APPENDIX

A Some Linear Algebra

A.1 Unitary, Hermitian and Positive Semi-Definite Matrices

We state some elementary facts about unitary, normal and Hermitian matrices; the corresponding proofs can be found in any standard linear-algebra textbook. We write $A^\dagger$ for the transpose complex-conjugate of a matrix $A \in \mathbb{C}^{m\times n}$, and we use the bra-ket notation for vectors (as introduced in Section 1.1).

Definition A.1. *A basis $\{|i\rangle\}_{i\in I}$ of $\mathbb{C}^n$ is* **orthonormal** *if $\langle i|i\rangle = 1$ and $\langle i|j\rangle = 0$ for all $i \neq j$.*

Definition A.2. *A matrix $U \in \mathbb{C}^{n\times n}$ is* **unitary** *if $U^\dagger U = \mathbb{I}$.*

Theorem A.3. *If $U \in \mathbb{C}^{n\times n}$ maps* some *orthonormal basis of $\mathbb{C}^n$ into an orthonormal basis of $\mathbb{C}^n$, then U is unitary, and if $U \in \mathbb{C}^{n\times n}$ is unitary, then U maps* every *orthonormal basis of $\mathbb{C}^n$ into an orthonormal basis of $\mathbb{C}^n$.*

Definition A.4. *A matrix $A \in \mathbb{C}^{n\times n}$ is* **normal** *if $A^\dagger A = AA^\dagger$, it is* **Hermitian** *if $A^\dagger = A$, and it is* **positive semi-definite**, *also denoted as $A \geq 0$, if $\langle\varphi|A|\varphi\rangle \geq 0$ for all $|\varphi\rangle \in \mathbb{C}^n$.*

Theorem A.5. *$A \in \mathbb{C}^{n\times n}$ is normal if and only if A is unitary diagonalizable, i.e., there exists a unitary matrix U such that $UAU^\dagger$ is a diagonal matrix, and $A \in \mathbb{C}^{n\times n}$ is Hermitian if and only if A is normal and all eigenvalues are real. Finally, $A \in \mathbb{C}^{n\times n}$ is positive semi-definite if and only if A is Hermitian and has non-negative eigenvalues.*

Theorem A.6 (Spectral decomposition). *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian. Then*

$$A = \sum_{i=1}^{n} \lambda_i |i\rangle\langle i|$$

where the λ_i's are A's real eigenvalues and the $|i\rangle$'s form a set of orthonormal eigenvectors of A.

Definition A.7. *The* **Euclidean norm** *of a vector $|\varphi\rangle \in \mathbb{C}^n$ is defined as $\||\varphi\rangle\| = \sqrt{\langle\varphi|\varphi\rangle}$.*

Theorem A.8 (Cauchy-Schwarz). *For any $|\varphi\rangle, |\psi\rangle \in \mathbb{C}^n$: $|\langle\varphi|\psi\rangle| \leq \||\varphi\rangle\| \cdot \||\psi\rangle\|$.*

A.2 The Tensor Product

Informally, the tensor product $V \otimes W$ of two arbitrary complex vector spaces (or, more generally, algebras) V and W is the complex vector space (or algebra) that consists of all formal finite sums $\sum_i v_i \otimes w_i$ with $v_i \in V$ and $w \in W$, where $v \otimes w$ can be understood as a formal non-commutative product, so that $(v + v') \otimes w = v \otimes w + v' \otimes w$, $v \otimes (w + w') = v \otimes w + v \otimes w'$ and $\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$ (and $(v \otimes w)(v' \otimes w') = vv' \otimes ww'$ in case of algebras) for all $v, v' \in V$, $w, w' \in W$ and $\alpha \in \mathbb{C}$. We refer to the literature for a more formal treatment.

If A and B are linear functions acting on V and W, respectively, then $A \otimes B$ is a linear function that naturally acts component-wise on $V \otimes W$.

In the special case where V and W are (complex) *matrix* vector spaces (or algebras), then there is a convenient interpretation of the tensor product: For matrices $A = (a_{ij})_{ij} \in \mathbb{C}^{k\times\ell}$ and $B \in \mathbb{C}^{m\times n}$, the tensor product $A \otimes B$ can be understood as the (block) matrix which has $a_{ij}B$ as (i, j)-th block. This allows for instance to talk about the trace of a tensor product of two (square) matrices, which then boils down to $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A) \cdot \mathrm{tr}(B)$, or about the eigenvalues of a tensor product $A \otimes B$.

