i^2)$ and $\mathrm{Var}(\eta) = \mathrm{E}(\eta^2) = \mathrm{E}((\sum \xi_i)^2) = \mathrm{E}(\sum_i \xi_i^2 + 2\sum_{i<j} \xi_i\xi_j) = \sum_i \mathrm{E}(\xi_i^2) + 2\sum_{i<j} \mathrm{E}(\xi_i\xi_j) = \sum_i \mathrm{Var}(\xi_i)$. □

Corollary 4.4.13. *Let $\xi_1, \ldots, \xi_n$ be random variables with the same standard deviation σ. Let us consider their average, $\eta := (1/n)\sum_{i=1}^{n} \xi_i$. If the ξ_i are pairwise independent then $\sigma(\eta) = \sigma/\sqrt{n}$.* □

Corollary 4.4.14 (Weak law of large numbers). *Let $\xi_1, \xi_2, \ldots$ be an infinite sequence of pairwise independent random variables each with expected value m and standard deviation σ. Let $\eta_n = (1/n)\sum_{i=1}^{n} \xi_i$. Then for any $\delta > 0$,*

$$\lim_{n\to\infty} \Pr(|\eta_n - m| > \delta) = 0. \tag{4.16}$$

Proof: Use Chebyshev's inequality and the preceding corollary. We obtain that the probability in question is $\leq \sigma^2/(\delta n) \to 0$ (as $n \to \infty$). □

Remark 4.4.15. Strictly speaking, we bent our rules here. An infinite sequence of non-constant, pairwise independent variables requires an infinite sample space. What we actually proved, then, is the following. Let us fix the values m and $\sigma \geq 0$. Assume that we are given an infinite sequence of finite probability spaces, and over the n^{th} space, we are given n independent random variables $\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,n}$. Let $\eta_n = (1/n)\sum_{i=1}^{n} \xi_{n,i}$. Then for any $\delta > 0$, the limit relation (4.16) holds.

Exercise 4.4.16. You and the bank play the following game: you flip n coins: if ξ of them come up "Heads," you receive 2^ξ dollars.

1. You have to buy a ticket to play this game. What is the fair price of the ticket? *Hint:* it is the expected amount you will receive.

2. Prove: the probability that you break even (receive at least your ticket's worth) is exponentially small. *Hint:* At least how many "heads" do you need for you to break even?

3. Calculate the standard deviation of the variable 2^ξ. Your answer should be a simple formula. Evaluate it asymptotically; obtain an even simpler formula.

4. State what the "weak law of large numbers" would say for the variable 2^ξ. *Hint.* This law talks about the probability that 2^ξ is not within $(1 \pm \epsilon)$-times its expectation.) Prove that the Law does NOT hold for this variable.

4.5 Chernoff's Bound

Although the bound in the proof of the Weak Law of Large Numbers tends to zero, it does so rather slowly. If our variables are fully independent and bounded, much stronger estimates can be obtained by a method due to Chernoff. The bounds will go to zero exponentially as a function of n, and this is what most combinatorial applications require.

For example, let us consider a sequence of n independent coin flips; let ψ denote the number of heads in this sequence. Then $\mathrm{E}(\psi) = n/2$ and $\mathrm{Var}(\xi) = n/4$ (by the additivity of the variance). Therefore Chebyshev's inequality tells us that

$$\Pr(|\psi - n/2| \geq r\sqrt{n}) < \frac{1}{4r^2}. \tag{4.17}$$

Below we shall prove the much stronger inequality

$$\Pr(|\psi - n/2| \geq r\sqrt{n}) < 2e^{-2r^2}. \tag{4.18}$$

under the same conditions.

The following corollary illustrates the power of inequality (4.18).

Corollary 4.5.1. *For any $\varepsilon > 0$, almost all graphs have no vertices of degree $< (1-\varepsilon)n/2$ or $> (1+\varepsilon)n/2$ where n is the number of vertices.*

Proof of the Corollary. Let $V = \{1, \ldots, n\}$ be the vertex set of our random graph. Let δ_i denote the degree of vertex i; so δ_i is the number of heads in a sequence of $(n-1)$ independent coin flips. Therefore, by inequality (4.18), we have that

$$\Pr(|\delta_i - (n-1)/2| \geq r\sqrt{n-1}) < 2e^{-2r^2}. \tag{4.19}$$

Let us now set $r = \varepsilon\sqrt{n-1}$. Then we obtain

$$\Pr(|\delta_i - (n-1)/2| \geq \varepsilon(n-1)) < 2e^{-2\varepsilon^2(n-1)}. \tag{4.20}$$

Therefore the probability that there exists an i such that $|\delta_i - (n-1)/2| \geq \varepsilon(n-1)$ is less than n times the right hand side, i. e., less than $2ne^{-2\varepsilon^2(n-1)}$. This quantity approaches zero at an exponential rate as $n \to \infty$.

The slight change in the statement (having changed n to $n-1$) can be compensated for by slightly reducing ε. □

Note that the same procedure using inequality (4.17) will fail. Indeed, setting $r = \varepsilon\sqrt{n-1}$ in inequality (4.17), the right hand side will be $1/(4\varepsilon^2(n-1))$, and if we multiply this quantity by n, the result will be greater than 1 (if $\varepsilon < 1/2$, a meaningless upper bound for a probability.

Now we turn to the proof of inequality (4.18). It will be convenient to state the main result in terms of random variables with zero expected value.

Theorem 4.5.2 (Chernoff). *Let ξ_i be independent random variables satisfying $\Pr(\xi_i = 1) = \Pr(\xi_i = -1) = 1/2$. Let $\eta = \sum_{i=1}^n \xi_i$. Then for any $a > 0$,*

$$\Pr(\eta \geq a) < e^{-a^2/2n} \tag{4.21}$$

and

$$\Pr(|\eta| \geq a) < 2e^{-a^2/2n}. \tag{4.22}$$

Exercise 4.5.3. Deduce inequality (4.18) from this theorem.

Hint. Represent ψ as $\sum_{i=1}^n \theta_i$ where θ_i is the indicator variable of the i-th coin flip. Set $\xi_i = 2\theta_i - 1$ and $\eta = \sum_{i=1}^n \xi_i$. Note that $\psi - n/2 = \eta/2$. Apply Theorem 4.5.2 to the ξ_i and translate the result back to ψ.

Exercise 4.5.4. Prove that the following is true for almost all graphs $\mathcal{G}_n$ on n vertices: the degree of every vertex is within the interval $[0.49n, 0.51n]$. In answering this question, be sure to clearly state the meaning of each variable occurring in your formulas. Also pay close attention to the logical connectives ("and," "if-then," and quantifiers).

Now we turn to the proof of Theorem 4.5.2.

Let t be a positive real number. We shall later suitably choose the value of t. Let us consider the random variables $\zeta_i := \exp(t\xi_i)$. (Notation: $\exp(x) = e^x$.) The ζ_i are again independent (for any fixed t) by Exercise 4.4.4. Therefore we can apply the multiplicativity of the expected value to them:

$$\mathrm{E}(e^{t\eta}) = \mathrm{E}(\exp(\sum_{i=1}^{n} t\xi_i)) = \mathrm{E}(\prod_{i=1}^{n} \zeta_i) == \prod_{i=1}^{n} \mathrm{E}(\zeta_i) = \prod_{i=1}^{n} \mathrm{E}(\exp(t\xi_i)). \tag{4.23}$$

Applying Markov's inequality to the variable $e^{t\eta}$, we conclude that

$$\Pr(\eta \geq a) = \Pr(e^{t\eta} \geq e^{ta}) \leq \prod_{i=1}^{n} \mathrm{E}(\exp(t\xi_i))e^{-ta}. \tag{4.24}$$

Recall that $\cosh(x) = (e^x + e^{-x})/2$ and observe that

$$\mathrm{E}(\exp(t\xi_i)) = \cosh(t). \tag{4.25}$$

Therefore the preceding inequality implies that

$$\Pr(\eta \geq a) < \frac{\cosh(t)^n}{e^{ta}}. \tag{4.26}$$

This is true for every $t > 0$. All we need to do is choose t appropriately to obtain the strongest possible result. To this end we need the following simple observation.

Lemma 4.5.5. *For all real numbers x,*

$$\cosh(x) \leq e^{x^2/2}.$$

Proof: Compare the Taylor series of the two sides. On the left hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + \ldots \tag{4.27}$$

On the right hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \ldots. \tag{4.28}$$

□

Consequently, from inequality (4.26) we infer that

$$\Pr(\eta \geq a) < \exp(t^2 n/2 - ta). \tag{4.29}$$

The expression $t^2n/2 - ta$ is minimized when $t = a/n$; setting $t := a/n$ we conclude that $\Pr(\eta \geq a) < \exp(-a^2/2n)$, as required.

Replacing each ξ_i by $-\xi_i$ we obtain the inequality $\Pr(\eta \leq -a) < \exp(-a^2/2n)$; adding this to the preceding inequality we obtain $\Pr(|\eta| \leq a) < 2\exp(-a^2/2n)$. □

We note that Chernoff's technique works under much more general circumstances. We state a useful and rather general case, noting that even this result does not exploit the full power of the method.

Theorem 4.5.6 (Chernoff). *Let ξ_i be independent random variables satisfying $|\xi_i| \leq 1$ and $\mathrm{E}(\xi_i) = 0$. Let $\eta = \sum_{i=1}^{n} \xi_i$. Then for any $a > 0$,*

$$\Pr(\eta \geq a) < e^{-a^2/2n} \tag{4.30}$$

and

$$\Pr(|\eta| \geq a) < 2e^{-a^2/2n}. \tag{4.31}$$

Proof: As before, we set $t = a/n$. Let

$$h(x) = \cosh(t) + x \cdot \sinh(t). \tag{4.32}$$

(Recall that $\sinh(t) = (e^t - e^{-t})/2$.) Observe that $h(x) \geq e^{tx}$ for all x in the interval $-1 \leq x \leq 1$. (The graph of $h(x)$ over the interval $[-1, 1]$ is the segment connecting the corresponding two points of the graph of the function e^{tx}, and e^{tx} is a convex function.)

Moreover, because of the linearity of the $h(x)$ function, we have $\mathrm{E}(h(\xi_i)) = h(E(\xi_i)) = h(0) = \cosh(t)$. Therefore

$$\mathrm{E}(e^{t\xi_i}) \leq \mathrm{E}(h(\xi_i)) = \cosh(t). \tag{4.33}$$

From here on the proof is identical with the proof of Theorem 4.5.2. □

4.6 Problems

Exercise 4.6.1. (Bipartite Ramsey) (Erdős) Let $n = 2^{t/2}$, where t is an even integer. Prove that it is possible to color the edges of $K_{n,n}$ red and blue (each edge receives one color) such that there will be no monochromatic $K_{t,t}$. *Hint.* Use the probabilistic method.

A *random graph on n vertices* is defined by fixing a set of n vertices, say $V = [n]$, and flipping a fair coin $\binom{n}{2}$ times to decide adjacency of the $\binom{n}{2}$ pairs of vertices. Let $\mathcal{G}_n$ denote a random graph on the vertex set $[n]$.

Exercise 4.6.2. (Diameter of a random graph)

(a) State the size of the sample space of the experiment which produces a random graph.

(b) What is the probability $\mathrm{diam}(\mathcal{G}_n) = 1$? Your answer should be a very simple closed-form expression. ($\mathrm{diam}(G)$ denotes the diameter of G. See the handout for the definition.)

(c) Prove that almost all graphs have diameter 2.

The meaning of this statement is the following. Let p_n denote the probability that a random graph on n vertices has diameter 2. Then $\lim_{n\to\infty} p_n = 1$.

Hint. Let $q_n = 1 - p_n$. Prove that $q_n \to 0$. Show this by proving that with large probability, every pair of vertices has a common neighbor. What is the probability that vertices x and y do not have a common neighbor? Give a precise answer to this question; it should be a simple formula. Now *estimate* the probability that there exist vertices x, y without a common neighbor.

Use without proof the following fact from calculus:

$$(\forall c, d > 0)(\lim_{x\to\infty} x^c \mathrm{e}^{-dx} = 0).$$

Exercise 4.6.3. (Chromatic number of a random graph) (Erdős) Recall from the graph theory handout that $\omega(G)$ denotes the size of the largest clique (complete subgraph) in the graph G; $\alpha(G)$ denotes the size of the largest independent set (anticlique) in G, and $\chi(G)$ denotes the chromatic number of G. Note that $\alpha(G) = \omega(\overline{G})$ where $\overline{G}$ denotes the complement of G. Note also (do!) that for every graph G, $\chi(G) \geq \omega(G)$.

1. prove: $\chi(G) \geq n/\alpha(G)$, where n is the number of vertices of G.
2. Show that the chromatic number can be *much* greater than the clique number by proving that there exists a constant $c > 0$ such that for all sufficiently large n there exists a graph G_n with n vertices such that
$$\frac{\chi(G_n)}{\omega(G_n)} \geq \frac{cn}{(\log n)^2}.$$

Estimate the value of c in your proof.

Hint. To prove the existence of these graphs, use the probabilistic method. To obtain a lower bound on $\chi(G_n)$, give an upper bound on $\alpha(G_n)$ for almost all graphs G_n.

3. Prove: for almost all graphs, $\chi(G) = \Theta(n/\log n)$. (The lower bound is easy; the upper bound is more challenging!)

Exercise 4.6.4. (Chromatic number of set systems) (Erdős) Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be an r-uniform set-system ($|A_i| = r$) over the universe $[n]$ (so $A_i \subset [n]$). Assume $m \leq 2^{r-1}$. Prove that $\mathcal{F}$ is 2-colorable, i. e., it is possible to color every vertex $v \in [n]$ red or blue such that none of the A_i is monochromatic (each A_i has both colors). *Hint.* Assign the colors at random. Compute the expected number of monochromatic sets A_i.

Exercise 4.6.5. (Error-correcting codes) Let X be a set of n elements. Let $\mathcal{B}(X)$ be the set of all subsets of X; we view $\mathcal{B}(X)$ as a uniform probability space. A "random subset of X" is an element of $\mathcal{B}(X)$ chosen from the uniform distribution.

(a) Prove: $E(|A \setminus B|) = n/4$, where A, B are two independent random subsets of X. What is the size of the sample space for this experiment?

(b) (Constant-rate, cn-error-correcting codes) Prove that there exists a constant $C > 1$ and there exists a family $\{A_1, \ldots, A_m\}$ of $m \geq C^n$ subsets of X such that $(\forall i, j)(i \neq j \Rightarrow |A_i \setminus A_j| \geq 0.24n)$. *Hint.* Take m random subsets, chosen independently. Use Chernoff's inequality to prove that $|A_i \setminus A_j| < 0.24n$ is exponentially unlikely. *Explanation of the title.* Suppose we want to send messages ($(0,1)$-strings) of length k through a noisy channel. Let $n = k/\log C$, so $2^k = C^n = m$ and we can think of the messages as integers from 1 to m. Rather than sending message i, we transmit the incidence vector of the set A_i. This increases the length of the message by a constant factor $(1/\log C)$. On the other hand, even if 23% of the transmitted bits get changed due to noise, the error can uniquely be corrected because the difference (Hamming distance) between any two valid codewords is at least $0.48n$. – Here we only prove the existence of such codes. Constructive versions exist (Justesen codes).

Exercise 4.6.6. (Strongly negatively correlated events) Let $A_1, \ldots, A_m$ be events with probability 1/2; suppose $(\forall i, j)(i \neq j \Rightarrow P(A_i \cap A_j) \leq 1/5)$. Prove: $m \leq 6$. *Hint.* Use the Cauchy– Schwarz inequality, Corollary 4.3.7.

Chapter 5

Finite Markov Chains

Exercises. The unmarked exercises are routine, the exercises marked with a "plus" (+) are creative, those marked with an asterisk (*) are challenging.

Recall that a **directed graph** (digraph, for short), is a pair $G = (V, E)$, where V is the set of "vertices" and E is a set of ordered pairs of vertices called "edges:" $E \subseteq V \times V$.

A *discrete system* is characterized by a set V of "states" and *transitions* between the states. V is referred to as the **state space**. We think of the transitions as occurring at each time beat, so the state of the system at time t is a value $X_t \in V$ ($t = 0, 1, 2, \dots$). The adjective "discrete" refers to discrete time beats.

A *discrete stochastic process* is a discrete system in which transitions occur randomly according to some probability distribution. The process is *memoryless* if the probability of an $i \to j$ transition does not depend on the history of the process (the sequence of previous states): $(\forall i, j, u_0, \dots, u_{t-1} \in V)(P(X_{t+1} = j \,|\, X_t = i, X_{t-1} = u_{t-1}, \dots, X_0 = u_0) = P(X_{t+1} = j \,|\, X_t = i))$. (Here the universal quantifier is limited to feasible sequences of states $u_0, u_1, \dots, u_{t-1}, i$, i. e., to sequences which occur with positive probability; otherwise the conditional probability stated would be undefined.) If in addition the transtion probability $p_{ij} = P(X_{t+1} = j \,|\, X_t = i\}$ does not depend on the time t, we call the process *homogeneous*.

A **finite Markov chain** is a memoryless homogeneous discrete stochastic process with a finite number of states.

Let $\mathcal{M}$ be a finite Markov chain with n states, $V = [n] = \{1, 2, \dots, n\}$. Let p_{ij} denote the probability of transition from state i to state j, i. e., $p_{ij} = P(X_{t+1} = j \,|\, X_t = i)$. (Note that this is a conditional probability: the question of $i \to j$ transition only arises if the system is in state i, i. e., $X_t = i$.)

The finite Markov chain $\mathcal{M}$ is characterized by the $n \times n$ **transition matrix** $T = (p_{ij})$ $(i, j \in [n])$ and an **initial distribution** $q = (q_1, \dots, q_n)$ where $q_i = P(X_0 = i)$.

Definition. An $n \times n$ matrix $T = (p_{ij})$ is **stochastic** if its entries are nonnegative real numbers and the sum of each row is 1:

$$(\forall i, j)(p_{ij} \geq 0) \text{ and } (\forall i)(\textstyle\sum_{j=1}^{n} p_{ij} = 1).$$

Exercise 5.1.1. The transition matrix of a finite Markov chain is a stochastic matrix. Conversely, every stochastic matrix can be viewed as the transition matrix of a finite Markov chain.

Exercise 5.1.2. Prove: if T is a stochastic matrix then T^k is a stochastic matrix for every k.

Random walks on digraphs are important examples of finite Markov chains. They are defined by hopping from vertex to neighboring vertex, giving equal chance to each out-neighbor. The state space will be V, the set of vertices. The formal definition follows.

Let $G = (V, E)$ be a finite digraph; let $V = [n]$. Assume $(\forall i \in V)(\deg^+(i) \geq 1)$. Set $p_{ij} = 1/\deg^+(i)$ if $(i, j) \in E$; $p_{ij} = 0$ otherwise.

Exercise 5.1.3. Prove that the matrix (p_{ij}) defined in the preceding paragraph is stochastic.

Conversely, all finite Markov chains can be viewed as *weighted* random walks on a digraph, the weights being the transition probabilities. The formal definition follows.

Let $T = (p_{ij})$ be an arbitrary (not necessarily stochastic) $n \times n$ matrix. We associate with T a digraph $G = (V, E)$ as follows. Let $V = [n]$ and $E = \{(i, j) : p_{ij} \neq 0\}$. We label the edge $i \to j$ with the number $p_{ij} \neq 0$ (the "weight" of the edge).

This definition makes sense for any matrix T; edges indicate nonzero entries. If T is the transition matrix of a finite Markov chain $\mathcal{M}$ then we call the associated digraph the **transition digraph** of $\mathcal{M}$. The **vertices** of the transition digraph represent the **states** of $\mathcal{M}$ and the **edges** the **feasible transitions** (transitions that occur with positive probability).

Exercise 5.1.4. Prove that in the transition digraph of a finite Markov chain, $(\forall i)(\deg^+(i) \geq 1)$.

Exercise 5.1.5. Draw the transition digraph corresponding to the stochastic matrix

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

Label the edges with the transition probabilities.

The principal subject of study in the theory of Markov chains is the **evolution** of the system.

The *initial distribution* $q = (q_1, \ldots, q_n)$ describes the probability that the system is in a particular state at time $t = 0$. So $q_i \geq 0$ and $\sum_{i=1}^{n} q_i = 1$.

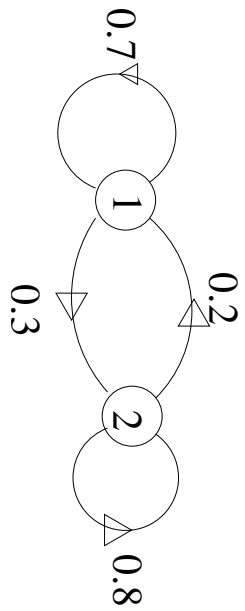


Figure 5.1: The solution to Exercise 5.1.5

Set $q(0) = q$ and let $q(t) = (q_{1t}, \ldots, q_{nt})$ be the distribution of the states at time t, i. e., the distribution of the random variable X_t:

$$q_{it} = P(X_t = i).$$

The following simple equation describes the evolution of a finite Markov chain.

Exercise 5.1.6. (Evolution of Markov chains) Prove: $q(t) = q(0)T^t$.

So the study of the *evolution of a finite Markov chain* amounts to studying the *powers of the transition matrix.*

Exercise 5.1.7. Experiment: study the powers of the matrix A defined in Exercise 5.1.5. Observe that the sequence $I, A, A^2, A^3, \ldots$ appears to converge. What is the limit?

Exercise$^+$ 5.1.8. Prove the convergence observed in the preceding exercise.

The study of the powers rests on the study of *eigenvalues* and *eigenvectors.*

Definition. A **left eigenvector** of an $n \times n$ matrix A is a $1 \times n$ vector $x \neq 0$ such that $xA = \lambda x$ for some (complex) number λ called the *eigenvalue* corresponding to x. A **right eigenvector** of A is an $n \times 1$ matrix $y \neq 0$ such that $Ay = \mu y$ for some (complex) number μ called the *eigenvalue* corresponding to y.

Remember that the zero vector is never an eigenvector.

The right action of a matrix. Note that if $x = (x_1, \ldots, x_n)$ is a $1 \times n$ vector, $A = (a_{ij})$ is an $n \times n$ matrix, and $z = (z_1, \ldots, z_n) = xA$ then

$$z_j = \sum_{i=1}^{n} x_i a_{ij}. \tag{5.1}$$

Note that if G is the digraph associated with the matrix A then the summation can be reduced to

$$z_j = \sum_{i:i\to j}^{n} x_i a_{ij}. \tag{5.2}$$

So the **left eigenvectors** to the eigenvalue λ is defined by the equation

$$\lambda x_j = \sum_{i:i\to j}^{n} x_i a_{ij}. \tag{5.3}$$

Exercise 5.1.9. State the equations for the left action and the right eigenvectors of the matrix A.

Theorem. The left and the right eigenvalues of a matrix are the same (but not the eigenvectors!).

Proof. Both the right and the left eigenvalues are the roots of the **characteristic polynomial** $f_A(x) = \det(xI - A)$ where I is the $n \times n$ identity matrix.

Exercise 5.1.10. Find the eigenvalues and the corresponding left and right eigenvectors of the matrix A from Exercise 5.1.5.

Hint. The characteristic polynomial is

$$f_A(x) = \begin{vmatrix} x - 0.7 & -0.3 \\ -0.2 & x - 0.8 \end{vmatrix} = x^2 - 1.5x + 0.5 = (x - 1)(x - 1/2).$$

So the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1/2$. Each eigenvalue gives rise to a system of linear equations for the coordinates of the corresponding (left/right) eigenvectors.

Exercise$^+$ 5.1.11. Prove: if λ is a (complex) eigenvalue of a stochastic matrix then $|\lambda| \leq 1$. *Hint.* Consider a right eigenvector to eigenvalue λ.

Exercise 5.1.12. Let A be an $n \times n$ matrix. Prove: if x is a left eigenvector to eigenvalue λ and y is a right eigenvector to eigenvalue μ and $\lambda \neq \mu$ then x and y are **orthogonal,** i. e., $xy = 0$. *Hint.* Consider the product xAy.

Definition. A **stationary distribution** (also called **equilibrium distribution**) for the Markov chain is a probability distribution $q = (q_1, \ldots, q_n)$ ($q_i \geq 0$, $\sum_{i=1}^{n} q_i = 1$) which is a left eigenvector to the eigenvalue 1: $qA = q$.

Exercise 5.1.13. If at time t, the distribution $q(t)$ is stationary then it will remain the same forever: $q(t) = q(t+1) = q(t+2) = \ldots$.

Exercise 5.1.14. Prove: if T is a stochastic matrix then $\lambda = 1$ is a right eigenvalue. *Hint.* Guess the (very simple) eigenvector.

Observe the consequence that $\lambda = 1$ is also a *left* eigenvalue. This is significant because it raises the possibility of having stationary distributions.

Exercise 5.1.15. Find a *left* eigenvector $x = (x_1, x_2)$ to the eigenvalue 1 for the stochastic matrix A defined in Exercise 5.1.5. Normalize your eigenvector such that $|x_1| + |x_2| = 1$. Observe that x is a stationary distribution for A.

Exercise 5.1.16. Let T be a stochastic matrix. Prove: **if** the limit $T^\infty = \lim_{t\to\infty} T^t$ **exists** then every row of T^∞ is a stationary distribution.

Exercise 5.1.17. Consider the stochastic matrix

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Prove that the sequence $I, B, B^2, B^3, \ldots$ does **not** converge, yet B does have a stationary distribution.

Exercise 5.1.18. Let $\vec{C}_n$ denote the directed cycle of length n. Prove that the powers of the transition matrix of the random walk on $\vec{C}_n$ do not converge; but a stationary distribution exists.

Exercise 5.1.19. Consider the following digraph: $V = [3]$, $E = \{1 \to 2, 1 \to 3, 2 \to 2, 3 \to 3\}$.

Write down the transition matrix of the random walk on the graph shown in Figure 5. Prove that the random walk on this graph has 2 stationary distributions.

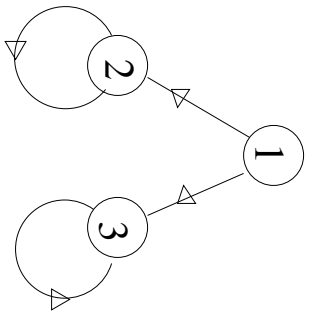


Figure 5.2: A graph with transition probabilities. FIX THIS!

Definition. A stochastic matrix $T = (p_{ij})$ is called **"doubly stochastic"** if its column sums are equal to 1: $(\forall j \in [n])(\sum_{i=1}^{n} p_{ij} = 1)$.

In other words, T is doubly stochastic if both T and its transpose are stochastic.

Exercise 5.1.20. Let T be the transition matrix for a finite Markov chain M. Prove that the uniform distribution is stationary if and only if T is doubly stochastic.

A matrix is called **non-negative** if all entries of the matrix are non-negative. The *Perron–Frobenius theory of non-negative matrices* provides the following fundamental result.

Theorem (Perron–Frobenius, abridged) If A is a non-negative $n \times n$ matrix then A has a non-negative left eigenvector.

Exercise 5.1.21. Prove that a non-negative matrix has a non-negative right eigenvector. (Use the Perron–Frobenius Theorem.)

Exercise 5.1.22. Let T be a stochastic matrix and x a non-negative left eigenvector to eigenvalue λ. Prove: $\lambda = 1$. *Hint.* Use Exercise 5.1.12.

Exercise 5.1.23. Prove: **every finite Markov chain has a stationary distribution.**

Exercise$^+$ 5.1.24. Let A be a non-negative matrix, x a non-negative left eigenvector of A, and G the digraph associated with A. Prove: if G is **strongly connected** then all entries of x are **positive.** *Hint.* Use equation (5.3).

