# Quantum computing, norms and polynomials

Escudero Gutiérrez, F.

# Chapter 2

# Preliminaries

## 2.1 Notation

**Vectors.** Given $z \in \mathbb{C}$, we use $z^*$ to refer to its complex conjugation. We will use $\{e_1, \ldots, e_m\}$ to refer to the canonical basis of $\mathbb{K}^n$. We will see $\mathbb{K}^n$ as a linear space equipped with the usual inner product $\langle z, z' \rangle = \sum_{i \in [n]} z_i^* z_i'$, where $(z_i)$ are the coordinates of $z$ in the canonical basis. We use $S^{n-1}$ to refer to the set of unit vectors of $\mathbb{K}^n$. For $p \in [1, \infty)$, the $\ell_p$ norms of such vectors are

$$\|z\|_p = \|z\|_{\ell_p} = \left( \sum_{i \in [n]} |z_i|^p \right)^{\frac{1}{p}}.$$

The $\ell_2$ norm is the norm induced by the mentioned inner product, and we will often simply call it $\|z\|$. The $L_p$ norms are

$$\|z\|_{L_p} = \left( \frac{1}{n} \sum_{i \in [n]} |z_i|^p \right)^{\frac{1}{p}}.$$

For $p = \infty$, $\|z\|_\infty = \max_i |z_i|$. Given a normed vector space $(V, \| \ \|)$ with $V \subseteq \mathbb{K}^d$, the dual norm of an element $v \in V$ is given by

$$\|v\|_* = \sup\{|\langle v, w \rangle| \mid w \in V,\ \|w\|_V \leq 1\}.$$

## 2.1. Notation

**Matrices.** Given $n \in \mathbb{N}$ and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, we use $M_n(\mathbb{K})$ to denote the space of $n \times n$ matrices with entries in $\mathbb{K}$. When $\mathbb{K}$ is clear from the context, we will simply write $M_n$. Given $Z \in M_n(\mathbb{C})$, $Z^\dagger$ to denote the adjoint matrix of $Z$. Given $X \in M_n(\mathbb{R})$, $X^\mathsf{T}$ to denote the transpose of X We will use $E_{ij}$ to refer to the matrix of $M_n$ whose $(i, j)$-entry is 1 and the rest are 0. We will see $M_n$ as a linear space equipped with the inner product $\langle A, B\rangle = \mathrm{Tr}[A^\dagger B]$. Given $n \in \mathbb{N}$, we will use $\mathrm{Id}_n$ to refer to the identity matrix of $M_n$. For $p \in [1, \infty)$ the Schatten-$p$ norms of a matrix $A \in M_n$, denoted as $\|A\|_{S_p}$, are the $\ell_p$ norms of their singular values (the singular values are the square roots of the eigengvalues of $A^\dagger A$). The Schatten infinity norm, $\|A\|_{S_\infty}$, is the largest singular value of $A$. We will often refer to $\|A\|_{S_\infty}$ as $\|A\|_{\mathrm{op}}$ or simply $\|A\|$, because it coincides with the operator norm of $A$ when regarding it as a linear map from $\ell_2$ to $\ell_2$, meaning that $\|A\|_{S_\infty} = \|A\|_{\mathrm{op}} = \sup_{z\neq 0} \|Az\|_{\ell_2}/\|z\|_{\ell_2}$. We will refer to the $S_1$ norm as the trace norm, and denote it as $\|\cdot\|_{\mathrm{tr}}$. We will refer to the $S_2$ norm as the Frobenius norm, and denote it as $\|\cdot\|_F$. We will say that a matrix $A \in M_n$ is a contraction if $\|A\|_{\mathrm{op}} \leq 1$. A matrix $U \in M_n(\mathbb{C})$ is unitary if $U^\dagger U = \mathrm{Id}_n$. A matrix $O \in M_n(\mathbb{R})$ is orthogonal if $O^\mathsf{T} O = \mathrm{Id}_n$.

**Indices.** We write $\mathbf{i}$ for a $t$-tuple $\mathbf{i} = (i_t, \ldots, i_t) \in [n]^t$ of indices. Given variables $x_1, \ldots, x_n$ and a $t$-tuple $\mathbf{i} \in [n]^t$, we use $x(\mathbf{i})$ to denote the monomial $x_{i_1} x_{i_2} \cdots x_{i_t}$. Similarly, given a matrix-valued map $A\colon [n] \to \mathbb{R}^{d\times d}$, we write $A(\mathbf{i}) := A(i_1)A(i_2)\ldots A(i_t)$.

**Quantum.** We write $I$, $X$, $Y$, $Z$, $H$ to refer to the following $2 \times 2$ matrices.

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

We will also use $\sigma_0$ to refer to $I$, $\sigma_1$ to $X$, $\sigma_2$ to $Y$ and $\sigma_3$ to $Z$. Given a matrix $A \in M_n$, its controlled version $cA$ is the matrix of $M_{2n}$ given by

$$cA = \begin{pmatrix} \mathrm{Id}_n & 0 \\ 0 & A \end{pmatrix}.$$

A qubit is a 2-dimensional vector space. We will often use $n$ to refer to the number of qubits, and $N$ to refer to $2^n$, which is the total dimension of the space of (the tensor product of) $n$ qubits.