B Some Probability and Information Theory

B.1 Random Variables and Distributions

A **discrete probability space** is given by a countable set Ω and a probability function $P : \Omega \to [0,1]$ with $\sum_{\omega\in\Omega} P(\omega) = 1$. An **event** $\mathcal{A}$ is a subset of Ω, and for any event $\mathcal{A}$ the probability $P[\mathcal{A}]$ of the event is given by $P[\mathcal{A}] := \sum_{\omega\in\mathcal{A}} P(\omega)$. For two events $\mathcal{A}$ and $\mathcal{B}$, the **conditional** probability $P[\mathcal{A}|\mathcal{B}]$ is defined as $P[\mathcal{A}|\mathcal{B}] := P[\mathcal{A}\cap\mathcal{B}]/P[\mathcal{B}]$. A **random variable** is a function $X : \Omega \to \mathcal{X}$. The **distribution** of X is the function $P_X : \mathcal{X} \to [0,1]$ defined as $P_X(x) = P[X{=}x]$, where $X{=}x$ is a shorthand for the event $\{\omega \in \Omega \,|\, X(\omega) = x\}$. Furthermore, we write P_{XY} for the joint distribution of two random variables X and Y, i.e. $P_{XY}(x,y) = P[X = x \wedge Y = y]$, and we write $P_{X|\mathcal{E}}(x) = P[X{=}x|\mathcal{E}]$ and $P_{X|Y}(x|y) = P[X{=}x|Y{=}y]$ for the **conditional distributions** (conditioned on an event $\mathcal{E}$ respectively a random variable Y).

In these lecture notes, we usually leave Ω and $P : \Omega \to [0,1]$ implicit, and understand it as defined by the joint distribution (and probabilities) of all the random variables (and events) involved, where a distribution (with domain $\mathcal{X}$) may be an arbitrary function $Q : \mathcal{X} \to [0,1]$ with $\sum_x Q(x) = 1$. Also, we sometimes abuse notation and write $X \in \mathcal{X}$ to denote that the range of the random variable X is $\mathcal{X}$.

Definition B.1. *The* **statistical distance** *between two distributions P and Q with common domain $\mathcal{X}$ is defined as*

$$\mathrm{SD}(P,Q) := \frac{1}{2}\sum_{x\in\mathcal{X}} \big|P(x) - Q(x)\big|$$

If the distributions describing two experiments have small statistical distance, then this can be interpreted as that the experiments behave in exactly the same way except with small "error" probability. This is formalized by the following lemma.

Lemma B.2. *Let Q and Q' be two probability distributions with common domain $\mathcal{X}$. Then there exists a joint distribution $P_{XX'}$ for random variables X and X' such that $P_X = Q$ and $P_{X'} = Q'$, and such that $P[X \neq X'] = \mathrm{SD}(Q,Q')$.*

B.2 Hoeffding's Inequality

Theorem B.3 (Hoeffding's inequality). *Let $v \in \{0,1\}^n$ be a bit string with relative Hamming weight $\mu = \omega(v) = \sum_i v_i/n$. Let the random variables $X_1, X_2, \ldots, X_k$ be obtained by sampling k random entries from v* with replacement*, i.e., the X_i's are independent and $P_{X_i}(1) = \mu$. Similarly, let $Y_1, Y_2, \ldots, Y_k$ be obtained by sampling k random entries from v* without replacement*. Then, for any $\delta > 0$, the random variables $\bar{X} := \frac{1}{k}\sum_i X_i$ and $\bar{Y} := \frac{1}{k}\sum_i Y_i$ satisfy*

$$\Pr\big[|\bar{Y} - \mu| \geq \delta\big] \leq \Pr\big[|\bar{X} - \mu| \geq \delta\big] \leq 2\exp(-2\delta^2 k)\,.$$

B.3 Jensen's Inequality

Proposition B.4 (Jensen's inequality). *Let $f : I \to \mathbb{R}$ be a convex function on some interval $I \subset \mathbb{R}$. Then, for any $x_1, \ldots, x_n \in I$ and any $0 \leq p_1, \ldots, p_n \in \mathbb{R}$ with $\sum_i p_i = 1$,*

$$\sum_i p_i f(x_i) \geq f\Big(\sum_i p_i x_i\Big)\,.$$

If f is a concave function then the inequality is reversed.

In probability-theoretic terms, Jensen's inequality can be expressed as follows.

Corollary B.5. *Let $f : I \to \mathbb{R}$ be a convex function on some interval $I \subset \mathbb{R}$. Then, for any random variable X over I, it holds that $E[f(X)] \leq f(E[X])$.*