Exercise 5.1.25. Let A be a non-negative matrix, x and x' two non-negative eigenvectors of A, and G the digraph associated with A. Prove: if G is **strongly connected** then x and x' belong to the same eigenvalue. *Hint.* Use the preceding exercise and Exercise 5.1.12.

Exercise$^+$ 5.1.26. Let A be a non-negative matrix; let x be a non-negative left eigenvector to the eigenvalue λ and let x' be another left eigenvector with real coordinates to the same eigenvalue. Prove: if G is **strongly connected** then $(\exists \alpha \in \mathbb{R})(x' = \alpha x)$. *Hint.* WLOG (without loss of generality we may assume that) all entries of x are positive (why?). Moreover, WLOG $(\forall i \in V)(x'_i \leq x_i)$ and $(\exists j \in V)(x'_j = x_j)$ (why?). Now prove: if $x_j = x'_j$ and $i \to j$ then $x_i = x'_i$. Use equation (5.3).

Finite Markov chains with a **strongly connected** transition digraph (every state is accessible from every state) are of particular importance. Such Markov chains are called **irreducible.** To emphasize the underlying graph theoretic concept (and reduce the terminology overload), we shall deviate from the accepted usage and use the term **strongly connected Markov chains** instead of the classical and commonly used term "irreducible Markov chains."

Our results are summed up in the following exercise, an immediate consequence of the preceding three exercises.

Exercise 5.1.27. Prove: **A strongly connected finite Markov chain (a) has exactly one stationary distribution; and (b) all probabilities in the stationary distribution are positive.**

As we have seen (which exercise?), strong connectivity is not sufficient for the powers of the transition matrix to converge. One more condition is needed.

Definition. The **period** of a vertex v in the digraph G is the g.c.d. of the lengths of all closed directed walks in G passing through v. If G has no closed directed walks through v, the period of v is said to be 0. If the period of v is 1 then v is said to be **aperiodic.**

Exercise 5.1.28. (a) Show that it is not possible for every state of a finite Markov chain to have period 0 (in the transition digraph). (b) Construct a Markov chain with n states, such that all but one state has period 0.

Note that a **loop** is a closed walk of length 1, so if G has a loop at v then v is automatically aperiodic. A **lazy random walk** on a digraph stops at each vertex with probability 1/2 and divides the remianing 1/2 evenly between the out-neighbors ($p_{ii} = 1/2$, and if $i \to j$ then $p_{ij} = 1/2 \deg +(i)$). So the lazy random walks are aperiodic at each vertex.

Exercise 5.1.29. Let $G = (V, E)$ be a digraph and $x, y \in V$ two vertices of G. Prove: if x and y belong to the same strong component of G (i. e., x and y are mutually accessible from one another) then the periods of x and y are equal.

It follows that **all states of a strongly connected finite Markov chain have the same period.** We call this common value the **period** of the strongly connected Markov chain. A Markov chain is **aperiodic** if every node has period 1.

Exercise 5.1.30. Recall that (undirected) graphs can be viewed as digraphs with each pair of adjacent vertices being connected in both directions. Let G be an undirected graph viewed as a digraph. Prove: every vertex of G has period 1 or 2. The period of a vertex v is 2 if and only the connected component of G containing v is bipartite.

Exercise 5.1.31. Suppose a finite Markov chain $\mathcal{M}$ is strongly connected and NOT aperiodic. (It follows that the period ≥ 2 (why?).)
Prove: the powers of the transition matrix do not converge.
Hint. If the period is d, prove that the transition graph is a "blown-up directed cycle of length d" in the following sense: the vertices of the transition graph can be divided into d disjoint subsets $V_0, V_1, \ldots, V_{d-1}$ such that $(\forall k)$ all edges starting at V_k end in V_{k+1}, where the subscript is read modulo d (wraps around). – Once you have this structure, observe that any t-step transition would take a state in V_k to a state in V_{k+t} (the subscript again modulo d).

Now we state the Perron–Frobenius Theorem in full.

Theorem (Perron–Frobenius, unabridged) Let A be a non-negative $n \times n$ matrix and G the associated digraph. Let $f_A(x) = \prod_{i=1}^{n}(x - \lambda_i)$ be the characteristic polynomial of A factored over the complex numbers. (So the λ_i are the eigenvalues, listed with multiplicity.) Then

(a) There is an eigenvalue λ_1 such that

  (a1) λ_1 is real and non-negative;

  (a2) $(\forall i)(\lambda_1 \geq |\lambda_i|)$;

  (a3) there exists a non-negative eigenvector to eigenvalue λ_1.

(b) If G is strongly connected and **aperiodic** then $(\forall i)(\lambda_1 > |\lambda_i|)$.

Definition. A **strongly connected aperiodic Markov chain** is called **ergodic.**

The significance of aperiodicity is illuminated by the following exercises.

Exercise 5.1.32. Prove that the eigenvalues of the random walk on the directed n-cycle are exactly the n-th roots of unity. (So all of them have unit absolute value.)

More generally, we have the following:

Exercise 5.1.33. Let A be a (not necessarily non-negative) $n \times n$ matrix and G the associated digraph. Suppose d is a common divisor of the periods of G. Let ω be a complex d-th root of unity (i. e., $\omega^d = 1$). Then, if λ is an eigenvalue of A then $\lambda\omega$ is also an eigenvalue of A. *Hint.* Equation (5.3).

The following consequence of the Perron–Frobenius Theorem is the fundamental result in the theory of finite Markov chains.

Exercise* 5.1.34. (Convergence of ergodic Markov chains.) Prove: if T is the transition matrix of an **ergodic Markov chain** then the powers of T **converge.** *Hint.* There exists an invertible complex matrix S such that $U = S^{-1}TS$ is an upper triangular matrix of which the first row is $[1, 0, 0, \ldots, 0]$. (This follows, for example, from the Jordan normal form.) Now the diagonal entries of U are the eigenvalues, starting with $\lambda_1 = 1$; all other eigenvalues satisfy $|\lambda_i| < 1$. Prove that as a consequence, the sequence U^t $(t \to \infty)$ converges to the matrix N which has a 1 in the top left corner and 0 everywhere else. Now $T^k \to M := SNS^{-1}$ (why?).

Exercise 5.1.35. Prove: if T is the transition matrix of an ergodic Markov chain and $\lim_{t\to\infty} T^t = M$ then all rows of M are equal.

Exercise 5.1.36. Prove: if a finite Markov chain is ergodic then from any initial distributrion, the process will approach the unique stationary distribution. In other words, let T be the transition matrix, s the stationary distribution, and q an arbitrary initial distribution. Then

$$\lim_{t\to\infty} qT^t = s.$$

The following example illuminates the kind of Markov chains encountered in combinatorics, theoretical computer science, and statistical physics.

Random recoloring: a class of large Markov chains. Let $G = (V, E)$ be a graph with n vertices and maximum degree Δ; and let $Q \geq \Delta + 1$. Let S be the set of all legal colorings of G with Q colors, i. e., S is the set of functions $f : V \to [Q]$ such that if $v, w \in V$ are adjacent then $f(v) \neq f(w)$. This "random recoloring process" is a Markov chain which takes S as its set of states (the "state space"). The transitions from a legal coloring are defined as follows. We pick a vertex $v \in V$ at random, and recolor it by one of the available colors (colors not used by the neighbors of v), giving each available color an equal chance (including the current color of v).

Exercise 5.1.37. Prove: if $Q \geq \Delta + 2$ then the random recoloring process is an ergodic Markov chain.

Exercise 5.1.38. Prove that the number of states of the random recoloring process is between $(Q - \Delta - 1)^n$ and Q^n. So if $Q \geq \Delta + 2$ then the state space is exponentially large.

Exercise 5.1.39. Prove: if $Q \geq \Delta + 2$ then the stationary distribution for the random recoloring process is uniform.

As a consequence, the random recoloring process will converge to a uniformly distributed random legal Q-coloring of G. Just how quickly the process approaches the uniform distribution is an open problem. While the state space is exponential, it is expected that the process distribution will be close to uniform within a polynomial (n^{const}) number of steps. This phenomenon is called **rapid mixing**. Marc Jerrum proved in 1995 that for $Q > 2\Delta$, the random recoloring process does indeed mix rapidly; Jerrum proved an $O(n \log n)$ bound on the mixing time. In a recent (2000) paper, published in the *Journal of Mathematical Physics*, Eric Vigoda showed that the 2Δ bound was not best possible; he proved that rapid mixing already occurs for $Q > (11/6)\Delta$; under this weaker condition Vigoda shows a somewhat less rapid, $O(n^2 \log n)$ mixing. The techniques leading to such improvements are expected to be widely applicable in combinatorics, theoretical computer science, and statistical physics.

Concluding remarks. Markov chains are widely used models in a variety of areas of theoretical and applied mathematics and science, including statistics, operations research, industrial engineering, linguistics, artificial intelligence, demographics, genomics. Markov chain models are used in performance evaluation for computer systems ("if the system goes down, what is the chance it will come back?"), in queuing theory (server queuing, intelligent transportation systems). Hidden Markov models (where the transition probabilities are not known) are a standard tool in the design of intelligent systems, including speech recognition, natural language modelling, pattern recognition, weather prediction.

In discrete mathematics, theoretical computer science, and statistical physics, we often have to consider finite Markov chains with an enormous number of states. Card shuffling is an example of a Markov chain with 52! states. The "random recoloring process," discussed above, is an example of a class of Markov chains which have exponentially many states compared to the length of the description of the Markov chain. (The description of an instance of the random recoloring process consists of specifying the graph G and the parameter Q.) We remark that the random recoloring process is but one instance of a class of Markov chains referred to as "Glauber dynamics," originating in statistical physics.

An example from computer science: if the state of a memory unit on a computer chip can be described by a bit-string of length k then the number of states of the chip is 2^k. (Transitions can be defined by changing one bit at a time.)

This exponential behavior is typical of combinatorially defined Markov chains.

Because of the exponential growth in the number of states, it is not possible to store the transition matrices and to compute their powers; the size of the matrices becomes prohibitive even for moderate values of the description length of the states. (Think of a $52! \times 52!$ matrix to study card shuffling!)

The evolution of such "combinatorially defined" Markov chains is therefore the subject of intense theoretical study. It is of great importance to find conditions under which the distribution is guaranteed to get **close** to the stationary distribution very fast (in a polynomial number of steps). As noted above, this circumstance is called **rapid mixing**. Note that rapid

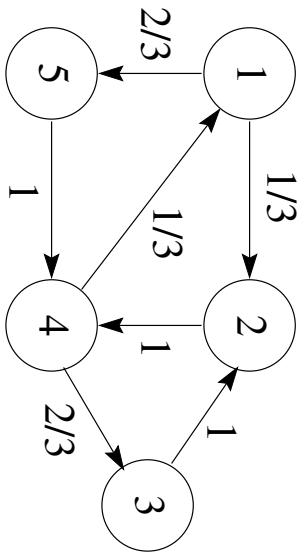


Figure 5.3: Transition graph for a Markov chain.

mixing takes place much faster than it would take to visit each state! (Why is this not a paradox?)

5.2 Problems

Exercise 5.2.1. Let $\mathcal{M}$ be the Markov chain shown in Figure 5.2.

1. Is $\mathcal{M}$ strongly connected?
2. Write down the transition matrix T for $\mathcal{M}$.
3. What is the period of vertex 1?
4. Find a stationary distribution for $\mathcal{M}$. You should describe this distribution as a 1×5 matrix.
5. Prove that $\lim_{t\to\infty} T^t$ does not exist. Prove this directly, do not refer to the Perron-Frobenius theorem.

Exercise 5.2.2. Consider the following digraph: $V = [3]$, $E = \{1 \to 2, 1 \to 3, 2 \to 2, 3 \to 3\}$. Write down the transition matrix of the random walk on this graph, with transition probabilities as shown in Figure 5.2. State two different stationary distributions for this Markov chain.

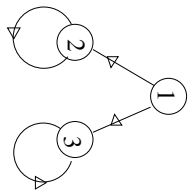


Figure 5.4: The transition graph for a Markov chain.

Chapter 6

Algebra Review

6.1 Groups

Definition 6.1.1. A **semigroup** $(G, \cdot)$ is a set G with a binary operation $\cdot$ such that:

Axiom 1 $(\forall a, b \in G)(\exists! a \cdot b \in G)$

Axiom 2 $(\forall a, b, c \in G)(a \cdot (b \cdot c) = (a \cdot b) \cdot c)$

Definition 6.1.2. A **group** $(G, \cdot)$ is a semigroup such that:

Axiom 3 (Identity element) $(\exists 1 \in G)(\forall a \in G)(1 \cdot a = a = a \cdot 1)$

Axiom 4 (Inverse) $(\forall a \in G)(\exists b \in G)(a \cdot b = b \cdot a = 1)$

Multiplicative Notation:

- $ab = a \cdot b$
- In Axiom 4, $b = a^{-1}$

Additive Notation:

- Binary operation '+'
- Identity becomes '0'
- Additive inverse '$-a$'

The size of G as a set, which is denoted $|G|$, is called the **order** of G.

Definition 6.1.3. G is an **abelian group** if G is a group such that $(\forall a, b \in G)(ab = ba)$.

Definition 6.1.4. $H \subseteq G$ is a **subgroup** of G (denoted $H \leq G$) if

1. $1 \in H$
2. H is closed under the binary operation
3. H is closed under inverses

Definition 6.1.5. Let $H \leq G$. The sets of the form $a \cdot H := \{ah : h \in H\}$ for $a \in G$ are the **left cosets** of H. The left cosets partition G. **Right cosets** are defined analogously.

Definition 6.1.6. $|G : H|$ = number of left cosets of H in G is called the **index** of H in G.

Exercise 6.1.7. Prove that the number of left cosets is the same as the number of right cosets, even if G is infinite. (*Hint:* construct a bijection between the left and the right cosets.)

Exercise$^+$ 6.1.8. Prove: if G is finite then the left and the right cosets have a common system of representatives, i. e., there exists a set T of size $|T| = |G : H|$ such that T contains exactly one element from every left coset as well as from every right coset.

Exercise 6.1.9 (Lagrange). If $H \leq G$ then $|G| = |H| \cdot |G : H|$. Therefore, if $|G| < \infty$ then $|H| \Big| |G|$.

Exercise 6.1.10. Prove: the intersection of subgroups is a subgroup.

Definition 6.1.11. Let $S \subset G$. We define the subgroup of G **generated** by S by

$$\langle S \rangle = \bigcap_{\substack{H \leq G \\ H \supseteq S}} H.$$

A group is **cyclic** if it is generated by an element ($|S| = 1$).

Exercise 6.1.12. $\langle S \rangle$ is the set of all (finite) products of elements of S and inverses of elements of S.

Example 6.1.13. Let $S = \{a, b\}$. Then $aba^{-4}bab^6 \in \langle S \rangle$.

Example 6.1.14. If $|S| = 1$ and $S = \{g\}$ then $\langle S \rangle = \{g^n : n \in \mathbb{Z}\}$.

Exercise 6.1.15. If G is cylic then

1. if $|G| = \infty$ then $G \cong (\mathbb{Z}, +)$ ($\cong$ denotes isomorphism)
2. if $|G| = n$ then $G \cong (\mathbb{Z}/n\mathbb{Z}, +)$

Definition 6.1.16. The **order** of an element $g \in G$ is the order of the cyclic group generated by g: $|g| := |\langle g \rangle|$.

Exercise 6.1.17. $g^k = 1 \Leftrightarrow |g| \;\Big|\; k$

Exercise 6.1.18. If G is finite then $g^{|G|} = 1$

Exercise 6.1.19 (Euler - Fermat). $(\forall a, n \in \mathbb{Z})(\text{g.c.d.}\,(a,n) = 1 \Rightarrow a^{\varphi(n)} \equiv 1 \bmod n)$

Exercise 6.1.20. If G is an abelian group then

$$\frac{\text{l.c.m.}\,[|a|,|b|]}{\text{g.c.d.}\,[|a|,|b|]} \;\Big|\; |ab| \;\Big|\; \text{l.c.m.}\,[|a|,|b|].$$

This shows that if g.c.d. $[|a|,|b|] = 1$ then $|ab| = $ l.c.m. $[|a|,|b|]$.

Definition 6.1.21. F_k is a **free group of rank k on free generators** $\{a_1, \ldots, a_k\}$ if the products of the a_i and the a_i^{-1} give 1 only by explicit cancellation.

Example 6.1.22. $ab^{-3}a^4a^{-2}a^{-2}b^5b^{-2}a^{-1} = 1$

Exercise$^+$ 6.1.23. $F_3 \leq F_2$. In fact, $F_\infty \leq F_2$.

Definition 6.1.24. For a commutative ring R (see Definition 6.2.3), the **special linear group** $SL(n,R)$ is the group of those $n \times n$ matrices $A \in M_n(R)$ with $\det(A) = 1$. (More about rings below; we assume all rings have an identity element.)

Exercise* 6.1.25. (Sanov) $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and A^T (A transpose) freely generate a free group $F_2 \leq SL(2,\mathbb{Z})$. (*Hint:* for $T = (t_{ij}) \in SL(2,\mathbb{Z})$, let $m(T) = \max|t_{ij}|$. Show that $\forall T$ there is at most one $X \in \{A, A^T, A^{-1}, A^{-T}\}$ such that $m(T) \geq m(TX)$.)

Definition 6.1.26. Let G be a group and $S \subseteq G \setminus 1$. The **Cayley graph** $\Gamma(G,S)$ has G for its vertex set; elements $g, h \in G$ are adjacent if $gh^{-1} \in S \cup S^{-1}$ (where $S^{-1} = \{s^{-1} : s \in S\}$).

Exercise 6.1.27. Prove: $\Gamma(G,S)$ is connected if and only if S generates G.

Exercise 6.1.28. Suppose $G = \langle S \rangle$. Then $\Gamma(G,S)$ is bipartite if and only if G has a subgroup N of index 2 such that $S \cap N = \emptyset$.

Exercise 6.1.29. Let S be a minimal set of generators of G, i. e., no proper subset of S generates G. Prove: $K_{3,5} \not\subset \Gamma(G,S)$.

A theorem of Erdős and Hajnal states that if an (infinite) graph X does not contain $K_{m,\aleph_1}$ as a subgraph (for some $m \in \mathbb{N}$) then $\chi(X) \leq \aleph_0$. As a consequence of the preceding exercise, if S is a minimal set of generators then $\chi(\Gamma(G,S)) \leq \aleph_0$.

Exercise 6.1.30. Prove that every group G has a set S of generators such that $\chi(G,S) \leq \aleph_0$. *Hint.* Not every group has a minimal set of generators (e. g., $(\mathbb{Q},+)$ does not). But every group has a *sequentially non-redundant* set of generators, $\{s_\alpha : \alpha \in I\}$, where I is a well-ordered set and $(\forall \alpha \in I)(s_\alpha \notin \langle s_\beta : \beta < \alpha\rangle$. Prove that if S is sequentially non-redundant then $K_{5,17} \not\subset \Gamma(G,S)$.

Exercise 6.1.31. If a regular graph of degree r with n vertices has girth g then

$$n \geq 1 + r + r(r-1) + \ldots r(r-1)^{\lfloor (g-3)/2 \rfloor} > (r-1)^{g/2-1}.$$

Consequently, $g < 1 + 2\log n/\log(r-1)$.

On the other hand, Erdős and Sachs proved for every $r \geq 3$ there exist r-regular graphs of girth $g \geq \log n/\log(r-1)$. The following problem addresses the question of **explicit construction** of a 4-regular graph with large girth. The girth will be optimal within a constant factor.

Exercise$^+$ 6.1.32. (Margulis) Let $G = SL(2,p) := SL(2,\mathbb{Z}/p\mathbb{Z})$. Let $S = \{A, B\}$ where $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $B = A^T$ (A transpose). Note that $|G| < p^3$ and $\Gamma(G,S)$ has degree 4. Prove that the girth of $\Gamma(G,S)$ is $\Omega(\log p)$. (*Hint.* Use Sanov's Theorem and the submultiplicativity of matrix norm.

6.2 Rings

Definition 6.2.1. A **ring** $(R,+,\cdot)$ is an abelian group $(R,+)$ and semigroup $(R,\cdot)$ such that:

- (Distributivity) $(\forall a,b,c \in R)(a(b+c) = ab + ac)$ and $((b+c)a = ba + ca)$

Exercise 6.2.2. In a ring R, $(\forall a \in R)(a \cdot 0 = 0 = 0 \cdot a)$

Definition 6.2.3. $(R,+,\cdot)$ is a **commutative** ring if $(R,\cdot)$ is abelian.

Definition 6.2.4. R is a **ring with identity** if $(R,\cdot)$ satisfies Axiom 3 (semigroup with identity) and $1 \neq 0$.

CONVENTION. By "rings" we shall always mean **rings with identity.**

Definition 6.2.5. $a \in R$ is a **unit** if $\exists a^{-1} \in R$.

Exercise 6.2.6. The units of R form a multiplicative group denoted $R^\times$.

Example 6.2.7. Let R be a ring.

- $M_n(R) :=$ set of $n \times n$ matrices over R is a ring

Exercise 6.2.8. Let R be a commutative ring. $GL(n, R)$ denotes the group of units of $M_n(R)$. Prove: $A \in M_n(R)$ belongs to $GL(R)$ if and only if $\det(A) \in R^\times$.

Example 6.2.9. mod m residue classes form a ring, denoted $\mathbb{Z}/m\mathbb{Z}$.

Exercise 6.2.10. What is the order of the group of units of $\mathbb{Z}/m\mathbb{Z}$?

Definition 6.2.11. $a \in R$ is a **left zero divisor** if $a \neq 0$ and $(\exists b \in R, b \neq 0)(ab = 0)$. **Right zero divisors** are defined analogously.

Definition 6.2.12. $a \in R$ is a **zero-divisor** if a is a left OR a right zero-divisor.

Exercise 6.2.13.

1. If $\exists a^{-1}$ then a is not a zero-divisor.
2. The converse is false.
3. The converse is true if R is finite.
4. The converse is true if $R = M_n(F)$ where F is a field. In this case, $A \in R$ is a zero-divisor if and only if $\det(A) = 0$.

Definition 6.2.14. An **integral domain** is a commutative ring with no zero-divisors.

Definition 6.2.15. A **division ring** is a ring where all nonzero elements are units, i. e., $R^\times = R \setminus \{0\}$.

6.3 Gaussian integers and quaternions; sums of two squares and four squares

Definition 6.3.1. The **Gaussian Integers** are complex numbers of the form $\{a+bi : a, b \in \mathbb{Z}\}$ where $i = \sqrt{-1}$. They form the ring $\mathbb{Z}[i]$. The *norm* of $z \in \mathbb{Z}[i]$ is $N(z) = a^2 + b^2 = z\overline{z}$.

Exercise 6.3.2. Define divisibility among Gaussian integers. Observe that $z \mid w \Rightarrow N(z) \mid N(w)$. Show that the units among the Gaussian integers are $\pm 1, \pm i$.

Exercise 6.3.3. Use Gaussian integers to show that $(a^2 + b^2)(c^2 + d^2) =$ sum of two squares. *Hint.* Observe that $N(zw) = N(z)N(w)$.

Exercise$^+$ 6.3.4. Define division with remainder among Gaussian integers. Show the existence of g.c.d. 's. Use this to establish unique prime factorization in $\mathbb{Z}[i]$.

Exercise 6.3.5. Show: if z is a prime in $\mathbb{Z}[i]$ then $N(z)$ is either p or p^2 for some prime $p \in \mathbb{Z}$. In the former case $p = N(z) = a^2 + b^2$; in the latter case, $p = z$.

Exercise$^+$ 6.3.6. Let $p \in \mathbb{Z}$ be a prime. Prove: p is a prime in $\mathbb{Z}[i]$ if and only if $p \equiv -1 \pmod 4$. *Hint.* "If:" if $p \equiv -1 \pmod 4$ then $p \neq a^2 + b^2$. "Only if:" if $p \equiv 1 \pmod 4$ then $(\exists a \in \mathbb{Z})(p \mid a^2 + 1)$. Let $w = a + bi \in \mathbb{Z}[i]$. Let $z =$ g.c.d. (p, w).

Exercise 6.3.7. Infer from the preceding exercise: if p is a prime (in $\mathbb{Z}$) and $p \equiv 1 \pmod 4$ then p can be written as $a^2 + b^2$.

Exercise$^+$ 6.3.8. The positive integer $n = \prod p_i^{\alpha_i}$ can be written as a sum of two squares if and only if $(\forall i)(p_i \equiv -1 \pmod 4 \Rightarrow 2 \mid \alpha_i)$.

Exercise$^+$ 6.3.9. Show that the number of ways to write n as $a^2 + b^2$ in $\mathbb{Z}$ is

$$\epsilon + \prod_{i:p_i \equiv 1 \bmod 4} (\alpha_i + 1)$$

where $\epsilon = 1$ if n is a square and 0 otherwise.

Exercise 6.3.10. Let n be a product of primes $\equiv 1 \pmod 4$ and suppose n is not a square. Prove: the number of ways to write n as $a^2 + b^2$ is $d(n)$ (the number of positive divisors of n).

Definition 6.3.11. The **quaternions** form a 4-dimensional division algebra $\mathbb{H}$ over $\mathbb{R}$, i. e., a division ring which is a 4-dimensional vector space over $\mathbb{R}$. The standard basis is denoted by $\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}$, so a quaternion is a formal expression of the form $z = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$. Multiplication is performed using distributivity and the following rules:

- $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$;
- $\mathbf{ij} = -\mathbf{ji} = \mathbf{k}$;
- $\mathbf{jk} = -\mathbf{kj} = \mathbf{i}$;
- $\mathbf{ki} = -\mathbf{ik} = \mathbf{j}$.

It is clear that $\mathbb{H}$ is a ring. We need to find inverses.

Exercise 6.3.12. For $z = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, we define the **norm** of z by $N(z) = a^2 + b^2 + c^2 + d^2$. Prove: $N(z) = z\overline{z} = \overline{z}z$, where $\overline{z} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$ is the **conjugate** quaternion.

Exercise 6.3.13. Let $z, w \in \mathbb{H}$. Prove: $N(zw) = N(z)N(w)$.

Exercise 6.3.14.

$$(a^2 + b^2 + c^2 + d^2)(k^2 + l^2 + m^2 + n^2) = t^2 + u^2 + v^2 + w^2$$

where t, u, v, w are bilinear forms of (a, b, c, d) and (k, l, m, n) with integer coefficients. Calculate the coefficients.

Exercise* 6.3.15. (Lagrange) Every integer is a sum of 4 squares. *Hint.* By the preceding exercise, it suffices to prove for primes. First prove that for every prime p there exist $x_1, \ldots, x_4 \in \mathbb{Z}$ such that $p \mid \sum x_i^2$ and g.c.d. $(x_1, \ldots, x_4) = 1$. Now let $m > 0$ be minimal such that $mp = x_1^2 + \cdots + x_4^2$; note that $m < p$. If $m \geq 2$, we shall reduce m and thereby obtain

a contradiction (Fermat's method of infinite descent; Fermat used it to prove that if $p \equiv 1 \pmod 4$ then p is the sum of 2 squares). If m is even, halve m by using $(x_1 \pm x_2)/2$ and $(x_3 \pm x_4)/2$ (after suitable renumbering). If m is odd, take $y_i = x_i - mt_i$ such that $|y_i| < m/2$. Observe that $0 < \sum y_i^2 < m^2$ and $m \mid \sum y_i^2$, so $\sum y_i^2 = md$ where $0 < d < m$. Now represent $m^2dp = (\sum x_i^2)(\sum y_i^2)$ as a sum of four squares, $\sum z_i^2$, using the preceding exercise. Analyzing the coefficients, verify that $(\forall i)(m \mid z_i)$. Now $dp = \sum (z_i/m)^2$, the desired contradiction.

6.4 Fields

Definition 6.4.1. A **field** is a commutative division ring.

