

Computational Complexity: A Modern Approach

Sanjeev Arora and Boaz Barak
Princeton University

<http://www.cs.princeton.edu/theory/complexity/>
complexitybook@gmail.com

Not to be reproduced or distributed without the authors' permission

Appendix A

Mathematical Background.

This appendix reviews the mathematical notions used in this book. However, most of these are only used in few places, and so the reader might want to only quickly review Sections A.1 and A.2, and come back to the other sections as needed. In particular, apart from probability, the first part of the book essentially requires only comfort with mathematical proofs and some very basic notions of discrete math.

The topics described in this appendix are covered in greater depth in many texts and online sources. Almost all of the mathematical background needed is covered in a good undergraduate “discrete math for computer science” course as currently taught at many computer science departments. Some good sources for this material are the lecture notes by Papadimitriou and Vazirani [PV06], and the book of Rosen [Ros06].

The mathematical tool we use most often is discrete probability. Alon and Spencer [AS00b] is a great resource in this area. Also, the books of Mitzenmacher and Upfal [MU05] and Motwani and Raghavan [MR95] cover probability from a more algorithmic perspective.

Although knowledge of algorithms is not strictly necessary for this book, it would be quite useful. It would be helpful to review either one of the two recent books by Dasgupta et al [DPV06] and Kleinberg and Tardos [KT06] or the earlier text by Cormen et al [CLRS01]. This book does not require prior knowledge of computability and automata theory, but some basic familiarity with that theory could be useful: see Sipser’s book [Sip96] for an excellent introduction. See Shoup’s book [Sho05] for a computer-science introduction to algebra and number theory.

Perhaps *the* mathematical prerequisite needed for this book is a certain level of comfort with *mathematical proofs*. The fact that a mathematical proof has to be absolutely convincing does not mean that it has to be overly formal and tedious. It just has to be clearly written, and contain no logical gaps. When you write proofs try to be clear and concise, rather than using too much formal notation. Of course, to be absolutely convinced that some statement is true, we need to be certain of what that statement means. This is why there is a special emphasis in mathematics (and this book) on very precise *definitions*. Whenever you read a definition, try to make sure you completely understand it, perhaps by working through some simple examples. Oftentimes, understanding the meaning of a mathematical statement is more than half the work to prove that it is true.

A.1 Sets, Functions, Pairs, Strings, Graphs, Logic.

Sets. A *set* contains a finite or infinite number of elements, without repetition or respect to order, for example $\{2, 17, 5\}$, $\mathbb{N} = \{1, 2, 3, \dots\}$ (the set of natural numbers), $[n] = \{1, 2, \dots, n\}$ (the set of natural numbers from 1 to n), \mathbb{R} (the set of real numbers). For a finite set A , we denote by $|A|$ the number of elements in A . Some operations on sets are: (1) *union*: $A \cup B = \{x : x \in A \text{ or } x \in B\}$, (2) *intersection*: $A \cap B = \{x : x \in A \text{ and } x \in B\}$, and (3) *set difference*: $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$.

Functions. We say that f is a *function* from a set A to B , denoted by $f : A \rightarrow B$, if it maps any element of A into an element of B . If B and A are finite, then the number of possible functions from A to B is $|B|^{|A|}$. We say that f is *one to one* if for every $x, w \in A$ with $x \neq w$, $f(x) \neq f(w)$. If A, B are finite, the existence of such a function implies that $|A| \leq |B|$. We say that f is *onto* if for every $y \in B$ there exists $x \in A$ such that $f(x) = y$. If A, B are finite, the existence of such a function implies that $|A| \geq |B|$. We say that f is a *permutation* if it is both one-to-one and onto. For finite A, B , the existence of a permutation from A to B implies that $|A| = |B|$.

Pairs and tuples. If A, B are sets, then the $A \times B$ denotes the set of all ordered pairs $\langle a, b \rangle$ with $a \in A, b \in B$. Note that if A, B are finite then $|A \times B| = |A| \cdot |B|$. We can define similarly $A \times B \times C$ to be the set of ordered triples $\langle a, b, c \rangle$ with $a \in A, b \in B, c \in C$. For $n \in \mathbb{N}$, we denote by A^n the set $A \times A \times \cdots \times A$ (n times). We will often use the set $\{0, 1\}^n$, consisting of all length- n sequences of bits (i.e., length n strings), and the set $\{0, 1\}^* = \cup_{n \geq 0} \{0, 1\}^n$ ($\{0, 1\}^0$ has a single element: a binary string of length zero, which we call the empty word and denote by ε). As mentioned in Section 0.1 we can represent various objects (numbers, graphs, matrices, etc...) as binary strings, and use $\lfloor x \rfloor$ (not to be confused with the floor operator $\lfloor x \rfloor$) to denote the representation of x . Moreover, we often drop the $\lfloor \cdot \rfloor$ symbols and use x to denote both the object and its representation.

Graphs. A *graph* G consists of a set V of *vertices* (which we often assume is equal to the set $[n] = \{1, \dots, n\}$ for some $n \in \mathbb{N}$) and a set E of *edges*, which consists of unordered pairs (i.e., size two subsets) of elements in V . We denote the edge $\{u, v\}$ of the graph by \overline{uv} . For $v \in V$, the *neighbors* of v are all the vertices $u \in V$ such that $\overline{uv} \in E$. In a *directed graph*, the edges consist of *ordered pairs* of vertices, and to stress this we sometimes denote the edge $\langle u, v \rangle$ in a directed graph by \overrightarrow{uv} . One can represent an n -vertex graph G by its *adjacency matrix* which is an $n \times n$ matrix A such that $A_{i,j}$ is equal to 1 if the edge \overrightarrow{ij} is present in G and is equal to 0 otherwise. One can think of an undirected graph as a directed graph G that satisfies that for every u, v , G contains the edge \overrightarrow{uv} if and only if it contains the edge \overrightarrow{vu} . Hence, one can represent an undirected graph by an adjacency matrix that is *symmetric* ($A_{i,j} = A_{j,i}$ for every $i, j \in [n]$).

Boolean operators. A *Boolean variable* is a variable that can be either TRUE or FALSE (we sometimes identify TRUE with 1 and FALSE with 0). We can combine variables via the logical operations AND (\wedge), OR (\vee) and NOT (\neg , sometimes also denoted by an overline), to obtain *Boolean formulae*. For example, the following is a Boolean formulae on the variables u_1, u_2, u_3 : $(u_1 \wedge \overline{u_2}) \vee \neg(u_3 \wedge \overline{u_1})$. The definitions of the operations are the usual: $a \wedge b = \text{TRUE}$ if $a = \text{TRUE}$ and $b = \text{TRUE}$ and is equal to FALSE otherwise; $\overline{a} = \neg a = \text{TRUE}$ if $a = \text{FALSE}$ and is equal to FALSE otherwise; $a \vee b = \neg(\overline{a} \wedge \overline{b})$. We sometimes use other Boolean operators such as the XOR (\oplus) operator, but they can be always replaced with the equivalent expression using \wedge, \vee, \neg (e.g., $a \oplus b = (a \wedge \overline{b}) \vee (\overline{a} \wedge b)$). If φ is a formulae in n variables u_1, \dots, u_n , then for any *assignment* of values $u \in \{\text{FALSE}, \text{TRUE}\}^n$ (or equivalently, $\{0, 1\}^n$), we denote by $\varphi(u)$ the value of φ when its variables are assigned the values in u . We say that φ is *satisfiable* if there exists a u such that $\varphi(u) = \text{TRUE}$.

Quantifiers. We will often use the *quantifiers* \forall (for all) and \exists (exists). That is, if φ is a condition that can be TRUE or FALSE depending on the value of a variable x , then we write $\forall x \varphi(x)$ to denote the statement that φ is TRUE for *every* possible value that can be assigned to x . If A is a set then we write $\forall_{x \in A} \varphi(x)$ to denote the statement that φ is TRUE for every assignment for x from the set A . The quantifier \exists is defined similarly. Formally, we say that $\exists x \varphi(x)$ holds if and only if $\neg(\forall x \neg \varphi(x))$ holds.

Big-Oh Notation. We will often use the big-Oh notation (i.e., $O, \Omega, \Theta, o, \omega$) as defined in Section 0.3.

A.2 Probability theory

A *finite probability space* is a finite set $\Omega = \{\omega_1, \dots, \omega_N\}$ along with a set of numbers $p_1, \dots, p_N \in [0, 1]$ such that $\sum_{i=1}^N p_i = 1$. A random element is selected from this space by choosing ω_i with probability p_i . If x is chosen from the sample space Ω then we denote this by $x \in_r \Omega$. If no distribution is specified then we use the uniform distribution over the elements of Ω (i.e., $p_i = \frac{1}{N}$ for every i).