**Miscellanea.** Given $n \in \mathbb{N}$, $[n]$ stands for the set $\{1, 2, \ldots, n\}$. $\mathcal{S}_n$ is the symmetric group, which is the group of permutations of $[n]$ elements. Given $z \in \mathbb{K}^n$ and $\pi \in \mathcal{S}_n$, we define $z \circ \pi \in \mathbb{K}^n$ as $(z \circ \pi)_i = z_{\pi(i)}$. Throughout this thesis we will consider different constants, all of which will be denoted by $C$ and their value will be clear from context. We will use $C_d$ to refer to quantities that only depend on $d$ and are constant with respect to other parameters. We will use $\delta_{i,j}$ to denote the indicator of the event $i = j$. Given a vector $z \in \mathbb{K}^n$, $\mathrm{Diag}(z)$ is the diagonal matrix of $M_n$ whose diagonal entries are given by $z$.

## 2.2 Quantum mechanics

A $n$-qubit state $\rho$ is an element of $M_N$ that is positive semidefinite and has trace one. A state $\rho$ is pure if it has rank 1, in which case $\rho = |\psi\rangle\langle\psi|$ for some unit vector of $M_N$ and we will also call $|\psi\rangle$ a state. A $n$-qubit channel $\Phi : M_N \to M_N$ is a completely positive trace preserving linear map. A measurement is a set $\{M_x\}_x$ of positive semidefinite matrices that sum to the identity. A projector operator valued measurement (POVM) is a measurement where $M_x$ are projectors. By the postulates of quantum mechanics, measuring a quantum state $\rho$ with $\{M_x\}_x$ outputs $x$ with probability $\mathrm{Tr}[\rho M_x]$.

We will often use the Choi-Jamiolkowski isomorphism to encode a quantum channel as a quantum state. We call the resulting state as the Choi-Jamiolkowski state (or CJ state for short). The CJ representation is given by

$$J(\Phi) = \sum_{i,j \in [N]} \Phi\left(|i\rangle\langle j|\right) \otimes |i\rangle\langle j| = (\Phi \otimes I)\left(\sum_{i,j \in [N]} |i\rangle\langle j| \otimes |i\rangle\langle j|\right), \tag{2.1}$$

which is an element in $M_N \otimes M_N = M_{N^2}$. The CJ state $v(\Phi)$ is defined to be

$$v(\Phi) = \frac{J(\Phi)}{\mathrm{Tr}[J(\Phi)]} = \frac{J(\Phi)}{N}. \tag{2.2}$$

According to (2.1), the CJ state $v(\Phi)$ can be prepared by first preparing $n$ EPR pairs (over $2n$ qubits) and then applying $\Phi$ to the $n$ qubits coming from the first half of each of the $n$ EPR pairs.

Given an $d$ dimensional quantum system, the dynamics of the system are described by a Hamiltonian $H$, which is an self-adjoint matrix of $M_d(\mathbb{C})$. For every time $t \in [0, \infty)$, a Hamiltonian $H$ defines a time evolution operator $U(t) = e^{-iHt}$ that determines the time evolution of the quantum system in the following way. If the

system at time $t = 0$ is described by state $\rho$, then at time $t' > 0$ it will be described by $U^\dagger(t')\rho U(t')$.

## 2.3 Quantum query complexity

We will mainly focus on the query complexity of decision problems, those whose answer is binary: YES or NO, -1 or 1, 0 or 1 ... These problems can be represented by Boolean functions $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$. In the setting of query complexity, we are given a known $f$ and the goal is to compute $f$ on an unknown input $x \in \{-1,1\}^n$ owned by an oracle. However, we can access this $x$ by making queries/questions to the oracle. The goal of a *good* query algorithm is to make as few queries as possible and compute $f(x)$. We will briefly introduce two models of query complexity, the classical and the quantum. The interest of quantum query complexity relies on the fact that in it the strengths and weaknesses of quantum computers can be rigorously studied with currently-available techniques (see e.g., [Amb18, Aar21, Ham25] for recent surveys). On the one hand, many of quantum computing's best-known algorithms, such as for unstructured search [Gro96], period finding (the core of Shor's algorithm for integer factoring) [Sho97] and element distinctness [Amb07], are most naturally described in the query model. On the other hand, the model admits powerful lower-bound techniques such as the polynomial method [BBC$^+$01], to which we will devote the first part of this thesis, and the adversary method, which we will revisit in Section 8.2.

### Classical query algorithms

In the classical query model, the queries consist on the most basic questions one could imagine asking about $x$, which are asking for entries of $x$. Formally, a classical query is an evaluation of the function

$$o_x : [n] \to \{-1,1\} : i \to x_i.$$

A classical query algorithm is allowed to do any computation in between queries. When finished, the algorithm should output $-1$ or 1. Thus, (deterministic) classical query algorithms can be represented as decision trees (see Fig. 2.1). On top of this, a classical query algorithm is also allowed to use randomness, i.e., choosing a decision tree at random.