Example 6.4.2. Let F be a field.

- $M_n(F) :=$ set of $n \times n$ matrices over F is a ring
- $GL_n(F) :=$ group of units of $M_n(F)$ is called the "General Linear Group"

Exercise 6.4.3. A finite ring with no zero divisors is a division ring. (*Hint:* use Exercise 6.2.13.)

Theorem 6.4.4 (Wedderburn). *A finite division ring is a field.*

Exercise 6.4.5. If F is a field and $G \leq F^\times$ is a finite multiplicative subgroup then G is cyclic.

Definition 6.4.6. Let R be a ring and for $x \in R$ let n_x be the g.c.d. of all n such that $nx = 0$ where

$$\begin{aligned} nx &:= x + \cdots + x \ (n \text{ times}) \text{ when } n > 0 \\ nx &:= -x - \cdots - x \ (|n| \text{ times}) \text{ when } n < 0 \\ nx &:= 0 \text{ when } n = 0. \end{aligned}$$

Exercise 6.4.7. $n_x \cdot x = 0$

Exercise 6.4.8. If R has no zero divisors then $(\forall x, y \neq 0)(n_x = n_y)$.

Definition 6.4.9. The common value n_x is called the **characteristic** of R.

Exercise 6.4.10. If R has no zero divisors then $\mathrm{char}(R) = 0$ or it is prime. In particular, every field has 0 or prime characteristic.

Exercise 6.4.11. If R is a ring without zero-divisors, of characteristic p, then $(a+b)^p = a^p + b^p$.

Exercise 6.4.12.

1. If R has characteristic 0 then $R \supseteq \mathbb{Z}$
2. If R has characteristic p then $R \supseteq \mathbb{Z}/p\mathbb{Z}$.

Exercise 6.4.13. If F is a field of characteristic 0 then $F \supseteq \mathbb{Q}$.

Definition 6.4.14. A **subfield** of a ring is a is a subset which is a field under the same operations. If K is a subfield of L then we say that L is an extension of K; the pair (K, L) is referred to as a **field extension** and for reasons of tradition is denoted L/K.

Definition 6.4.15. A **prime field** is a field without a proper subfield.

Exercise 6.4.16. The prime fields are $\mathbb{Q}$ and $\mathbb{Z}/p\mathbb{Z}$ (p prime).

Definition 6.4.17. Observe: if L/K is a field extension then L is a vector space over K. The **degree** of the extension is $[L : K] := \dim_K L$. A **finite extension** is an extension of finite degree.

Exercise 6.4.18. The order of a finite field is a prime power. *Hint.* Let L be a finite field and K its prime field, so $|K| = p$; let $[L : K] = k$. Prove: $|L| = p^k$.

Exercise 6.4.19. The degree of the extension $\mathbb{C}/\mathbb{R}$ is 2. The degree of the extension $\mathbb{R}/\mathbb{Q}$ is uncountably infinite (continuum).

Exercise$^+$ 6.4.20. Prove that $\sqrt{2}, \sqrt{3}, \sqrt{5}, \ldots, \sqrt{p}$ are linearly independent over $\mathbb{Q}$.

Exercise 6.4.21. If $K \subset L \subset M$ are fields then $[M : L][L : K] = [M : K]$.

6.5 Polynomials over Rings and Fields

Definition 6.5.1. Let R be a commutative ring. (As always we assume R has an identity.) $R[x]$ denotes the ring of polynomials in the variable x with coefficients in R.

Exercise 6.5.2. If R is an integral domain then $R[x]$ is an integral domain.

Definition 6.5.3. A **unique factorization domain** (UFD) is an integral domain in which every element can be written uniquely as a product of irreducible elements. The factorization is unique up to the order of the factors and multiplying each factor by units.

Example 6.5.4. Every field is a UFD. The rings $\mathbb{Z}$ and $\mathbb{Z}[i]$ are UFDs.

Exercise 6.5.5. Prove: if R is a UFD then $R[x]$ is a UFD.

Exercise 6.5.6. For an integral domain R, define the field of quotients (or field of fractions) $\left\{ \frac{a}{b} : a, b \in R, b \neq 0 \right\}$ by copying how $\mathbb{Q}$ arises from $\mathbb{Z}$. (*Hint:* "rational numbers" are equivalence classes of fractions).

Definition 6.5.7. Let F be a field and $F(x)$ be the field of quotients of $F[x]$. $F(x)$ is called the **function field** over F.

Exercise 6.5.8. $\dim_{\mathbb{R}} \mathbb{R}[x]$ is countably infinite and $\dim_{\mathbb{R}} \mathbb{R}(x)$ is uncountably infinite. (*Hint:* Show $\left\{ \frac{1}{x-\alpha} : \alpha \in \mathbb{R} \right\}$ is linearly independent over $\mathbb{R}$.)

Definition 6.5.9. Let F be a field. $f \in F[x]$ is **irreducible** if f is not constant (i. e., $\deg f \geq 1$) and $(\forall g, h \in F[x])(f = gh \Rightarrow \deg g = 0 \text{ or } \deg h = 0)$.

Definition 6.5.10. Let L/K be a field extension; let $\alpha \in L$. We say that α is **algebraic** over K if $(\exists f \in K[x])(f \neq 0 \text{ and } f(\alpha) = 0)$. We define $m_\alpha(x)$ as the g.c.d. of all such polynomials and we call it the **minimal polynomial** of α (over K). If all elements of L are algebraic over K then we call L/K an algebraic extension.

Exercise 6.5.11. $m_\alpha(\alpha) = 0$.

Exercise 6.5.12. m_α is irreducible over K.

Exercise 6.5.13. Every finite extension is algebraic.

Definition 6.5.14. Let R be a ring. $I \subseteq R$ is a **left ideal** of R if I is an additive subgroup of R and $(\forall r \in R)(rI \subseteq I)$. Right ideals are defined analogously. I is an **ideal** if it is both a left- and a right-ideal.

Definition 6.5.15. Let R be a ring and I and ideal of R. The additive quotient group R/I with elements $a + I$ is a ring under the multiplication rule $(a + I)(b + I) = ab + I$. It is called the **quotient ring**.

Exercise 6.5.16. Let F be a field and $f \in K[x]$. The ring $K[x]/(f)$ if a field if and only if f is irreducible.

Definition 6.5.17. A **simple extension** $K(\alpha)$ is the smallest field containing K and α.

Exercise 6.5.18. If α is algebraic over K then $K(\alpha) = K[\alpha] = \{f(\alpha) : f \in K[x]\} \simeq K[x]/(m_\alpha)$.

Exercise 6.5.19. Let $\mathbb{F}_q$ be a finite field of order q. Then

$$x^q - x = \prod_{\alpha \in \mathbb{F}_q} (x - \alpha)$$

$$x^{q-1} - 1 = \prod_{\alpha \in \mathbb{F}_q^\times} (x - \alpha)$$

Exercise$^+$ 6.5.20. Let $q = p^n$ be a prime power. Let $F_d(x)$ be the product of of all monic irreducible polynomials of degree d over $\mathbb{F}_p$. Prove that $x^q - 1 = \prod_{d|n} F_d(x)$. (For this exercise, do not assume the existence of $\mathbb{F}_q$. The field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ of course exists.)

Exercise$^+$ 6.5.21. Let N_d be the number of monic irreducible polynomials of degree d over $\mathbb{F}_p$. Observe from the preceding exercise that $p^k = \sum_{d|n} dN_d$. Infer:

$$N_n = (1/n)\sum_{d|n} \mu(n/d)p^d.$$

Conclude that $N_n \neq 0$.

Exercise 6.5.22. Prove that there exists a field of order p^n. *Hint.* The preceding exercise shows that there exists an irreducible polynomial f of degree n over $\mathbb{F}_p$. Take the field $\mathbb{F}_p[x]/(f)$.

Exercise 6.5.23. Prove: the field of order p^k is unique (up to isomorphism).

6.6 Irreducibility over $\mathbb{Z}$, Gauss lemma, cyclotomic polynomials

Exercise 6.6.1. (Gauss Lemma) A polynomial $f \in \mathbb{Z}[x]$ is **primitive** if the g.c.d. of its coefficients is 1. Prove: the product of primitive polynomials is primitive. (*Hint.* Assume $fg = ph$ where $f, g, h \in \mathbb{Z}[x]$ and p is a prime. Look at this equation modulo p and use the fact that $\mathbb{F}_p[x]$ is an integral domain.)

Exercise 6.6.2. If $f \in \mathbb{Z}[x]$ splits into factors of lower degree over $\mathbb{Q}[x]$ then such a split occurs over $\mathbb{Z}[x]$. In fact, if $f = gh$ where $g, h \in \mathbb{Q}[x]$ then $(\exists r \in \mathbb{Q})(rg \in \mathbb{Z}[x]$ and $h/r \in \mathbb{Z}[x])$.

Exercise 6.6.3. Let $f(x) = \prod_{i=1}^{n}(x - a_i) - 1$ where the a_i are distinct integers. Then $f(x)$ is irreducible over $\mathbb{Q}$. *Hint.* Let $f = gh$ where $g, h \in \mathbb{Z}[x]$. Observe that $(\forall i)(g(a_i) + h(a_i) = 0)$.

Exercise 6.6.4. Let $f(x) = \left(\prod_{i=1}^{n}(x - a_i)\right)^2 + 1$ where the a_i are distinct integers. Then $f(x)$ is irreducible over $\mathbb{Q}$. *Hint.* Let $f = gh$ where $g, h \in \mathbb{Z}[x]$. Observe that $(\forall i)(g(a_i) = h(a_i) = \pm 1)$. Observe further that g never changes sign; nor does h.)

Exercise 6.6.5. Let $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{Z}[x]$. If p is a prime and $p \nmid a_n, p \mid a_0, \ldots, p \mid a_{n-1}, p^2 \nmid a_0$, then f is irreducible over $\mathbb{Q}$. (*Hint:* Unique factorization in $\mathbb{F}_p[x]$).

Exercise 6.6.6. If p is a prime then $\Phi_p(x) := x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible over $\mathbb{Q}$. (*Hint:* Introduce the variable $z = x - 1$.)

Exercise 6.6.7. g.c.d. $(x^k - 1, x^\ell - 1) = x^d - 1$ where $d =$ g.c.d. (k, ℓ).

Definition 6.6.8. The **n-th cyclotomic polynomial** is defined as

$$\Phi_n(x) = \prod_{\omega}(x - \omega) \in \mathbb{C}[x]$$

where the product extends over all complex primitive n-th roots of unity.

Exercise 6.6.9.

$$\prod_{d\,|\,n} \Phi_d(x) = x^n - 1$$

Exercise 6.6.10. Let $f, g \in \mathbb{Z}[x]$ with the leading coefficient of g equal to 1. If $\frac{f}{g} \in \mathbb{Q}[x]$ then $\frac{f}{g} \in \mathbb{Z}[x]$.

Exercise 6.6.11.

$$\Phi_n(x) \in \mathbb{Z}[x]$$

Exercise 6.6.12.

$$\Phi_n(x) = \prod_{d|n}(x^d - 1)^{\mu(n/d)}$$

Exercise 6.6.13. Let f, g be polynomials over the field F. Prove: if $g^2 \,|\, f$ then $g \,|\, f'$, where f' is the (formal) derivative of f.

Exercise 6.6.14. Prove: if $p \nmid n$ then $x^n - 1$ has no multiple factors over $\mathbb{F}_p$.

Exercise$^+$ 6.6.15. Let $a \neq b$ and n be positive integers. Let p be a prime. Assume $p \,|\, \text{g.c.d.}\,(\Phi_a(n), \Phi_b(n))$. Prove: $p \,|\, \text{g.c.d.}\,(a, b)$.

Exercise 6.6.16.

1. Prove: if f is a polynomial over $\mathbb{F}_p$ then $f(x^p) = (f(x))^p$.
2. Prove: if f is a polynomial over $\mathbb{F}_q$ where q is a power of the prime p then there exists a polynomial g over $\mathbb{F}_q$ such that $f(x^p) = (g(x))^p$.
3. Find an infinite field F of characteristic p such that part (b) is false if $\mathbb{F}_q$ is replaced by F.

Exercise$^+$ 6.6.17. Let ω be a complex primitive n-th root of unity. Prove: if p is a prime and $p \nmid n$ then the minimal polynomials of ω and ω^p (over $\mathbb{Q}$) coincide. (*Hint.* Let f and g be the minimal polynomials of ω and ω^p, respectively. Assume $f \neq g$; then $fg \,|\, x^n - 1$. Observe that $f(x) \,|\, g(x^p)$. Look at this equation over $\mathbb{F}_p$ and conclude that $x^n - 1$ has a multiple factor over $\mathbb{F}_p$, a contradiction.)

Exercise 6.6.18. A major result is now immediate: Φ_n is irreducible over $\mathbb{Q}$.

Chapter 7

Finite Projective Planes

7.1 Basics

Definition 7.1.1. An *incidence geometry* is a set P of "points," a set L of "lines," and an *incidence relation* $I \subseteq P \times L$.

Notation 7.1.2. If $(p, \ell) \in I$ then we say that p is *incident* with ℓ and we write $p \multimap \ell$. If $(p, \ell) \notin I$, we write $p \overset{\bullet}{—} \ell$.

Definition 7.1.3. The *dual* of the incidence geometry (P, L, I) is the incidence geometry (L, P, I^{-1}) (we switch the roles of points and lines; the same pairs remain incident).

Definition 7.1.4. A *projective plane* is an incidence geometry satisfying the following three axioms:

Axiom 1. $(\forall p_1 \neq p_2 \in P)(\exists! \ell \in L)(p_1 \multimap \ell \text{ and } p_2 \multimap \ell)$.

Axiom 2. $(\forall \ell_1 \neq \ell_2 \in L)(\exists! p \in P)(p \multimap \ell_1 \text{ and } p \multimap \ell_2)$.

Axiom 3. (Non-degeneracy) $\exists p_1, p_2, p_3, p_4 \in P$ such that no three are on the same line.

Exercise 7.1.5. Prove that the dual of a projective plane is a projective plane. (Note the dual of Axiom 1 is Axiom 2 and vice versa. State the dual of Axiom 3 and prove that it follows from Axioms 1–3.)

Unless expressly stated otherwise, all projective planes considered here will be finite. For $p \in P$, let $\deg(p)$, the *degree of* p, be the number of lines incident with p. For $\ell \in L$, let $\mathrm{rk}(\ell)$, the *rank of* ℓ, be the number of points incident with ℓ.

Exercise 7.1.6. If $p \overset{\bullet}{—} \ell$, then $\deg(p) = \mathrm{rk}(\ell)$.

This result is an immediate consequence of Axioms 1 and 2 and does not require Axiom 3. (Prove!)

Exercise 7.1.7. $\forall p_1, p_2 \in P$, $\exists \ell \in L$, $p_1 \overset{\bullet}{—} \ell$ and $p_2 \overset{\bullet}{—} \ell$.

The proof of this requires some care.

The following is immediate from the preceding two exercises.

Exercise 7.1.8. In a projective plane, all points have the same degree.

Use the fact that the dual of a projective plane is a projective plane to infer:

Exercise 7.1.9. In a projective plane, all lines have the same rank.

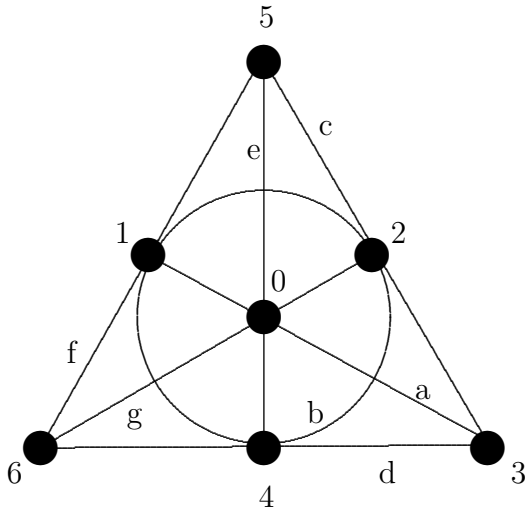
Exercise 7.1.10. In a projective plane, the degree of every point and the rank of each line is the same.

In other words, projective planes are *regular* and uniform, and their degree and rank are equal.

For reasons of tradition, we denote this common value by $n+1$. Every point of the plane is thus incident with $n+1$ lines and every line is incident with $n+1$ points. The number n is called the *order* of the projective plane.

Proposition 7.1.11. $|P| = |L| = n^2 + n + 1$.

The smallest projective plane is the *Fano plane*, which has order $n = 2$.



The incidence matrix for the Fano plane is as follows:

	0	1	2	3	4	5	6
a	1	1	0	1	0	0	0
b	0	1	1	0	1	0	0
c	0	0	1	1	0	1	0
d	0	0	0	1	1	0	1
e	1	0	0	0	1	1	0
f	0	1	0	0	0	1	1
g	1	0	1	0	0	0	1

7.2 Galois Planes

A class of projective planes called *Galois planes* is constructed as follows. Let F be a finite field of order q. Let F^3 be the 3-dimensional space over F. We define the *inner product* over F^3 in the usual way: for $u = (\alpha_1, \alpha_2, \alpha_3)$ and $v = (\beta_1, \beta_2, \beta_3)$ we set $u \cdot v = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3$. We say that u and v are *perpendicular* if $u \cdot v = 0$.

Let us say that two nonzero vectors $u, v \in F^3$ are equivalent if $u = \lambda v$ for some $\lambda \in F$.

Let S be the set of equivalence classes on $F^3 - 0$. Note that each equivalence class has $q-1$ elements and therefore the number of equivalence classes is $(q^3 - 1)/(q - 1) = q^2 + q + 1$.

Set $P = L = S$ and let us say that $p \in P$ and $\ell \in L$ are incident if $u \cdot v = 0$ where $u \in p$ (u is a vector in the equivalence class p) and $v \in \ell$. The coordinates of u are called *homogeneous coordinates* of p (they are not unique–every point has $q-1$ triples of homogeneous coordinates); similarly, the coordinates of v are called homogeneous coordinates of ℓ.

Exercise 7.2.1. Prove that this definition gives a projective plane of order q. It is called a *Galois plane* after Évariste Galois (1811–1832), the discoverer of finite fields and of modern algebra.

Exercise 7.2.2. Prove that the Fano plane is a Galois plane (necessarily over the field of order 2) by assigning homogeneous coordinates to the points and lines of the Fano plane.

A set of points is *collinear* if there is a line with which all of them are incident.

We say that four points are in *general position* if no three of them are collinear.

A *collineation* is a transformation of the projective plane consisting of a permutation of the points and a permutation of the lines which preserves incidence.

Theorem 7.2.3 (Fundamental Theorem of Projective Geometry.). *If $a_1, \ldots, a_4$ and $b_1, \ldots, b_4$ are two quadruples of points in general position in a Galois plane then there exists a collineation φ such that $(\forall i)(\varphi(a_i) = b_i)$.*

Exercise 7.2.4. (Prove!) Use Theorem 7.2.3 to show that the Fano plane has 168 collineations.

Exercise 7.2.5. Consider the projective plane $\Pi = PG(2, F)$ over the field F. for $i = 1, 2, 3$, let $p_i = (a_i, b_i, c_i)$ be three points in Π, given by homogeneous coordinates ($a_i, b_i, c_i \in F$, not all zero). Prove: the three points are collinear (lie on a line) if and only if the 3×3 determinant $|a_i\ b_i\ c_i|$ is zero.

Exercise 7.2.6. A projective plane $\Pi_1 = (P_1, L_1, I_1)$ is a *subplane* of the projective plane $\Pi_2 = (P_2, L_2, I_2)$ if $P_1 \subset P_2$, $L_1 \subset L_2$, and the incidence relation I_1 is the restriction of I_2 to $P_1 \times L_1$. Prove: if Π_1 is a proper subplane of Π_2 then $n_1 \leq \sqrt{n}_2$ (where n_i is the order of Π_i).

Exercise 7.2.7. Let $P(n)$ be the number of projective planes of order n. Prove: $P(n) < (ne)^{(n+1)^3}$. *Hint.* first prove that

$$P(n) \leq \binom{\binom{n^2+n+1}{n+1}}{n^2+n+1}.$$

Exercise 7.2.8. Let us consider the Galois plane PG$(2, 5)$ (over the field of 5 elements).

1. How many points does this plane have, and what is the number of points per line?
2. Points are given by "homogeneous coordinates." Determine whether or not the points given by $a = [1, 4, 0]$, $b = [3, 2, 2]$, and $c = [4, 1, 2]$ are collinear (belong to the same line). (Coordinates are mod 5.) Prove your answer.

Chapter 8

Matroids and Geometric Lattices

8.1 Matroids

Definition. Let P be a poset and $a, b \in P$. We say that a *covers* b if $b < a$ and there is no c satisfying $b < c < a$.

Matroids (also called "combinatorial geometries") are a combinatorial abstraction of the concept of "linear independence" of a finite set of vectors. There are several related concepts which can be used to produce equivalent sets of axioms that describe matroids. We start with the description via "flats." We consider finite matroids only.

Definition 1 (via "flats") A *matroid* is a pair $\mathcal{M} = (X, \mathcal{F})$ where X is a set of points and $\mathcal{F}$ is a family of subsets of X called *flats*, such that

(1) $\mathcal{F}$ is closed under intersection,

(2) $\mathcal{F}$ contains the empty set, all singletons $\{x\}, x \in X$ and the set X itself,

(3) for every flat $E \in \mathcal{F}$, $E \neq X$, the union of all flats that cover E in $\mathcal{F}$ ($\mathcal{F}$ is a poset ordered by inclusion) is equal to X.

Exercise 8.1.1. Prove that the flats that cover E partition $X \backslash E$.

The *closure* $\overline{Y}$ of a subset $Y \subseteq X$ is defined as the intersection of all flats that contain Y. The subset $Y \subseteq X$ is *closed* if $Y = \overline{Y}$. The subset $Y \subseteq X$ is *independent* if for each $x \in Y$ we have $x \notin \overline{Y \backslash \{x\}}$. The *rank* $\rho(Y)$ of $Y \subseteq X$ is the maximum size of an independent subset of Y. The rank of X is the rank of the matroid $\mathcal{M}$.

One can define matroids directly through independent sets.

Definition 2 (via "independent sets") A matroid is a pair $(X, \mathcal{I})$, where $\mathcal{I}$ is a family of subsets of X, called *independent sets*, having the following three properties:

(A) $\emptyset \in \mathcal{I}$

(B) If $Y \in \mathcal{I}$ and $Z \subseteq Y$ then $Z \in \mathcal{I}$,

(C) (Exchange principle) If $Y, Z \in \mathcal{I}$ and $|Z| > |Y|$ then $(\exists z \in Z)(Y \cup \{z\} \in \mathcal{I})$.

It follows that any two maximal independent sets have the same cardinality, the rank of the matroid; therefore every maximal independent set is maximum. A maximal independent set is called a *basis*, and a minimal dependent set is called a *cycle*.

For $Y \subseteq X$, let $\mathcal{I}_Y$ be the set of those members of $\mathcal{I}$ contained in Y. Then $(Y, \mathcal{I}_Y)$ satisfies axioms (A)–(C) and therefore defines a matroid on the point set Y. We define $\rho(Y)$ to be the rank of this matroid. This defines a rank function $\rho : \mathcal{P}(X) \to \{$ nonnegative integers $\}$.

A subset Y of X is now said to be closed if $(\forall x \notin Y)(\rho(Y \cup \{x\}) > \rho(Y))$. The *closure* $\overline{Y}$ of an arbitrary set Y is the smallest closed set containing Y.

Exercise 8.1.2. Prove: (a) The closed subsets are exactly the maximal subsets of a given rank. (b) The closure of a closed set is itself. (c) Intersection of closed sets is closed.

The *flats* of a matroid defined via independent sets are defined as the closed sets.

Having completed the translation of the concepts of each definition in terms of the fundamental concept of the other definition, we can state that the two classes of structures are identical. In other words:

Exercise 8.1.3. Prove that the two sets of axioms (via flats and via independent sets) are equivalent (after translation of the concepts).

Matroids can also be axiomatized by making any of the following the basic concept: bases, the rank function, the cycles, the closure operator on $\mathcal{P}(X)$.

8.2 Examples of matroids

We list some examples of matroids.

Exercise 8.2.1. For each class of examples below, prove that they define matroids.

1. (Linear independence: the principal example) Let $V = \{v_1, ..., v_k\}$ be elements of a vector space. A subset of V is independent if it is linearly independent.

   A matroid is *representable* over a field F if it is isomorphic to a matroid defined by a set of vectors over F as above.

2. (Algebraic independence) Let F_1, F_2 be fields with $F_1 \subseteq F_2$. The elements $a_1, \ldots, a_k \in F_2$ are *algebraically independent* over F_1 if there is no nonzero polynomial in k variables with coefficients in F_1 which has $(a_1, \ldots, a_k)$ as a root. Take any finite subset $X \subseteq F_2$; call $Y \subseteq X$ "independent" if it is algebraically independent.

3. (Graphic matroids) Let X be the set of *edges* of a graph G. A subset of edges is "independent" if it is a forest (cycle free).

   Note that the points of this matroid $\mathcal{M}(G)$ are the *edges* of the graph G.

   Exercise 8.2.2. Prove that the rank of $\mathcal{M}(G)$ is $n-k$, where n is the number of vertices of G and k is the number of connected components.

   Exercise 8.2.3. Prove that the cycles in $\mathcal{M}(G)$ are exactly the cycles of the graph G (hence the term).

   Exercise 8.2.4. A *connected partition* of the graph G is a partition $V = V_1 \dot\cup \ldots \dot\cup V j$ of the vertex set of G such that the subgraph of G induced by each V_i is connected. Show that the flats of $\mathcal{M}(G)$ correspond to the connected partitions of the graph G.

   Exercise 8.2.5. Prove that a graphic matroid is representable over any field.

4. (Transversal matroids) Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a set-system. We define the matroid $\mathcal{T} = (\mathcal{F}, \mathcal{I})$ by calling a subfamily $\mathcal{G} \subseteq \mathcal{F}$ independent if $\mathcal{G}$ has a SDR (system of distinct representatives).

   Exercise 8.2.6. Prove that a transversal matroid is representable over every sufficiently large field.

5. (Dual matroid) Let $\mathcal{M} = (X, \mathcal{B})$ be matroid defined by its set of bases, $\mathcal{B}$. We define the *dual matroid* $\mathcal{M}^d = (X, \mathcal{B}^d)$ to be the matroid with the same point set whose bases are the complements in X of the bases of $\mathcal{M}$, i. e., $\mathcal{B}^d = \{X \setminus B : B \in \mathcal{B}\}$.

   Exercise 8.2.7. Prove: if $\mathcal{M}$ is representable over a field F, then $\mathcal{M}^d$ is representable over F.

   Exercise 8.2.8. Prove: if G is a *planar* graph then the dual of $\mathcal{M}(G)$ is also graphic (and corresponds to a planar graph). (Note: the converse also holds: if the dual of a graphic matroid is also graphic then the corresponding graphs are planar.)

6. (Contraction) Let $\mathcal{M} = (X, \mathcal{I})$ be a matroid defined by its set of independent sets. For $x \in X$, we define the *contraction* $\mathcal{M}/x$ to have point set $X' = X \setminus \{x\}$; a subset $Y \subseteq X'$ is "independent" in $\mathcal{M}/x$ if $Y \cup \{x\} \in \mathcal{I}$.

   Exercise 8.2.9. Prove: for a graphic matroid $\mathcal{M}(G)$ and an edge $x \in E(G)$, the matroid $\mathcal{M}(G)/x$ is the same as the matroid $\mathcal{M}(G/x)$, where G/x is the graph obtained by contracting the edge x (hence the term).

   Exercise 8.2.10. Prove: if the matroid $\mathcal{M}$ is representable over the field F then so is $\mathcal{M}/x$ (for any $x \in X$).