An *event* over the space Ω is a subset $A \subseteq \Omega$ and the *probability* that A occurs, denoted by $\Pr[A]$, is equal to $\sum_{i:\omega_i \in A} p_i$. To give an example, the probability space could be that of all 2^n possible outcomes of n tosses of a fair coin (i.e., $\Omega = \{0, 1\}^n$ and $p_i = 2^{-n}$ for every $i \in [2^n]$) and the event A can be that the number of coins that come up “heads” (or, equivalently, 1) is even. In this case, $\Pr[A] = 1/2$ (exercise). The following simple bound—called the *union bound*—is often used in the book. For every set of events A_1, A_2, \dots, A_n ,

$$\Pr[\cup_{i=1}^n A_i] \leq \sum_{i=1}^n \Pr[A_i]. \quad (1)$$

Inclusion exclusion principle. The union bound is a special case of a more general principle. Indeed, note that if the sets A_1, \dots, A_n are not *disjoint* then the probability of $\cup_i A_i$ could be smaller than $\sum_i \Pr[A_i]$ since we are overcounting elements that appear in more than one set. We can correct this by subtracting $\sum_{i < j} \Pr[A_i \cap A_j]$ but then we might be undercounting, since we subtracted elements that appear in at least 3 sets too many times. Continuing this process we get

Claim A.1 (*Inclusion-Exclusion principle*) For every A_1, \dots, A_n ,

$$\Pr[\cup_{i=1}^n A_i] = \sum_{i=1}^n \Pr[A_i] - \sum_{1 \leq i < j \leq n} \Pr[A_i \cap A_j] + \dots + (-1)^{n-1} \Pr[A_1 \cap \dots \cap A_n].$$

Moreover, this is an alternating sum which means that if we take only the first k summands of the right hand side, then this upper bounds the left-hand side if k is odd, and lower bounds it if k is even. \diamond

We sometimes use the following corollary of this claim, known as the Bonefforni Inequality:

Corollary A.2 For every events A_1, \dots, A_n ,

$$\Pr[\cup_{i=1}^n A_i] \geq \sum_{i=1}^n \Pr[A_i] - \sum_{1 \leq i < j \leq n} \Pr[A_i \cap A_j] \quad \diamond$$

A.2.1 Random variables and expectations.

A *random variable* is a mapping from a probability space to \mathbb{R} . For example, if Ω is as above (i.e., the set of all possible outcomes of n tosses of a fair coin), then we can denote by X the number of coins that came up heads.

The *expectation* of a random variable X , denoted by $E[X]$, is its weighted average. That is, $E[X] = \sum_{i=1}^N p_i X(\omega_i)$. The following simple claim follows from the definition:

Claim A.3 (*Linearity of expectation*) For X, Y random variables over a space Ω , denote by $X + Y$ the random variable that maps ω to $X(\omega) + Y(\omega)$. Then,

$$E[X + Y] = E[X] + E[Y] \quad \diamond$$

This claim implies that the random variable X from the example above has expectation $n/2$. Indeed $X = \sum_{i=1}^n X_i$ where X_i is equal to 1 if the i^{th} coin came up heads and is equal to 0 otherwise. But clearly, $E[X_i] = 1/2$ for every i .

For a real number α and a random variable X , we define αX to be the random variable mapping ω to $\alpha \cdot X(\omega)$. Note that $E[\alpha X] = \alpha E[X]$.

Example A.4

Suppose that we choose k random numbers x_1, \dots, x_k independently in $[n]$. What is the expected number of *collisions*: unordered pairs $\{i, j\}$ such that $x_i = x_j$? For every $i \neq j$, define the random variable $Y_{i,j}$ to equal 1 if $x_i = x_j$ and 0 otherwise. Since for every choice of x_i , the probability that $x_j = x_i$ is $1/n$, we have that $E[Y_{i,j}] = 1/n$. The number of collisions is the sum of $Y_{i,j}$ over all $i \neq j$ in $[k]$. Thus, by linearity of expectation the expected number of collisions is

$$\sum_{1 \leq i < j \leq k} E[Y_{i,j}] = \binom{k}{2} \frac{1}{n}.$$

This means that we expect at least one collision once $\binom{k}{2} \geq n$, which happens once k is larger than roughly $\sqrt{2n}$. This fact is often known as the *birthday paradox* because it explains the seemingly strange phenomenon that a class of more than 27 or so students is quite likely to have a pair of students sharing the same birthday, even though there are 365 days in the year.

Note that in contrast, if $k \ll \sqrt{n}$ then by the union bound, the probability there will be even one collision is at most $\binom{k}{2}/n \ll 1$.

Notes: (1) We sometimes also consider random variables whose range is not \mathbb{R} , but other sets such as \mathbb{C} or $\{0, 1\}^n$. (2) Also, we often identify a random variable X over the sample space Ω with the distribution $X(\omega)$ for $\omega \in_{\mathbb{r}} \Omega$. For example, we may use both $\Pr_{x \in_{\mathbb{r}} X}[x^2 = 1]$ and $\Pr[X^2 = 1]$ to denote the probability that for $\omega \in_{\mathbb{r}} \Omega$, $X(\omega)^2 = 1$.

A.2.2 The averaging argument

The following simple fact can be surprisingly useful:

The Averaging Argument: If a_1, a_2, \dots, a_n are some numbers whose average is c then some $a_i \geq c$.

Equivalently, we can state this in probabilistic terms as follows:

Lemma A.5 (“The Probabilistic Method”) If X is a random variable which takes values from a finite set and $E[X] = \mu$ then the event “ $X \geq \mu$ ” has nonzero probability. \diamond

The following two facts are also easy to verify

Lemma A.6 If $a_1, a_2, \dots, a_n \geq 0$ are numbers whose average is c then the fraction of a_i 's that are at least kc is at most $1/k$. \diamond

Lemma A.7 (“Markov’s inequality”) Any non-negative random variable X satisfies

$$\Pr(X \geq k E[X]) \leq \frac{1}{k}. \quad \diamond$$

Can we give any meaningful upper bound on the probability that X is much *smaller* than its expectation? Yes, if X is bounded.

Lemma A.8 If a_1, a_2, \dots, a_n are numbers in the interval $[0, 1]$ whose average is ρ then at least $\rho/2$ of the a_i 's are at least as large as $\rho/2$. \diamond

PROOF: Let γ be the fraction of i 's such that $a_i \geq \rho/2$. Then the average of the a_i 's is bounded by $\gamma \cdot 1 + (1 - \gamma)\rho/2$. Hence, $\rho \leq \gamma + \rho/2$, implying $\gamma \geq \rho/2$. ■

More generally, we have

Lemma A.9 *If $X \in [0, 1]$ and $E[X] = \mu$ then for any $c < 1$ we have*

$$\Pr[X \leq c\mu] \leq \frac{1 - \mu}{1 - c\mu}.$$

Example A.10

Suppose you took a lot of exams, each scored from 1 to 100. If your average score was 90 then in at least half the exams you scored at least 80.

A.2.3 Conditional probability and independence

If we already know that an event B happened, this reduces the space from Ω to $\Omega \cap B$, where we need to scale the probabilities by $1/\Pr[B]$ so they will sum up to one. Thus, the probability of an event A *conditioned* on an event B , denoted $\Pr[A|B]$, is equal to $\Pr[A \cap B]/\Pr[B]$ (where we always assume that B has positive probability).

We say that two events A, B are *independent* if $\Pr[A \cap B] = \Pr[A] \Pr[B]$. Note that this implies that $\Pr[A|B] = \Pr[A]$ and $\Pr[B|A] = \Pr[B]$. We say that a set of events A_1, \dots, A_n are *mutually independent* if for every subset $S \subseteq [n]$,

$$\Pr[\bigcap_{i \in S} A_i] = \prod_{i \in S} \Pr[A_i]. \quad (2)$$

We say that A_1, \dots, A_n are *k-wise independent* if (2) holds for every $S \subseteq [n]$ with $|S| \leq k$.

We say that two random variables X, Y are *independent* if for every $x, y \in \mathbb{R}$, the events $\{X = x\}$ and $\{Y = y\}$ are independent. We generalize similarly the definition of mutual independence and k -wise independence to sets of random variables X_1, \dots, X_n . We have the following claim:

Claim A.11 *If X_1, \dots, X_n are mutually independent then*

$$E[X_1 \cdots X_n] = \prod_{i=1}^n E[X_i] \quad \diamond$$

PROOF:

$$\begin{aligned} E[X_1 \cdots X_n] &= \sum_x x \Pr[X_1 \cdots X_n = x] = \\ &= \sum_{x_1, \dots, x_n} x_1 \cdots x_n \Pr[X_1 = x_1 \text{ and } X_2 = x_2 \cdots \text{ and } X_n = x_n] = \text{(by independence)} \\ &= \sum_{x_1, \dots, x_n} x_1 \cdots x_n \Pr[X_1 = x_1] \cdots \Pr[X_n = x_n] = \\ &= \left(\sum_{x_1} x_1 \Pr[X_1 = x_1] \right) \left(\sum_{x_2} x_2 \Pr[X_2 = x_2] \right) \cdots \left(\sum_{x_n} x_n \Pr[X_n = x_n] \right) = \prod_{i=1}^n E[X_i] \end{aligned}$$

where the sums above are over all the possible real numbers that can be obtained by applying the random variables or their products to the finite set Ω . ■

A.2.4 Deviation upper bounds

Under various conditions, one can give better upper bounds on the probability of a random variable “straying too far” from its expectation. These upper bounds are usually derived by clever use of Markov’s inequality.

