Appendices
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 the section on probability, 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. A gentle introduction to these notions appears in Chapter 0 of Sipser’s book [SIP96].

The topics described in the appendix are covered in greater depth in many places. The lecture notes by Lehman and Leighton [LL06] cover the majority of the math topics needed for this book (and some more). The books of Mitzenmacher and Upfal [MU05] and Prabhakar and Raghavan [?] cover both algorithmic reasoning and probability. For more insight on discrete probability, see the book by Alon and Spencer [AS00]. A fuller coverage of algorithms appears in the recent book of Kleinberg and Tardos [KT06] or the earlier text of Cormen at 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.

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

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, \ldots\} (the set of natural numbers), \{n\} = \{1, 2, \ldots, 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: \(A \cup B = \ldots\)
We denote the edge which consists of unordered pairs (i.e., size two subsets) of elements in a single element: a binary string of length zero, which we call the empty word and denote by $\varepsilon$.

A graph $G$ consists of a set $V$ of vertices (which we often assume is equal to the set $[n] = \{1, \ldots, n\}$ 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 $uv$. For $v \in V$, the neighbors of $v$ are all the vertices $u \in V$ such that $uv \in E$. In a directed graph, the edges consist of ordered pairs of vertices, to stress this we sometimes denote the edge $(u, v)$ in a directed graph by $u \rightarrow v$. 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 $ij$ is present in $G$ $i^{th}$ 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 $u \rightarrow v$ if and only if it contains the edge $v \rightarrow u$.

Hence, one can represent an undirected graph by an adjacency matrix that is symmetric ($A_{i,j} = A_{j,i}$ for every $i, j \in [n]$).

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 ($\land$), OR ($\lor$) and NOT ($\neg$, sometimes also denoted by an overline), to obtain Boolean formulae. For example, the following is a Boolean formulae on the variables $u_1, u_2, u_3$: $(u_1 \land \overline{u}_2) \lor \neg(u_3 \land \overline{u}_1)$. The definitions of the operations are the usual: $a \land b = \text{True}$ if $a = \text{True}$ and $b = \text{True}$ and is equal to False oth-
A.2. Probability theory 519

otherwise; \( \overline{a} = \neg a = \text{True} \) if \( a = \text{False} \) and is equal to \( \text{False} \) otherwise; \( a \lor b = \neg (\overline{a} \lor \overline{b}) \). If \( \varphi \) is a formulae in \( n \) variables \( u_1, \ldots, u_n \), 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 \( \varphi \) when its variables are assigned the values in \( u \).

We say that \( \varphi \) is satisfiable if there exists a \( u \) such that \( \varphi(u) = \text{True} \).

We will often use the quantifiers \( \forall \) (for all) and \( \exists \) (exists). That is, if \( \varphi \) is a condition that can be \( \text{True} \) or \( \text{False} \) depending on the value of a variable \( x \), then we write \( \forall x \varphi(x) \) to denote the statement that \( \varphi \) is \( \text{True} \) 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 \( \varphi \) is \( \text{True} \) 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.

A.2 Probability theory

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

An event over the space \( \Omega \) 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 \( A_1, A_2, \ldots, A_n \),

\[
\Pr[\cup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]. \tag{1}
\]

Inclusion exclusion principle. The union bound is a special case of a more general principle. Indeed, note that if the sets \( A_1, \ldots, A_n \) 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
A.2. PROBABILITY THEORY

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

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

Moreover, this is an alternating sum which means that if we take only the first $k$ summands of the right hand side, then this upperbounds the left-hand side if $k$ is odd, and lowerbounds it if $k$ is even.

We sometimes use the following corollary of this claim:

Claim A.2
For every events $A_1, \ldots, A_n$,

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

Random subsum principle. The following fact is used often in the book:

Claim A.3 (The random subsum principle)
For $x, y \in \{0, 1\}^n$, denote $x \circ y = \sum_{i=1}^n x_i y_i \pmod{2}$ (that is, $x \circ y$ is equal to 1 if the number of $i$'s such that $x_i = 1$ and $y_i = 1$ is odd and equal to 0 otherwise). Then for every $y \neq 0^n$,

$$\Pr_{x \in \{0, 1\}^n} [x \circ y = 1] = \frac{1}{2}.$$ 

Proof: Suppose that $y_j$ is nonzero. We can think of choosing $x$ as follows: first choose all the coordinates of $x$ other than the $j^{th}$ and only choose the $j^{th}$ coordinate last. After we choose all the coordinates of $x$ other than the $j^{th}$, the value $\sum_{i \neq j} x_i y_i \pmod{2}$ is fixed to be some $c \in \{0,1\}$. Regardless of what $c$ is, with probability $1/2$ we choose $x_j = 0$, in which case $x \circ y = c$ and with probability $1/2$ we choose $x_j = 1$, in which case $x \circ y = 1-c$. We see that in any case $x \prod y$ will be equal to 1 with probability $1/2$. ■

A.2.1 Random variables and expectations.