Given that for every $x$ the outcome of the algorithm is a binary random variable, it is characterized by its bias (the difference between the probability of outputting 1

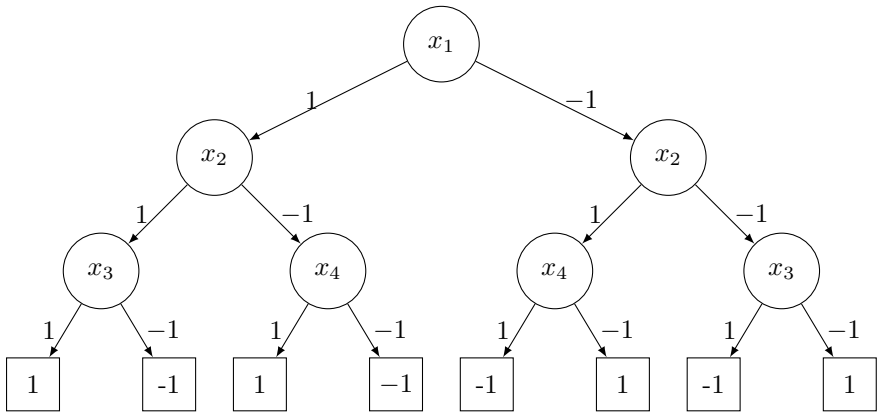


**Figure 2.1:** Decision tree representing a 3-query classical algorithm that computes the function $f(x_1, x_2, x_3, x_4) = (x_1 + x_2)x_3/2 + (x_1 - x_2)x_4/2$.

and the probability of outputting -1). We will thus identify an algorithm $\mathcal{A}$ with the function $\mathcal{A} : \{-1, 1\}^n \to [-1, 1]$ that maps $x$ to the bias of $\mathcal{A}$ on $x$. Now, we are ready to define classical query complexity.

**Definition 2.1.** Given $f : D \subseteq \{-1, 1\}^n \to \{-1, 1\}$ and $\varepsilon > 0$, the *randomized classical query complexity of $f$ with error $\varepsilon$* is the minimum number of queries made by a classical algorithm $\mathcal{A}$ such that $|\mathcal{A}(x) - f(x)| \leq \varepsilon$ for every $x \in D$. We use $R_\varepsilon(f)$ to refer to this quantity. We also use $R(f)$ to refer to $R_{2/3}(f)$ and $D(f)$ to refer to $R_0(f)$.

*Remark* 2.2. The number $2/3$ appearing in the definition of $R_{2/3}$ is somehow arbitrary, as for any constant $0 < c < 1$ we have that $R_c = \Theta(R_{2/3}(d))$. Indeed, say that $c < 2/3$. By definition, we have that $R_{2/3}(f) \leq R_c(f)$. On the other hand, $R_c(f) = O(R_{2/3}(f))$ because one can take an algorithm that $2/3$-approximates $f$, run it $O(\log(1/c))$ times and take the majority outcome, resulting in an algorithm that $c$-approximates $f$ and makes $O(\log(1/c))R_{2/3}(f)$ queries.

### Quantum query algorithms

In a quantum world, the queries to $x \in \{-1, 1\}^n$ are evaluations of the controlled version of the unitary map

$$\mathbb{C}^n \to \mathbb{C}^n : |i\rangle \to x_i|i\rangle.$$

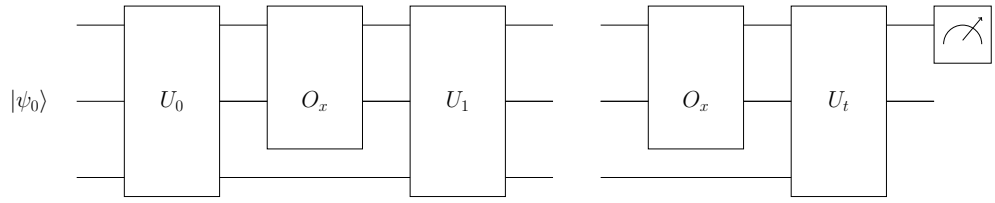


**Figure 2.2:** Quantum query algorithm.

Thus, it maps $|b\rangle|i\rangle \to (1 + \delta_{b,1}x_i)|b\rangle|i\rangle$ for $b \in \{0,1\}$ and $i \in [n]$, and it can be represented as the matrix

$$O_x = \mathrm{Diag}(1^n, x).$$

A quantum query algorithm is allowed to use extra quantum memory and to perform $x$-independent unitary operations in between queries. Finally, it should perform a binary measure and output $-1$ or 1. Thus, before the measurement the state of a $t$-query quantum algorithm on input $x$ looks like

$$|\psi_t\rangle = U_t(O_x \otimes \mathrm{Id}_d)U_{t-1} \dots U_1(O_x \otimes \mathrm{Id}_d)U_0|\psi_0\rangle, \tag{2.3}$$

where $U_t, \dots, U_0$ are $(2nd)$-dimensional unitaries and $|\psi_0\rangle$ is a fixed $(2nd)$-dimensional pure state. Fig. 2.2 Again, we identify a quantum algorithm with its bias. Now, we can define quantum query complexity.