The *variance* of a random variable X is defined to be $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}(X))^2]$. Note that since it is the expectation of a non-negative random variable, $\text{Var}[X]$ is always non-negative. Also, using linearity of expectation, we can derive that $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. The *standard deviation* of a variable X is defined to be $\sqrt{\text{Var}[X]}$.

The first bound is Chebyshev’s inequality, useful when only the variance is known.

Lemma A.12 (*Chebyshev inequality*) *If X is a random variable with standard deviation σ , then for every $k > 0$,*

$$\Pr[|X - \mathbb{E}[X]| > k\sigma] \leq 1/k^2 \quad \diamond$$

PROOF: Apply Markov’s inequality to the random variable $(X - \mathbb{E}[X])^2$, noting that by definition of variance, $\mathbb{E}[(X - \mathbb{E}[X])^2] = \sigma^2$. ■

Chebyshev’s inequality is often useful in the case that X is equal to $\sum_{i=1}^n X_i$ for pairwise independent random variables X_1, \dots, X_n . This is because of the following claim, that is left as an exercise:

Claim A.13 *If X_1, \dots, X_n are pairwise independent then*

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) \quad \diamond$$

The next inequality has many names, and is widely known in theoretical computer science as the *Chernoff bound* (see also Note 7.11). It considers scenarios of the following type. Suppose we toss a fair coin n times. The expected number of heads is $n/2$. How tightly is this number concentrated? Should we be very surprised if after 1000 tosses we have 625 heads? The bound we present is slightly more general, since it concerns n different coin tosses of possibly different expectations (the expectation of a coin is the probability of obtaining “heads”; for a fair coin this is $1/2$). These are sometimes known as Poisson trials.

Theorem A.14 (*“Chernoff” bounds*) *Let X_1, X_2, \dots, X_n be mutually independent random variables over $\{0, 1\}$ (i.e., X_i can be either 0 or 1) and let $\mu = \sum_{i=1}^n \mathbb{E}[X_i]$. Then for every $\delta > 0$,*

$$\Pr\left[\sum_{i=1}^n X_i \geq (1 + \delta)\mu\right] \leq \left[\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}}\right]^\mu. \quad (3)$$

$$\Pr\left[\sum_{i=1}^n X_i \leq (1 - \delta)\mu\right] \leq \left[\frac{e^{-\delta}}{(1 - \delta)^{(1 - \delta)}}\right]^\mu. \quad (4)$$

◇

Often, we will only use the following corollary:

Corollary A.15 *Under the above conditions, for every $c > 0$*

$$\Pr\left[\left|\sum_{i=1}^n X_i - \mu\right| \geq c\mu\right] \leq 2 \cdot e^{-\min\{c^2/4, c/2\}\mu}.$$

In particular this probability is bounded by $2^{-\Omega(\mu)}$ (where the constant in the Ω notation depends on c). ◇

PROOF: Surprisingly, the Chernoff bound is also proved using the Markov inequality. We only prove the first inequality; the second inequality can be proved similarly. We introduce a positive dummy variable t , and observe that

$$\mathbb{E}[\exp(tX)] = \mathbb{E}\left[\exp\left(t \sum_i X_i\right)\right] = \mathbb{E}\left[\prod_i \exp(tX_i)\right] = \prod_i \mathbb{E}[\exp(tX_i)], \quad (5)$$

where $\exp(z)$ denotes e^z and the last equality holds because the X_i r.v.s are independent. Now,

$$\mathbb{E}[\exp(tX_i)] = (1 - p_i) + p_i e^t,$$

therefore,

$$\begin{aligned} \prod_i \mathbb{E}[\exp(tX_i)] &= \prod_i [1 + p_i(e^t - 1)] \leq \prod_i \exp(p_i(e^t - 1)) \\ &= \exp\left(\sum_i p_i(e^t - 1)\right) = \exp(\mu(e^t - 1)), \end{aligned} \tag{6}$$

as $1 + x \leq e^x$. Finally, apply Markov's inequality to the random variable $\exp(tX)$, viz.

$$\Pr[X \geq (1 + \delta)\mu] = \Pr[\exp(tX) \geq \exp(t(1 + \delta)\mu)] \leq \frac{\mathbb{E}[\exp(tX)]}{\exp(t(1 + \delta)\mu)} = \frac{\exp((e^t - 1)\mu)}{\exp(t(1 + \delta)\mu)},$$

using (5), (6) and the fact that t is positive. Since t is a dummy variable, we can choose any positive value we like for it. Simple calculus shows that the right hand side is minimized for $t = \ln(1 + \delta)$ and this leads to the theorem statement. ■

So, if all n coin tosses are fair (Heads has probability $1/2$) then the the probability of seeing N heads where $|N - n/2| > a\sqrt{n}$ is at most $2e^{-a^2/4}$. In particular, the chance of seeing at least 625 heads in 1000 tosses of an unbiased coin is less than 5.3×10^{-7} .

A.2.5 Some other inequalities.

Jensen's inequality.

The following inequality, generalizing the inequality $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$, is also often useful:

Lemma A.16 (*Jensen's Inequality*) A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is convex if for every $p \in [0, 1]$ and $x, y \in \mathbb{R}$, $f(px + (1 - p)y) \leq p \cdot f(x) + (1 - p) \cdot f(y)$. For every random variable X and convex function f , $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$. ◇

Approximating the binomial coefficient

Of special interest is the *Binomial* random variable B_n denoting the number of coins that come up "heads" when tossing n fair coins. For every k , $\Pr[B_n = k] = 2^{-n} \binom{n}{k}$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ denotes the number of size- k subsets of $[n]$. Clearly, $\binom{n}{k} \leq n^k$, but sometimes we will need a better estimate for $\binom{n}{k}$ and use the following approximation:

Claim A.17 For every $n, k < n$, $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$ ◇

The best approximation can be obtained via Stirling's formula:

Lemma A.18 (*Stirling's formula*) For every n ,

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \quad \diamond$$

It can be proven by taking natural logarithms and approximating $\ln n! = \ln(1 \cdot 2 \cdots n) = \sum_{i=1}^n \ln i$ by the integral $\int_1^n \ln x \, dx = n \ln n - n + 1$. It implies the following corollary:

Corollary A.19 For every $n \in \mathbb{N}$ and $\alpha \in [0, 1]$,

$$\binom{n}{\alpha n} = (1 \pm O(n^{-1})) \frac{1}{\sqrt{2\pi n \alpha(1-\alpha)}} 2^{H(\alpha)n}$$

where $H(\alpha) = \alpha \log(1/\alpha) + (1 - \alpha) \log(1/(1 - \alpha))$ and the constants hidden in the O notation are independent of both n and α . ◇

More useful estimates.

The following inequalities can be obtained via elementary calculus:

- For every $x \geq 1$, $(1 - \frac{1}{x})^x \leq \frac{1}{e} \leq (1 - \frac{1}{x+1})^x$
- For every k , $\sum_{i=1}^n i^k = \Theta\left(\frac{n^{k+1}}{k+1}\right)$
- For every $k > 1$, $\sum_{i=1}^{\infty} n^{-k} < O(1)$.
- For every $c, \epsilon > 0$, $\sum_{i=1}^{\infty} \frac{n^c}{(1+\epsilon)^n} < O(1)$.
- For every n , $\sum_{i=1}^n \frac{1}{n} = \ln n \pm O(1)$

A.2.6 Statistical distance

The following notion of when two distributions are close to one another is often very useful.

Definition A.20 (*Statistical Distance*) Let Ω be some finite set. For two random variables X and Y with range Ω , their *statistical distance* (also known as *variation distance*) is defined as $\Delta(X, Y) = \max_{S \subseteq \Omega} \{|\Pr[X \in S] - \Pr[Y \in S]|\}$. \diamond

Some texts use the name *total variation distance* for the statistical distance. The next lemma gives some useful properties of this distance:

Lemma A.21 *Let X, Y, Z be any three distributions taking values in the finite set Ω . Then,*

1. $\Delta(X, Y) \in [0, 1]$ where $\Delta(X) = \Delta(Y)$ iff X is identical to Y .
2. (*Triangle inequality*) $\Delta(X, Z) \leq \Delta(X, Y) + \Delta(Y, Z)$.
3. $\Delta(X, Y) = \frac{1}{2} \sum_{x \in \Omega} |\Pr[X = x] - \Pr[Y = x]|$.
4. $\Delta(X, Y) \geq \epsilon$ iff there is a Boolean function $f: \Omega \rightarrow \{0, 1\}$ such that $|E[f(X)] - E[f(Y)]| \geq \epsilon$.
5. For every finite set Ω' and function $f: \Omega \rightarrow \Omega'$, $\Delta(f(X), f(Y)) \leq \Delta(X, Y)$. (*Here $f(X)$ is a distribution on Ω' obtained by taking a sample of X and applying f .*)

Note that Item 3 means that $\Delta(X, Y)$ is equal to the L_1 -distance of X and Y divided by 2 (see Section A.5.4 below). That is, if we think of X as a vector in \mathbb{R}^Ω where $X_\omega = \Pr[X = \omega]$, and define for every vector $\mathbf{v} \in \mathbb{R}^\Omega$, $|\mathbf{v}|_1 = \sum_{\omega \in \Omega} |\mathbf{v}_\omega|$, then $\Delta(X, Y) = 1/2 |X - Y|_1$.