A random variable is a mapping from a probability space to $\mathbb{R}$. For example, if $\Omega$ is as above, 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 $\mathbb{E}[X]$, is its weighted average. That is, $\mathbb{E}[X] = \sum_{i=1}^N p_i X(\omega_i)$. The following simple claim follows from the definition:
A.2. PROBABILITY THEORY

Claim A.4 (Linearity of expectation)
For $X,Y$ random variables over a space $\Omega$, denote by $X + Y$ the random variable that maps $\omega$ to $X(\omega) + Y(\omega)$. Then,

$$E[X + Y] = E[X] + E[Y]$$

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 $i^{th}$ coins came up heads and is equal to 0 otherwise. But clearly, $E[X_i] = 1/2$ for every $i$.

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

A.2.2 The averaging argument
We list various versions of the “averaging argument.” Sometimes we give two versions of the same result, one as a fact about numbers and one as a fact about probability spaces.

Lemma A.5
If $a_1, a_2, \ldots, a_n$ are some numbers whose average is $c$ then some $a_i \geq c$.

Lemma A.6 (“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.

Lemma A.7
If $a_1, a_2, \ldots, a_n \geq 0$ are numbers whose average is $c$ then the fraction of $a_i$’s that are greater than (resp., at least) $kc$ is less than (resp, at most) $1/k$.

Lemma A.8 (“Markov’s inequality”)
Any non-negative random variable $X$ satisfies

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

Corollary A.9
If $a_1, a_2, \ldots, a_n \in [0,1]$ are numbers whose average is $1 - \gamma$ then at least $1 - \sqrt{\gamma}$ fraction of them are at least $1 - \sqrt{\gamma}$.

Can we give any meaningful upperbound on $Pr[X < c \cdot E[X]]$ where $c < 1$? Yes, if $X$ is bounded.
Lemma A.10
If $a_1, a_2, \ldots, a_n$ are numbers in the interval $[0, 1]$ whose average is $\rho$ then at least $\rho/2$ of the $a_i$’s are at least as large as $\rho/2$.

Proof: Let $\gamma$ be the fraction of $i$’s such that $a_i \geq \rho/2$. Then $\gamma + (1 - \gamma)\rho/2$ must be at least $\rho/2$, so $\gamma \geq \rho/2$. □

More generally, we have

Lemma A.11
If $X \in [0, 1]$ and $\mathbb{E}[X] = \mu$ then for any $c < 1$ we have

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

Example A.12
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 $\Omega$ 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 $A_1, \ldots, A_n$ are mutually independent if for every subset $S \subseteq [n]$,

$$
\Pr[\bigcap_{i \in S} A_i] = \prod_{i \in S} \Pr[A_i].
$$

(2)

We say that $A_1, \ldots, A_n$ 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 $X_1, \ldots, X_n$. We have the following claim:
A.2. PROBABILITY THEORY

Claim A.13
If $X_1, \ldots, X_n$ are mutually independent then

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

Proof:

$$E[X_1 \cdots X_n] = \sum_{x} x \Pr[X_1 = x_1 \text{ and } X_2 = x_2 \cdots \text{ and } X_n = x_n] = \text{(by independence)}$$

$$\sum_{x_1,\ldots,x_n} x_1 \cdots x_n \Pr[X_1 = x_1] \cdots \Pr[X_n = x_n] = \left(\sum_{x_1} \Pr[X_1 = x_1]\right)\left(\sum_{x_2} \Pr[X_2 = x_2]\right) \cdots \left(\sum_{x_n} \Pr[X_n = x_n]\right) = \prod_{i=1}^{n} E[X_i]$$

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 $\Omega$. ■

A.2.4 Deviation upperbounds

Under various conditions, one can give upperbounds on the probability of a random variable “straying too far” from its expectation. These upperbounds 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.14 (Chebyshev inequality)
If $X$ is a random variable with standard deviation $\sigma$, then for every $k > 0$,

$$\Pr[|X - \mathbb{E}[X]| > k\sigma] \leq 1/k^2$$
A.2. PROBABILITY THEORY

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 \(X_1, \ldots, X_n\). This is because of the following claim, that is left as an exercise:

Claim A.15
If \(X_1, \ldots, X_n\) are pairwise independent then

\[
\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i)
\]

The next inequality has many names, and is widely known in theoretical computer science as the Chernoff bound. 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.16 (“Chernoff” bounds)
Let \(X_1, X_2, \ldots, X_n\) 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 \(\delta > 0\),

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

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

Often, what we use need is only the corollary that under the above conditions, for every \(c > 0\)

\[
\text{Pr}\left[\left|\sum_{i=1}^{n} X_i - \mu\right| \geq c\mu\right] \leq 2^{-c^2 n/2}
\]

Proof: Surprisingly, the Chernoff bound is also proved using the Markov inequality. We only prove the first inequality; a similar proof exists for the second. We introduce a positive dummy variable \(t\), and observe that

\[
\mathbb{E}[\exp(tX)] = \mathbb{E}[\exp(t \sum_{i} X_i)] = \mathbb{E}\left[\prod_{i} \exp(tX_i)\right] = \prod_{i} \mathbb{E}[\exp(tX_i)], \tag{5}
\]
where \( \exp(z) \) denotes \( e^z \) and the last equality holds because the \( X_i \) r.v.s are independent. Now,

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

therefore,

\[
\prod_i E[\exp(tX_i)] = \prod_i [1 + p_i(e^t - 1)] \leq \prod_i \exp(p_i(e^t - 1)) = \exp(\sum_i p_i(e^t - 1)) = \exp(\mu(e^t - 1)),
\]

(6)
as \( 1 + x \leq e^x \). Finally, apply Markov’s inequality to the random variable \( \exp(tX) \), viz.

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

using lines (5) and (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. ■

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

A.2.5 Some other inequalities.

Jensen’s inequality.

The following inequality, generalizing the inequality \( E[X^2] \geq E[X]^2 \), is also often useful:

**Claim A.17**

We say that \( f : \mathbb{R} \to \mathbb{R} \) is convex if for every \( p \in [0, 1] \) and \( x, y \in \mathbb{R} \),

\[
f(px + (1-p)y) \leq p \cdot f(x) + (1-p) \cdot f(y).
\]

Then, for every random variable \( X \) and convex function \( f \),

\[
f(E[X]) \leq E[f(X)].
\]
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.18**
For every $n$, $k < n$,

$$\binom{n}{k}^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$$

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

**Lemma A.19 (Stirling’s formula)**
For every $n$,

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

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.20**
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 $\alpha$.

More useful estimates.

The following inequalities can be obtained via elementary calculus:

- For every $x \geq 1$, $(1 - \frac{1}{x})^x \leq \frac{1}{e} \leq \left(1 - \frac{1}{x+1}\right)^x$
- 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 $c, \epsilon > 0$, $\sum_{i=1}^{\infty} \frac{n^c}{(1+c)^n} < O(1)$.
- For every $n$, $\sum_{i=1}^{n} \ln n = O(1)$.
A.3 Finite fields and groups

A field is a set \( F \) 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. 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.

If \( q \) is a prime, then we denote by \( \text{GF}(q) \) the field consisting of the elements \( \{0, \ldots, q-1\} \) with addition and multiplication performed modulo \( q \). For example, the numbers \( \{0, \ldots, 6\} \) yield a field if addition and multiplication are performed modulo 7. We leave it to the reader to verify \( \text{GF}(q) \) is indeed a field for every prime \( q \). 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 \( F \) has a number \( \ell \) such that for every \( x \in F \), \( x + x + \cdots + x \) (\( \ell \) times) is equal to the zero element of \( F \) (exercise). This number \( \ell \) is called the characteristic of \( F \). For every prime \( q \), the characteristic of \( \text{GF}(q) \) is equal to \( q \).

A.3.1 Non-prime fields.

One can see that if \( n \) is not prime, then the set \( \{0, \ldots, n-1\} \) with addition and multiplication modulo \( n \) is not a field, as there exist two non-zero elements \( x, y \) in this set such that \( x \cdot y = n = 0 \) (mod \( n \)). 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 poly(q, k)-time algorithm to obtain an irreducible polynomial of degree k over GF(q). There are also probabilistic algorithms (and deterministic algorithms whose analysis relies on unproven assumptions) that obtain such a polynomial in poly(log q, k) time.