8.3 Lattices

Definition: A *finite lattice* L is a partially ordered set with the property that any subset S has a *join* (or *least upper bound*), i. e., an element $b \in L$ such that

$(\forall a \in S)\ (a \leq b)$ and $(\forall c \in S)(((\forall a \in S)(a \leq c)) \Rightarrow (b \leq c))$.

It follows (why?) that every subset has a *meet* (or *greatest lower bound*), that is an element $b \in L$ such that

$(\forall a \in S)\ (b \leq a)$ and $(\forall c \in S)(((\forall a \in S)(c \leq a)) \Rightarrow (c \leq b))$.

The join of two elements a and b is denoted by $a \bigvee b$ and the meet by $a \bigwedge b$.

A finite lattice L always has a minimum element 0_L and a maximum element 1_L.

A *point* in L is an element that covers 0_L.

Definition (Geometric lattice) A lattice is called *geometric* if:

(α) L is *atomic* (or a *point lattice*), that is, each element of L is the join of points of L, and

(β) L is *semimodular*, that is, if $a \neq b \in L$ and both a and b cover c in L then $a \bigvee b$ covers both a and b.

Exercise 8.3.1. (Jordan–Dedekind property) Prove: if the lattice L is semimodular then it has the *Jordan-Dedekind* (JD) property: if $a < b$ $(a, b \in L)$ then any two maximal chains between a and b have the same length.

In particular,if L is semimodular then for $a \in L$ all maximal chains between 0_L and a have the same length; this length is called the *rank* of a. The *points* of a lattice are the elements of rank 1.

The relation between matroids and geometric lattices is given by the following exercise.

Exercise 8.3.2. (Equivalence of matroids and geometric lattices) The set of flats of a matroid, ordered by inclusion, is a geometric lattice. Conversely, given a geometric lattice L with point set X, then $(X, \{\mathcal{F}_y : y \in L\})$ is a matroid, where $\mathcal{F}_y = \{x \in X : x \leq y\}$ are the flats of the matroid.

Exercise 8.3.3. (Lattice of divisors) Let n be a positive integer. Consider the lattice $\mathcal{D}(n)$ of divisors of n. The points of the lattice are the positive integers dividing n, ordered by divisibility. Determine, for what values of n is this lattice geometric.

Bibliography

[1] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, 1998, Chapter 25 (READ!)

[2] M. Aigner, *Combinatorial Theory*, 1979.

[3] P. J. Cameron , *Combinatorics: Topics, Techniques, Algorithms*, 1996.

[4] L. Babai, P. Frankl, *Linear Algebra Methods in Combinatorics*, 1992, Chapter 9.3 (READ!)

Chapter 9

Linear Algebra and Applications to Graphs

9.1 Basic Linear Algebra

Exercise 9.1.1. Let V and W be linear subspaces of $\mathbb{F}^n$, where $\mathbb{F}$ is a field, $\dim V = k$, $\dim W = \ell$. Show that $\dim(V \cap W) \geq k + \ell - n$.

Exercise 9.1.2. Let A be an $n \times n$ matrix over the field $\mathbb{F}$ and $\mathbf{x} \in \mathbb{F} \setminus \{0\}$. Then $(\exists \mathbf{x})(A\mathbf{x} = 0) \Leftrightarrow \det(A) = 0$, where $\det(A)$ is the determinant of A.

Definition 9.1.3. Let A be an $n \times n$ matrix over the field $\mathbb{F}$ and $\mathbf{x} \in \mathbb{F}^n \setminus \{0\}$. We say that $\mathbf{x}$ is an **eigenvector** for A with **eigenvalue** λ if

$$A\mathbf{x} = \lambda\mathbf{x}.$$

Exercise 9.1.4. Show that if $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{F}^n$ are eigenvectors with distinct eigenvalues then they are linearly independent.

Definition 9.1.5. The **characteristic polynomial** of the $n \times n$ matrix A is

$$f_A(x) := \det(xI - A).$$

Exercise 9.1.6. λ is an eigenvalue of A if and only if it is a root of $f_A(x)$, i.e. $f_A(\lambda) = 0$.

Exercise 9.1.7. Let $f_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ (why is it monic?). Show that $a_0 = (-1)^n \det(A)$ and $a_{n-1} = -\operatorname{tr}(A)$, where the **trace** of A is defined as $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$ (sum of the diagonal elements).

Definition 9.1.8. If λ is an eigenvalue of A then the **geometric multiplicity** of λ is $\dim \ker(A - \lambda I)$ (the number of linearly independent eigenvectors for eigenvalue λ). The **algebraic multiplicity** of λ is its multiplicity as a root of $f_A(x)$.

CONVENTION. By the multiplicity of the eigenvalue (without adjective) we always mean the **algebraic multiplicity**.

Exercise 9.1.9. The algebraic multiplicity of λ is greater than or equal to its geometric multiplicity.

Exercise 9.1.10. If A is an $n \times n$ matrix then the algebraic multiplicity of the eigenvalue λ equals $\dim \ker(A - \lambda I)^n$.

Definition 9.1.11. The $n \times n$ matrices A and B are **similar**, $A \sim B$, if there exists an invertible matrix S s.t. $A = S^{-1}BS$.

Exercise 9.1.12. Show that if A and B are similar then $f_A(x) = f_B(x)$.

Definition 9.1.13. An **eigenbasis** for A is a basis of $\mathbb{F}^n$ consisting of eigenvectors of A.

Definition 9.1.14. A is **diagonalizable** if it is similar to a diagonal matrix.

Exercise 9.1.15. A is diagonalizable if and only if it has an eigenbasis.

Exercise 9.1.16. If A is an upper triangular matrix then its eigenvalues, with proper algebraic multiplicity, are its diagonal elements.

Exercise 9.1.17. Every matrix over $\mathbb{C}$ is similar to an upper triangular matrix. More generally, a matrix over the field $\mathbb{F}$ is similar to a triangular matrix if and only if all roots of f_A belong to $\mathbb{F}$.

Exercise 9.1.18. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the $n \times n$ matrix A (listed with their algebraic multiplicities). Then $\det(A) = \prod_i \lambda_i$ and $\mathrm{tr}(A) = \sum_i \lambda_i$.

Exercise 9.1.19. Show that $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is not similar to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Verify that their characteristic polynomials are identical. Show that the second matrix is not diagonalizable.

Exercise 9.1.20. Let $\mathbb{F} = \mathbb{C}$ (or any algebraically closed field). Show that A is diagonalizable if and only if each eigenvalue of A has its geometric multiplicity equal to its algebraic multiplicity.

Exercise 9.1.21. If A is a diagonal matrix, $A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$, and f is a polynomial then $f(A) = \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix}$.

Definition 9.1.22. $m_A(x)$, the **minimal polynomial** of A, is the monic polynomial of lowest degree such that $m_A(A) = 0$.

Exercise 9.1.23. Show that $m_A(x)$ exists and $\deg m_A \leq n^2$.

Exercise 9.1.24. Show that if $f \in \mathbb{F}[x]$ is a polynomial then $(f(A) = 0) \Leftrightarrow (m_A \mid f)$.

Theorem 9.1.25 (Cayley-Hamilton Theorem).

$$m_A | f_A \quad \textit{or, equivalently,} \quad f_A(A) = 0.$$

Consequently $\deg m_A \leq n$.

Exercise 9.1.26. A proof of the Cayley-Hamilton theorem over $\mathbb{C}$ is outlined in this series of exercises:

1. Prove Cayley-Hamilton for diagonal matrices.
2. Prove the theorem for diagonalizable matrices.
3. Show that if A_i is a sequence of matrices, $\lim_{i\to\infty} A_i = A$, and f_i is a sequence of polynomials of the same degree, and $\lim_{i\to\infty} f_i = f$ (coefficientwise convergence) then $\lim_{i\to\infty} f_i(A_i) = f(A)$. In other words, polynomials of matrices are continuous functions of the matrix entries and the coefficients of the polynomials.
4. Show that for any matrix A there exists a sequence of diagonalizable matrices A_i, such that $\lim_{i\to\infty} A_i = A$. In other words diagonalizable matrices form a dense subset of the set of all matrices.
   (Hint: prove it first for upper triangular matrices.)
5. Complete the proof of Cayley-Hamilton theorem over $\mathbb{C}$.

Exercise 9.1.27. Complete the proof of the Cayley-Hamilton Theorem (over any field) by observing that if an identity of (multivariate) polynomials holds over $\mathbb{Z}$ then it holds over any commutative ring with identity.

9.2 Euclidean Spaces, Gram–Schmidt orthogonalization

In this section the field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. If $z = a + bi \in \mathbb{C}$ ($i = \sqrt{-1}$), we will denote by $\overline{z}$ the complex conjugate $\overline{z} = a - bi$. Note that $z \in \mathbb{R}$ if and only if $\overline{z} = z$. If $A = (a_{ij})$ is a matrix then each entry of $\overline{A} = (\overline{a_{ij}})$ is the complex conjugate of the corresponding entry of A.

Let V be a vector space over $\mathbb{F}$, i. e., and abelian group under addition which permits multiplication by scalars (members of $\mathbb{F}$). Multiplication by scalars is an $\mathbb{F} \times V \to V$ function satisfying to the rules of associativity $((\forall a, b \in \mathbb{F})(\forall \mathbf{x} \in V)((ab)\mathbf{x} = a(b\mathbf{x})))$, both rules of distributivity ($(a+b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$ and $a(\mathbf{x}+\mathbf{y}) = a\mathbf{x} + a\mathbf{y}$) and the normalization rule $1\mathbf{x} = \mathbf{x}$.

Definition 9.2.1. A **Hermitian bilinear form** over V is a function $f : V \times V \to \mathbb{F}$ satisfying the indentities $f(\mathbf{x}, \mathbf{y} + \mathbf{z}) = f(\mathbf{x}, \mathbf{y}) + f(\mathbf{x}, \mathbf{z})$; $f(\mathbf{x}, a\mathbf{y}) = af(\mathbf{x}, \mathbf{y})$ and $f(\mathbf{x}, \mathbf{y}) = \overline{f(\mathbf{y}, \mathbf{x})}$. Consequently we also have $f(\mathbf{y} + \mathbf{z}, \mathbf{x}) = f(\mathbf{y}, \mathbf{x}) + f(\mathbf{z}, \mathbf{x})$ and $f(a\mathbf{y}, \mathbf{x}) = \overline{a} f(\mathbf{y}, \mathbf{x})$.

Definition 9.2.2. If f is a Hermitian bilinear form then the function $Q_f(\mathbf{x}) := f(\mathbf{x}, \mathbf{x})$ is called a Hermitian quadratic form.

Exercise 9.2.3. Prove that the values of a Hermitian quadratic form are always real.

Definition 9.2.4. A Hermitian bilinear form f and its corresponding quadratic form Q_f are called **positive semidefinite** if $f(\mathbf{x}, \mathbf{x}) \geq 0$ for all $\mathbf{x} \in V$; if in addition $f(\mathbf{x}, \mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$ then we call f and Q_f **positive definite.**

Definition 9.2.5. A **Eulidean space** is a pair (V, f) where V is vector space over $\mathbb{F}$ and f is a positive definite Hermitian form over V. We refer to f as the **inner product**.

Definition 9.2.6. In a Euclidean space (V, f), the **norm** of $\mathbf{x} \in V$ is defined as

$$\|\mathbf{x}\| := \sqrt{f(\mathbf{x}, \mathbf{x})}.$$

The distance of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is defined as

$$\operatorname{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

Exercise 9.2.7. Prove that the distance defined above is a *metric*, i. e., it satisfies the triangle inequality.

Definition 9.2.8. An **isometry** between two Euclidean spaces (V, f) and (W, g) over the same field $\mathbb{F}$ is a linear isomorphism $V \to W$ which preserves the norm. (V, f) and (W, g) are **isometric** if there exists an isometry between them.

Definition 9.2.9. Two vectors $\mathbf{x}, \mathbf{y}$ are **orthogonal** if their inner product is zero. Notation: $\mathbf{x} \perp \mathbf{y}$. A set of vectors is orthogonal if every pair among them is orthogonal.

Exercise 9.2.10. Prove: if a set of nonzero vectors is orthogonal then the vectors are linearly independent.

Definition 9.2.11. The **Gram matrix** of a sequence $\mathbf{x}_1, \ldots, \mathbf{x}_m$ of vectors is the $m \times m$ matrix

$$G = G(\mathbf{x}_1, \ldots, \mathbf{x}_m) = (f(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^n.$$

The **Gramian** is the *determinant* of the Gram matrix.

Exercise 9.2.12. Prove: the Gramian of a sequence of vectors is zero if and only if the vectors are linearly dependent.

Exercise 9.2.13. Prove: the Gramian of a sequence of vectors is always real and nonnegative. *Hint.* Use Exercise 9.3.11.

Our next goal is to turn any (finite or infinite) sequence $\mathbf{b}_1, \mathbf{b}_2, \ldots$ of vectors into an orthogonal sequence while preserving the chain of subspaces generated by the initial segments of the sequence of vectors and also preserving the Gramian of the initial segments.

Theorem 9.2.14. (Gram–Schmidt orthogonalization) *Let $\mathbf{b}_1, \mathbf{b}_2, \ldots$ be a (finite or infinite) sequence of vectors; for every i, let U_i be the span of $\mathbf{b}_1, \ldots, \mathbf{b}_i$. Then there exists a sequence of orthogonal vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots$ such that for all i, U_i is the span of of $\mathbf{e}_1, \ldots, \mathbf{e}_i$ and* $\det G(\mathbf{e}_1, \ldots, \mathbf{e}_i) = \det G(\mathbf{b}_1, \ldots, \mathbf{b}_i)$.

Proof. We construct the $\mathbf{e}_j$ inductively. To construct $\mathbf{e}_i$, assume we have the orthogonal sequence $\mathbf{e}_1, \ldots, \mathbf{e_{i-1}}$ which satisfies the Theorem with respect to the sequence $\mathbf{b}_1, \ldots, \mathbf{b_{i-1}}$. Guided by the requirement that the the span of U_{i-1} and $\mathbf{e}_i$ must agree with the span of U_{i-1} and $\mathbf{b}_i$, we try to find $\mathbf{e}_i$ in the form

$$\mathbf{e}_i = \mathbf{b}_i + \sum_{j=1}^{i-1} \alpha_{i,j} \mathbf{e}_j.$$

To determine the coefficients $\alpha_{i,j}$, we take the inner product of each side of the equation by vve_j from the left ($j < i$). The condition $f(\mathbf{e}_j, \mathbf{e}_i) = 0$ then becomes equivalent to

$$0 = f(\mathbf{e}_j, \mathbf{b}_i) + \alpha_{i,j} \|\mathbf{e}_j\|^2.$$

If $\mathbf{e}_j = 0$ then this condition is automatically satisfied; otherwise there is a unique $\alpha_{i,j}$ which satisfies it:

$$\alpha_{i,j} = -\frac{f(\mathbf{e}_j, \mathbf{b}_i)}{\|\mathbf{e}_j\|^2}.$$

Exercise 9.2.15. Prove that the sequence $\{\mathbf{e}_i\}$ constructed indeed satisfies the conclusions of the Theorem.

Exercise 9.2.16. Assuming the $\mathbf{b}_j$ are linearly independent, prove that the sequence $\{\mathbf{e}_i\}$ constructed is the unique solution, up to $\pm$ signs, satisfying the requirements stated in the Theorem.

Exercise 9.2.17. Prove that the following holds for any sequence $\{\mathbf{e}_j\}$ satisfying the conditions of the Theorem: $\mathbf{e}_i = 0$ if and only if $\mathbf{b}_i$ belongs to the span of $\{\mathbf{b}_j : j < i\}$. In particular, zero occurs among the $\mathbf{e}_j$ if and only if the $\mathbf{b}_j$ are linearly dependent.

Definition 9.2.18. A sequence of vectors is **orthonormal** if the vectors are orthogonal and have unit norm.

Exercise 9.2.19. A sequence of vectors is orthonormal if and only if their Gram matrix is the identity matrix.

Corollary 9.2.20. *Every finite dimensional Euclidean space has an* **orthonormal basis** *(ONB).*

Exercise 9.2.21. Prove: two finite-dimensional Euclidean spaces over the same filed are isometric if and only if they have the same dimension. *Hint.* Use ONBs to construct an isometry.

Exercise 9.2.22. (Shortest distance to a subspace) Let $U \leq V$ be a subspace and $\mathbf{x} \in V$ a vector. Show that there is a unique $\mathbf{y} \in U$ such that $\mathbf{y} \perp \mathbf{x} - \mathbf{y}$. Show that $\|\mathbf{x} - \mathbf{y}\| = \min_{\mathbf{z} \in U} \|\mathbf{x} - \mathbf{z}\|$. Give an algorithm to find $\mathbf{y}$. (We call $\mathbf{y}$ the component of $\mathbf{x}$ parallel to U; and $\mathbf{x} - \mathbf{y}$ the component of $\mathbf{x}$ perpendicular to U.)

Definition 9.2.23. Let V be an n-dimenisional real vector space and $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$. The **parallelepiped** spanned by the vectors $\mathbf{x}_i$ is the set $\{\sum a_i \mathbf{x}_i : 0 \leq a_i \leq 1\}$. The parallelepiped is non-degenerate if the $\mathbf{x}_i$ are linearly independent.

Definition 9.2.24. Let V be an n-dimensional real vector space and $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be a sequence of vectors in V. We define the k-dimensional volume of the parallelepiped spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_k$ inductively as follows:

(a) For $k = 1$, $\mathrm{vol}(\mathbf{x}_1) := \|\mathbf{x}_1\|$.

(b) For $k \geq 2$, $\mathrm{vol}(\mathbf{x}_1, \ldots, \mathbf{x}_k) := h_k \cdot \mathrm{vol}(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1})$, where h_k is the distance of $\mathbf{x}_k$ from the subspace spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$.

Exercise 9.2.25. Prove: $\mathrm{vol}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = 0$ if and only if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent.

Exercise$^+$ 9.2.26. Let $G = G(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ be the Gram matrix. Prove: $\det G = (\mathrm{vol}(\mathbf{x}_1, \ldots, \mathbf{x}_k)^2$.

Exercise 9.2.27. Let $V = \mathbb{R}^n$ and consider n vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Prove: $\mathrm{vol}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = |\det(\mathbf{x}_1, \ldots, \mathbf{x}_n)|$.

EXAMPLES of Euclidean spaces follow. The first example is $\mathbb{F}^n$, the set of n-tuples over $\mathbb{F}$, usually represented as column vectors, with their "standard" inner product:

Definition 9.2.28. Let $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$. Their **standard inner product** is

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{n} \overline{x}_i y_i.$$

Definition 9.2.29. If A is a matrix, then the **adjoint** matrix is $A^* = \overline{A}^T$ (conjugate-transpose).

Exercise 9.2.30. We think of vectors in $\mathbb{F}^n$ as column matrices. Verify the following:

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \mathbf{y}$;
2. $\|\mathbf{x}\| = \sqrt{\mathbf{x}^* \mathbf{x}}$.

Exercise 9.2.31. Orthogonalize the sequence of columns of this matrix:

$\begin{pmatrix} 3 & -1 & 2 & 0 \\ 1 & 2 & 3 & 1 \\ -1 & 0 & -1 & 5 \end{pmatrix}$. Your output is a sequence of 4 orthogonal vectors in $\mathbb{R}^3$. Naturally, one of them must be zero. Which one? Why? Switch the first two columns and orthogonalize again. Observe what changes and what does not change in the output.

Last update: October 12, 2004

The next example: ORTHOGONAL POLYNOMIALS

The field is $\mathbb{R}$. Let (a, b) be a finite or infinite interval and $w : (a, b) \to \mathbb{R}$ a nonnegative continuous function, not everywhere zero, such that for all n, $\int_a^b |x^n| w(x) < \infty$. We call w the *weight function*. We define the following inner product with respect to w over the space $\mathbb{R}[x]$ of real polynomials: for $p(x), q(x) \in \mathbb{R}[x]$, we set

$$\langle p, q \rangle := \int_a^b p(x)q(x)w(x)dx.$$

Exercise 9.2.32. Verify that this is a positive definite bilinear form.

Definition 9.2.33. The classical *Legendre polynomials* $P_n(x)$ are defined as follows:

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left[(x^2 - 1)^n \right].$$

Exercise$^+$ 9.2.34. Orthogonalize the infinite sequence $1, x, x^2, \ldots$ with respect to the function $w(x) = 1$ over the interval $(-1, 1)$. Describe the spaces U_i. Prove that the orthogonalized sequence is the sequence of Legendre polynomials $P_0, P_1, \ldots$.

"Orthogonal polynomials" are sequences of polynomials arising by orthogonalizing the sequence $1, x, x^2, \ldots$ under various weight functions. For more about the remarkable properties of orthogonal polynomials and their surprising connections to graph theory, see Chapter 12. polynomials

9.3 Normal matrices and the Spectral Theorem

Exercise 9.3.1. For an $n \times n$ matrix A, verify the following:

1. $\langle A^*\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle$
2. $((\forall \mathbf{x}, \mathbf{y})(\langle B\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle)) \Leftrightarrow B = A^*$

Exercise 9.3.2. $(AB)^* = B^*A^*$, where A, B are not necessarily square matrices. (What dimensions should they have so that we can multiply them?)

Definition 9.3.3. We say that a matrix A is **Hermitian** if $A^* = A$. A real Hermitian matrix is **symmetric**.

Exercise 9.3.4. If $A = A^*$ then all eigenvalues of A are real.

Exercise 9.3.5. Show that the characteristic polynomial of a Hermitian matrix has real coefficients.

Definition 9.3.6. The **Hermitian bilinear form** associated with a Hermitian matrix is the $\mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ function defined by

$$B_A(\mathbf{x}, \mathbf{y}) = \mathbf{x}^* A \mathbf{y} = \sum_{i,j} a_{ij} \overline{x}_i y_j.$$

quadratic form associated with a A is the function $Q_A(\mathbf{x}) : \mathbb{F}^n \to \mathbb{F}$ defined by

$$Q_A(\mathbf{x}) = \mathbf{x}^* A \mathbf{x} = \sum_{i,j} a_{ij} \overline{x}_i x_j.$$

Exercise 9.3.7. Prove that if A is Hermitian then B_A is indeed a Hermitian bilinear form in the sense of Definition 9.2.1.

Definition 9.3.8. A Hermitian matrix is **positive (semi)definite** if the corresponding quadratic form is positive (semi)definite.

Exercise 9.3.9. (a) Prove that every matrix of the form A^*A is Hermitian, positive semidefinite. A does not need to be a square matrix for this exercise. (b) Prove that A^*A is positive definite if and only if the columns of A are linearly independent.

Exercise 9.3.10. Prove that every positive semidefinite Hermitian matrix B can be written as $B = A^*A$. *Hint.* Use the Spectral Theorem. Prove that B in fact has a "square root," i. e., there exists a positive semidefinite Hermitian matrix A such that $B = A^2$.

Exercise 9.3.11. Let (V, F) be a Euclidean space and let $G = G(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ be the Gram matrix of a sequence of vectors. Prove: There exists A such that $G = A^*A$. Consequently G is positive semidefinite; and G is positive definite if and only if the $\mathbf{x}_i$ are linearly independent.

Definition 9.3.12. The **operator norm** of a matrix A is defined as

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

Exercise 9.3.13. Show that $\|A\| = \sqrt{\lambda_1(A^*A)}$, where $\lambda_1(A^*A)$ denotes the largest eigenvalue of A^*A. (Note that A^*A is Hermitian. A does not need to be a square matrix for this exercise.)

Definition 9.3.14. A Hermitian matrix A is called **positive semidefinite** if $(\forall \mathbf{x} \in \mathbb{F})(Q_A(\mathbf{x}) \geq 0)$. A is called **positive definite** if $(\forall \mathbf{x} \in \mathbb{F} \setminus \{0\})(Q_A(\mathbf{x}) > 0)$.

Exercise 9.3.15. Show that a Hermitian matrix is positive definite (resp. semidefinite) if and only if all its eigenvalues are positive (resp. nonnegative).

Exercise 9.3.16. Show that a Hermitian matrix A is positive definite if and only if all of its upper left corner determinants $\det(a_{11})$, $\det \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix} \right)$, etc, are positive.
Hint. Use the Interlacing Theorem, given in Exercise 9.3.39 below.

Definition 9.3.17. A is a **unitary** matrix if $A^*A = I$. A real unitary matrix is called an **orthogonal** matrix.

Exercise 9.3.18. Show that the following conditions on an $n \times n$ matrix are equivalent:

(a) $(\forall \mathbf{x}, \mathbf{y} \in \mathbb{F}^n)(\langle Ax, Ay\rangle = \langle x, y\rangle)$;

(a) $(\forall x \in \mathbb{F})(\|Ax\| = \|x\|)$ (i. e., A is an $\mathbb{F}^n \to \mathbb{F}^n$ isometry);

(b) A is unitary.

Exercise 9.3.19. Let A be an $n \times n$ matrix. Prove that the following are equivalent:

1. A is unitary;
2. $AA^* = I$;
3. the columns of A form an orthonormal basis of $\mathbb{F}^n$;
4. the rows of A form an orthonormal basis of $\mathbb{F}^n$.

Exercise 9.3.20. Show that if A is unitary and λ is an eigenvalue of A then $|\lambda| = 1$.

Exercise 9.3.21. Prove: the isometries of the real plane which fix the origin are the rotations (about the origin) and the reflections (in axes passing through the origin). Verify that the following are the matrices of the rotations and reflections with respect to a fixed orthonormal basis:

$R_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ and $T_\alpha = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. What is the angle of rotation for R_α? What is the position of the axis of reflection for T_α?

Exercise 9.3.22. Find the eigenvalues and orthonormal eigenbases for the 2×2 matrices given in the preceding exercise. Note that all matrices R_α share an orthogonal eigenbasis.

Exercise 9.3.23. Given a complex number λ with $|\lambda| = 1$, construct a 2×2 orthogonal matrix with eigenvalues λ and $\overline{\lambda}$.

Exercise 9.3.24. Prove: if A is an $n \times n$ orthogonal matrix and n is odd then one of the eigenvalues of A is ± 1. Prove that this statement is false in all even dimensions.

Exercise$^+$ 9.3.25. (a) Let $\mathbb{F} = \mathbb{C}$. Prove that to any matrix A there exists a unitary matrix S such that the matrix S^*AS is upper triangular (all entries below the diagonal are zero). Observe that the diagonal entries of S^*AS are the eigenvalues of A.

(b) Let $\mathbb{F} = \mathbb{R}$. Let A be a real matrix and suppose all eigenvalues of A are real. Prove that there exists an orthogonal matrix S such that the matrix S^*AS is upper triangular (all entries below the diagonal are zero). (Note that this statement is false if not all eigenvalues of A are real.)

Definition 9.3.26. The $n \times n$ matrix A is **normal** if $AA^* = A^*A$.

Exercise 9.3.27. Prove: if an upper trianguilar matrix T is normal then T is diagonal.

Exercise 9.3.28. Prove: if A is normal and S is unitary then S^*AS is normal.

Exercise$^+$ 9.3.29. (Characterization of complex normal matrices) Prove that the following three conditions on a complex $n \times n$ matrix are equivalent:

(a) A has an orthonormal eigenbasis;

(b) there exists a unitary matrix S such that S^*AS is a diagonal matrix; if so, the eigenvalues of A are the diagonal elements of S^*AS;

(c) A is normal.

Exercise$^+$ 9.3.30. (Characterization of real symmetric matrices) Prove that the following three conditions on a real $n \times n$ matrix are equivalent:

(a) A has an orthonormal eigenbasis in $\mathbb{R}^n$;

(b) there exists a orthogonal matrix S such that S^*AS is a diagonal matrix; if so, the eigenvalues of A are the diagonal elements of S^*AS;

(c) A is normal and has real eigenvalues;

(d) A is a real symmetric matrix.