**Definition 2.3.** Given $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon > 0$, the *quantum query complexity of $f$ with error $\varepsilon$* is the minimum number of queries made by a quantum algorithm $\mathcal{A}$ such that $|\mathcal{A}(x) - f(x)| \leq \varepsilon$ for every $x \in D$. We use $Q_\varepsilon(f)$ to refer to this quantity. We also use $Q(f)$ to refer to $Q_{2/3}(f)$.

*Remark* 2.4. Because of the same reasons as in the classical case, we have that $Q(f) = \Theta(Q_c(f))$ for any constant $0 < c < 1$.

*Remark* 2.5. Although complex numbers are necessary to describe quantum physics [RTW+21], the quantum query complexity of a function does not change if we assume that the underlying Hilbert space is real, thanks to the construction in [MMG09]. Furthermore, every real square matrix with operator norm at most 1 (largest singular value at most 1) is a convex combination of orthogonal matrices. Putting both things together, we have that for the purpose of quantum query complexity we may assume that $|\psi_0\rangle$ is a unit vector of a real Hilbert space and that $U_0, \dots, U_t$ are real square matrices with operator norm at most 1.

We will also analyze the smallest additive error that a $t$-query quantum algorithm can achieve when computing a function $f : D \subseteq \{-1,1\}^n \to \mathbb{R}$, which is given by

$$\mathcal{E}(f,t) := \inf\big\{\varepsilon \geq 0 \mid \exists\ t\text{-query quantum algorithm } \mathcal{A} \text{ with } |f(x) - \mathcal{A}(x)| \leq \varepsilon \quad \forall x \in D\big\}. \tag{2.4}$$

Note that $\mathcal{E}(f,t)$ and $Q_\varepsilon(f)$ are similar quantities conceptually, as they both encapsulate a notion of optimal quantum algorithm, but they do it in different ways. On the one hand, $Q_\varepsilon(f)$ refers to optimal quantum algorithms to approximate up to a given error $\varepsilon$. On the other hand, $\mathcal{E}(f,t)$ refers to the best $t$-query quantum algorithm.

## 2.4 Learning theory

The meta question of learning theory is the following. Given an unknown object from which we can access *expensive* units of information, how many of these units do we need in order to obtain an approximation of the object? This is a broad question, that has many variants depending on: *i*) the object to learn; *ii*) the access model; *iii*) the distance with respect to which we can measure what is a good approximation.

In addition, we will also consider the second-most important meta-question of learning theory, which is the problem of testing. In some cases, the number of units of information required to learn is prohibitive, but we may only be interested on whether the unknown object satisfies a certain property or it is far from it, i.e., to test whether the object satisfies the property.

In the second part of this thesis, we will mainly focus on learning quantum objects: quantum query algorithms, quantum channels and Hamiltonians. We will also consider the problem of testing properties of Hamiltonians.

We will need the following well-known result about distribution learning theory. See [Can20, Theorem 9] for a proof.

**Lemma 2.6.** *Let $p = \{p(x)\}_x$ be a probability distribution over some set $\mathcal{X}$. Let $p' = (p'(x))_x$ be the empirical probability distribution obtained after sampling $T$ times from $p$. Then, for $T = O((1/\varepsilon)^2 \log(1/\delta))$, with probability at least $1-\delta$, we have that $|p(x) - p'(x)| \leq \varepsilon$ for every $x \in \mathcal{X}$.*

## 2.5 Fourier and Pauli analysis

In this section, we describe Fourier expansion of Boolean functions and of different quantum objects (states, unitaries, channels) that we consider throughout this work. Note that the terms *Pauli expansion* and *Fourier expansion* will often be used interchangeably in the context of quantum objects .

**Fourier expansion.** In this section we will talk about the space of functions defined on the Boolean hypercube $f : \{-1,1\}^n \to \mathbb{R}$ endowed with the inner product $\langle f, g\rangle = \mathbb{E}_x[f(x)g(x)]$, where the expectation is taken with respect to the uniform measure of probability. For $s \subseteq \{0,1\}^n$, the Fourier characters, defined by $\chi_s(x) = \prod_{i\in \mathrm{supp}(s)} x_i$, constitute an orthonormal basis of this space. Hence, every $f$ can be identified with a multilinear polynomial (a polynomial that is affine on every variable) via the Fourier expansion

$$f = \sum_{s\in\{0,1\}^n} \widehat{f}(s)\chi_s, \tag{2.5}$$

where $\widehat{f}(s)$ are the Fourier coefficients given by

$$\widehat{f}(s) = \langle \chi_s, f\rangle = \mathbb{E}_x[f(x)\chi_s(x)]. \tag{2.6}$$

The degree of $f$ is the minimum $d$ such that $\widehat{f}(s) = 0$ if $|s| > d$. We will often use Parseval's identity:

$$\|f\|_2^2 := \langle f, f\rangle = \sum_{s\in[n]} \widehat{f}(s)^2. \tag{2.7}$$

We will also consider the $\ell_p$-norms of the Fourier spectrum, which are defined as

$$\|\widehat{f}\|_p = \left(\sum_{s\in\{0,1\}^n} |\widehat{f}(s)|^p\right)^{1/p}.$$