PROOF OF LEMMA A.21: We start with Item 3. For every pairs of distributions X, Y over $\{0, 1\}^n$ let S be the set of strings x such that $\Pr[X = x] > \Pr[X = y]$. Then it is easy to see that this choice of S maximizes the quantity $b(S) = \Pr[X \in S] - \Pr[Y \in S]$ and in fact $b(S) = \Delta(X, Y)$ since if we had a set T with $b(T) < -b(S)$ then the complement \bar{T} of T would satisfy $b(\bar{T}) > b(S)$. But,

$$\begin{aligned} \sum_{x \in \{0,1\}^n} |\Pr[X = x] - \Pr[Y = x]| &= \\ \sum_{x \in S} \Pr[X = x] - \Pr[Y = x] + \sum_{x \notin S} \Pr[Y = x] - \Pr[X = x] &= \\ \Pr[X \in S] - \Pr[Y \in S] + (1 - \Pr[Y \in S]) - (1 - \Pr[X \in S]) &= \\ 2\Pr[X \in S] - 2\Pr[Y \in S], & \end{aligned}$$

establishing Item 3.

The triangle inequality (Item 2) follows immediately from Item 3 since $\Delta(X, Y) = 1/2|X - Y|_1$ and the L_1 norm satisfies the triangle inequality. Item 3 also implies Item 1 since $|X - Y|_1 = 0$ iff $X = Y$ and $|X - Y|_1 \leq \|X\| + \|Y\|_1 = 1 + 1$.

Item 4 is just a rephrasing of the definition of statistical distance, identifying a set $S \subseteq \{0, 1\}^n$ with the function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ such that $f(x) = 1$ iff $x \in S$. Item 5 follows from Item 4 noting that if $\Delta(X, Y) \leq \epsilon$ then $|\mathbb{E}[g(f(X))] - \mathbb{E}[g(f(Y))]| \leq \epsilon$ for every function g . ■

A.3 Number theory and groups

The *integers* are the set $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ while the natural numbers are the subset $\mathbb{N} = \{0, 1, 2, \dots\}$.¹ A basic fact is that we can *divide* any integer n by a nonzero integer k to obtain ℓ, r such that $n = k\ell + r$ and $r \in \{0, \dots, k-1\}$. If $r = 0$ then we say that k *divides* n and denote this by $k|n$. The *factors* of n are the set of positive integers that divide n .

The *greatest common divisor* of two integers n, m , denoted by $\gcd(n, m)$ is the largest integer d such that $d|n$ and $d|m$. We say that n and m are *co-prime* if their greatest common divisor is equal to 1. The following basic facts are not hard to verify:

- If a nonzero integer c divides both n and m then $c|d$.
- The greatest common divisor of n and m is the smallest positive integer d such that there exist integers x, y satisfying $nx + my = d$.
- There is a polynomial-time (i.e., $\text{polylog}(n, m)$ -time) algorithm that on input n, m outputs the greatest common divisor d of n, m and the integers x, y satisfying $nx + my = d$. (This algorithm is known as *Euclid's Algorithm*.)

A number $p > 1$ is *prime* if its only factors are 1 and p . The following basic facts are known about prime numbers:

- Every positive integer n can be written uniquely (up to ordering) as a product of prime numbers. This is called the *prime factorization* of n .
- If $\gcd(p, a) = 1$ and $p|ab$ then $p|b$. In particular, if a prime p divides $a \cdot b$ then either $p|a$ or $p|b$.

A fundamental question in number theory is how many primes exist. A celebrated result is:

Theorem A.22 (*The Prime Number Theorem (Hadamard, de la Vallée Poussin 1896)*) For $n > 1$, let $\pi(n)$ denote the number of primes between 1 and n then

$$\pi(n) = \frac{n}{\ln n} (1 \pm o(1)) \quad \diamond$$

The original proofs of the prime number theorem used rather deep mathematical tools, and in fact people have conjectured that this is *inherently* the case. But in 1949 both Erdős and Selberg (independently) found elementary proofs for this theorem. For most computer science applications, the following weaker statement proven by Chebychev suffices:

Theorem A.23 $\pi(n) = \Theta\left(\frac{n}{\log n}\right)$ ◇

PROOF: Consider the number $\binom{2n}{n} = \frac{2n!}{n!n!}$. By Stirling's formula we know that $\log \binom{2n}{n} = (1 - o(1))2n$ and in particular $n \leq \log \binom{2n}{n} \leq 2n$. Also, all the prime factors of $\binom{2n}{n}$ are

¹Some texts exclude 0 in \mathbb{N} ; in most cases this does not any difference.

between 0 and $2n$, and each factor p cannot appear more than $k = \left\lfloor \frac{\log 2n}{\log p} \right\rfloor$ times. Indeed, for every n , the number of times p appears in the factorization of $n!$ is $\sum_i \left\lfloor \frac{n}{p^i} \right\rfloor$, since we get $\left\lfloor \frac{n}{p} \right\rfloor$ times a factor p in the factorizations of $\{1, \dots, n\}$, $\left\lfloor \frac{n}{p^2} \right\rfloor$ times a factor of the form p^2 , etc... Thus the number of times p appears in the factorization of $\binom{2n}{n} = \frac{(2n)!}{n!n!}$ is equal to $\sum_i \left\lfloor \frac{2n}{p^i} \right\rfloor - 2 \left\lfloor \frac{n}{p^i} \right\rfloor$: a sum of at most k elements (since $p^{k+1} > 2n$) each of which is either 0 or 1.

Thus, $\binom{2n}{n} \leq \prod_{\substack{1 \leq p \leq 2n \\ p \text{ prime}}} p^{\left\lfloor \frac{\log 2n}{\log p} \right\rfloor}$. Taking logs we get that

$$n \leq \log \binom{2n}{n} \leq \sum_{\substack{1 \leq p \leq 2n \\ p \text{ prime}}} \left\lfloor \frac{\log 2n}{\log p} \right\rfloor \log p \leq \sum_{\substack{1 \leq p \leq 2n \\ p \text{ prime}}} \log 2n = \pi(2n) \log 2n,$$

establishing $\pi(n) = \Omega\left(\frac{n}{\log n}\right)$.

To prove that $\pi(n) = O\left(\frac{n}{\log n}\right)$, we define the function $\vartheta(n) = \sum_{\substack{1 \leq p \leq n \\ p \text{ prime}}} \log p$. It suffices to prove that $\vartheta(n) = O(n)$ (exercise!). But since all the primes between $n+1$ and $2n$ divide $\binom{2n}{n}$ at least once, $\binom{2n}{n} \geq \prod_{\substack{n+1 \leq p \leq 2n \\ p \text{ prime}}} p$. Taking logs we get

$$2n \geq \log \binom{2n}{n} \geq \sum_{\substack{n+1 \leq p \leq 2n \\ p \text{ prime}}} \log p = \vartheta(2n) - \vartheta(n),$$

thus getting a recursive equation $\vartheta(2n) \leq \vartheta(n) + 2n$ which solves to $\vartheta(n) = O(n)$. ■

A.3.1 Groups.

A *group* is an abstraction that captures some properties of mathematical objects such as the integers, matrices, functions and more. Formally, a group is a set that has a binary operation, say \star , defined on it that is associative and has an inverse. That is, (G, \star) is a group if

1. For every $a, b, c \in G$, $(a \star b) \star c = a \star (b \star c)$
2. There exists a special element $id \in G$ such that $a \star id = a$ for every $a \in G$, and for every $a \in G$ there exists $b \in G$ such that $a \star b = b \star a = id$. (This element b is called the *inverse* of a , and is often denote as a^{-1} or $-a$.)

Examples for groups are the integers, with addition being the group operation (and zero the identity element), the non-zero real numbers with multiplication being the group operation (and one the identity element), and the set of functions from a domain A to itself, with function composition being the group operation.

Often, it is natural to use additive (+) or multiplicative (\cdot) notation to denote the group operation rather than \star . In these cases we will use la (or respectively a^ℓ) to denote the result of applying the operation to a ℓ times.