Comments. In each of the two preceding problems, the equivalence of parts (a) and (b) is easy to prove, as is either of the implications (a) $\Rightarrow$ (c) and (b) $\Rightarrow$ (c). The converses of each of these last implications are fundamental results which include the Spectral Theorem (complex and real versions, resp.). *Hint.* For the (c) $\Rightarrow$ (b) implications, use the exercises leading up to the characterizations of the normal matrices. For the equivalence with part (d) in the last exercise, prove (b) $\Rightarrow$ (d) $\Rightarrow$ (c).

Exercise 9.3.31. Show that the $n \times n$ matrix A is

1. Hermitian if and only if A is normal and all its eigenvalues are real;
2. unitary if and only if A is normal and all its eigenvalues have unit absolute value.

NOTATION: For the rest of this section we use the following notation: $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$; A is a Hermitian matrix over $\mathbb{F}$. The eigenvalues of A (with multiplicity) are $\lambda_1 \geq \cdots \geq \lambda_n$.

Theorem 9.3.32 (Spectral Theorem). *The eigenvalues of A are real and A has an orthonormal eigenbasis.*

Note that the Spectral Theorem follows from Exercises 9.3.29 and 9.3.30.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be an orthonormal eigenbasis for A.

Exercise 9.3.33. Show that if $\mathbf{x} = \sum_i x_i \mathbf{e}_i$, then

$$Q_A(\mathbf{x}) = \sum_i \lambda_i |\mathbf{x}_i|^2.$$

Definition 9.3.34. The **Rayleigh quotient** is the function $R : \mathbb{F}^n \setminus \{0\} \to \mathbb{R}$ defined by

$$R(\mathbf{x}) = \frac{\mathbf{x}^* A \mathbf{x}}{\mathbf{x}^* \mathbf{x}} = \frac{Q_A(\mathbf{x})}{\|\mathbf{x}\|^2}.$$

Exercise 9.3.35. Show that

$$\lambda_1 = \max_{\|\mathbf{x}\|=1} R(\mathbf{x}).$$

Exercise 9.3.36. Show that

$$\lambda_2 = \max_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x} \perp \mathbf{e}_1}} R(\mathbf{x});$$

$$\lambda_3 = \max_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x} \perp \mathbf{e}_1, \mathbf{e}_2}} R(\mathbf{x});$$

and so on.

Exercise 9.3.37. Show that if $\lambda_1 = \mathbf{x}^* A \mathbf{x}$, $\|\mathbf{x}\| = 1$, then $A\mathbf{x} = \lambda_1 \mathbf{x}$.

Exercise 9.3.38 (Fischer-Courant Theorem).

$$\lambda_i = \max_{\substack{U \leq \mathbb{F}^n \\ \dim U = i}} \min_{x \in U} R(x)$$

where the maximum runs over all linear subspaces $U \leq \mathbb{F}^n$ of dimension i.

Exercise 9.3.39 (Interlacing Theorem). Let A be an $n \times n$ Hermitian matrix. We can construct a new $(n-1) \times (n-1)$ matrix by removing the ith row and the ith column of A. The resulting matrix B is Hermitian. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A and $\mu_1 \geq \cdots \geq \mu_{n-1}$ be the eigenvalues of B (with multiplicity). Show that $\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \cdots \geq \mu_{n-1} \geq \lambda_n$.

9.4 Applications to Graph Theory

There are two important square matrices commonly associated to graphs – the adjacency matrix of the graph, and the (finite or combinatorial) Laplacian. This allows us to apply the theory of eigenvalues to graphs, and it turns out that a great deal of information about the graph is carried in the spectra of these matrices.

For graph theory terminology please refer to "Graph Theory Terminology" handout.

9.4.1 The Adjacency Matrix

Definition 9.4.1. Let $G = (V, E)$ be a graph; assume $V = [n] = \{1, 2, \ldots, n\}$. The **adjacency matrix** $A_G = (a_{ij})$ of G is the $n \times n$ $(0, 1)$-matrix defined by $a_{ij} = 1$ if $\{i, j\} \in E$ (vertices i and j are adjacent); and $a_{ij} = 0$ otherwise. Note that $a_{ii} = 0$.

Exercise 9.4.2. Show that the (i, j) entry of $(A_G)^k$ gives the number of walks of length k between vertex i and vertex j. Give an interpretation for the (i, i) entry of $(A_G)^k$ and for $\sum_{j=1}^{n} (A_G)_{ij}$.

The adjacency matrix acts on functions on the graph. That is, if $f : V \to \mathbb{R}$ is a function on the vertices of the graph (which can also be considered a column matrix), then

$$Af(i) = \sum_{\substack{j \\ \{i,j\} \in E}} f(j).$$

Notice that this action is just matrix multiplication.

Exercise 9.4.3. Isomorphic graphs have similar adjacency matrices.

This allows us to make the following definitions:

Definition 9.4.4. κ is an **eigenvalue** of G if it is an eigenvalue of A_G. The **characteristic polynomial** of G is the characteristic polynomial of A_G. The **spectrum** of G is the ordered set of all eigenvalues of A_G (with multiplicities).

As before we will assume that eigenvalues of G are always ordered $\kappa_1 \geq \cdots \geq \kappa_n$.

Exercise 9.4.5. Compute the spectrum of each of the following graphs: K_n (the complete graph on n vertices), the star on n vertices (a tree with a vertex of degree $n - 1$, denoted $K_{n-1,1}$), $K_{k,\ell}$ (the complete bipartite graph).

Exercise 9.4.6. Let $G = (V, E)$ be a graph. Let G_i be the graph obtained by deleting the ith vertex (and the edges incident with it) from G. Show that eigenvalues of G and G_i interlace.

Exercise 9.4.7. $(\forall i)$ $(|\kappa_i| \leq \kappa_1)$.

Exercise 9.4.8. If G is connected then $\kappa_1 > \kappa_2$.

Exercise 9.4.9. If G is bipartite , then $\kappa_{n-i} = -\kappa_{i+1}$.

Exercise 9.4.10. If G is connected and $\kappa_1 = -\kappa_n$ than G is bipartite. Thus if G is connected and not bipartite then $(\forall i > 1)$ $(|\kappa_i| < \kappa_1)$.

Exercise 9.4.11. $\kappa_1 \leq \max_i \deg_G(i)$.

Exercise 9.4.12. $\kappa_1 \geq \frac{2|E|}{n} = \frac{1}{n}\sum_i \deg_G(i)$ (average degree).

Exercise 9.4.13. If G is k-regular, i.e. $(\forall i)(\deg_G(i) = k)$, then $\kappa_1 = k$.

Exercise 9.4.14. If $\kappa_1 = \max_i \deg_G(i)$ and G is connected then G is regular.

Exercise 9.4.15. If $\kappa_1 = \frac{1}{n}\sum_i \deg_G(i)$ then G is regular.

Exercise 9.4.16.
1. Upper bounds on the maximal eigenvalue are hereditary; that is, if $H \subset G$ is a subgraph, then $\kappa_1(H) \leq \kappa_1(G)$.
2. Show that upper bounds on the second eigenvalue κ_2 fail to be hereditary in general, but are hereditary in the special case that H is an *induced* subgraph.
(Hint: for the first part, consider the spectrum of K_n. For the second part, recall the Interlacing Theorem.)

Exercise 9.4.17. If $\mathrm{diam}(G) = d$, then the number of distinct eigenvalues of A_G is at least $d + 1$.
(Hint: Prove that under the diameter hypothesis, $I, A, \ldots, A^d$ are linearly independent. To show this, recall the significance of the (i, j) entry of A^k from Exercise 9.4.2.)

9.4.2 The Laplacian and Expansion of a Graph

Definition 9.4.18. We define the **Laplacian** of the graph G to be

$$\Delta_G = D_G - A_G$$

where A_G is the adjacency matrix and D_G is a diagonal matrix, $D_G(i, i) = \deg_G(i)$.

Exercise 9.4.19. Verify that for $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$,

$$\mathbf{x}^* \Delta_G \mathbf{x} = \sum_{\{i,j\} \in E} |x_i - x_j|^2.$$

Exercise 9.4.20. Show that Δ_G is positive semidefinite.

However, Δ_G is *not* positive definite:

Exercise 9.4.21. Check that $\Delta_G \mathbf{j} = 0$, where $\mathbf{j} = (1, \ldots, 1)^T$.

Exercise 9.4.22. Show that if G is connected, then 0 is a simple eigenvalue.

Exercise 9.4.23. Prove that the multiplicity of 0 as an eigenvalue of Δ_G is equal to the number of connected components of G.

Therefore if $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are eigenvalues of Δ_G then $\lambda_2 = 0$ if and only if G is disconnected.

Definition 9.4.24 (Fiedler). λ_2 is the **algebraic connectivity** of G.

Exercise 9.4.25. If G is a k-regular graph (every vertex has degree k) and $k = \kappa_1 \geq \cdots \geq \kappa_n$ are eigenvalues of A_G and $0 = \lambda_1 \leq \cdots \leq \lambda_n$ are eigenvalues of Δ_G the $\lambda_i + \kappa_i = k$. In particular, $\lambda_2 = \kappa_1 - \kappa_2$. λ_2 is also referred to as the **eigenvalue gap** or **spectral gap**.

Definition 9.4.26. If $A \subseteq G$ we denote by $\delta(A)$ the number of edges between A and $\overline{A} = V \setminus A$. The **isoperimetric ratio** for A is $\frac{\delta(A)}{|A|}$. The **isoperimetric constant** of G is

$$i_G = \min_{\substack{A \neq \emptyset \\ |A| \leq \frac{n}{2}}} \frac{\delta(A)}{|A|}.$$

The next result shows the important fact that if λ_2 is large then G is "highly expanding".

Exercise* 9.4.27. $\lambda_2 \leq 2\ \frac{\delta(A)}{|A|}$.

Later we will state a companion result which shows that in some sense if λ_2 is small then G has a small isoperimetric constant.

9.4.3 More basic properties of the eigenvalues of graphs

Recall the notation that, for a graph G, the adjacency matrix is denoted A_G and has eigenvalues $\kappa_1 \geq \ldots \geq \kappa_n$, while the Laplacian is denoted Δ_G and has eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$.

Exercise 9.4.28. Prove: $\kappa_1 + \cdots + \kappa_n = 0$.

Exercise 9.4.29. Prove: $\lambda_1 + \cdots + \lambda_n = 2m$ where $m = |E|$ is the number of edges.

Exercise 9.4.30. Prove: $\kappa_1^2 + \cdots + \kappa_n^2 = 2m$.

Exercise 9.4.31. Prove: $\kappa_1^3 + \cdots + \kappa_n^3 = 6t$ where t is the number of triangles in G.

Exercise 9.4.32. Prove: $(\forall s \geq 0)(\kappa_1^s + \cdots + \kappa_n^s$ is an integer.) The same holds for the λ_i.

Exercise 9.4.33. Prove: the following numbers cannot occur as eigenvalues of a graph: $\sqrt{-1},\ \pi,\ 3/5,\ \sqrt{3/2},\ 2^{1/3}$. They cannot occur as eigenvalues of the Laplacian either.

Exercise$^+$ 9.4.34. Prove: if the characteristic polynomial of a graph G is irreducible (over $\mathbb{Q}$) then G has no nontrivial automorphisms. (An *automorphism* is a self-isomorphism, i. e., a permutation of the vertices which preserves adjacency.)

OPEN PROBLEM. The characteristic polynomial of almost every graph is irreducible (over $\mathbb{Q}$). ("Almost every graph" means that if we create a *random graph* on a given set of n vertices by flipping $\binom{n}{2}$ coins to decide adjacency then the probability of the event in question is $1-o(1)$.)

9.4.4 Eigenvalues and chromatic number

The chromatic number of a graph is the smallest number of "colors" needed for an assignment of colors to the vertices of G such that no pair of adjacent vertices receives the same color. This is one of the most important graph invariants. Here we consider the connections of chromatic number and the spectral theory of graphs.

Recall the notation that $[k] = \{1, 2, \ldots, k\}$.

Definition 9.4.35. A **legal k-coloring** of a graph G is a map $c : V \to [k]$ such that $\{i, j\} \in E \Rightarrow c(i) \neq c(j)$. The **chromatic number** of G, denoted $\chi(G)$, is the smallest value of k for which a legal k-coloring exists.

Exercise 9.4.36. Compute the chromatic number k of the Petersen graph and present a legal k-coloring.

Exercise 9.4.37. Show that $\chi(G) \leq \deg_{\max} +1$.

The famous four-color theorem asserts that the chromatic number of a planar graph is at most four. But the weaker result that planar graphs are all six-colorable can be derived in a very elementary way, which we describe here.

Exercise 9.4.38. We will show, in several parts, that $\chi(G) \leq 6$ for planar graphs.

1. Show: If G is planar, then $|E| \leq 3n - 6$, where $n = |V|$.
   (Hint: recall *Euler's formula:* if a connected graph is embedded in the plane, then $|V| - |E| + |F| = 2$, where F is the set of faces, or regions; the outside counts as one of the regions.)
2. Show: If G is planar, then $\deg_{\min} \leq 5$.
3. Conclude that $\chi(G) \leq 6$.
   (Hint: use induction, setting aside a vertex of smallest degree at each stage.)

Eigenvalue bounds on the chromatic number are given below.

Theorem 9.4.39 (H. Wilf). $\chi(G) \leq 1 + \kappa_1$.

This is an improvement over the result that $\chi \leq 1 + \deg_{\max}$ (Exercise 9.4.37) because $\kappa_1 \leq \deg_{\max}$ (Exercise 9.4.11).

Theorem 9.4.40 (Hoffman). $\chi(G) \geq 1 + \dfrac{\kappa_1}{-\kappa_n}$.

This is our first lower bound on chromatic number. While an upper bound on the chromatic number requires presenting a coloring, in order to prove a lower bound, we need to show that *all attempted colorings* with fewer colors fail. So the question of lower bounds is more profound and accordingly leads to deeper results.

Exercise 9.4.41. We prove Wilf's theorem, in steps.

1. For a graph G, let G_v denote the graph obtained by deleting vertex v; that is, the induced subgraph on $V \setminus \{v\}$. Show that $\deg v < \chi(G) - 1 \Rightarrow \chi(G_v) = \chi(G)$.
2. Conclude that G has an induced subgraph H with $\chi(G) = \chi(H)$ and $\deg_{\min}(H) \geq \chi(G) - 1$.
3. Use this to finish the proof of the theorem.
   (Hint: recall Exercise 9.4.12.)

Exercise$^+$ 9.4.42 (Biggs). We show Hoffman's theorem, in steps.

1. If a real symmetric matrix A is in block form

$$A = \left[\begin{array}{c|c} P & Q \\ \hline Q^t & R \end{array}\right],$$

   where P, Q, R are $n \times n$ matrices, then

$$\lambda_{\max}(A) + \lambda_{\min}(A) \leq \lambda_{\max}(P) + \lambda_{\max}(R).$$

   (Hint: a somewhat delicate application of Rayleigh quotients, see Def. 9.3.34.)

2. Show by induction: if A is a real symmetric matrix in block form with t^2 submatrices A_{ij} $(1 \leq i, j \leq t)$ such that the diagonal submatrices A_{ii} are square, then

$$\lambda_{\max}(A) + (t-1)\lambda_{\min}(A) \leq \sum_{i=1}^{t} \lambda_{\max}(A_{ii}).$$

3. Deduce Hoffman's theorem.
(Hint: consider a partition of the vertices by color to apply the previous part of the exercise. Then observe that $\lambda_{\min} < 0$ to complete the proof.)

Next, we consider a special class of graphs for which the spectral gap is as big as possible.

Definition 9.4.43. A **Ramanujan graph** G is a regular graph of degree r such that $(\forall i \geq 2)(|\kappa_i| \leq \sqrt{2r-1})$.

Note that $\kappa_1 = \deg G = r$, so there is a large gap between κ_1 and κ_2.

It follows, by Hoffman's theorem, that $\chi(G) \geq 1 + \frac{r}{\sqrt{2r-1}} = \Omega(\sqrt{r})$.

In fact, this bound on the eigenvalues is asymptotically tight; that is, the $\sqrt{2r-1}$ bound in the definition of Ramanujan graphs cannot be replaced with any smaller value. This fact is quite difficult to show, and can be found in the work of Lubotzky-Phillips-Sarnak.

While it is hard to get a good upper bound on κ_2, we can obtain lower bounds with less work. The next exercise provides one such bound.

Exercise 9.4.44. We will show that, for all r-regular graphs on n vertices with sufficiently large n, the second eigenvalue is at least $\sqrt{r}$.

1. Show that for r-regular graphs G, we have

$$\operatorname{diam}(G) = 3 \quad \Rightarrow \quad n \leq r^3 - r^2 + r + 1.$$

2. Conclude that for such graphs with $r \geq 2$, $n \geq r^3 \Rightarrow \operatorname{diam}(G) \geq 4$.

3. Show that, if n is sufficiently large relative to r then we have $\kappa_2(G) \geq \sqrt{r}$.
(Hint: use the diameter information to find induced subgraphs which are disjoint, and reason from there using the results of Exercise 9.4.16.)

Chapter 10

Hadamard Matrices

10.1 Introduction

Notation. $[n]$ denotes the set $\{1, 2, \ldots, n\}$

Exercise 10.1.1. Let $4 \mid n$ and let $A_1, \ldots, A_m \subset [n]$ such that for all i, $|A_i| = n/2$ and for all $i \neq j$, $|A_i \cap A_j| = n/4$. Prove: $m \leq n - 1$. *Hint.* Use linear algebra.

Exercise 10.1.2. Prove that the inequality $m \leq n - 1$ in the preceding exercise is tight, i. e., for infinitely many values of n, set systems as described in the preceding exercise exist with $m = n - 1$. *Hint.* Does this problem belong in these notes?

Definition 10.1.3. A (± 1)-matrix is a matrix whose entries are 1 and -1.

An $n \times n$ (± 1)-matrix is called an **Hadamard matrix** if the rows are orthogonal.

Remark. In Hadamard's name, the "H" and the final "d" are silent.

Exercise 10.1.4. Prove that an $n \times n$ (± 1)-matrix H is Hadamard $\Leftrightarrow H \cdot H^t = nI_n$, where I_n denotes the $n \times n$ identity matrix.

Definition 10.1.5. An $n \times n$ real matrix is *orthogonal* if $AA^t = I_n$.

Exercise 10.1.6. A real $n \times n$ matrix A is orthogonal $\Leftrightarrow (\forall x \in \mathbb{R}^n)(\|Ax\| = \|x\|)$, where $\|x\| = \sqrt{x \cdot x^t}$ denotes the Euclidean norm.

Exercise 10.1.7. Prove: if H is an $n \times n$ Hadamard matrix then $\frac{1}{\sqrt{n}} H$ is an orthogonal matrix.

Exercise 10.1.8. If H is an $n \times n$ Hadamard matrix then $(\forall x \in \mathbb{R}^n)(\|Hx\| = \sqrt{n}\|x\|)$.

Exercise 10.1.9. Prove that the columns of an Hadamard matrix are also orthogonal, i. e., $H^t \cdot H = nI_n$.

Exercise 10.1.10. Prove: all (complex) eigenvalues of an $n \times n$ Hadamard matrix have absolute value $\sqrt{n}$.

Exercise 10.1.11. Prove: if H is an $n \times n$ Hadamard matrix then $\det(H) = \pm n^{n/2}$.

Exercise 10.1.12. Prove: if A is an $n \times n$ (± 1)-matrix then $|\det(A)| \leq n^{n/2}$. Equality holds if and only if A is an Hadamard matrix. *Hint.* Prove Hadamard's Inequality: if A is an $n \times n$ real matrix then $|\det(A)| \leq N_1 \cdot \cdots \cdot N_n$ where N_i is the Euclidean norm of the i^{th} row of A. Equality holds exactly when either a row is zero or the rows are orthogonal. Use the geometric meaning of the determinant (volume of the paralellopiped spanned by the rows).

Example 10.1.13. $S_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$,

$$S_{k+1} = \begin{bmatrix} S_k & S_k \\ S_k & -S_k \end{bmatrix} \quad (k \geq 1).$$

The matrix S_k is called the $2^k \times 2^k$ Sylvester matrix.

Exercise 10.1.14. Prove that S_k is an Hadamard matrix.

Exercise 10.1.15. Let $a_{v,w} = (-1)^{v \cdot w}$, where $v, w \in \mathbb{F}_2^k$. Prove that the $2^k \times 2^k$ matrix $(a_{v,w})$ is S_k (after suitable renumbering of the rows and columns).

Definition 10.1.16. The group $\mathbb{Z}_2^n = \mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2$ is called an **elementary Abelian 2-group**.

Remark. This group is the additive group of the n-dimensional vector space over $\mathbb{F}_2$. ($\mathbb{F}_2$ is the field of two elements.) $\mathbb{Z}_2^n$ is also the additive group of the field $\mathbb{F}_{2^n}$.

Exercise 10.1.17. Prove that the Sylvester matrix S_k is the character table of $\mathbb{Z}_2^k$.

Exercise 10.1.18. Let p be a prime number, $p \equiv -1 \pmod 4$. Construct a

$(p+1) \times (p+1)$ Hadamard matrix using the quadratic character of the field $\mathbb{F}_p$.

Hint. Consider the $p \times p$ matrix (a_{ij}) where $a_{ij} = \left(\frac{i+j}{p}\right)$ (Legendre symbol). Modify this matrix by adding a row and column and suitably changing the zeros to ± 1.

Exercise 10.1.19. Prove: if $\exists n \times n$ Hadamard matrix, then $n = 2$ or $4 \,|\, n$.

Exercise 10.1.20. Prove: if $\exists k \times k$ Hadamard matrix and $\exists l \times l$ Hadamard matrix, then $\exists kl \times kl$ Hadamard matrix.
Hint. Kronecker product.

Comment. The Sylvester matrices are Kronecker powers of S_1: $S_k = S_1 \otimes \cdots \otimes S_1$.

Conjecture 10.1.21. *If $4 \,|\, n$, then $\exists$ an $n \times n$ Hadamard matrix.*

Comment. Let $\mathcal{H}=\{n \mid \exists n \times n$ Hadamard matrix $\}$ and let $h_n = |\mathcal{H} \cap [n]|$. If the conjecture is true, then $h_n = \Omega(n)$. But even this weak consequence of the conjecture remains unsolved.

OPEN PROBLEM 10.1.22. *Prove that* $h_n \neq o(n)$.

Exercise 10.1.23. Prove that $h_n = \Omega\left(\frac{n}{\log(n)}\right)$.

10.2 Discrepancy and Ramsey Theory for (± 1)-Matrices

Lemma 10.2.1. (Lindsey's Lemma) *Let $H = (h_{ij})$ be a Hadamard matrix. Let $S, T \subseteq [n]$ and $s = |S|$, $t = |T|$. Then*

$$\left| \sum_{i \in S} \sum_{j \in T} h_{ij} \right| \leq \sqrt{stn}.$$

Definition 10.2.2. We call the submatrix on the entries corresponding to $S \times T$ an $s \times t$ **rectangle** in H. We call the sum $\left|\sum_{i \in S} \sum_{j \in T} h_{ij}\right|$ the **discrepancy** of this rectangle.

Discrepancy measures the deviation from uniform distribution.

Exercise 10.2.3. Prove Lindsey's Lemma.

Hint. Let $v_S \in \{0,1\}^n$ denote the incidence vector of $S \subseteq [n]$, i. e., the $(0,1)$-vector indicating membership in S. Observe that

$$\left| \sum_{i \in S} \sum_{j \in T} h_{ij} \right| = v_S H v_T^t.$$

Now use Exercise 10.1.8 and the Cauchy-Schwarz inequality:

$$(\forall a, b \in \mathbb{R}^n)(|a \cdot b^t| \leq \|a\| \cdot \|b\|).$$

Definition 10.2.4. A rectangle is **homogeneous** if all of its entries are equal.

Exercise 10.2.5. If H is an $n \times n$ Hadamard matrix, then H has no homogeneous rectangles of area $(= st)$ greater than n.

Exercise 10.2.6. (**Erdős**)
Prove: For all sufficiently large n, $\exists (n \times n)$ (± 1) matrices without homogeneous $t \times t$ rectangles such that $t \geq 1 + 2\log_2 n$.

Hint. Use the Probabilistic Method. Flip a coin for each entry. Show that the probability that a random matrix is "bad" is less than 1. In fact it will be $o(1)$ (almost all matrices are "good").

Exercise 10.2.7. Construct an explicit family of $(n \times n)$ (± 1) matrices A_n (for infinitely many values of n) such that A_n has no homogeneous $t \times t$ rectangles for $t > \sqrt{n}$.

OPEN PROBLEM 10.2.8. *Construct an explicit family of $(n \times n)$ (± 1) matrices A_n (for infinitely many values of n) such that A_n has no homogeneous $t \times t$ rectangles for $t > n^{0.49}$.*

10.3 Gale–Berlekamp Switching Game

Let $A = (a_{i,j})$ be a matrix with entries ± 1. The first player sets the initial entries of A. Subsequently the second player may switch any row or column (multiply the row or column by -1) and repeat this operation any number of times. The second player's "score" is the quantity $|\sum_{i,j \in [n]} a_{i,j}|$ which the second player wishes to maximize. The second player's gain is the first player's loss (zero-sum game), so the first player's goal is to keep the second player's score low.

Let $m(n)$ denote the score an optimal Player 2 can achieve against an optimal Player 1.

Exercise 10.3.1. Prove that $m(n) = \Theta(n^{3/2})$.

Hint 1. $m(n) = O(n^{3/2})$ requires Player 1 to be clever. Use an Hadamard matrix and Lindsey's Lemma (Lemma 10.2.1). Warning: an $n \times n$ Hadamard matrix may not exist (but a slightly larger one will be just as good).

Hint 2. $m(n) = \Omega(n^{3/2})$). Player 2 needs a good strategy.

Let Player 2 flip a coin for each row to decide whether or not to switch that row. Subsequently, Player 2 should switch those columns whose sum is negative. Use the Central Limit Theorem for the analysis.

Chapter 11

Character sums, Weil's Estimates and Paradoxical Tournaments

11.1 Characters of finite fields

Definition 11.1.1. A *character* of a finite field F is a function $\chi : F \to \mathbb{C}$, satisfying the following conditions:

1. $\chi(0) = 0$
2. $\chi(1) = 1$
3. $(\forall a, b \in F)(\chi(ab) = \chi(a)\chi(b))$.

Note that a character is a homomorphism from the multiplicative group $F^\times = F \setminus \{0\}$ to the multiplicative group $\mathbb{C}^\times$.

Example 11.1.2. For any field F, we define the *principal character*, χ_0, by $\chi_0(0) = 0$ and $(\forall a \neq 0)(\chi_0(a) = 1)$.

Notation. For a prime power $q = p^k$, $\mathbb{F}_q$ denotes the field of order q (i. e., the field $\mathbb{F}_q$ has q elements). For $k = 1$, the field $\mathbb{F}_p$ is the field of mod p residue classes. Note that for $k \geq 2$, the mod p^k residue classes do *not* form a field, so for $k \geq 2$, the field $\mathbb{F}_q$ is not the same as the ring of residue classes mod q. It is known, however, that for every prime power q there exists a field $\mathbb{F}_q$ and this field is unique up to isomorphism. If you are not familiar with finite fields, you may still read this note, always replacing q by p.

Example 11.1.3. When $F = \mathbb{F}_p$ for an odd prime p, we define the *quadratic character* $\chi(a) := \left(\frac{a}{p}\right)$, where $\left(\frac{a}{p}\right)$ is 0 when $a = 0$, 1 when a is a quadratic residue, and -1 when a is a quadratic nonresidue. $\left(\frac{a}{p}\right)$ is called the *Legendre symbol.*

Exercise 11.1.4. Show that, for all a, $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$.