The supremum, infinity or $\ell_\infty$ norm of such an $f$ is $\|f\|_\infty = \max_x |f(x)|$. The variance of $f$ is given by

$$\mathrm{Var}[f] = \sum_{|S|\geq 1} \widehat{f}^2(S),$$

and the influence of the $i$-th variable by

$$\mathrm{Inf}_i[f] = \sum_{S\ni i} \widehat{f}^2(S) = \mathbb{E}_x\left[\left(\frac{f(x) - f(x^{\oplus i})}{2}\right)^2\right],$$

where given $x \in \{-1,1\}^n$ and $i \in [n]$, $x^{\oplus i}$ is the element of $\{-1,1\}^n$ obtained by flipping the $i$th entry of $x$. The maximum influence of $f$ is $\mathrm{MaxInf}[f] := \max_{i\in[n]} \mathrm{Inf}_i[f]$. One may interpret $\mathrm{Var}[f]$ as the deviation of $f$ from its expectation and $\mathrm{Inf}_i[f]$ as the deviation of $f$ from its expectation that is due to varying the $i$-th variable.

*Remark* 2.7. We will also use different notation for the indexing of the characters. Namely, given $s \in \{0,1\}^n$ we will identify it with its support $S$, so, for example, $\chi_S(x)$ will be given by $\prod_{i\in S} x_i$.

**Pauli expansion of operators.** Here, we introduce the Pauli analysis for operators, which was first explored by Montanaro and Osborne [MO08]. We consider $M_N$ endowed with the usual inner product $\langle A, B\rangle = \frac{1}{N}\mathrm{Tr}[A^\dagger B]$. The tensor product of Pauli operators form an orthonormal basis for this space. The Pauli expansion of a matrix $M$ of $M_N$ is given by

$$M = \sum_{x\in\{0,1,2,3\}^n} \widehat{M}(x)\sigma_x, \tag{2.8}$$

where $\widehat{M}(x) = \langle \sigma_x, M\rangle$are Pauli coefficients of $M$. We will refer to the collection of non-zero Pauli coefficients $\{\widehat{M}(x)\}_x$ as the Pauli spectrum of $M$ with the set of corresponding strings denoted by $\mathrm{spec}(M)$. As $\{\sigma_x\}_x$ is an orthonormal basis, we have a version of Parseval's identity for operators.

$$\|M\|_2^2 := \langle M, M\rangle = \sum_{x\in\{0,1,2,3\}^n} |\widehat{M}(x)|^2. \tag{2.9}$$

In particular, for $U \in \mathcal{U}_N$, this implies that $(|\widehat{U}(x)|^2)_x$ is a probability distribution. We will also consider the $p$-norms of the Pauli spectrum, which are defined as

$$\|\widehat{M}\|_p = \left(\sum_{x\in\{0,1,2,3\}^n} |\widehat{M}(x)|^p\right)^{1/p}.$$

We now define a notion of degree for states and unitaries that generalizes the classical notion of Fourier degree (see [MO08, Section 5]).

**Definition 2.8** (Degree of a matrix)**.** Given $M \in M_N$ its degree is the minimum $d$ such that $\widehat{M}(x) = 0$ for any $x \in \{0,1,2,3\}^n$ with $|x| > d$. Here, $|x|$ is the cardinality of the set $\{i \in [n] : x_i \neq 0\}$.

**Pauli expansion of superoperators.** Here, we introduce the Pauli analysis for superoperators, which was first explored by Bao and Yao [BY23]. We consider the space of superoprators (linear maps from $M_N$ to $M_N$) endowed with the inner product $\langle \Phi, \Psi \rangle = \langle J(\Phi), J(\Psi) \rangle / N^2$. An orthonormal basis for superoperators is defined using characters

$$\Phi_{x,y}(\rho) = \sigma_x \rho \sigma_y, \tag{2.10}$$

for any $x, y \in \{0,1,2,3\}^n$. The Pauli expansion of superoperators and hence quantum channels is then defined as

$$\Phi = \sum_{x,y \in \{0,1,2,3\}^n} \widehat{\Phi}(x,y) \Phi_{x,y}, \tag{2.11}$$

where $\widehat{\Phi}(x,y) = \langle \Phi_{x,y}, \Phi \rangle$ are the Pauli coefficients of the superoperator. As $\{\Phi_{x,y}\}_x$ is an orthonormal basis, we have a version of Parseval's identity for superoperators

$$\|\Phi\|_2^2 := \langle \Phi, \Phi \rangle = \sum_{x,y \in \{0,1,2,3\}^n} |\widehat{\Phi}(x,y)|^2.$$

We will also consider the $p$-norms of the Pauli spectrum of superopertors, which are defined as

$$\|\widehat{\Phi}\|_p = \left( \sum_{x,y \in \{0,1,2,3\}^n} |\widehat{\Phi}(x,y)|^p \right)^{1/p}.$$

If $\Phi$ is a channel, then $\widehat{\Phi} = (\widehat{\Phi}(x,y))_{x,y}$ has a couple of important properties [BY23, Lemma 8].