For us, the most important example of a finite field is 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 fine in poly(k) time. In fact, we will mostly not even be interested in the multiplicative structure of GF(2^k) and only use the addition operation (i.e., use it as the vector space GF(2)^k, see below).

### A.3.2 Groups.

A group is a set that only has a single operation, say \(*\), that is associative and has an inverse. That is, \((G, *)\) is a group if

1. For every \(a, b, c \in G\), \((a * b) * c = a * (b * c)\)

2. There exists a special element \(id \in G\) such that \(a * id = a\) for every \(a \in G\), and for every \(a \in G\) there exists \(b \in G\) such that \(a * b = b * a = id\).

If \(G\) is a finite group, it is known that for every \(a \in G\), \(a * a * \cdots * a\) (\(|G|\) times) is equal to the element \(id\). A group is called commutative or Abelian if its operation satisfies \(a * b = b * a\) for every \(a, n \in G\). For every number \(n \geq 2\), the set \(\{0, \ldots, n - 1\}\) with the operation being addition modulo \(n\) is an Abelian group. Also, the set \(\{k : k \in [n - 1], \gcd(k, n) = 1\}\) with the operation being multiplication modulo \(n\) is an Abelian group.

If \(F\) is a field and \(k \geq 1\), then the set of \(k\)-dimensional vectors of \(F\) (i.e., \(F^k\)) together with the operation of componentwise addition, yields an Abelian group. As mentioned above, the most interesting special case for us is the group GF(2)^k for some \(k\). Note that in this group the identity element is the vector \(0^k\) and for every \(x \in GF(2)^k\), \(x + x = 0^k\). This group is often referred to as the Boolean cube.

### A.4 Vector spaces and Hilbert spaces

### A.5 Polynomials

We list some basic facts about univariate polynomials.
Theorem A.21
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 $\alpha_1, \ldots, \alpha_{d+1}$ in some field $\mathbb{F}$. Then
\[
\sum_{i=0}^{d} \alpha_j^i \cdot c_i = p(\alpha_j) = 0,
\]
for $j = 1, \ldots, d + 1$. This means that the system $A y = 0$ with
\[
A = \begin{pmatrix}
1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^d \\
1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^d \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
1 & \alpha_{d+1} & \alpha_{d+1}^2 & \cdots & \alpha_{d+1}^d
\end{pmatrix}
\]
has a solution $y = c$. The matrix $A$ is a Vandermonde matrix, and it can be shown that
\[
\det A = \prod_{i>j} (\alpha_i - \alpha_j),
\]
which is nonzero for distinct $\alpha_i$. Hence $\text{rank} A = d + 1$. The system $A y = 0$ has therefore only a trivial solution — a contradiction to $c \neq 0$. □

Theorem A.22
For any set of pairs $(a_1, b_1), \ldots, (a_{d+1}, b_{d+1})$ there exists a unique polynomial $g(x)$ of degree at most $d$ such that $g(a_i) = b_i$ for all $i = 1, 2, \ldots, d + 1$.

Proof: The requirements are satisfied by Lagrange Interpolating Polynomials:
\[
\sum_{i=1}^{d+1} b_i \cdot \frac{\prod_{j \neq i}(x - a_j)}{\prod_{j \neq i}(a_i - a_j)}.
\]
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 $x = a_1, \ldots, a_{d+1}$. 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 [?]).
Lemma A.23

If a polynomial \( p(x_1, x_2, \ldots, x_m) \) over \( F = GF(q) \) is nonzero and has total degree at most \( d \), then

\[
\Pr[p(a_1, a_2, \ldots, a_m) \neq 0] \geq 1 - \frac{d}{q},
\]

where the probability is over all choices of \( a_1 \ldots a_m \in F \).

Proof: We use induction on \( m \). If \( m = 1 \) the statement follows from Theorem A.21. Suppose the statement is true when the number of variables is at most \( m - 1 \). Then \( p \) can be written as

\[
p(x_1, x_2, \ldots, x_m) = \sum_{i=0}^{d} x_1^i p_i(x_2, \ldots, x_m),
\]

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,

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

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

\[
\Pr[p(a_1, a_2, \ldots, a_m) \neq 0] \geq (1 - \frac{k}{q})(1 - \frac{d - k}{q}) \geq 1 - \frac{d}{q},
\]

and the induction is completed. □
Bibliography


BIBLIOGRAPHY