Next, we extend the concept of the **quadratic character** to all finite fields of odd order.

Example 11.1.5. Let $\mathbb{F}_q$ be a finite field of odd order q. The *quadratic character* χ of $\mathbb{F}_q$ is defined as follows: for $a \in \mathbb{F}_q$,

$$\chi(a) = \begin{cases} 1 & \text{if } (\exists b \in \mathbb{F}_q)(a = b^2 \neq 0); \\ -1 & \text{if } (\forall b \in \mathbb{F}_q)(a \neq b^2); \\ 0 & \text{if } a = 0. \end{cases}$$

Exercise 11.1.6. Let q be an odd prime power and χ the quadratic character of $\mathbb{F}_q$. Prove: if $q \equiv -1 \pmod{4}$ then $\chi(-1) = -1$; and if $q \equiv 1 \pmod{4}$ then $\chi(-1) = 1$.

Exercise 11.1.7. For any prime power q, prove: $(\forall a \in \mathbb{F}_q)(a^{q-1} = 1)$.

(Note that for $q = p$ a prime, this is Fermat's Little Theorem.) *Hint.* Use Lagrange's theorem from group theory (the order of a subgroup divides the order of the group).

The *order* of a nonzero element $a \in \mathbb{F}_q$ is the smallest positive k such that $a^k = 1$. It follows from the preceding exercise that $k \mid q-1$ ("k divides $q-1$).

Corollary 11.1.8. $(\forall a \in \mathbb{F}_q)(\chi(a)$ *is a complex root of unity*$)$.

Indeed, if $a^k = 1$ then $(\chi(a))^k = \chi(a^k) = \chi(1) = 1$.

Definition 11.1.9. The *order* of a character is the least positive integer s such that $\chi(a)^s = 1$ for all $a \in F$, $a \neq 0$.

Note that, for any character of $\mathbb{F}_q$, the order s must divide $q-1$.

The following is a basic fact about the structure of finite fields.

Theorem 11.1.10. *For any prime power q, the multiplicative group $\mathbb{F}_q^\times$ is cyclic. Equivalently, there exists some $g \in \mathbb{F}_q^\times$ such that $\mathbb{F}_q^\times = \{g, g^2, \ldots, g^{q-1} = 1\}$.*

Such an element g is called a *generator* of $\mathbb{F}_q^\times$, or a *primitive root* of the field $\mathbb{F}_q$.

Exercise 11.1.11. Prove the Theorem. *Hint.* Use Sylow's Theorem from group theory and the fact that a polynomial of degree n has at most n roots in a field.

Corollary 11.1.12. *If χ is a character of $\mathbb{F}_q$ of order s, and g is a primitive root of $\mathbb{F}_q$, then $\chi(g)$ is a primitive s^{th} root of unity. Conversely, for any $\omega \in \mathbb{C}$ such that $\omega^{d-1} = 1$, there exists a unique character χ of $\mathbb{F}_q$ with $\chi(g) = \omega$.*

Exercise 11.1.13. Prove the Corollary.

Note that if we take $\omega = 1$ we get the principal character, and, for q odd, if we take $\omega = -1$, we get the quadratic character.

11.2 Character Sums: Weil's Theorem

In this section we describe one of the most beautiful results of 20th century mathematics.

First we consider the sum of characters over all elements of a field.

Exercise 11.2.1. If $\chi \neq \chi_0$, then $\sum_{a \in \mathbb{F}_q} \chi(a) = 0$.

Let now f be a polynomial of degree d over $\mathbb{F}_q$. We wish to estimate the sum

$$S(\chi, f) = \sum_{a \in \mathbb{F}_q} \chi(f(a))$$

Clearly, since $|\chi(f(a))|$ is 0 or 1 for all a, we have $|S(\chi, f)| \leq q$. This is the best possible upper bound; for example, if f is identically 1 then $S(\chi, f) = q$; if χ is the quadratic character and $f(x) = x^2$, then $S(\chi, f) = q - 1$.

Amazingly, once the trivial exceptions have been eliminated, a much stronger bound holds on the magnitude of $S(\chi, f)$: the values of the character tend to cancel each other out roughly by the same amount as if they were chosen to be ± 1 by coin flips.

Theorem 11.2.2 (André Weil). *Let $\mathbb{F}_q$ be a finite field, and let χ be a character of $\mathbb{F}_q$ of order s. Let $f(x)$ be a polynomial of degree d over $\mathbb{F}_q$ such that $f(x)$ cannot be written in the form $c\,(h(x))^s$, where $c \in \mathbb{F}_q$. Then*

$$\left| \sum_{a \in \mathbb{F}_q} \chi(f(a)) \right| \leq (d-1)\sqrt{q}.$$

Thus, in a sense, the values of a character over the range of a polynomial behave as "random" values, even though they are fully "deterministic." This feature is the key to a large number of applications to combinatorics and the theory of computing where the goal is "derandomization": the elimination of random choice from the proof of existence of a combinatorial object, i. e., replacing a probabilistic proof of existence by an explicit construction.

11.3 Paradoxical tournaments: proof of existence

Let $X = (V, E)$ be a digraph. Let $x \in V$ and $A \subseteq V$. We say that x **dominates** A if $(\forall a \in A)((x, a) \in E)$. We write $x \to A$ to denote this statement.

Definition 11.3.1. A digraph $X = (V, E)$ is **k-paradoxical** if $(\forall A \subset V)(|A| = k \Rightarrow \exists x \in V)(x \to A)$.

Definition 11.3.2. A *tournament* is a digraph $T = (V, E)$ in which for every pair $\{x, y\}$ of vertices, exactly one of the following holds: $x = y$ or $(x, y) \in E$ or $(y, x) \in E$.

Note that this concept corresponds to diagrams of round-robin tournaments without draws and without rematches. An edge (arrow) from a to b indicates that player a beat player b.

In a 1-paradoxical tournament, every player is beaten by someone. In a 2-paradoxical tournament, every pair of players is beaten by someone. Even 2-paradoxical tournaments are not straightforward to construct.

Exercise 11.3.3. Construct a 2-paradoxical tournament on 7 players. *Hint.* Make your diagram have a symmetry of order 7.

So it is quite surprising that k paradoxical tournaments actually do exist for every k. Constructing such tournaments even for $k = 3$ is quite hard. However, Paul Erdős, in one of the gems of his Probabilistic Method, demonstrated the *existence* of such tournaments without telling us how to construct them.

Theorem 11.3.4 (Erdős). *If $n > ck^2 2^k$ then there exists a k-paradoxical tournament on n vertices. (c is an absolute constant.)*

What Erdős has shown is not just that such tournaments *exist,* but they *abound:* almost every tournament on a given set of n vertices (players) is k-paradoxical. The model of "random tournaments" is very simple: flip a coin to decide the outcome of each match.

Exercise 11.3.5. Let $A \subset V$ be a set of k players (out of the set V of n players) and let x be a player, not in A. Calculate the probability that $x \to A$.

Exercise 11.3.6. Let A be as before. Show that the probability that none of the remaining $n - k$ players dominates A is exactly $(1 - 2^{-k})^{n-k}$.

Exercise 11.3.7. Infer from the preceding exercise that the probability that our random tournament is not k-paradoxical is less than

$$\binom{n}{k} \left(1 - 2^{-k}\right)^{n-k}. \tag{11.1}$$

Exercise 11.3.8. Conclude that if $\binom{n}{k} \left(1 - 2^{-k}\right)^{n-k} \leq 1$ then there exists a k-paradoxical tournament on n vertices.

Exercise 11.3.9. Prove that if $k \geq 3$ and $n > 4k^2 2^k$ then the inequality in the preceding exercise will hold. (A constant $c > 4$ works for $k = 2$; smaller constants work for larger values of k. As $k \to \infty$, the value of a suitable constant $\to 1$.) *Hint.* Use the following facts: $\binom{n}{k} < n^k/k!$; $1 - x < \mathrm{e}^{-x}$; and the monotonicity of the function $x/\ln x$.

This concludes the proof of Erdős's Theorem.

Exercise 11.3.10. Prove that if $n > ck^2 2^k$ (for some absolute constant c) then *almost all tournaments* on a given set of n players are k-paradoxical.

Here "almost all" means that for every $\epsilon > 0$ there exists n_0 such that if $n > n_0$ and $n > ck^2 2^k$ then the probability that the random tournament is k-paradoxical is greater than $1 - \epsilon$. *Hint.* Revisit the same calculations done for the previous exercises. Only minimal modifications are needed.

11.4 Paley tournaments: an explicit construction

We describe an explicit construction of k-paradoxical tournaments for arbitrarily large k.

Definition 11.4.1. Let $q \equiv -1$ be a prime power and let χ denote the quadratic character of $\mathbb{F}_q$. The *Paley tournament of order* q is defined as a digraph $P(q) = (V, E)$ where $V = \mathbb{F}_q$; we have a directed edge $a \to b$ iff $\chi(a - b) = 1$.

Note that because $q \equiv -1 \pmod 4$, we have $\chi(-1) = -1$ (Exercise 11.1.6). Since the character is multiplicative, this ensures that $\chi(a - b) = -\chi(b - a)$, so there is exactly one directed edge between any two distinct vertices. This shows that $P(q)$ is a tournament. (We also need to note that $\chi(0) = 0$, so there are no loops in the digraph.)

Theorem 11.4.2 (Graham-Spencer). *If $q \equiv -1 \pmod 4$ and $q \geq k^2 4^k$, then $P(q)$ is a k-paradoxical tournament.*

Proof: Let $A = \{a_1, \ldots, a_k\} \subset V$ be an arbitrary k-subset. Let $N = \#\{x \in V : x \to A\}$ be the number of vertices which dominate the set A. We seek to show that $N > 0$. In fact, we will show that $N \approx \frac{q}{2^k}$.

Consider the following three cases:

- $x \to A \quad \Rightarrow \quad (\forall i)(\chi(x - a_i) = 1)$.
- $x \not\to A$ and $x \notin A \quad \Rightarrow \quad (\forall i)(\chi(x - a_i) = \pm 1)$ and $(\exists i)(\chi(x - a_i) = -1)$.
- $x \in A \quad \Rightarrow \quad (\exists! i)(\chi(x - a_i) = 0)$.

Now let $\psi(x) := \prod_{i=1}^{k}(\chi(x - a_i) + 1)$. Considering the cases above, we have

$$\psi(x) = \begin{cases} 2^k, & x \to A \\ 0, & x \not\to A,\ x \notin A \\ 0 \text{ or } 2^{k-1}, & x \in A \end{cases}$$

The case $\psi(x) = 2^{k-1}$ occurs for at most one $x \in A$; namely, if and only if x dominates the rest of A.

Thus, we can compute the sum $S := \sum_{x\in\mathbb{F}_q} \psi(x) = 2^k N + \epsilon 2^{k-1}$, where $\epsilon \in \{0,1\}$. We will have succeeded in showing that $N > 0$ if we can prove that S is large ($S > 2^{k-1}$ will suffice).

Using the notation $[k] := \{1,\dots,k\}$, we obtain the expansion

$$S = \sum_{x\in\mathbb{F}_q} \prod_{i=1}^{k} (\chi(x-a_i)+1) = \sum_{x\in\mathbb{F}_q} \sum_{I\subseteq[k]} \prod_{i\in I} \chi(x-a_i).$$

Letting $f_I(x) := \prod_{i\in I}(x-a_i)$ and using the multiplicativity of χ we see that

$$S = \sum_{x\in\mathbb{F}_q} \sum_{I\subseteq[k]} \chi(f_I(x)) = \sum_{I\subseteq[k]} \sum_{x\in\mathbb{F}_q} \chi(f_I(x)) = \sum_{x\in\mathbb{F}_q} \chi(f_\emptyset(x)) + \sum_{I\neq\emptyset} \sum_{x\in\mathbb{F}_q} \chi(f_I(x)).$$

Let us denote by R the rest of the sum: $R := \sum_{I\neq\emptyset} \sum_{x\in\mathbb{F}_q} \chi(f_I(x))$. Since the empty product is 1 and $\chi(1) = 1$, we have $S = (\sum_{\mathbb{F}_q} 1) + R = q + R$. If we can show that q dominates R then we shall be done since then $N \approx S/2^k \approx q/2^k$, as desired. Now

$$|R| = \left| \sum_{I\neq\emptyset} \sum_{x\in\mathbb{F}_q} \chi(f_I(x)) \right| \leq \sum_{I\neq\emptyset} \left| \sum_{x\in\mathbb{F}_q} \chi(f_I(x)) \right| \leq \sum_{I\neq\emptyset} (|I|-1)\sqrt{q} \quad \text{(by Weil)}.$$

Note we can apply Weil because f_I, by definition, has no multiple roots, so in particular $f_I(x) \neq c(h(x))^2$. There are 2^k choices for $I \subseteq [k]$ and, for each choice, $|I| \leq k$. Thus, we have shown that $|R| < 2^k \cdot (k-1)\sqrt{q}$.

From above, $S = 2^k N + \epsilon 2^{k-1} = q + R$, so

$$N > \frac{q}{2^k} - (k-1)\sqrt{q} - \frac{1}{2} > \frac{q}{2^k} - k\sqrt{q}.$$

So for $N > 0$ it suffices that $\frac{q}{2^k} \geq k\sqrt{q}$, i. e., $q \geq k^2 4^k$. □

Chapter 12

Zeros of Matching Polynomials

12.1 Orthogonal Polynomials

Let $\mathcal{P}$ denote the vector space of all polynomials in one variable with real coefficients, and let the "weight function" $w(x)$ be a continuous non-negative function over a (finite or infinite) open interval, which is not identically zero (in the interval). Then we can define an inner product in $\mathcal{P}$ as:

$$(f, g) := \int w(x)f(x)g(x)\,dx$$

where the integration is over the given interval in $\mathbb{R}$[1]. By the hypothesis on w, we see that for all $f \in \mathcal{P}$, $(f, f) = 0$ exactly when $f = 0$ (Why?).

Definition 12.1.1. A sequence of polynomials, $\{p_n\}_{n\geq 0}$, is called an orthogonal family with respect to the weight function w if the following hold:

(1) $\deg(p_n) = n$

(2) $(p_n, p_m) = 0$, if $m \neq n$

Note that (p_n, p_n) is strictly positive. Also we may assume that all the p_n's are monic (the leading coefficient is 1) by scaling them. Scaling does not affect orthogonality.

The following is straightforward; we leave the proof to the reader.

Lemma 12.1.2. *If the sequence of polynomials $\{p_n\}_{n\geq 0}$ satisfies condition (1) then $\{p_n\}_{n\geq 0}$ is a basis of $\mathcal{P}$.*

[1] If the interval is infinite then w has to go to zero at infinity fast enough, to make these integrals finite. If the interval is finite then $w(x)$ does not need to be bounded but its integral must be finite

Example 12.1.3. Some examples of families of orthogonal polynomials, and their corresponding weight function.

Family of Polynomials	**Defining Interval**	**Weight function**
Chebyshev Polynomial of first kind	(-1,1)	$(1-x^2)^{-1/2}$
Chebyshev Polynomial of second kind	(-1,1)	$\sqrt{1-x^2}$
Hermite Polynomials	$(-\infty,+\infty)$	e^{-x^2}
Laguerre Polynomials	$(0,\infty)$	e^{-x}
Legendre Polynomials	$(-1,1)$	1

The actual polynomials and their norms $c_n = (p_n, p_n)$ are given below. Here $D := \frac{d}{dx}$ is the derivative operator.

Type	c_n	Polynomial
Chebychev 1	$\pi/2(n=0), \pi(n>0)$	$T_n(\cos\theta) = \cos(n\theta)$
Chebychev 2	$\pi/2$	$U_n(\cos\theta) = \sin((n+1)\theta)/\sin\theta$
Hermite	$(n!)^2$	$H_n(x) = (-1)^n e^{x^2} D^n[e^{-x^2}]$
Laguerre	$(n!)^2$	$L_n(x) = e^x D^n[x^n e^{-x}]$
Legendre	$2/(2n+1)$	$P_n(x) = \frac{1}{2^n n!} D^n[(x^2-1)^n]$

Lemma 12.1.4. *Let $\{p_n\}_{n\geq 0}$ be a sequence of orthogonal polynomials. If $f(x)$ is a non-negative factor of p_n for some n, then $f(x)$ must be a constant.*

Proof: Suppose $\deg(f) \geq 1$. Then $p_n(x) = f(x)q(x)$, where $\deg(q) < n$. Since p_n is orthogonal to all polynomials of degree $< n$, we have

$$0 = (q, p_n) = \int w(x)q(x)p_n(x)\,dx = \int w(x)f(x)q(x)^2\,dx$$

But since f is non-negative and not identically zero, the integral is strictly positive, a contradition.

Corollary 12.1.5. *With the notation as above:*

- *All zeros of $p_n(x)$ are real.*
- *p_n has no multiple zeros.*

Proof: If $p_n(a+ib) = 0$, then since complex zeros of real polynomials come in conjugate pairs, we have that $(x-a)^2 + b^2$ is a positive factor of p_n. Similarly, if θ is a multiple real zero of p_n, then $(x-\theta)^2$ is a non-negative factor of p_n.

Proposition 12.1.6. *If $\{p_n\}$ is a sequence of monic orthogonal polynomials, then it satisfies a three-term recurrence of the form $p_{n+1} = (x-a_n)p_n - b_n p_{n-1}$, where $a_n, b_n \in \mathbb{R}$ and $b_n > 0$.*

Proof: Since xp_n is a polynomial of degree $n+1$, we have that $xp_n = \sum_{i=0}^{n+1} c_i p_i$. Taking the inner product on both sides with p_j gives $c_j(p_j, p_j) = (xp_n, p_j) = (p_n, xp_j)$. So we have that $c_j = 0$ if $j < n-1$, i.e. $xp_n = c_{n+1}p_{n+1} + c_n p_n + c_{n-1}p_{n-1}$. Comparing the coefficients of the leading terms gives $c_{n+1} = 1$. Hence $p_{n+1} = (x - a_n)p_n - b_n p_{n-1}$, where $a_n = c_n, b_n = c_{n-1}$. To show that $b_n > 0$, note that

$$\begin{aligned} b_n(p_{n-1}, p_{n-1}) &= (xp_n, p_{n-1}) \\ &= (p_n, xp_{n-1}) \\ &= (p_n, p_n + f) \\ &= (p_n, p_n). \end{aligned}$$

Here f is a polynomial of degree $< n$, and hence $(p_n, f) = 0$.

Remark 12.1.7. One corollary to this recurrence is that if p_n and p_{n+1} have a common factor q, then q divides all $p_n, n \geq 0$. This is impossible because p_0 is a nonzero constant.

Consequently, for all $n \geq 1$, g.c.d.$(p_n, p_{n-1}) = 1$; in other words, adjacent members of a sequence of orthogonal polynomials have no common zeros.

We shall prove that in fact the zeros of p_n and p_{n-1} interlace.

Definition 12.1.8. Let f and g be two polynomials of degree n and $n-1$ respectively. Assume all zeros of these polynomials are real. The zeros of f and g are said to **interlace** if there is a root of g between any two roots of f, and vice-versa.

This definition permits f and g to have have multiple zeros.

Proposition 12.1.9. *Let f and g be monic polynomials of degree n and $n-1$ respectively. Assume all zeros of each polynomial are real. Then the zeros of f and g interlace iff g/f is a decreasing function on each interval between conscutive zeros of f.*

Proof: We prove the result only when each of f and g have distinct roots, and they have no common roots. The result holds otherwise also, but requires a little more work.

($\Longleftarrow$) Let $\alpha < \beta$ be two consecutive zeros of f. Then $(g/f)' = (fg' - gf')/f^2$ implies $fg' - gf' \leq 0$, in the interval (α, β). Hence $g(\alpha)f'(\alpha)$ and $g(\beta)f'(\beta)$ have the same sign. Since α and β are consecutive zeros, we see that $f'(\alpha)$ and $f'(\beta)$ have opposite signs. Hence $g(\alpha)$ and $g(\beta)$ differ in sign and hence g has a zero between α and β. Now if α and β are two consecutive zeros of g and f has no zero in the interval $[\alpha, \beta]$, then $fg' - gf' \leq 0$ in $[\alpha, \beta]$. Hence $f(\alpha)g'(\alpha)$ and $f(\beta)g'(\beta)$ have the same sign. Hence f changes sign in $[\alpha, \beta]$ contradicting assumption, that f has no zero in $[\alpha, \beta]$.

($\Longrightarrow$) Now suppose that the zeros of f and g interlace. We can write g/f in partial fractions as:

$$\frac{g(x)}{f(x)} = \sum_{i=1}^{n} \frac{a_i}{x - \alpha_i}$$

where the α_i are the zeros of f, and the a_i are real numbers. Multiplying both sides by $(x-\alpha_j)$, and taking the limit as $x \to \alpha_j$, we see that $a_j = g(\alpha_j)/f'(\alpha_j)$. To show that g/f is decreasing, it is enough to show that the a_i's are all positive. Since the polynomials are monic, we see that $f(+\infty) = g(+\infty) = f'(+\infty) = +\infty$[2]. This shows that $f'(\alpha_n) > 0$. Since $\alpha_n > \beta_{n-1}$ and $g(\infty) = \infty$, $g(\alpha_n) > 0$. Hence we have that $a_n > 0$. Since $f/(x-\alpha_n)$ and $g/(x-\beta_{n-1})$ satisfy the hypothesis, we can apply induction to complete the rest. Alternatively, one can show that $\operatorname{sgn}(g(\alpha_i)) = \operatorname{sgn}(f'(\alpha_i)) = (-1)^{n-i}$ and conclude that $a_i > 0$ for all i.

Exercise 12.1.10. Identify the exact places in the proof, where we used the additional hypothesis that f and g have distinct zeros, and no common zeros.

Corollary 12.1.11. *Suppose f and g are monic polynomials of degree n and $n-1$ respectively, and g/f is a decreasing function between every pair of consecutive roots of f. If one of f, g has all real zeros, then so does the other and their zeros interlace.*

Proof: If f has all real zeros, then the proof shows g has $n-1$ real zeros. Hence all its zeros are real. Hence the result holds.

If g has all real zeros, then the proof shows that f has exactly $n-2$ real zeros in the interval $[\alpha, \beta]$, where α is the smallest zero of g and β is its largest zero. We need to show that f has one zero $< \alpha$ and another $> \beta$. Let γ be largest zero of f in the interval $[\alpha, \beta]$. Since β is the largest zero of g and $g(\infty) = \infty$, $g > 0$ to the right of β, and hence $g(\gamma) < 0$ (Why?). Since $fg' - gf'(\gamma) \leq 0$ (Why?), we have $g(\gamma)f'(\gamma) \geq 0$. But $g(\gamma) < 0$ implies $f'(\gamma) < 0$. This together with $f(\gamma) = 0$ shows that $f < 0$ immediately after γ. But f is monic implies $f(\infty) = \infty$. Hence f has a zero after γ. So f has at least $n-1$ real zeros. But f is degree n and complex zeros come in pairs. So f has all real zeros.

Corollary 12.1.12. *Let $g_1, \ldots, g_k$ be polynomials of degree $n-1$ and f a polynomial of degree n. Suppose that for some $i = t, g_t$ has all real zeros. Let $\alpha_i \geq 0$ be a set of k real numbers. If the zeros of f and g_i interlace for all i, then the zeros of f and $\sum \alpha_i g_i$ interlace.*

Proof: Applying the preceding corollary to g_t and f, shows that f has all real zeros. Hypothesis implies g_i/f is a decreasing function, off the zeros of f. Hence $(\sum \alpha_i g_i)/f$ is a decreasing function and the result follows.

Proposition 12.1.13. *Fix a sequence of monic orthogonal polynomials $\{p_n\}$. Let $\alpha_1, \ldots, \alpha_n$ be the zeros of p_n in increasing order, and $\beta_1, \ldots, \beta_{n-1}$ be the zeros of p_{n-1} in increasing order. Then the α's and the β's interlace, i.e.*

$$\alpha_1 < \beta_1 < \alpha_2 < \cdots < \alpha_{n-1} < \beta_{n-1} < \alpha_n$$

Proof: By 12.1.7, we know that the $2n-1$ numbers α_i, β_j are all distinct. So by 12.1.5 and 12.1.9, it is enough to show that p_{n-1}/p_n is a decreasing function, whenever it is defined. But by 12.1.6,

[2] $f(\infty)$ is a shorthand for $\lim_{x\to\infty} f(x)$

$$\begin{aligned} p_n &= (x-a_n)p_{n-1} - b_n p_{n-2} \\ \Rightarrow \frac{p_n}{p_{n-1}} &= (x-a_n) - b_n \frac{p_{n-2}}{p_{n-1}} \end{aligned}$$

But induction hypothesis implies p_{n-2}/p_{n-1} is a decreasing function. This together with $b_n > 0$ gives that p_n/p_{n-1} is increasing function, or p_{n-1}/p_n is a decreasing function.

12.2 Matching Polynomial of a graph

All graphs we consider will be undirected with no loops.

Definition 12.2.1. Let $G = (V, E)$ be a graph. A k-matching is a set of k disjoint edges. Let $p(G, k)$ denote the number of k-matchings of G. Then the matching polynomial $\mu(G, x)$ is defined as

$$\sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k p(G,k) x^{n-2k}$$

We will also denote the matching polynomial by $\mu(G)$, if the variable is clear from the context.

At first the definition may seem a bit strange, but note that $n-2k$ is the number of vertices of the graph which are not covered by a k-matching. The matching polynomial is a variant of the generating function of the number of k-matchings.

Example 12.2.2. $\mu(\overline{K}_n) = x^n$.

Example 12.2.3. $\mu(P_n) = \sum_r (-1)^r \binom{n-r}{r} x^{n-2r}$. Think of the vertices of P_n as going from left to right. If we contract the edges of an r-matching towards the left, we get a path of length $n-r$ with r distinguished vertices. Hence $P(P_n, r) = \binom{n-r}{r}$.

Lemma 12.2.4.

$$\mu(K_n) = \sum_r (-1)^r \binom{n}{2r} \frac{(2r)!}{2^r r!} x^{n-2r}.$$

Proof: We need to show that $P(K_n, r) = \binom{n}{2r} 2^{-r}(2r)!/r!$. First choose the $2r$ vertices of the matching and then a perfect matching of K_{2r}. Hence $P(K_n, r) = \binom{n}{2r} T(r)$, where $T_r = P(K_{2r}, r)$.

Now $T(r) = (2r-1) * T(r-1)$ and $T(1) = 1$, gives $T(r) = 2^{-r}(2r)!/r!$.

Many well known families of orthogonal polynomials arise as matching polynomials of graphs. For example,

$\mu(C_n, 2x)$	$= 2T_n(x)$	Chebyshev first kind
$\mu(P_n, 2x)$	$= U_n(x)$	Chebyshev 2nd kind
$\mu(K_n, x)$	$= H_n(x)$	Hermite Polynomial
$\mu(K_{n,n}, x)$	$= (-1)^n L_n(x^2)$	Laguerre Polynomial

12.3 Characteristic Polynomial of a graph

Definition 12.3.1. The *adjacency matrix* of a graph with n vertices is an $n \times n$ $(0,1)$-matrix (a_{ij}) where $a_{ij} = 1$ if vertex i and vertex j are adjacent; and $a_{ij} = 0$ otherwise.

Definition 12.3.2. Let $G = (V, E)$ be an undirected simple graph with no loops, and let A be its adjacency matrix. The *characteristic polynomial* $\phi(G, x)$ is defined to be the characteristic polynomial of A, i.e. $\phi(G, x) = \det(A - xI)$, where I is the identity matrix of the appropriate size.

Exercise 12.3.3. Isomorphic graphs have the same characteristic polynomial. In particular, $\phi(G, x)$ does not depend on the numbering of the vertices.