*Fact* 2.9. If $\Phi$ is a channel, then $\widehat{\Phi}$ is a state unitarily equivalent to $v(\Phi)$. In particular, $(\widehat{\Phi}(x,x))_x$ is a probability distribution.

The degree of a superoperator is defined in the analogue way to operators.

**Definition 2.10** (Degree of a superoperator)**.** Given a superoperator $\Phi$ its degree is the minimum $d$ such that $\widehat{\Phi}(x,y) = 0$ for any $x, y \in \{0,1,2,3\}^n$ with $|x| + |y| > d$.

## 2.6 Polynomials

For $p \in \mathbb{R}[x_1, \ldots, x_n]$ we define the following quantities, which are seminorms and norms when restricted to the space of multilinear polynomials,

$$\|p\|_\infty := \sup_{x \in \{-1,1\}^n} |p(x)|, \tag{2.12}$$

$$\|p\|_q := (\mathbb{E}_{x \in \{-1,1\}^n} |p(x)|^q)^{\frac{1}{q}}, \quad \text{for } q \in [1, \infty), \tag{2.13}$$

We will say that a a polynomial $p$ is *bounded* if its restriction to the Boolean hypercube takes values in the interval $[-1, 1]$.

### Univariate polynomials

We list a few well-known results about univariate polynomials. For $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, we will use $\mathbb{K}[x_1, \ldots, x_n]$ to denote the space of polynomials with coefficients in $\mathbb{K}$ depending on variables $x_1, \ldots, x_n \in \mathbb{K}$. We will use $\mathbb{K}[x_1, \ldots, x_n]_{=t}$ to refer to the space of forms of degree $t$ (or homogeneous polynomials of degree $t$), which are those whose only non-zero coefficients correspond to monomials of degree $t$. We will use $\mathbb{K}[x_1, \ldots, x_n]_{\leq t}$ to refer to the space of polynomials of degree at most $t$. A polynomial is multilinear if it is affine on every variable. Given $f : \{-1, 1\}^n \to \mathbb{R}$, we will identify $f$ with the unique multilinear polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ such that $f(x) = p(x)$ for every $x \in \{-1, 1\}^n$. The latter polynomial is given by the Fourier expansion of $f$.

**Proposition 2.11** (Markov brothers' inequality)**.** *Let $p \in \mathbb{R}[x]$ have degree at most $d$. Then, $\sup_{x \in [-1,1]} |p'(x)| \leq d^2 \sup_{x \in [-1,1]} |p(x)|$.*

**Definition 2.12.** Let $p \in \mathbb{R}[x_1, \ldots, x_n]$. Then, $p$ is symmetric if for every $\pi \in \mathcal{S}_n$ and every $x \in \mathbb{R}^n$ we have that $p(x) = p(\pi \circ x)$.

**Proposition 2.13** (Minsky-Papert symmetrization principle [MS69])**.** *Consider a symmetric multilinear polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $d$. Then, there is a univariate polynomial $q : \mathbb{R} \to \mathbb{R}$ of degree $d$ such that $\sup_{y \in [-1,1]} |q(y)| = \sup_{x \in \{-1,1\}^n} |p(x)|$ and $q(\sum_i x_i/n) = p(x)$ for every $x \in \{-1, 1\}^n$.*

## 2.7 Completely bounded norms

We will introduce the *completely bounded* norm of a multilinear form. Informally, it is a variation of the infinity norm where the supremum is not only evaluated on scalar inputs, but also on matrix inputs.

**Definition 2.14** (Multilinear forms)**.** Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. A map $T : \mathbb{K}^n \times \cdots \times \mathbb{K}^n \to \mathbb{K}$ is a *t-linear form* if it is linear with respect to every copy of $\mathbb{K}^n$. We will also use *multilinear forms* to refer to these functions. We will also identify every $t$-linear form with a tensor $T \in (\mathbb{K}^n)^t$ such that

$$T(x_1, \ldots, x_t) = \sum_{\mathbf{i} \in [n]^t} T_{\mathbf{i}} x_1(i_1) \ldots x_t(i_t)$$

for every $x_1, \ldots, x_t \in \mathbb{K}^n$. This tensor is uniquely determined by

$$T_{\mathbf{i}} = T(e_{i_1}, \ldots, e_{i_t})$$

for every $\mathbf{i} \in [n]^t$.

Throughout this thesis, we will use the notions of multilinear form and tensor interchangeably.

**Definition 2.15.** Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Let $m \in \mathbb{N}$ and $T : \mathbb{K}^n \times \cdots \times \mathbb{K}^n = (\mathbb{K}^n)^t \to \mathbb{K}$ be a $t$-linear form. We define the $t$-linear form $T_m : M_m^n \times \cdots \times M_m^n \to M_m$ by

$$T_m(X_1, \ldots, X_t) = \sum_{\mathbf{i} \in [n]^t} T_{\mathbf{i}} X_1(i_1) \ldots X_t(i_t)$$

for every $X_1, \ldots, X_t \in M_m^n$. We define its norm as

$$\|T_m\| := \sup \|T(X_1, \ldots, X_t)\|_{\mathrm{op}},$$

where the supremum runs over all $X_1, \ldots, X_t \in M_m^n$ with $\|X_1(i_1)\|_{\mathrm{op}}, \ldots, \|X_t(i_t)\|_{\mathrm{op}} \leq 1$.