A.3.2 Finite groups

A group is *finite* if it has a finite number of elements. We denote by $|G|$ the number of elements of G . Examples for finite groups are the following:

- The group \mathbb{Z}_n of the integers from 0 to $n-1$ with the operation being addition modulo n . In particular \mathbb{Z}_2 is the set $\{0, 1\}$ with the XOR operation.

- The group S_n of the permutations on $[n]$, with the operation being function composition.
- The group $(\mathbb{Z}_2)^n$ of n -bit strings with the operation being bitwise XOR. More generally for every two groups G and H , we can define the group $G \times H$ to be a group whose elements are pairs $\langle g, h \rangle$ with $g \in G$ and $h \in H$ and with the group operation corresponding to applying the group operations of G and H componentwise. Similarly, we define G^n to be the group $G \times G \times \cdots \times G$ (n times).
- For every n , the group \mathbb{Z}_n^* consists of the set $\{k : 1 \leq k \leq n - 1, \gcd(k, n) = 1\}$ and the operation of multiplication modulo n . Note that if $\gcd(k, n) = 1$ then there exist x, y such that $kx + ny = 1$ or in other words $kx = 1 \pmod{n}$, meaning that x is the inverse of k modulo n . This also means that we can find this inverse in polynomial time using Euclid's algorithm. The size of \mathbb{Z}_n^* is denoted by $\varphi(n)$ and the function φ is known as *Euler's Quotient function*. Note that if n is prime then $\varphi(n) = n - 1$. It is known that for every $n > 6$, $\varphi(n) \geq \sqrt{n}$.

A *subgroup* of G is a subset of G that is itself a group (i.e., closed under the group operation and taking inverses). The following result is often quite useful

Theorem A.24 *If G is a finite group and H is a subgroup of G then $|H|$ divides $|G|$.* \diamond

PROOF: Consider the family of sets of the form $aH = \{ah : h \in H\}$ for all $a \in G$ (we're using here multiplicative notation for the group). It is easy to see that the map $x \mapsto ax$ is one-to-one and hence $|aH| = |H|$ for every a . Hence it will suffice to show that we can partition G into disjoint sets from this family. Yet this family clearly covers G (as $a \in aH$ for every $a \in G$) and hence it suffices to show that for every a, b either $aH = bH$ or aH and bH are disjoint. Indeed, suppose that there exist $x, y \in H$ such that $ax = by$ then for every element $az \in aH$, we have that $az = (byx^{-1})z$ and since $yx^{-1}z \in H$ we get that $az \in bH$. \blacksquare

Corollary A.25 (*Fermat's Little Theorem*) *For every n and $x \in \{1, \dots, n - 1\}$, $x^{\varphi(n)} = 1 \pmod{n}$. In particular, if n is prime then $x^{n-1} = 1 \pmod{n}$.* \diamond

PROOF: Consider the set $H = \{x^\ell : \ell \in \mathbb{Z}\}$. This is clearly a subgroup of \mathbb{Z}_n^* and hence $|H|$ divides $\varphi(n)$. But the size of H is simply the smallest number k such that $x^k = 1 \pmod{n}$. Indeed, there must be such a number since, because \mathbb{Z}_n^* , if we consider the sequence of numbers $1, x, x^2, x^3, \dots$ then eventually we get i, j such that $x^i = x^j$ for $i < j$, meaning that $x^{i-j} = 1 \pmod{n}$. Thus, the above sequence looks like $1, x, x^2, \dots, x^{k-1}, 1, x, x^2, \dots$, meaning that $|H| = k$.

Since $x^{|H|} = 1 \pmod{n}$, obviously taking x to the power $\varphi(n)$ (which is a multiple of $|H|$) yields also 1 modulo n . \blacksquare

The *order* of an element x of a group G is the smallest integer k such that x^k is equal to the identity element. The proof above shows that in a finite group G , every element has a finite order and furthermore this order divides the size of G . An element x of G with order $|G|$ is called a *generator* of G , since in this case the subgroup $\{x, x^1, x^2, \dots\}$ is all of G .² If a group G has a generator then we say that G is *cyclic*. An example for a simple cyclic group is the group \mathbb{Z}_n of the numbers $\{0, \dots, n - 1\}$ with addition modulo n , that is generated by the element 1 (and also by any other element that is co-prime to n — exercise).

A.3.3 The Chinese Remainder Theorem

Let $n = pq$ where p, q are co-prime. The Chinese Remainder Theorem (CRT) says that the group \mathbb{Z}_n^* (multiplicative group modulo n) is isomorphic to the group $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$ (pairs of numbers with multiplication done componentwise modulo p and q respectively).

²A more general definition (that works also for infinite groups) is that x is a generator of G if the subgroup $\{x^\ell : \ell \in \mathbb{Z}\}$ is equal to G .

Theorem A.26 *If $n = pq$ where p, q coprime then function f that maps x to $\langle x \pmod{p}, x \pmod{q} \rangle$ is one-to-one on \mathbb{Z}_n^* . Furthermore f is an isomorphism in the sense that $f(xy) = f(x)f(y)$ (where multiplication on the left hand side is modulo n and on the right hand side is componentwise modulo p and q respectively). \diamond*

PROOF: The furthermore part can be easily verified and so we focus on showing that f is one-to-one. We need to show that if $f(x) = f(x')$ then $x = x'$. Since $f(x - x') = f(x) - f(x')$, it suffices to show that if $x = 0 \pmod{p}$ (i.e., $p|x$) and $x = 0 \pmod{q}$ (i.e., $q|x$) then $x = 0 \pmod{n}$ (i.e., $pq|x$). Yet, assume that $p|x$ and write $x = pk$. Then since $\gcd(p, q) = 1$ and $q|x$ we know that $q|k$, meaning that $pq|x$. ■

The Chinese Remainder Theorem can be easily generalized to show that for every $n = p_1 p_2 \dots p_k$, where all the p_i 's are co-prime, there is an isomorphism between \mathbb{Z}_n^* to $\mathbb{Z}_{p_1}^* \times \dots \times \mathbb{Z}_{p_k}^*$, meaning that for every n , the group \mathbb{Z}_n^* is isomorphic to a product of groups of the form \mathbb{Z}_q^* for q a prime power (i.e., number of the form p^ℓ for prime p). In fact, it can be generalized even further to show that every Abelian group G is isomorphic to a product $G_1 \times G_2 \times \dots \times G_k$ where all the G_i 's are cyclic. (This can be viewed as a generalization of the CRT because all the groups of the form \mathbb{Z}_q^* for q a power of an odd prime are cyclic, and all groups of the form $\mathbb{Z}_{2^k}^*$ are either cyclic or products of two cyclic groups.)

A.4 Finite fields

A *field* is a set \mathbb{F} that has an addition (+) and multiplication (\cdot) operations that behave in the expected way: satisfy associative, commutative and distributive laws, have both additive and multiplicative inverses, and neutral elements 0 and 1 for addition and multiplication respectively. In other words, \mathbb{F} is a field if it is an Abelian group with the operation + and an identity element 0, and has an additional operation \cdot such that $\mathbb{F} \setminus \{0\}$ and \cdot forms an Abelian group, and furthermore the two operation satisfy the distributive rule $a(b + c) = ab + ac$.

Familiar fields are the real numbers (\mathbb{R}), the rational numbers (\mathbb{Q}) and the complex numbers (\mathbb{C}), but there are also *finite* fields. Recall that for a prime p , the set $\{0, \dots, p - 1\}$ is an Abelian group with the addition modulo p operation and the set $\{1, \dots, p - 1\}$ is an Abelian group with the multiplication modulo p operation. Hence $\{0, \dots, p - 1\}$ form a field with these two operations, which we denote by $\text{GF}(p)$. The simplest example for such a field is the field $\text{GF}(2)$ consisting of $\{0, 1\}$ where multiplication is the AND (\wedge) operation and addition is the XOR operation.

Every finite field \mathbb{F} has a number ℓ such that for every $x \in \mathbb{F}$, $x + x + \dots + x$ (ℓ times) is equal to the zero element of \mathbb{F} (exercise). This number ℓ is called the *characteristic* of \mathbb{F} . For every prime q , the characteristic of $\text{GF}(q)$ is equal to q .

A.4.1 Non-prime fields.

One can see that if n is not prime, then the set $\{0, \dots, n - 1\}$ with addition and multiplication modulo n is not a field, as there exist two non-zero elements x, y in this set such that $x \cdot y = n = 0 \pmod{n}$. Nevertheless, there are finite fields of size n for non-prime n . Specifically, for every prime q , and $k \geq 1$, there exists a field of q^k elements, which we denote by $\text{GF}(q^k)$. We will very rarely need to use such fields in this book, but still provide an outline of their construction below.