Definition 12.3.4. A principal $(n - r) \times (n - r)$ minor of a matrix A is the determinant obtained by deleting rows $i_1, \ldots, i_r$ and columns $i_1, \ldots, i_r$, for some $i_1 < \cdots < i_r$.

Lemma 12.3.5. *Let $G = (V, E)$ be a graph with adjacency matrix A. Then*

$$\phi(G, x) = \sum_{r=0}^{n} (-1)^r a_r x^{n-r}$$

where a_r is the sum of all the principal $r \times r$ minors of A.

Proof: [Sketch] $\phi(G, x) = \det(xI - A)$. The coefficient of x^{n-r} corresponds to choosing x in r diagonal elements and not choosing x for any of the others.

Proposition 12.3.6. *Fix a graph $G = (V, E)$, and let $\phi(G)$ be its charecteristic polynomial. Then the coefficient of x^{n-r} is*

$$(-1)^r a_r = \sum_{C} (-1)^{Comp(C)} 2^{Cyc(C)}$$

where the sum is taken over all subgraphs C of G consisting of disjoint edges and cycles, $|C| = r$. $Comp(C)$ is the number of components of C and $Cyc(C)$ = number of cycles in C.

Proof: Let B be any $r \times r$ principal submatrix of A. Then B is the adjacency matrix of a subgraph C^3 of G, where $|C| = r$.

$$\det(B) = \sum_{\sigma \in Sym(r)} \left[\text{sgn}(\sigma) \prod_i b_{i\sigma(i)} \right]$$

Since B is a $(0,1)$-matrix we see that each σ contributes $\text{sgn}(\sigma)$ to $\det(B)$ iff $(i, \sigma(i))$ is an edge for each i. Suppose that

$$\sigma = \theta_1 \dots \theta_l \tau_1 \dots \tau_p$$

be the decomposition of σ into cycles, where the τ's are 2-cycles and θ's are cycles of length ≥ 3. Hence the subgraph C is the union of l cycles and p disjoint edges.

Suppose $\gamma \in Sym(r)$ is another permutation which gives rise to the same subgraph, then the transpositions occurring in the cycle structure of γ should be the same as that in σ. If δ is a cycle occurring in the cycle structure of γ, then either δ or δ^{-1} occurs in the cycle structure of σ. Moreover, since σ and γ have the same cycle structure, they are conjugates in $Sym(r)$ and hence $\text{sgn}(\sigma) = \text{sgn}(\gamma)$. So the subgraph C contributes exactly $\text{sgn}(\sigma) 2^{Cyc(C)}$ towards $\det(B)$.

Since odd cycles are even permutations, we have that $\text{sgn}(\sigma) = (-1)^\alpha$, where α is the number of even cycles in σ. Let β denote the number of odd cycles in σ. Since $\sum_{i=1}^{l} length(\theta_i) + 2p = r$, we have that $\beta \equiv r \mod 2$ and hence $\alpha \equiv p + l - r \equiv p + l + r \mod 2$.

Hence $\text{sgn}(\sigma) 2^l = (-1)^r (-1)^{p+l} 2^l = (-1)^r (-1)^{Comp(C)} 2^{Cyc(C)}$. Hence the result.

Lemma 12.3.7. *If G is a forest (graph without cycles) then $\phi(G, x) = \mu(G, x)$[4].*

Proof: By assumption G has no cycles, so coeffcient of $x^{n-r} = \sum_C (-1)^{Comp(C)}$, where C runs over all sub-graphs of size r composed of disjoint edges (no cycles here). So we have that r is even and $Comp(C) = r/2$, and that C is an r-matching of T. So coefficient of $x^{n-r} = (-1)^{r/2} p(G, r/2)$.

$$\phi(G, x) = \sum_{r=0}^{\lfloor n/2 \rfloor} (-1)^r p(G, r) x^{n-2r} = \mu(G, x)$$

Corollary 12.3.8. *If G is a forest then $\mu(G)$ has real zeros.*

Proof: By previous lemma, $\mu(G, x) = \phi(G, x)$. But $\phi(G, x)$, is the characteristic polynomial of a real symmetric matrix. Since real symmetric matrices have all real eigenvalues, $\phi(G)$ has all real zeros, and hence $\mu(G)$ has all real zeros.

[3] not necessarily induced

[4] Infact $\phi(G) = \mu(G)$ iff G is a forest

Lemma 12.3.9 (Interlacing Theorem). *Let A be a $n \times n$ real symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Let B be the matrix obtained by deleting the first row and column of A, with eigenvalues $\nu_1 \geq \cdots \geq \nu_{n-1}$. Then $\lambda_1 \geq \nu_1 \geq \lambda_2 \geq \cdots \geq \nu_{n-1} \geq \lambda_n$.*

Proof: This result is an application of the spectral theorem in Linear Algebra. Let P_i denote the projection onto the eigenspace corresponding to λ_i. Then from the spectral theorem, we have

$$(xI - A)^{-1} = \sum_i \frac{1}{x - \lambda_i} P_i.$$

Let f, g denote the charecteristic polynomials of A and B, respectively. Then

$$\frac{g(x)}{f(x)} = \frac{\det(xI_{n-1} - B)}{\det(xI_n - A)} = \left[(xI_n - A)^{-1}\right]_{11} = \sum_i \frac{1}{x - \lambda_i}[P_i]_{11}$$

But the diagonal entries of P_i are the inner product of some eigenvector of A with itself. So the partial fraction expansion of g/f has non-negative numerators. Hence by 12.1.9, the zeros of f and g interlace.

Corollary 12.3.10. *Let G be a forest, and v a vertex of G. Then the zeros of $\mu(G)$ and $\mu(G - v)$ interlace.*

Proof: Apply previous result to the adjacency matrix of G.

12.4 Matching Polynomials have real zeros

We have computed the matching polynomial for several kinds of graphs. Many of them give rise to families of orthogonal polynomials. We have seen that orthogonal polynomials always have real zeros, and the zeros of adjacent orthogonal polynomials interlace. If we consider the matching polynomials of trees, then they equal their characteristic polynomials and hence the matching polynomials of trees have real zeros. The interlacing theorem for real symmetric matrices, shows that even in this case, the zeros of $\mu(T)$ and $\mu(T-v)$ interlace. In this section, we prove that the matching polynomials of all graphs have real zeros, and that the zeros of $\mu(G)$ and that of $\mu(G - v)$ always interlace.

These central results were proved by *Heilmann* and *Lieb*, two statistical physicists, in a paper about monomer-dimer systems[4]. They gave three proofs of this result. One of them uses the theory of orthogonal polynomials, which we have developed. Another proof, due to Godsil, reduces the problem to that for trees.

Before we get to any of these proofs, we need to establish some properties of these polynomials.

Proposition 12.4.1. *Let G, H be two graphs and $G+H$ be their disjoint union. Then $\mu(G+H) = \mu(G)\mu(H)$*

Proof: Any matching of $G+H$ is a (possibly empty) matching of G together with a (possibly empty) matching of H. So $p(G+H,k) = \sum_{r=0}^{k} p(G,r)p(H,k-r)$.

Proposition 12.4.2. *Let $G = (V,E)$ be a graph and $e = \{v,w\} \in E$. Then $\mu(G) = \mu(G-e) - \mu(G-v-w)$.*

Proof: Every k-matching of G either includes e, or does not include e. In the first case, it gives a $(k-1)$-mathcing of $G-v-w$. In the other case, it gives a k-matching of $G-e$. Hence

$$\begin{aligned} p(G,k) &= p(G-e,k) + p(G-v-w,k-1) \\ \sum_k (-1)^k p(G,k) x^{n-2k} &= \sum_k (-1)^k p(G-e,k) x^{n-2k} \\ &\quad - \sum_k (-1)^{k-1} p(G-v-w,k-1) x^{(n-2)-2(k-1)} \\ \mu(G) &= \mu(G-e) - \mu(G-v-w). \end{aligned}$$

Proposition 12.4.3. *$G = (V,E)$, $v \in V$, and $N(v) = \{w_1, \ldots, w_d\}$, where $d = \deg(v)$. Then $\mu(G) = x\mu(G-v) - \sum_{i=1}^{d} \mu(G-v-w_i)$.*

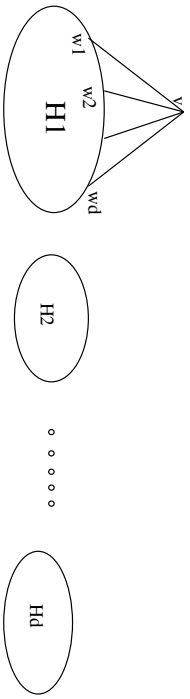
Proof: Every k-matching either covers v or its does not. If it does not cover v, then we get a k-matching of $G-v$. Otherwise the matching has to contain (v,w_i) for some i, and hence gives a $k-1$-matching of $G-v-w_i$. So

$$\begin{aligned} p(G,k) &= p(G-v,k) + \sum_{i=1}^{d} p(G-v-w_i,k-1) \\ (-1)^k p(G,k) x^{n-2k} &= x(-1)^k p(G-v,k) x^{n-1-2k} \\ &\quad - \sum_i (-1)^{k-1} p(G-v-w_i,k-1) x^{(n-2)-2(k-1)} \\ \mu(G) &= x\mu(G-v) - \sum_i \mu(G-v-w_i) \end{aligned}$$

Theorem 12.4.4. *If G is any graph, $\mu(G)$ has real zeros. Moreover, if v is any vertex of G, then the zeros of $\mu(G)$ and those of $\mu(G-v)$ interlace.*

Proof: Induction on the number of vertices. With the notation, as before, we have

Figure 12.1: The graph F1



$$\frac{\mu(G)}{\mu(G-v)} = x - \sum_i \frac{\mu(G-v-w_i)}{\mu(G-v)}$$

By induction hypothesis the zeros of $\mu(G-v-w_i)$ and that of $\mu(G-v)$, interlace and $\mu(G-v)$ has all real zeros. By 12.1.9, we have $\mu(G-v-w)/\mu(G-v)$ is a decreasing function and hence $\mu(G)/\mu(G-v)$ is an increasing function. Hence, by 12.1.12, we have that the zeros of $\mu(G)$ are real, and interlace those of $\mu(G-v)$.

Proposition 12.4.5. *Let G be a graph, $v \in V$, and $N(v) = \{w_1, \ldots, w_d\}$. Let $H = G - v$. Let $H_i (i = 1 \ldots d)$ be graphs isomorphic to H. Let $F_1 = G + H_2 + \cdots + F_d$. Let $w_i(H_i)$ denote the vertex corresponding to w_i in H_i. Let F_2 be the graph obtained by adding the edges $(v, w_i(H_i))$ to the graph $v + H_1 + \cdots + H_d$. Then $\mu(F_1) = \mu(F_2)$.*

Proof: By 12.4.3 we have,

$$\mu(F_2) = x\mu(H_1)\ldots\mu(H_d) - \mu(H_1)\ldots\mu(H_d)\sum_i \frac{\mu(H_i - w_i(H_i))}{\mu(H_i)}$$

Again by 12.4.3 and 12.4.1 we have,

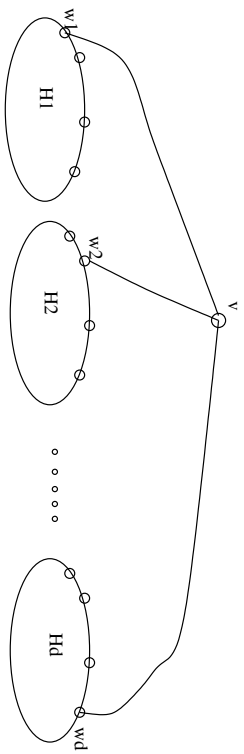


Figure 12.2: The graph F2

$$\mu(F_1) = \mu(H_2)\dots\mu(H_d)\left[x\mu(H1) - \mu(H_1 - w_i)\right]$$

Since isomorphic graphs have the same matching polynomial, the result follows.

As a corollary, we have that $\mu(G)$ divides $\mu(F_2)$. In the process of going from G to F_2, we have eliminated all cycles based at v. If we could repeat this process and unwind all the cycles, then we would have constructed a tree T, such that $\mu(G)$ divides $\mu(T)$. Since $\mu(T)$ has real zeros, we can conclude that $\mu(G)$ has real zeros.

If we unwind the graph all the way starting at v, we get a tree $T(G, v)$ called the path-tree of G.

Definition 12.4.6. Let G be an undirected graph, and $v \in V$. By a path π in G, we mean a sequence of *distinct* vertices such that consecutive members are adjacent in G. $T(G, v)$ is the graph, whose vertices are paths in G, which start at v. Two paths π_1 and π_2 are adjacent if one is a maximal proper subpath of the other.

Clearly $T(G, v)$ depends only on the component of G containing v. We often identify the path in $T(G, v)$ consisting of only the vertex v, with the vertex v.

Proposition 12.4.7. *Fix any graph G, and $v \in V$.*

- *$T(G, v)$ is a tree.*
- *if G is a tree then $T(G, v)$ is isomorphic to G.*

Proof:

- If π is any vertex of $T(G, v)$, then there is a unique path in $T(G, v)$ from v to π, consisting of sub-paths of π.
- The map from G to $T(G, v)$ taking w to the unique path from v to w, gives the required isomorphism.

Proposition 12.4.8. *Let G be a graph, and $v \in V$. Put $T = T(G, v)$. Then*

$$\frac{\mu(G)}{\mu(G - v)} = \frac{\mu(T)}{\mu(T - v)}$$

and more over $\mu(G)$ divides $\mu(T)$.

Proof: By induction on the number of vertices. If G is a tree then the result trivially holds. So the result holds if the $n \leq 2$. Let N be the set of neighbours of v in G, and $H = G - v$. Then by 12.4.3

$$\begin{aligned}\frac{\mu(G)}{\mu(H)} &= \frac{x\mu(H) - \sum_{w\in N}\mu(H-w)}{\mu(H)} \\ &= x - \sum_{w\in N}\frac{\mu(H-w)}{\mu(H)} \\ &= x - \sum_{w\in N}\frac{\mu(T(H,w)-w)}{\mu(T(H,w))}\end{aligned}$$

Now $T(H,w) = T(G-v,w)$ is isomorphic to the component of $T(G,v)-v$ which contains the path (u,v) in G. Therefore we have,

$$\frac{\mu(T(H,w)-w)}{\mu(T(H,w))} = \frac{T(G,v)-uv}{\mu(T(G,v))}$$

Hence we have

$$\begin{aligned}\frac{\mu(G)}{\mu(H)} &= x - \sum_{w\in N}\frac{\mu(T(G,v)-vw)}{\mu(T(G,v)-v)} \\ &= \frac{\mu(T(G,v))}{\mu(T(G,v)-v)}\end{aligned}$$

This proves the first part of the theorem. Since $T(G-v,w)$ is isomorphic to a component of $T(G,v)-v$, $\mu(T(G-v,w))$ divides $\mu(T(G,v)-v)$. Induction hypothesis gives $\mu(G-v)$ divides $\mu(T(G-v,w))$. Hence we have $\mu(G-v)$ divides $\mu(T-v)$. This together with the first part of the theorem proves the second part.

Bibliography

[1] C.D. Godsil, *Algebraic Combinatorics*. Chapman and Hall Mathematics.

[2] C.D.Godsil, I. Gutman, *On the theory of the Matching Polynomial*. Journal of Graph Theory, Vol 5 (1981), p 137-144.

[3] C.D. Godsil *Matchings and Walks in Graphs*. Journal of Graph Theory, Vol 5 (1981), p. 285-297.

[4] O.J. Heilmann, E.H. Leib, *Theory of monomer-dimer systems*. Comm. Math. Phys. 25(1972) p 190-232.

Chapter 13

Set Systems

TO BE WRITTEN.

13.1 Problems

Definition 13.1.1. Let $\mathcal{F}$ be a family of subsets of a universe of size n. We say that $\mathcal{F}$ is a *Sperner family* if there do not exist $A_1, A_2 \in \mathcal{F}$ such that $A_1 \subsetneq A_2$. In other words, the elements of $\mathcal{F}$ are pairwise incomparable with respect to inclusion.

The next result falls under the heading of Extremal Set Theory.

Theorem 13.1.2 (Sperner's Theorem). *The largest Sperner family in a universe of size n has size $\binom{n}{\lfloor n/2 \rfloor}$. Moreover, there are at most two Sperner families of this size: all subsets of size $\lfloor n/2 \rfloor$, and all subsets of size $\lceil n/2 \rceil$.*

Exercise 13.1.3. We say that a Boolean function f is *monotone* if

$$(\forall x_1, \ldots, x_n, y_1, \ldots, y_n)(\text{ if } (\forall i)(x_i \geq y_i) \text{ then } f(x_1, \ldots, x_n) \geq f(y_1, \ldots, y_n)).$$

Let $M(n)$ denote the number of monotone Boolean functions in n variables. Let $S(n)$ denote the number of Sperner families of subsets of a universe of size n (ordered by inclusion). Prove: $M(n) = S(n)$.

Exercise 13.1.4. Let $a_1, \ldots a_n, b$ be given positive real numbers. Let p denote the probability that $\sum x_i a_i = b$, where each x_i is 0 or 1, chosen at random by independent unbiased coin flips.

Prove: $p = O(1/\sqrt{n})$, i. e., prove that there exists a constant C such that $p \leq C/\sqrt{n}$, regardless of the values of the a_i and b.

Determine, moreover, the smallest value of c such that the following is true: there exists a sequence $c_n \to c$ such that $p \leq c_n/\sqrt{n}$ regardless of the values of the a_i and b. (So c and the c_n must not depend on the a_i or b.) *Hint.* Sperner's Theorem.

Exercise 13.1.5. ("Littlewood-Offord problem") Let $a_1, \dots a_n$ be given unit vectors in the plane (i. e., vectors of unit length). Let p denote the probability that the vector $\sum x_i a_i$ has length $\leq 1/10$, where each x_i is 0 or 1, chosen at random by independent unbiased coin flips.

Prove: $p = O(1/\sqrt{n})$. – You are not requested to find the smallest constant implied by the big-Oh notation.

Definition 13.1.6. A set T is a *transversal* of a set system $\mathcal{F} = \{A_1, \dots, A_m\}$, if T "hits" each A_i, i. e., if $(\forall i)(T \cap A_i \neq \emptyset)$. ($T$ is also called a "hitting set" or a "cover" for $\mathcal{F}$.) The *transversal number* (a.k.a. hitting number, covering number) $\tau(\mathcal{F})$ is the size of the smallest transversal. (Note that if $\emptyset \in \mathcal{F}$ then no transversal exists and $\tau(\mathcal{F}) = \infty$.) – "$\tau$" is the Greek letter "tau."

Definition 13.1.7. A *matching* in a set system $\mathcal{F}$ is a subset of $\mathcal{F}$ whose elements A_i are pairwise disjoint sets (so that each element of the universe belongs to at most one element of the matching). The *matching number* $\nu(\mathcal{F})$ is the cardinality of the largest matching in $\mathcal{F}$. – "ν" is the Greek letter "nu."

Recall that $\mathcal{F}$ is *k-uniform* if $(\forall i)(|A_i| = k)$. Graphs can be viewed as 2-uniform set-systems (the set of edges); the above concepts can then be applied to graphs. For instance, if C_n denotes the cycle of length n, then $\tau(C_n) = \lceil n/2 \rceil$ and $\nu(C_n) = \lfloor n/2 \rfloor$; $\tau(K_n) = n - 1$, $\nu(K_n) = \lfloor n/2 \rfloor$ (DO: verify these statements!).

Exercise 13.1.8. Prove: if M is a *maximal* matching in a (not necessarily bipartite) graph G then $|M| \geq \nu/2$, where ν denotes the size of a *maximum* matching in G. (A matching is "maximal" if no further edge can be added to it; it is "maximum" if it has largest size among all matchings in G. The "size" $|M|$ of a matching M is the number of edges in M.) What can be proven for k-uniform set systems?

Exercise 13.1.9. (a) Let $K_n^{(k)}$ be the "complete k-uniform set-system" on n points, i. e., the set of all k-subsets of an n-set (so $m = \binom{n}{k}$). Determine $\tau(\mathcal{F})$.

(b) Prove: $\nu(\mathcal{F}) \leq \tau(\mathcal{F})$.

(c) Prove: if $\mathcal{F}$ is k-uniform then $\tau(\mathcal{F}) \leq k\nu(\mathcal{F})$.

(d) For every positive integer ℓ, construct a graph G_ℓ such that $\nu(G_\ell) = \ell$ and $\tau(G_\ell) = 2\ell$. Give a one-line solution to this problem, even though it is a special case of the next one.

(e) For every pair of positive integers k, ℓ, construct a k-uniform set-system $\mathcal{F}_\ell$ such that $\nu(\mathcal{F}_\ell) = \ell$ and $\tau(\mathcal{F}_\ell) = k\ell$. Prove your answer.

Exercise 13.1.10. (König's Theorem) Prove: if G is a bipartite graph then $\nu(G) = \tau(G)$.

Exercise 13.1.11. Let $\mathcal{L}$ be a projective plane of order n. Prove: $\tau(\mathcal{L}) = n + 1$. (Here we view $\mathcal{L}$ as the set-system consisting of the lines.) (2 points for $\tau \leq n + 1$ and 5 points for $\tau > n$.) *Hint* to the 5-point part: counting.

Exercise 13.1.12. Let $\mathcal{F}$ be a k-uniform family of subsets of an n-set X. Prove: $\tau(\mathcal{F}) \leq \lceil (n/k) \ln m \rceil$, where $m = |\mathcal{F}|$. *Hint.* Let $t = \lceil (n/k) \ln m \rceil$. Pick a sequence of t points independently at random (permitting repetitions). Prove that with positive probability, this sequence will hit each A_i. Along the way, use the inequality $1 + x < e^x$ (which is true for every real number $x \neq 0$).

Definition 13.1.13. Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a family of subsets of a universe of size n. A *system of distinct representatives, or SDR* for $\mathcal{F}$ is a set $\{x_1, \ldots, x_m\}$, where $x_i \in A_i$ for all $1 \leq i \leq m$, and such that $x_i \neq x_j$ for all $1 \leq i < j \leq m$.

Exercise 13.1.14. Let $\mathcal{F} = \{A_1, \ldots, A_n\}$ be a family of n subsets of $[n]$. Let M denote the incidence matrix of this family (so M is an $n \times n$ (0,1)-matrix.) True or false (prove each answer):

1. If $\det(M) \neq 0$ then $\mathcal{F}$ has an SDR.
2. If $\det(M) = 0$ then $\mathcal{F}$ has no SDR.

Exercise 13.1.15. (Systems of Distinct Representatives) Beyond the seven seas there is a tiny island, 6 square miles in all. The island is inhabited by six native tribes and by six turtle species. Each tribe and each turtle species occupies one square mile of territory; the territories of the tribes don't overlap with one another; nor do the territories of the different turtle species.

Each tribe wishes to select a totem animal from among the turtle species found in the tribe's territory; and each tribe must have a different totem animal. Prove that such a selection is always possible.

Comments. Your solution should be clear and simple, with reference to a result stated in class. State the result, then define the variables in terms of the problem on hand. If your solution refers to certain finite sets, make sure you clearly specify what the elements of each set are. WARNING: the territory of a tribe is *not* a finite set.

Let $L(n)$ denote the number of $n \times n$ Latin squares. Use the result that $\log L(n) \sim n^2 \log n$ To solve the following problem:

Exercise 13.1.16. Two Latin Squares are *isomorphic* if one of them can be obtained from the other by permuting the rows, columns, and relabeling the entries. Prove: almost all Latin Squares are not isomorphic to a symmetric matrix.

Hint. Give an easy upper bound on the number of symmetric matrices with entries from $[n]$, and another easy upper bound on the number of Latin Squares isomorphic to a given Latin Square. All these numbers should be dwarfed by $L(n)$.

Definition 13.1.17. Let $A = (a_{i,j})$ be an $n \times n$ matrix. The *permanent* of A, $\mathrm{per}(A)$, is defined as

$$\mathrm{per}(A) = \sum_{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)},$$

where σ ranges over all permutations of $[n] = \{1, \ldots, n\}$ (i. e., all one-to-one functions from $[n]$ to $[n]$).

Remark 13.1.18. Note that this is just like the definition of the determinant, except without the sign terms (± 1).

Definition 13.1.19. Let $A = (a_{i,j})$ be a real $n \times n$ matrix. If all the entries are non-negative and all the row sums and column sums equal 1, then we call A a *doubly stochastic* matrix.

Theorem 13.1.20 (Permanent Inequality) (Egorychev-Falikman Theorem, formerly van der Waerden's Permanent Conjecture). *Let A be a doubly stochastic matrix. Then*

$$\mathrm{per}(A) \geq \frac{n!}{n^n}$$

Exercise 13.1.21. (a) Prove: $\frac{n!}{n^n} > \exp(-n)$. Do not use Sterling's formula. *Hint:* Use the power series of $\exp(x)$.

(b) Use the Permanent Inequality to prove that a regular bipartite graph of degree $r \geq 3$ has at least $(r/\mathrm{e})^n$ perfect matchings, where n is half the number of vertices.

(c) Use part (b) to show that $\log L(n) \sim n^2 \log n$, where $L(n)$ is the number of Latin squares of order n.

(d) Two Latin Squares are *isomorphic* if one of them can be obtained from the other by permuting the rows, columns, and relabeling the entries. Prove: almost all Latin Squares are not isomorphic to a symmetric matrix.

Definition 13.1.22. A *sunflower with s petals* is a family of (not necessarily distinct) sets $B_1, \ldots, B_s$ such that $(\forall i \neq j)(B_i \cap B_j = \bigcap_{k=1}^{s} B_k)$.

Exercise 13.1.23. Let $A_1, \ldots, A_m$ be *not necessarily distinct* sets of size $|A_i| \leq r$. Prove: if $m > r!(s-1)^{r+1}$ then there is a sunflower with s petals among the A_i, i. e., $(\exists i_1, \ldots, i_s)(1 \leq i_1 < i_2 < \cdots < i_s)$ such that $A_{i_1}, \ldots, A_{i_s}$ form a sunflower.

Definition 13.1.24. A *Steiner triple system* ...

Fix or add def. Right place?

Exercise 13.1.25. Let $S = (P, L, I)$ be a STS (Steiner triple system) with $n = |P|$ points. Let $S_1 = (P_1, L_1, I_1)$ be a sub-STS, i. e., $P_1 \subseteq P$, $L_1 \subseteq L$, and $I_1 = I \cap (P_1 \times L_1)$; and S_1 is also a STS. (So if two points of a line belong to S_1 then the third point also belongs to S_1.) Prove: if $P_1 \neq P$ then $|P_1| \leq (n-1)/2$.

Exercise 13.1.26. Let $f(x_1, \ldots, x_n) = C_1 \wedge \cdots \wedge C_m$ be a 3-CNF formula, i. e., each clause C_i is an "OR" ("$\vee$") of three literals (Boolean variables and their negations). Prove: there exists a substitution (assignment of $(0,1)$-values to the x_i) which will satisfy at least $7m/8$ of the m clauses. *Hint.* Try a random substitution (flip a coin for each x_i). What is the expected number of clauses that are satisfied by this substitution? (A clause is "satisfied" if it evaluates to 1.)

Chapter 14

Miscellaneous Exercises

14.1 2002 Midterm 1

Exercise 14.1.1. Let T_n denote the set of strings of length n over the alphabet $\{A, B\}$ (so $|T_n| = 2^n$). Consider the following subsets of T_n:

(a) E_n : strings with an even number of A's.

(b) M_n : strings with at least as many A's as B's.

(c) C_n: strings without consecutive A's.

(d) $D_{n,k}$: strings without consecutive A's, having exactly k A's.

(e) $A_{n,k}$: strings with k alternations. (The string $AABBBBABBBBAA$ has 4 alternations, the string BBA has one alternation, the string $AAAA$ has zero alternations.)