*Remark* 2.16. The supremum of $\|T_m\|$ does not change if $X_s(i_s)$ are not only contractions but also orthogonal matrices in the real case, or unitary matrices in the complex case. This follows from the Krein-Milman theorem and the fact that orthogonal (unitary) matrices are the extreme points of the set of real (complex) contractions.

**Definition 2.17** (Completely bounded norm of multilinear form)**.** Let $T : \mathbb{K}^n \times \cdots \times \mathbb{K}^n = (\mathbb{K}^n)^t \to \mathbb{K}$ be a $t$-linear form. Its *completely bounded norm* is given by

$$\|T\|_{\mathrm{cb}} := \sup_{m \in \mathbb{N}} \|T_m\|.$$

Notably, for the supremum in the completely bounded norm, one can take $X_1 =$

$\cdots = X_t$. Thus, one could say that *polarization constant* for the completely bounded norm is 1.[1]

**Proposition 2.18.** *Let $T \in \mathbb{R}^{n\times\cdots\times n}$ be a $t$-tensor. Then,*

$$\|T\|_{\mathrm{cb}} = \sup\big\{\|T(X,\dots,X)\|_{\mathrm{op}},\ d\in\mathbb{N}\big\},$$

*where the supremum runs over all contractions $X(1),\dots,X(n)\in M_m$ and all $m\in\mathbb{N}$.*

*Proof.* Let $|||T|||$ be the expression in the right-hand side of the statement. Note that it is the same as the expression of $\|T\|_{\mathrm{cb}}$, but now the contraction-valued maps $X_1,\dots,X_t$ are all equal. This shows that $|||T||| \le \|T\|_{\mathrm{cb}}$. To prove the other inequality, let $X_1,\dots,X_t\colon [n]\to B_{M_d}$ and $u,v\in S^{d-1}$. Now, define the contraction-valued map $X$ by $X(i) := \sum_{s\in[t]} e_s e_{s+1}^{\mathsf{T}}\otimes X_s(i)$ for $i\in[n]$, and define the unit vectors $u' := e_1\otimes u$ and $v' := e_{t+1}\otimes v$. They satisfy

$$\langle u, X_1(i_1)\dots X_t(i_t)v\rangle = \langle u', X(\mathbf{i})v'\rangle \quad \text{for all } \mathbf{i}\in[n]^t,$$

so in particular

$$\sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}}\langle u, X_1(i_1)\dots X_t(i_t)v\rangle = \sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}}\langle u', X(\mathbf{i})v'\rangle.$$

Taking the supremum over all maps $X_s$ and $u,v$ shows that $\|T\|_{\mathrm{cb}}\le |||T|||$, which concludes the proof. □

### 2.7.1 Grothendieck inequality

Let $\mathbb{K}\in\{\mathbb{R},\mathbb{C}\}$. Given a bilinear form $A:\mathbb{K}^n\times\mathbb{K}^n\to\mathbb{K}$, we can write its infinity norm as

$$\|A\|_\infty = \sup_{|x_i|=|y_i|\le 1}\left|\sum_{i,j\in[n]} A_{ij}x_iy_j\right|.$$

[1] In Banach space theory a $t$-linear map $T: X\times\cdots\times X\to Y$ determines a homogeneous degree-$t$ polynomial $P: X\to Y: A\to T(A,\dots,A)$. The operator norms of $T$ and $P$ are equivalent if $T$ is symmetric: $\|T\|\le\|P\|\le K(t)\|T\|$, where $K(t)$ is the polarization constant of degree $t$. For a survey on the topic see [MMFPSS22, Section 5.1].

For $\mathbb{K} = \mathbb{R}$ the absolute value inside the supremum is not necessary, and by linearity we have that the supremum is attained in the extreme points, so

$$\|A\|_\infty = \sup_{x_i, y_i \in \{-1,1\}} \sum_{i,j \in [n]} A_{ij} x_i y_j.$$

For $\mathbb{K} = \mathbb{C}$ by the maximum modulus principle we have that

$$\|A\|_\infty = \sup_{|x_i|=|y_i|=1} \left| \sum_{i,j \in [n]} A_{ij} x_i y_j \right|.$$

Also, note that $\|A\|_\infty \leq \|A\|_{\mathrm{cb}}$.