For every prime q and k there exists an *irreducible* degree k polynomial P over the field $\text{GF}(q)$ (P is irreducible if it cannot be expressed as the product of two polynomials P', P'' of lower degree). We then let $\text{GF}(q^k)$ be the set of all $k - 1$ -degree polynomials over $\text{GF}(q)$. Each such polynomial can be represented as a vector of its k coefficients. We perform both addition and multiplication modulo the polynomial P . Note that addition corresponds to standard vector addition of k -dimensional vectors over $\text{GF}(q)$, and both addition and multiplication can be easily done in $\text{poly}(n, \log q)$ time (we can reduce a polynomial S

modulo a polynomial P using a similar algorithm to long division of numbers). It turns out that no matter how we choose the irreducible polynomial P , we will get the same field, up to renaming of the elements. There is a deterministic $\text{poly}(q, k)$ -time algorithm to obtain an irreducible polynomial of degree k over $\text{GF}(q)$. There are also probabilistic algorithms (and deterministic algorithms whose analysis relies on unproven assumptions) that obtain such a polynomial in $\text{poly}(\log q, k)$ time (see the book [Sho05]).

For us, the most important example of a finite field is $\text{GF}(2^k)$, which consists of the set $\{0, 1\}^k$, with addition being component-wise XOR, and multiplication being polynomial multiplication via some irreducible polynomial which we can find in $\text{poly}(k)$ time. In fact, we will mostly not even be interested in the multiplicative structure of $\text{GF}(2^k)$ and only use the addition operation (i.e., use it as the vector space $\text{GF}(2)^k$, see below).

A.5 Basic facts from linear algebra

For \mathbb{F} a field and $n \in \mathbb{N}$, we denote by \mathbb{F}^n the set of n -length tuples (or *vectors*) of elements of \mathbb{F} . If $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ and $x \in \mathbb{F}$ then we denote by $\mathbf{u} + \mathbf{v}$ the vector obtained by componentwise addition of \mathbf{u} and \mathbf{v} and by $x\mathbf{u}$ the vector obtained by multiplying each entry of \mathbf{u} by x .

A set of vectors $\mathbf{u}^1, \dots, \mathbf{u}^k$ in \mathbb{F}^n is *linearly independent* if the only solution to the equation $x_1\mathbf{u}^1 + \dots + x_k\mathbf{u}^k = \mathbf{0}$ (where $\mathbf{0}$ denotes the all-zero vector) is $x_1 = x_2 = \dots = x_k = 0$. It can be shown that if $\mathbf{u}^1, \dots, \mathbf{u}^k$ are linearly independent then $k \leq n$ (exercise). A set of n linearly independent vectors in \mathbb{F}^n is called a *basis* of \mathbb{F}^n . It is not hard to see that if $\mathbf{u}^1, \dots, \mathbf{u}^n$ is a basis of \mathbb{F}^n then every vector $\mathbf{v} \in \mathbb{F}^n$ can be expressed as a linear combination $\mathbf{v} = \sum_i x_i \mathbf{u}^i$ of the vectors $\mathbf{u}^1, \dots, \mathbf{u}^n$ and furthermore this expression is unique. The *standard basis* of \mathbb{F}^n is the set $\mathbf{e}^1, \dots, \mathbf{e}^n$, where \mathbf{e}_j^i is equal to 1 if $j = i$ and 0 otherwise.

A subset $S \subseteq \mathbb{F}^n$ is called a *subspace* if it is closed under addition and scalar multiplication (i.e., $\mathbf{u}, \mathbf{v} \in S$ and $x, y \in \mathbb{F}$ implies that $x\mathbf{u} + y\mathbf{v} \in S$). The *dimension* of S , denoted by $\dim(S)$ is defined to be the maximum number k such that there are k linearly independent vectors in S . Such a set of $\dim(S)$ linearly independent vectors in S is called a *basis* and one can see that every vector in S can be expressed as a linear combination of the vectors in the basis.

A function $f : \mathbb{F}^n \rightarrow \mathbb{F}^m$ is *linear* if $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$. It's not hard to verify that the following hold for every linear function f :

- If $\mathbf{u}^1, \dots, \mathbf{u}^n$ is a basis for \mathbb{F}^n then for every $\mathbf{v} \in \mathbb{F}^n$, $f(\mathbf{v}) = \sum_i x_i f(\mathbf{u}_i)$ where x_1, \dots, x_n are the elements such that $\mathbf{v} = \sum x_i \mathbf{u}^i$. Thus, to know f 's value at every point it suffices to know its value on the basis elements.
- The set $\text{Im}(f) = \{f(\mathbf{v}) : \mathbf{v} \in \mathbb{F}^n\}$ is a subspace of \mathbb{F}^m .
- The set $\text{Ker}(f) = \{\mathbf{v} : f(\mathbf{v}) = \mathbf{0}\}$ is a subspace of \mathbb{F}^n .
- $\dim(\text{Im}(f)) + \dim(\text{Ker}(f)) = n$

A linear function $f : \mathbb{F}^n \rightarrow \mathbb{F}^m$ is often described by an $m \times n$ matrix A whose i^{th} column is $f(\mathbf{e}^i)$. The *multiplication* of an $m \times n$ matrix A and an $n \times k$ matrix B is the $m \times k$ matrix $C = AB$ where $C_{i,j} = \sum_{\ell \in [n]} A_{i,\ell} B_{\ell,j}$. One can verify that if A describes a function $f : \mathbb{F}^n \rightarrow \mathbb{F}^m$ and B describes a function $g : \mathbb{F}^k \rightarrow \mathbb{F}^n$ then C describes the function $h : \mathbb{F}^k \rightarrow \mathbb{F}^m$ mapping \mathbf{v} to $f(g(\mathbf{v}))$. It can also be verified that if we identify members of \mathbb{F}^n with $n \times 1$ matrices (i.e., column vectors) then $f(\mathbf{v}) = A\mathbf{v}$.

The *determinant* of an $n \times n$ matrix A , denoted by $\det(A)$ is equal to $\sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} \prod_{i=1}^n A_{i,\sigma(i)}$ where S_n denotes the group of permutations over $[n]$ and $\text{sgn}(\sigma)$ is equal to 1 if the number of pairs $\langle i, j \rangle$ such that $i < j$ but $\sigma(i) > \sigma(j)$ is odd, and is equal to 0 otherwise. We have the following two facts:

- $\det(AB) = \det(A) \det(B)$. This can be verified by direct computation.

- If A is an upper triangular matrix (i.e., $A_{i,j} = 0$ whenever $i > j$) then $\det(A) = A_{1,1}A_{2,2}\cdots A_{n,n}$. Indeed, for a permutation σ to give a non-zero contribution to the determinant in this case it must satisfy $\sigma(i) \geq i$ for every i , which means that it is the identity permutation.

Together these two rules give a polynomial-time algorithm to compute the determinant of a matrix A by following the well known Gaussian elimination algorithm to express A as $E_1E_2\cdots E_mD$ where the E_i 's are elementary matrices (multiplication by which corresponds to switching two columns, multiplying a column by a field element, or adding one column to another) and the D is upper diagonal. Since the determinant is easy to compute for all these matrices, we can compute the determinant of A as well.

The following lemma relates the determinant of a matrix to the function it represents:

Lemma A.27 For a function $f : \mathbb{F}^n \rightarrow \mathbb{F}^n$ represented by an $n \times n$ matrix A , the following conditions are equivalent:

- The columns of A are a basis for \mathbb{F}^n .
- f is one-to-one.
- $\dim(\text{Im}(f)) = n$.
- $\dim(\text{Ker}(f)) = 0$.
- $\det(A) \neq 0$.
- There exists $\mathbf{v} \in \mathbb{F}^n$ such that the equation $A\mathbf{x} = \mathbf{v}$ has exactly one solution.
- For every $\mathbf{v} \in \mathbb{F}^n$, the equation $A\mathbf{x} = \mathbf{v}$ has exactly one solution.

Furthermore, if f is one-to-one then the mapping f^{-1} is linear and is represented by an $n \times n$ matrix A^{-1} whose $(i, j)^{\text{th}}$ entry is $\frac{\det(A_{-(i,j)})}{\det(A)}$, where $A_{-(i,j)}$ denotes the $(n-1) \times (n-1)$ matrix obtained by removing the i^{th} row and j^{th} column from A . \diamond

A.5.1 Inner product

The vector spaces \mathbb{R}^n and \mathbb{C}^n have an additional structure that is often quite useful.³ An *inner product* over \mathbb{C}^n to be a function mapping two vectors \mathbf{u}, \mathbf{v} to a complex number $\langle \mathbf{u}, \mathbf{v} \rangle$ satisfying the following conditions:

- $\langle x\mathbf{u} + y\mathbf{w}, \mathbf{v} \rangle = x\langle \mathbf{u}, \mathbf{v} \rangle + y\langle \mathbf{w}, \mathbf{v} \rangle$
- $\langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}$ where \bar{z} denotes complex conjugation (i.e., if $z = a + ib$ then $\bar{z} = a - ib$).
- For every \mathbf{u} , $\langle \mathbf{u}, \mathbf{u} \rangle$ is a non-negative real number with $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ iff $\mathbf{u} = \mathbf{0}$.