Count each set. Your answers should be a simple **closed-form expressions** (no summation or product symbols, no dot-dot-dots) using binomial coefficients and Fibonacci numbers. Indicate the proofs of your answers.

Exercise 14.1.2. A *monomial* is a product of powers of the variables, for instance x^2y^6z is a monomial of degree 9. Count the monomials of degree k over a set of n variables. The monomials you count need not involve all variables. For instance, if $k = 2$ and $n = 3$ then we have the 6 monomials x^2, xy, xz, y^2, yz, z^2. Your answer should be a very simple formula.

Definition 14.1.3. Let $\mathbb{F}$ be a field, and $x, y \in \mathbb{F}^n$. The *inner product* of x and y, often denoted $\langle x, y\rangle$ or $x \cdot y$, is an element of $\mathbb{F}$ defined by $\sum_{i=1}^{n} x_i y_i$. We say x and y are *orthogonal* if their inner product equals the zero element of $\mathbb{F}$.

Definition 14.1.4. Let U be a subset of $\mathbb{F}^n$. The *perpendicular subspace* $U^\perp$ is defined as $U^\perp = \{y \in \mathbb{F}^n \colon (\forall x \in U)(x \cdot y = 0)\}$. (Check that $U^\perp$ is indeed a subspace of $\mathbb{F}^n$!)

Exercise 14.1.5. Let U, V, W be pairwise perpendicular subspaces of F^n where F is a field.

(a) Prove: $\dim(U) + \dim(V) + \dim(W) \leq 3n/2$.

(b) Prove that equality is possible for every even value of n. (For this part, you choose F. Name the field you choose.)

(c) Prove: if $F = \mathbb{R}$ then $\dim(U) + \dim(V) + \dim(W) \leq n$.

Exercise 14.1.6. (Reverse Oddtown) A town has n citizens and m clubs. Each club has an even number of members, and each pair of clubs shares an odd number of members.

1. Prove: $m \leq n + 1$.
2. Prove: if n is odd then $m \leq n$.

Exercise 14.1.7. Let $\mathbb{R}^{\mathbb{R}}$ denote the space of $\mathbb{R} \to \mathbb{R}$ functions. Let $S = \{\sin(x + a) : a \in \mathbb{R}\} \subset \mathbb{R}^{\mathbb{R}}$. Prove that the rank of S is 2.

Exercise 14.1.8. Consider the space $\mathbb{R}[x]$ of polynomials in one variable over $\mathbb{R}$. Let $f(x) = \binom{x-1}{n}$. Note that $f(x)$ is a polynomial of degree n.

(a) Prove that the polynomials $f_i(x) = f(x)/(x - i)$ are linearly independent ($i = 1, \ldots, n$).

(b) Prove that every polynomial of degree $\leq n - 1$ can be written as a linear combination of the f_i.

Exercise 14.1.9. (Fisher's inequality) If $A_1, \ldots, A_m \subseteq [n]$ and for all $i \neq j$, $|A_i \cap A_j| = \lambda$ then $m \leq n$. Assume $\lambda \geq 1$ and $(\forall i)(|A_i| > \lambda)$.

14.2 2002 Midterm 2

Exercise 14.2.1. (a) Prove: for all graphs, $\tau(G) \leq 2\nu(G)$. (Recall Definitions 13.1.6 and 13.1.7: $\tau(G)$ is the minimum cover [minimum size of a subset of the vertices which hits every edge]; $\nu(G)$ is the size of a maximum matching, i. e., the maximum number of disjoint edges in G). This result is actually a special case of part (c) of Exercise 13.1.9.

(b) Prove: for every $\epsilon > 0$, for almost all graphs G, $\tau(G) > (2 - \epsilon)\nu(G)$. *Hint.* Relate τ to α. What inequality did we prove in class about $\alpha(G)$ for almost all graphs G?

Exercise 14.2.2. Recall: $S \subset \mathbb{R}^n$ is called a "2-distance set" if at most two numbers occur as distances between pairs of distinct points in S. Let $m_2(n)$ denote the maximum size of a 2-distance set in $\mathbb{R}^n$.

(a) Prove: $m_2(n) \geq \binom{n}{2}$.

(b) Prove: $m_2(n) \geq \binom{n+1}{2}$.

(c) Prove: $m_2(n) \leq (n+1)(n+4)/2$.

(d) Define $m_3(n)$ analogously. Prove: $m_3(n) = \Theta(n^3)$.

Exercise 14.2.3. (a) For even n, show that there exists a *maximal* set of Oddtown clubs consisiting of only 2 clubs. (Of course for $n > 2$ such a set will not be *maximum*.) How large is the smallest maximal set of Oddtown clubs for odd n?

(b) Show that every maximal set of clubs in Eventown is maximum.

14.3 2002 Midterm 3

Exercise 14.3.1. Let A be a $k \times n$ matrix over $\mathbb{R}$; let $b \in \mathbb{R}^k$, $c \in \mathbb{R}^n$, and let $x = (x_1, \ldots, x_n)$ be an n-tuple of real variables. Consider the linear programming problem

$$\max_x \{c^T x \mid Ax \leq b,\ x \geq 0\}.$$

(a) We refer to the above problem as the "primal" problem. How many variables does the *dual* problem have? Denote the vector of dual variables by $y = (y_1, \ldots)$. State the dual problem.

(b) Prove the easy part of the Duality Theorem, namely, that if both the primal and the dual problems are feasible (the constraints are satisfiable) then the primal optimum is $\leq$ the dual optimum.

(c) State a sufficient condition discussed in class, under which the primal problem is guaranteed to have an integral optimum (where all the x_i are integers).

(d) Prove that the signed (± 1) incidence matrix of a digraph is totally unimodular (i. e., the determinant of every square submatrix is 0 or ± 1).

(e) State two important "combinatorial duality theorems" (where the max of a quantity is equal to the min of a "dual" quantity). The first result should be about bipartite graphs, the second about network flows. State the theorems, not just their names. Define the concepts involved. You do not need to define "bipartite graph" but do define "network flow."

Exercise 14.3.2. Let G be a graph. Let $m_k(G)$ denote the number of k-matchings in G, i. e., the number of ways one can select k independent edges in G. Example: if $G = tK_2$ (the graph is a set of t disjoint edges) then $m_k = \binom{t}{k}$.

(a) Let $G = P_n$ (the path of length $n-1$; so P_n has n vertices and $n-1$ edges). Find the value of $m_k(P_n)$. Your answer should be a very simple expression involving a binomial coefficient.

(b) Find the quantity $a_n = \sum_{k=0}^{\lfloor n/2 \rfloor} m_k(P_n)$. Your answer should be a very simple expression involving a well-known sequence.

(c) The polynomial

$$\mu_G(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k m_k(G) x^{n-2k}$$

is called the *matchings polynomial* of G. A major result, published in 1972 by physicists O. J. Heilmann and E. H. Lieb, states that all roots of the matchings polynomial are real. Use this result and a result stated in a homework to prove that the sequence $m_0(G), m_1(G), \ldots$ is *unimodal* (increases until its maximum and then decreases).

Exercise 14.3.3. Prove that the expected number of k-cycles in a random permutation is $1/k$.

14.4 2003 Midterm 1

Exercise 14.4.1. Find the number of solutions to the equation $x_1 + \cdots + x_k = n$ in integers x_i satisfying the constraints $(\forall i)(x_i \geq 2)$. Your answer should be a very simple expression in terms of n and k (a binomial coefficient).

Exercise 14.4.2. Calculate the chromatic polynomials of the following graphs. Prove your answers. Your solutions should be *very simple*, just a couple of lines.

1. The complete graph K_n.
2. The path P_n of length $n-1$ (P_n has n vertices).
3. The graph K_n^-, the graph obtained from K_n by removing an edge. (K_n^- has n vertices.)

Exercise 14.4.3. 1. Let Ω denote the set of strings of length n over the alphabet $\{A, B, C\}$. What is the probability that the letter A occurs exactly k times in a random string from Ω? Your answer should be a simple closed-form expression (no summation or product symbols, no ellipses (dot-dot-dots).

2. What is the probability that the number of occurrences of A in a random string from Ω is divisible by 3? Your answer should be a very simple closed-form expression involving complex numbers. You do not need to get rid of the complex numbers.

14.5 2003 Midterm 2

14.6 2003 Midterm 3

Exercise 14.6.1. 1. Construct $n+1$ vectors in general position in $\mathbb{F}_2^n$. (Every n of the vectors must be linearly independent over $\mathbb{F}_2$.)

2. Prove: no set of $n+2$ vectors in $\mathbb{F}_2^n$ is in general position.

Exercise 14.6.2. Prove: for almost all graphs G on n vertices, $\alpha(G) \leq 1+2\log_2 n$. Here $\alpha(G)$ denotes the size of the largest independent set in G.

14.7 Misc Misc

Exercise 14.7.1. Let $v_1, \ldots, v_m \in \mathbb{Z}^n$ be integral vectors. Prove: if the v_i are linearly independent over $\mathbb{F}_2$ then they are linearly independent over $\mathbb{R}$.

Chapter 15

Solutions

15.1 2003 Midterm 1 Solutions

Solution to Exercise 14.4.1. Find the number of solutions to the equation $x_1 + \cdots + x_k = n$ in integers x_i satisfying the constraints $(\forall i)(x_i \geq 2)$. Your answer should be a very simple expression in terms of n and k (a binomial coefficient).

Answer. Let $y_i = x_i - 1$. Now $y_i \geq 1$ and $\sum_{i=1}^{k} y_i = n - k$. The integral solutions $(y_1, \ldots, y_k)$ of this system are in 1-1 correspondence with the solutions $(x_, \ldots, x_k)$ of the original system. The solutions $(y_1, \ldots, y_k)$ can be represented by $k-1$ dividers placed in the $n-k-1$ slots between $n-k$ balls; so the number of solutions is $\binom{n-k-1}{k-1}$.

Solution to Exercise 1.6.26. Let a_n, b_n be sequences of real numbers.

1. Define the relation $a_n \sim b_n$ (asymptotic equality).

   Answer. $\lim_{n\to\infty} a_n/b_n = 1$. For the purposes of this definition, we consider $0/0$ to be 1.

2. Define the relation $a_n \gtrsim b_n$ (greater than or asymptotically equal).

   Answer. $a_n \sim \max\{a_n, b_n\}$.

   Comment 15.1.1. The handouts are your most important reading. Please review the "Asymptotic notation" handout (and all other handouts). Here are some incorrect answers. "$\lim_{n\to\infty} a_n/b_n \geq 1$." One problem with this definition is that this limit may not exist even while $a_n > b_n$. This difficulty is eliminated by using the concept "lim inf" instead of "lim." The resulting definition will be correct if $a_n, b_n > 0$. But if we permit negative terms in the sequence, we run into trouble: the constant sequences $a_n = -2$, $b_n = -1$ satisfy this definition, even though we certainly do not want to say that "-2 is greater than or asymptotically equal to -1."

3. Based on your definitions, prove: if $a_n \gtrsim b_n$ and $b_n \gtrsim a_n$ then $a_n \sim b_n$.

   Answer. Let $c_n = \max\{a_n, b_n\}$. Then by definition we have $a_n \sim c_n \sim b_n$, therefore $a_n \sim b_n$ by the transitivity of the $\sim$ relation.

4. Assume $b_n \to \infty$ and $a_n \geq b_n^2 \ln b_n$. Prove: $b_n \lesssim c\sqrt{a_n/\ln a_n}$, where c is a constant. Determine the smallest value of c for which this statement follows from the assumptions.

   Answer. Since $a_n \to \infty$, we have $a_n \geq 0$ for $n \geq n_0$. For $n \geq n_0$, let x_n be the unique solution to the equation $a_n = x_n^2 \ln x_n$. The solution exists because the function $f(x) = x^2 \ln x$ is continuous and goes from 0 to ∞ while $0 \leq x < \infty$. The solution is unique because $f(x)$ is strictly increasing. For the same reason, $b_n \leq x_n$ and therefore $x_n \to \infty$.

   Taking logarithms, we obtain $\ln a_n = 2 \ln x_n + \ln \ln x_n \sim 2 \ln x_n$ (because $\ln x_n \to \infty$). Therefore $a_n \sim x_n^2 \ln a_n/2$, hence $b_n \leq x_n \sim \sqrt{2a_n/\ln a_n}$ and therefore $b_n \lesssim \sqrt{2a_n/\ln a_n}$. So $c = \sqrt{2}$ works, and nothing less could work since b_n can be chosen to equal x_n.

Solution to Exercise 13.1.23. Recall that a *sunflower with s petals* is a family of (not necessarily distinct) sets $B_1, \ldots, B_s$ such that $(\forall i \neq j)(B_i \cap B_j = \bigcap_{k=1}^{s} B_k)$.

Let $A_1, \ldots, A_m$ be *not necessarily distinct* sets of size $|A_i| \leq r$. Prove: if $m > r!(s-1)^{r+1}$ then there is a sunsflower with s petals among the A_i, i. e., $(\exists i_1, \ldots, i_s)(1 \leq i_1 < i_2 < \cdots < i_s)$ such that $A_{i_1}, \ldots, A_{i_s}$ form a sunflower.

Answer. In class we proved the Erdős–Rado Theorem which states that if all the A_i are *distinct* and $m > r!(s-1)^r$ then there exists a sunflower with s petals among the A_i. We outline two solutions to the midterm problem: one that *adapts the proof method* given in class; the other *uses the result* proved in class.

First solution. As in class, we proceed by induction on r. The base case is different: for $r = 1$ we now have $m > (s-1)^2$ but not all the singletons A_i are different. By the pigeon hole principle, either there is an A_i which is repeated $\geq s$ times (so we have a sunflower of s identical sets), or there are s distinct and therefore disjoint singletons, again giving us a sunflower with s petals. – The inductive step is identical with that given in class.

Second solution. We do *not* use induction, so r is any positive integer. If there are s identical sets among the A_i then they form a subflower with s petals. If every set occurs at most $s-1$ times then (again by the pigeon hole principle) there are at least $m/(s-1) > r!(s-1)^r$ *distinct* sets among the A_i, so a sunflower with s petals exists among these distinct sets by the Erdős–Rado Theorem.

Solution to Exercise 4.2.11. What is the expected number of runs of k heads in a string of n coin-flips? (A "run of k heads" means a string of k consecutive heads. Example: the string HHTHTTHHHT has 3 runs of 2 heads.) Prove your answer! *Hint.* Indicator variables.

Answer. (To be discussed in class.)

Solution to Exercise 14.4.2. Calculate the chromatic polynomials of the following graphs. Prove your answers. Your solutions should be *very simple*, just a couple of lines.

1. The complete graph K_n.

   Answer. Once the first j vertices have been colored (with j distinct colors), the number of choices for the color of the next vertex is $n-j$. So $p_{K_n}(x) = x(x-1)\cdot\cdots\cdot(x-n+1) = n!\binom{x}{n}$.

2. The path P_n of length $n-1$ (P_n has n vertices).

   Answer. There are x choices for the color of the first vertex; all subsequent vertices will have one color to avoid, so $p_{P_n}(x) = x(x-1)^{n-1}$.

3. The graph K_n^-, the graph obtained from K_n by removing an edge. (K_n^- has n vertices.)

   Answer. Let e be an edge of K_n; note that $K_n/e = K_{n-1}$. Now, applying the recurrence $p_G = p_{G-e} - p_{G/e}$ to $G = K_n$, we obtain

$$p_{K_n^-} = p_{K_n} + p_{K_{n-1}} = x(x-1)\cdot\cdots\cdot(x-n+1) + x(x-1)\cdot\cdots\cdot(x-n+2) =$$

$$x(x-1)\cdot\cdots\cdot(x-n+3)(x-n+2)^2.$$

Solution to Exercise 13.1.3. 1. Define Boolean functions in n variables.

   Answer. A Boolean function in n variables is a function $f:\ \{0,1\}^n \to \{0,1\}$.

2. Count the Boolean functions in n variables.

   Answer. 2^{2^n}.

3. (G only, 4 points) We say that a Boolean function f is *monotone* if

$$(\forall x_1,\ldots,x_n,y_1,\ldots,y_n)(\text{ if } (\forall i)(x_i \geq y_i) \text{ then } f(x_1,\ldots,x_n) \geq f(y_1,\ldots,y_n)).$$

   Let $M(n)$ denote the number of monotone Boolean functions in n variables. Let $S(n)$ denote the number of Sperner families of subsets of a universe of size n. Prove: $M(n) = S(n)$.

   Answer. For $A \subseteq [n]$ let us set $f(A) = f(x_A)$ where x_A is the indicator vector of A, i. e., $(x_A)_i = 1$ if $i \in A$ and $(x_A)_i = 0$ otherwise. A *min-term* of a monotone Boolean function is a set $A \subseteq [n]$ such that $f(A) = 1$ but $f(B) = 0$ for all proper subsets of A. It is clear by definition that the set of min-terms of f is a Sperner family; and every Sperner family is the set of min-terms of a unique Boolean function, namely, the function defined by setting $f(A) = 1$ if and only if A contains a min-term.

Solution to Exercise 14.4.3. 1. Let Ω denote the set of strings of length n over the alphabet $\{A, B, C\}$. What is the probability that the letter A occurs exactly k times in a random string from Ω? Your answer should be a simple closed-form expression (no summation or product symbols, no ellipses (dot-dot-dots).

Answer.

$$\binom{n}{k} 2^{n-k}/3^n.$$

2. What is the probability that the number of occurrences of A in a random string from Ω is divisible by 3? Your answer should be a very simple closed-form expression involving complex numbers. You do not need to get rid of the complex numbers.

Answer. $(3^n + (2+\omega)^n + (2+\omega^2)^n)/3^{n+1}$, where ω is a primitive 3rd root of unity. (Why is this so?)

15.2 2003 Midterm 2 Solutions

Solution to Exercise 4.4.16. You and the bank play the following game: you flip n coins: if ξ of them come up "Heads," you receive 2^ξ dollars.

1. You have to buy a ticket to play this game. What is the fair price of the ticket? *(Hint:* it is the expected amount you will receive.

Answer. $\xi = \sum_{i=1}^n \theta_i$ where θ_i indicates the event that the i-th die came up "Heads." Therefore $2^\xi = \prod_{i=1}^n 2^{\theta_i}$. Now the variables 2^{θ_i} are independent and $E(\theta_i) = (2^0+2^1)/2 = 3/2$; therefore $E(\xi) = \prod_{i=1}^n E(2^{\theta_i}) = (3/2)^n$, using the fact that the expected value of independent variables is multiplicative.

2. Prove: the probability that you break even (receive at least your ticket's worth) is exponentially small. *Hint:* At least how many "heads" do you need for you to break even?

Answer. To break even, you need $2^\xi \geq (3/2)^n$. Taking logarithms, we obtain $\xi \geq n(\log 3 - 1)$. Now $\log 3 - 1 > 1/2$ because $\log 3 > 3/2$ because $2 \log 3 = \log 9 > \log 8 = 3$. Let $\log 3 - 1 = 1/2 + c$; so $c > 0$. ($c \approx 0.08496$.) Now $P(\xi \geq n/2 + cn)$ decreases exponentially by Chernoff's inequelity, as seen in class. Here is a review. Let $\zeta_i = 2\theta_i - 1$, so $\zeta_i = 1$ if the i-th coin shows "Heads," and $\zeta_i = -1$ otherwise. Now $\sum_{i=1}^n \zeta_i = 2\xi - n$, so the event "$\xi \geq n/2 + cn$" is the same as the event $\sum_{i=1}^n \zeta_i \geq 2cn$; by Chernoff's bound, the probability of this event is less than $\mathrm{e}^{-(2cn)^2/2n} = \mathrm{e}^{-2c^2 n}$.

3. Calculate the standard deviation of the variable 2^ξ. Your answer should be a simple formula. Evaluate it asymptotically; obtain an even simpler formula.

Last update: October 12, 2004

Answer.

$$E((2^\xi)^2) = E(2^{2\xi}) = \prod_{i=1}^{n} E(2^{2\theta_i}) = \left(\frac{5}{2}\right)^n.$$

Moreover, $(E(2^\xi)^2) = ((3/2)^n)^2 = (9/4)^n$. Therefore $\mathrm{Var}(2^\xi) = (5/2)^n - (9/4)^n$ and the standard deviation of 2^ξ is $\sqrt{(5/2)^n - (9/4)^n} \sim (5/2)^{n/2}$.

4. State what the "weak law of large numbers" would say for the variable 2^ξ. Prove that the Law does NOT hold for this variable. *Hint.* This law talks about the probability that 2^ξ is not within $(1 \pm \epsilon)$-times its expectation.)

   Answer. The law would say that for all $\epsilon > 0$, the quantity $P(|2^\xi - (3/2)^n| \geq \epsilon(3/2)^n)$ goes to zero as $n \to \infty$. But in fact in our case this quantity goes to 1 at an exponential rate for *every* $\epsilon < 1$ because $2^{n/2} = o((3/2)^n)$ (why?) and the values of ξ are (exponentially) concentrated around $n/2$ by Chernoff's bound, so the event "$2^\xi > (1-\epsilon)(3/2)^n$" is exponentially unlikely (why?).

Solution to Exercise 13.1.26. Let $f(x_1, \ldots, x_n) = C_1 \wedge \cdots \wedge C_m$ be a 3-CNF formula, i. e., each clause C_i is an "OR" ("$\vee$") of three literals (Boolean variables and their negations). Prove: there exists a substitution (assignment of $(0,1)$-values to the x_i) which will satisfy at least $7m/8$ of the m clauses. *Hint.* Try a random substitution (flip a coin for each x_i). What is the expected number of clauses that are satisfied by this substitution? (A clause is "satisfied" if it evaluates to 1.)

Answer. The probability that a random substitution does not satisfy an "OR" of k literals is $1/2^k$ (because each literal must take value 0). Therefore, each C_i is satisfied with probability $7/8$. Let θ_i denote the indicator variable of the event that the clause C_i is satisfied; then $E(\theta_i) = 7/8$. If ξ denotes the number of clauses satisfied by a random substitution then $\xi = \sum_{i=1}^m \theta_i$ and therefore $E(\xi) = \sum_{i=1}^m E(\theta_i) = 7m/8$. Consequently the event "$\xi \geq 7m/8$" is not empty (otherwise $E(\xi)$ would be less than $7m/8$). In other words, "$\xi \geq 7m/8$" is *possible*, i. e., there exists an elementary event (assigment of the variables) which makes $\xi \geq 7m/8$.

Comment 15.2.1. This is a probabilistic proof of the *existence* of an appropriate assignment. The proof does not tell us how to *find* such an assignment more efficiently than by brute force (try all possibilities). DO: find such an assignment in polynomial time, i. e., the number of steps must be bounded by m^{const}.

Solution to Exercise 4.3.9. A vertex z is a "common neighbor" of vertices x and y in a graph G if both x and y are adjacent to z in G. Let $N(x,y)$ denote the number of common neighbors of x and y. Prove that the following statement is true for *almost all* graphs $G = (V, E)$ with n vertices:

$$(\forall x \neq y \in V)(0.24n < N(x,y) < 0.26n).$$

In other words, if p_n denotes the probability of the event described by the displayed formula then $\lim_{n\to\infty} p_n = 1$.

Answer. Let us first fix x and y and for all other vertices z, let θ_z indicate the event that z is a common neighbor of x and y. Then $N(x,y) = \sum_{z\neq x,y} \theta_z$. The variables θ_z are independent, therefore, by the standard Chernoff argument (see details below), for every $\epsilon > 0$, the quantity $q_n := P(|N(x,y) - (n-2)/4| \geq \epsilon n)$ approaches zero at an exponential rate. Consequently $\binom{n}{2} q_n \to 0$ (exponential decay beats polynomial growth). But for any $\epsilon < 0.01$, we have $1 - p_n < \binom{n}{2} q_n \to 0$, so $p_n \to 1$.

Here is the "standard Chernoff argument." $E(\theta_z) = 1/4$, therefore we consider the variable $\zeta_z = (4/3)(\theta_z - 1/4)$. Then the ζ_z are independent, $E(\zeta_z) = 0$ and $|\zeta_z| \leq 1$ so Chernoff's estimate applies to the sum $\psi = \sum_z \zeta_z$. It follows that for every $\delta > 0$, $P(|\psi| \geq \delta(n-2)) < 2\mathrm{e}^{-\delta^2(n-2)/2}$. (Why $(n-2)$?) Now $\psi = \sum_z \zeta_z = (4/3)\sum_z(\theta_z - 1/4) = (4/3)(N(x,y) - (n-2)/4)$. Now if $N(x,y)$ is not between $0.24n$ and $0.26n$ then for any $\epsilon < 0.01$, for sufficiently large n, we have $|N(x,y) - (n-2)/4| > \epsilon(n-2)$ and therefore $|\psi| > (4/3)\epsilon(n-2)$. Setting $\delta = (4/3)\epsilon$ we obtain the bound $1 - p_n < 2\mathrm{e}^{-(16/9)\epsilon^2(n-2)/2}$, which goes to zero at an exponential rate, as desired.

Solution to Exercise 7.2.7. Let $P(n)$ be the number of projective planes of order n. Prove: $P(n) < (n\mathrm{e})^{(n+1)^3}$. *Hint.* first prove that

$$P(n) \leq \binom{\binom{n^2+n+1}{n+1}}{n^2+n+1}.$$

Answer. The formula given in the Hint should be clear: to specify a projective plane of order n, we fix a set of n^2+n+1 points; each line is a subset of size $n+1$; so we select n^2+n+1 out of the $\binom{n^2+n+1}{n+1}$ subsets of size $n+1$. Using the estimate $\binom{a}{b} < (\mathrm{e}a/b)^b$, we obtain $\binom{n^2+n+1}{n+1} < (\mathrm{e}(n+1/n))^{n+1}$ and therefore, using the straightforward estimate $\binom{a}{b} < a^b$, we obtain $P(n) < (\mathrm{e}(n+1/n))^{(n+1)(n^2+n+1)}$. We claim that this number is less than $(\mathrm{e}n)^{(n+1)^3}$. Taking $(n+1)$-st roots, the inequality claimed is

$$(\mathrm{e}(n+1/n))^{n^2+n+1} < (\mathrm{e}n)^{n^2+2n+1} = (\mathrm{e}n)^n(\mathrm{e}n)^{n^2+n+1}.$$

Dividing the left hand side by the right hand side we obtain

$$(1+n^{-2})^{n^2+n+1}(\mathrm{e}n)^{-n} < (1+n^{-2})^{2n^2}(\mathrm{e}n)^{-n}.$$

Now $(1+n^{-2})^{2n^2} < \mathrm{e}^2$; so the quotient above is less than $\mathrm{e}^{-n+2}n^{-n} \leq n^{-n} < 1$.

Solution to Exercise 7.2.8. Let us consider the Galois plane $\mathrm{PG}(2,5)$ (over the field of 5 elements).

1. How many points does this plane have, and what is the number of points per line?

   Answer. The number of points is $5^2+5+1 = 31$; the number of points per line is $5+1 = 6$.

2. Points are given by "homogeneous coordinates." Determine whether or not the points given by $a = [1, 4, 0]$, $b = [3, 2, 2]$, and $c = [4, 1, 2]$ are collinear (belong to the same line). (Coordinates are mod 5.) Prove your answer.

   Answer. They are collinear.

   Proof. Three points are collinear if and only if the corresponding vectors are linearly dependent. Notice that in our case $a + b = c$ (computation mod 5), so the three points are collinear.

   . If you did not notice that $a + b = c$, you can try the determinant: the three vectors are linearly dependent if and only if their determinant is zero (mod 5). This can be done by Gaussian elimination.

   A simpler solution follows if we replace 4 by -1 and 3 by -2: $a = [1, -1, 0]$, $b = [-2, 2, 2]$, $c = [-1, 1, 2]$. Now we notice that the second column is (-1)-times the first column, so the determinant is zero.