**Theorem 2.19** (Grothendieck's theorem [Gro53])**.** *There exists a constant $K < \infty$ such that for any $n \in \mathbb{N}$ and any bilinear form $A : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$, we have*

$$\|A\|_{\mathrm{cb}} \leq K \|A\|_\infty. \tag{2.14}$$

*Equivalently,*

$$\max_{|x_i|,|y_i| \leq 1} \left| \sum_{i,j \in [n]} A_{ij} x_i y_j \right| \leq K \max_{\|u_i\|_2, \|v_j\|_2 \leq 1} \sum_{i,j \in [n]} A_{ij} \langle u_i, v_j \rangle,$$

*where the supremum runs over all $d \in \mathbb{N}$ and all vectors $u_i, v_j \in \mathbb{K}^d$.*

The smallest possible constant $K$ for which Theorem 2.19 holds is known as the Grothendieck constant, $K_G^{\mathbb{K}}$. Determining the precise value of $K_G^{\mathbb{K}}$ is a notorious open problem posed in [Gro53]. For $\mathbb{K} = \mathbb{R}$ the best-known lower and upper bounds place it in the interval $(1.676, 1.782)$ [Dav84, Ree91, BMMN13]. For $\mathbb{K} = \mathbb{C}$, we know that the constant lies in $(1.338, 1.405)$ [Haa87, Dav06].

## 2.8 Semidefinite programming

Semidefinite programming is an extension of linear programming that includes a bigger family of problems and can still be efficiently solved up to arbitrary precision (see [LR05] for an introduction to semidefinite programming). To be more precise, let $S_n$ be the space of symmetric matrices of $M_n$ and let $S_n^+$ be the cone of positive semidefinite matrices. A collection of matrices $C, B_1, \ldots, B_l \in S_n$ and a vector $b \in \mathbb{R}^l$

define a *primal semidefinite program (P)* and a *dual semidefinite program (D)*, which in their *canonical form* are given by

$$\begin{array}{llllll} (P) & \inf & \langle C, Y\rangle & (D) & \sup & \langle b, y\rangle \\ & \text{s.t.} & Y \in S_n^+ & & \text{s.t.} & y \in \mathbb{R}^l \\ & & \mathcal{B}(Y) = b & & & C - \mathcal{B}^*(y) \in S_n^+, \end{array} \tag{2.15}$$

where $\mathcal{B} : S_n \to \mathbb{R}^l$ is given by $\mathcal{B}(Y) := (\langle B_1, Y\rangle, \ldots, \langle B_l, Y\rangle)$, $\mathcal{B}^*(y) = \sum_{i\in[l]} y_i B_i$ and $\langle B, Y\rangle = \mathrm{Tr}(BY)$. A semidefinite program is feasible if there exists an instance satisfying its constraints.

Note that if all matrices $C, B_1, \ldots, B_l$ were diagonal, $(P)$ and $(D)$ would be linear programs. Indeed, in that case the value of $(P)$ would not change if we further impose that $Y$ is diagonal, which makes $(P)$ a linear program. Also, the constraint $C - \mathcal{B}^*(y) \in S_n^+$ is equivalent to saying that the diagonal entries of $C - \mathcal{B}^*(y)$ are non-negative, so $(D)$ is also a linear program.

It is always satisfied that the optimal value of $(P)$ is at least the optimal value of $(D)$, what is known as *weak duality*. In addition, under some mild assumptions provided by Slater's theorem, both values are equal, what is known as *strong duality*.

**Theorem 2.20** (Slater's theorem)**.** *Let $(P)$ and $(D)$ be a primal-dual pair of semidefinite programs, as in Eq.* (2.15)*. Assume that $(P)$ is feasible and there exists a strictly positive instance for $(D)$, i.e., there exists $y \in \mathbb{R}^l$ such that $C - \mathcal{B}^*(y)$ is strictly positive. Then the optimal values of $(P)$ and $(D)$ are equal.*

## 2.9 Concentration inequalities

We state a few concentration inequalities that we use often. All of them can be found in [BLM13].

**Lemma 2.21** (Hoeffding bound)**.** *Let $X_1, \ldots, X_m$ be independent-random variables that satisfy $-a_i \le |X_i| \le a_i$ for some $a_i > 0$. Then, for any $\tau > 0$, we have*

$$\Pr\Big[\Big|\sum_{i\in[m]} X_i - \sum_{i\in[m]} \mathbb{E}[X_i]\Big| > \tau\Big] \le 2\exp\left(-\frac{\tau^2}{2(a_1^2 + \cdots + a_m^2)}\right).$$

**Lemma 2.22** (Bernstein inequality)**.** *Let $X_1, \ldots, X_m$ be independent-random vari-*

*ables with* $|X_i| \leq M$ *for some* $M > 0$*. Then,*

$$\Pr\Big[\Big|\sum_{i\in[m]} X_i - \sum_{i\in[m]} \mathbb{E}[X_i]\Big| > \tau\Big] \leq 2\exp\left(-\frac{\tau^2/2}{\sum_{i\in[m]} \mathrm{Var}[X_i] + \tau M/3}\right).$$

**Lemma 2.23** (McDiarmid's inequality)**.** *Let* $f : \{-1,1\}^n \to \mathbb{R}$ *such that* $|f(x) - f(x^{\oplus i})| \leq c$ *for every* $x \in \{-1,1\}^n$ *and every* $i \in [n]$*. Then, over a uniformly random* $x$ *and for any* $\varepsilon > 0$ *we have that*

$$\Pr_x[|f(x) - \mathbb{E}_y f(y)| \geq \varepsilon] \leq \exp\left(-\frac{2\varepsilon^2}{nc^2}\right).$$