The two examples for inner products we will use are the standard inner product mapping $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ to $\sum_{i=1}^n \mathbf{x}_i \bar{\mathbf{y}}_i$ and the expectation or normalized inner product mapping $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ to $\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \bar{\mathbf{y}}_i$. We can also define inner products over the space \mathbb{R}^n , in which case we drop the conjugation.

If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ we say that \mathbf{u} and \mathbf{v} are *orthogonal* and denote this by $\mathbf{u} \perp \mathbf{v}$. We have the following result:

Lemma A.28 If non-zero vectors $\mathbf{u}^1, \dots, \mathbf{u}^k$ satisfy $\mathbf{u}^i \perp \mathbf{u}^j$ for all $i \neq j$ then they are linearly independent. \diamond

³The reason we restrict ourselves to these fields is that they have characteristic zero which means that there does not exist a number $k \in \mathbb{N}$ and nonzero $a \in \mathbb{F}$ such that $ka = 0$ (where ka is the result of adding a to itself k times). You can check that if there is such a number for a field \mathbb{F} then there will not be an inner product over \mathbb{F}^n .

PROOF: Suppose that $\sum_i x_i \mathbf{u}^i = \mathbf{0}$ and consider take an inner product of this vector with itself. We get that

$$0 = \left\langle \sum_i x_i \mathbf{u}^i, \sum_j x_j \mathbf{u}^j \right\rangle = \sum_{i,j} x_i \bar{x}_j \langle \mathbf{u}^i, \mathbf{u}^j \rangle = \sum_i |x_i|^2 \langle \mathbf{u}^i, \mathbf{u}^i \rangle, \quad (7)$$

where the last equality follows from the fact that $\langle \mathbf{u}^i, \mathbf{u}^j \rangle = 0$ for $i \neq j$. But unless all the x_i 's are zero, the righthand side of (7) is strictly positive. (Recall that for a complex number $x = a + ib$, $|x| = \sqrt{a^2 + b^2}$ and $|x|^2 = x\bar{x}$.) ■

A set $\mathbf{u}^1, \dots, \mathbf{u}^n$ of nonzero vectors in \mathbb{C}^n satisfying $\langle \mathbf{u}^i, \mathbf{u}^j \rangle = 0$ for $i \neq j$ is called an *orthogonal basis* of \mathbb{C}^n . If in addition $\langle \mathbf{u}^i, \mathbf{u}^i \rangle = 1$ for all i then we say this is an *orthonormal basis*. An orthonormal basis consists of n linearly independent vectors and hence as its name implies is a basis of \mathbb{C}^n , meaning that every vector \mathbf{v} can be expressed as $\mathbf{v} = \sum_i x_i \mathbf{u}^i$. By taking an inner product of this equality with \mathbf{u}^i , one can see that $x_i = \langle \mathbf{v}, \mathbf{u}^i \rangle$.

The following identity (that can be viewed as a generalization of the Pythagorean theorem) is often useful:

Lemma A.29 (*Parseval's identity*) If $\mathbf{u}^1, \dots, \mathbf{u}^n$ is an orthonormal basis for \mathbb{C}^n , then for every \mathbf{v} ,

$$\langle \mathbf{v}, \mathbf{v} \rangle = \sum_{i=1}^n |x_i|^2,$$

where x_1, \dots, x_n are the numbers such that $\mathbf{v} = \sum_i x_i \mathbf{u}^i$. ◇

PROOF: As in the proof of Lemma A.28,

$$\langle \mathbf{v}, \mathbf{v} \rangle = \left\langle \sum_i x_i \mathbf{u}^i, \sum_j x_j \mathbf{u}^j \right\rangle = \sum_i |x_i|^2 \langle \mathbf{u}^i, \mathbf{u}^i \rangle. \quad \blacksquare$$

Vector spaces with an inner product are known as *Hilbert spaces*.

A.5.2 Dot product

Even in a field \mathbb{F} that doesn't have an inner product, we can define the *dot product* of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$, denoted by $\mathbf{u} \odot \mathbf{v}$, as $\sum_{i=1}^n \mathbf{u}_i \mathbf{v}_i$. For every subspace $S \subseteq \mathbb{F}^n$, we define $S^\perp = \{\mathbf{u} : \mathbf{u} \odot \mathbf{v} = 0 \forall \mathbf{v} \in S\}$. We leave the following simple claim as an exercise:

Claim A.30 $\dim(S) + \dim(S^\perp) = n$ ◇

In particular for every nonzero vector $\mathbf{u} \in \mathbb{F}^n$, the subspace \mathbf{u}^\perp of vectors \mathbf{v} satisfying $\mathbf{u} \odot \mathbf{v} = 0$ has dimension $n - 1$ and hence cardinality $|\mathbb{F}|^{n-1}$. As a corollary we get the following very useful fact:

Claim A.31 (*The random subsum principle*) For every nonzero $\mathbf{u} \in \text{GF}(2)$ (the field $\{0, 1\}$ with addition and multiplication modulo 2):

$$\Pr_{\mathbf{v} \in_r \text{GF}(2)^n} [\mathbf{u} \odot \mathbf{v} = 0] = 1/2 \quad \diamond$$

A.5.3 Eigenvectors and eigenvalues

If A is an $n \times n$ complex matrix and $\mathbf{v} \in \mathbb{C}^n$ is a nonzero vector, we say that \mathbf{v} is an *eigenvector* of A if there exists $\lambda \in \mathbb{C}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. We say that A is *diagonalizable* if there is a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of eigenvectors for A . In other words, there is an invertible matrix P such that PAP^{-1} is a diagonal matrix.

Note that A has an eigenvector with eigenvalue λ if and only if the matrix $A - \lambda I$ is non-invertible, where I is the identity matrix. Thus in particular λ is a root of the polynomial $p(x) = \det(A - xI)$. Thus the fundamental Theorem of Algebra (that every complex polynomial has as many roots as the degree) that every square matrix has at least one eigenvector. (A non-invertible matrix has an eigenvector zero.)

For a matrix A , the *conjugate transpose* of A , denoted A^* , is the matrix such that for every i, j , $A_{i,j}^* = \overline{A_{j,i}}$ where $\overline{}$ denotes the complex conjugate operation. We say that an $n \times n$ matrix A is *Hermitian* if $A = A^*$. An Hermitian matrix with only real entries is called *symmetric*. That is, a real matrix is symmetric if $A = A^\dagger$ where \dagger is the transpose operation (i.e., $A_{i,j}^\dagger = A_{j,i}$). An equivalent condition (exercise) is that A is Hermitian if and only if

$$\langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle. \quad (8)$$

An important useful fact about Hermitian matrices is the following theorem:

Theorem A.32 *If A is an $n \times n$ Hermitian matrix then there exists an orthogonal basis of eigenvectors for A .* \diamond

PROOF: We prove this by induction on n . We know that A has one eigenvector \mathbf{v} with eigenvalue λ . Now let $S = \mathbf{v}^\perp$ be the $n - 1$ dimensional space of all vectors orthogonal to \mathbf{v} . We claim that for every $\mathbf{u} \in S$, $A\mathbf{u} \in S$. Indeed, if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then

$$\langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle = \overline{\lambda} \langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

Thus the restriction of A to S is an $n - 1$ dimensional linear operator satisfying (8) and hence by induction this restriction has an orthogonal basis of eigenvectors $\mathbf{v}^2, \dots, \mathbf{v}^n$. Adding \mathbf{v} to this set we get an n -dimensional orthogonal basis of eigenvectors for A . \blacksquare

Note that if A is real and symmetric then all its eigenvalues must be real also (with no imaginary components). Indeed, if $A\mathbf{v} = \lambda\mathbf{v}$ then

$$\lambda \langle \mathbf{v}, \mathbf{v} \rangle = \langle A\mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, A\mathbf{v} \rangle = \overline{\lambda} \langle \mathbf{v}, \mathbf{v} \rangle,$$

meaning that for a nonzero \mathbf{v} , $\lambda = \overline{\lambda}$. This implies that the eigenvectors, that are obtained by solving a linear equation with real coefficients, are also real.

A.5.4 Norms

A *norm* of a vector in \mathbb{C}^n is a function mapping a vector \mathbf{v} to a real number $\|\mathbf{v}\|$ satisfying:

- For every \mathbf{v} , $\|\mathbf{v}\| \geq 0$ with $\|\mathbf{v}\| = 0$ iff $\mathbf{v} = \mathbf{0}$.
- If $x \in \mathbb{C}$ then $\|x\mathbf{v}\| = |x|\|\mathbf{v}\|$.
- (Triangle inequality) For every \mathbf{u}, \mathbf{v} , $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

For every $\mathbf{v} \in \mathbb{C}^n$ and number $p \geq 1$, the L_p norm of \mathbf{v} , denoted $\|\mathbf{v}\|_p$, is equal to $(\sum_{i=1}^n |\mathbf{v}_i|^p)^{1/p}$. One particularly interesting case is $p = 2$, the so-called *Euclidean norm*, in which $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^n |\mathbf{v}_i|^2} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. Another interesting case is $p = 1$, where we use the single bar notation and denote $\|\mathbf{v}\|_1 = \sum_{i=1}^n |\mathbf{v}_i|$. Another case is $p = \infty$, where we denote $\|\mathbf{v}\|_\infty = \lim_{p \rightarrow \infty} \|\mathbf{v}\|_p = \max_{i \in [n]} |\mathbf{v}_i|$.

Some relations between the different norms can be derived from the *Hölder inequality*, stating that for every p, q with $\frac{1}{p} + \frac{1}{q} = 1$, $\|\mathbf{u}\|_p \|\mathbf{v}\|_q \geq \sum_{i=1}^n |\mathbf{u}_i \mathbf{v}_i|$. To prove it, note that by simple scaling, it suffices to consider norm one vectors, and so it enough to show that if $\|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1$ then $\sum_{i=1}^n |\mathbf{u}_i| |\mathbf{v}_i| \leq 1$. But $\sum_{i=1}^n |\mathbf{u}_i| |\mathbf{v}_i| = \sum_{i=1}^n |\mathbf{u}_i|^{p(1/p)} |\mathbf{v}_i|^{q(1/q)} \leq \sum_{i=1}^n \frac{1}{p} |\mathbf{u}_i|^p + \frac{1}{q} |\mathbf{v}_i|^q = \frac{1}{p} + \frac{1}{q} = 1$, where the last inequality uses the fact that for every $a, b > 0$ and $\alpha \in [0, 1]$, $a^\alpha b^{1-\alpha} \leq \alpha a + (1 - \alpha)b$.

The Hölder inequality implies the following relations between the L_2 , L_1 and L_∞ norms of every vector (see Exercise 21.2):

$$\|\mathbf{v}\|_1 / \sqrt{n} \leq \|\mathbf{v}\|_2 \leq \sqrt{\|\mathbf{v}\|_1 \|\mathbf{v}\|_\infty} \quad (9)$$

Vector spaces with a norm are sometimes known as *Banach spaces*.

A.5.5 Metric spaces

For any set Ω and $d : \Omega^2 \rightarrow \mathbb{R}$, we say that d is a *metric* on Ω if it satisfies the following conditions:

1. $d(x, y) \geq 0$ for every $x, y \in \Omega$ where $d(x, y) = 0$ if and only if $x = y$.
2. $d(x, y) = d(y, x)$ for every $x, y \in \Omega$.
3. (Triangle Inequality) For every $x, y, z \in \Omega$, $d(x, z) \leq d(x, y) + d(y, z)$.

That is, $d(x, y)$ denotes the *distance* between x and y according to some measure. If Ω is a vector space with a norm then the function $d(x, y) = \|x - y\|$ is a metric over Ω , but there are other examples for metrics that do not come from any norm. For example, for every graph G we can define a metric over the vertex set of G by letting the distance of x and y be the length of the shortest path between them. Various metric spaces and the relations between them have found recently many applications in theoretical computer science, see Chapter 15 of [Mat02] for a good survey.

A.6 Polynomials

We list some basic facts about univariate polynomials.

Theorem A.33 *A nonzero polynomial of degree d has at most d distinct roots.* ◇

PROOF: Suppose $p(x) = \sum_{i=0}^d c_i x^i$ has $d + 1$ distinct roots $\alpha_1, \dots, \alpha_{d+1}$ in some field \mathbb{F} . Then

$$\sum_{i=0}^d \alpha_j^i \cdot c_i = p(\alpha_j) = 0,$$

for $j = 1, \dots, d + 1$. This means that the system $\mathbf{A}\mathbf{y} = \mathbf{0}$ with

$$\mathbf{A} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^d \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^d \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \alpha_{d+1} & \alpha_{d+1}^2 & \dots & \alpha_{d+1}^d \end{pmatrix}$$

has a solution $\mathbf{y} = \mathbf{c}$. The matrix \mathbf{A} is a *Vandermonde* matrix, and it can be shown that

$$\det \mathbf{A} = \prod_{i>j} (\alpha_i - \alpha_j),$$

which is nonzero for distinct α_i . Hence $\text{rank} \mathbf{A} = d + 1$. The system $\mathbf{A}\mathbf{y} = \mathbf{0}$ has therefore only a trivial solution — a contradiction to $\mathbf{c} \neq \mathbf{0}$. ■

This theorem has an interesting corollary:

Corollary A.34 *For every finite field \mathbb{F} , the multiplicative group \mathbb{F}^* is cyclic.* ◇

PROOF: The fact that the polynomial $x^k - 1$ has at most k roots implies that the group \mathbb{F}^* has the property (*) that for every k the number of elements x satisfying $x^k = 1$ is always at most k . We will prove by induction that every group G satisfying (*) is cyclic.

Let $n = |G|$. We consider three cases:

- n is prime. In this case every element of G has either order 1 or order n . Since the only element with order 1 is the identity element, we see that G has an element of order n — G is cyclic.

- $n = p^c$ for some prime p and $c > 1$. In this case if there is no element of order n , then all the orders must divide p^{c-1} . We get $n = p^c$ elements x such that $x^{p^{c-1}} = 1$, violating (*).
- $n = pq$ for co-prime p and q . In this case let H and F be two subgroups of G defined as follows: $H = \{a : a^p = 1\}$ and $F = \{b : b^q = 1\}$. Then $|H| \leq p < n$ and $|F| \leq q < n$ and also as subgroups of G both H and F satisfy (*). Thus by the induction hypothesis both H and F are cyclic and have generators a and b respectively. We claim that ab generates the entire group G . Indeed, let c be any element in G . Since p, q are coprime, there are x, y such that $xq + yp = 1$ and hence $c = c^{xq+yp}$. But $(c^{xq})^p = 1$ and $(c^{yp})^q = 1$ and hence c is a product of an element of H and an element of F , and hence $c = a^i b^j$ for some $i \in \{0, \dots, p-1\}$ and $j \in \{0, \dots, q-1\}$. Thus, to show that $c = (ab)^z$ for some z all we need to do is to find z such that $z = i \pmod{p}$ and $z = j \pmod{q}$, but this can be done using the Chinese Remainder Theorem.

■

Theorem A.35 For any set of pairs $(a_1, b_1), \dots, (a_{d+1}, b_{d+1})$ there exists a unique polynomial $g(x)$ of degree at most d such that $g(a_i) = b_i$ for all $i = 1, 2, \dots, d+1$. \diamond

PROOF: The requirements are satisfied by *Lagrange Interpolating Polynomial*:

$$\sum_{i=1}^{d+1} b_i \cdot \frac{\prod_{j \neq i} (x - a_j)}{\prod_{j \neq i} (a_i - a_j)}.$$

If two polynomials $g_1(x), g_2(x)$ satisfy the requirements then their difference $p(x) = g_1(x) - g_2(x)$ is of degree at most d , and is zero for $x = a_1, \dots, a_{d+1}$. Thus, from the previous theorem, polynomial $p(x)$ must be zero and polynomials $g_1(x), g_2(x)$ identical. ■

The following elementary result is usually attributed to Schwartz and Zippel in the computer science community, though it was certainly known earlier (see e.g. DeMillo and Lipton [DLne]).

Lemma A.36 If a polynomial $p(x_1, x_2, \dots, x_m)$ over $F = GF(q)$ is nonzero and has total degree at most d , then

$$\Pr[p(a_1..a_m) \neq 0] \geq 1 - \frac{d}{q},$$

where the probability is over all choices of $a_1..a_m \in F$. \diamond

PROOF: We use induction on m . If $m = 1$ the statement follows from Theorem A.33. Suppose the statement is true when the number of variables is at most $m-1$. Then p can be written as

$$p(x_1, x_2, \dots, x_m) = \sum_{i=0}^d x_1^i p_i(x_2, \dots, x_m),$$

where p_i has total degree at most $d-i$. Since p is nonzero, at least one of p_i is nonzero. Let k be the largest i such that p_i is nonzero. Then by the inductive hypothesis,

$$\Pr_{a_2, a_3, \dots, a_m} [p_i(a_2, a_3, \dots, a_m) \neq 0] \geq 1 - \frac{d-k}{q}.$$

Whenever $p_i(a_2, a_3, \dots, a_m) \neq 0$, $p(x_1, a_2, a_3, \dots, a_m)$ is a nonzero univariate polynomial of degree k , and hence becomes 0 only for at most k values of x_1 . Hence

$$\Pr[p(a_1..a_m) \neq 0] \geq (1 - \frac{k}{q})(1 - \frac{d-k}{q}) \geq 1 - \frac{d}{q},$$

and the induction is completed. ■