Quantum computing, norms and polynomials

Escudero Gutiérrez, F.

Quantum computing, norms and polynomials

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 10 februari 2026
klokke 16:00 uur

door
Francisco Escudero Gutiérrez

geboren te Madrid

in 1997

Promotor:	Dr. J. Briët	(CWI & Universiteit Leiden)
Co-promotor:	Prof.dr. S.O. Fehr	(CWI & Universiteit Leiden)

Promotiecomissie:	Prof.dr.ir. G.L.A. Derks	
	Prof.dr. L. Ducas	(CWI & Universiteit Leiden)
	Prof.dr. M. Laurent	(Tilburg University)
	Prof.dr. C. Palazuelos	(Universidad Complutense de Madrid)
	Dr. H. Zhang	(University of South Carolina)

This thesis was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 945045, and by the NWO Gravitation project NETWORKS under grant no. 024.002.003.

Cover design by Marta Crespí Campomar and Francisco Escudero Gutiérrez.

Para vosotros: mamá, papá, Nacho y Raquel.

Acknowledgements

I want to start these acknowledgements by thanking the most important person in my PhD journey: my advisor, Jop Briët. Since the first moment, you did not care only about research, but also about my personal well-being and my mental health. You introduced me to the beautiful world of quantum query complexity and its interplay with functional analysis, making me a better researcher by showing in our meetings what are the questions to be asked. Furthermore, you made sure to connect me with other researchers that I ended up collaborating with. You also made of me a way better writer than the one I was four years ago. Last but not least, you have always been a role model, as a great researcher with an excellent work-personal-life balance, who is aware of his surroundings, and who cares about his colleagues in a personal way. I am certain that you are going to be a great group leader.

I also want to devote a special mention to my promotor, Serge Fehr. Thanks, Serge, for kindly agreeing to be my promotor at the last stage of my PhD, and for being available and actively involved in the process of graduation since then.

I want to continue by thanking three of my collaborators, who were more than collaborators: Srinivasan Arunachalam, Sander Gribling and Carlos Palazuelos. Thank you, Srini, Sander, and Carlos, for the support, encouragement, career advice and writing tips throughout all my PhD. I also want to thank Ronald de Wolf. Thanks, Ronald, for offering yourself to be on my committee when we were struggling to form it, for very useful feedback on my thesis, and for the constant support during these four years.

Now, it is turn to thank my colleagues for making the office a place where I enjoyed many, many hours of my life. To Marten, for the dinners at our places, the trips, the conversations, for being paranymphs together, and for your laugh that brings joy wherever you go. To Sebas, for the money-saving advice, for caring about the new colleagues, and for always being open to a conversation with whoever is feeling alone. To Randy, for the times we have gone out for dinner, for the parties, and for your brilliant jokes. To Toto, for the political discussions, for the cinema sessions, and for all the culture I have learnt from you. To Dyon, for the padel sessions, for bringing me to the top of the foosball championship, for the trips together, and for the honor of letting me be your paranymph. To Garazi, for sharing views, for being there for each other when we had to *desahogarnos,* and for showing me that things can be done in a different way. To Davi, for the fun dinners, parties, and foosball sessions. To Lynn, for organising the junior meeting together and the support you gave me during stressful times. To Nikhil, for the visit to Liverpool and for making foosball a way better experience. To Simona and Jelena, for the protests and parties together. To Léo, for making me a better foosball player, and for that great weekend in the South

of France. To Arjan, for the collaborations, the trips, and your house parties. To Jordi, for the constant opinion and ideas exchange, and for the fun week in Okinawa. To Adam, for the trip to Tenerife, for the nice time together in the office, and for always listening to feelings. To Yanlin, for always asking me how I was doing when you (accurately) thought that I was feeling blue. To Subha, for the dinners and for always giving me confidence when I asked for advice. To Quinten, for always being down for a conversation about any topic. To Luca and Pippo, for the dinners and the time spent sharing the office. To Francesco and Nunzia, for being the social catalyst during your stay at QuSoft. To Philip, for the company during the Networks weeks and the boat trips. To Harold and Joran, for the parties and the good conversations.

I want to continue by thanking all the friends I made outside the office during these four and a half years in Amsterdam. To Ade, Ake, Cris, Johny, Mateo, Nil and Prudi, for the football games, the losses and the trophies. To Chelsea, René, and Marion, for the evenings and trips. To my comrades Aday, Javi, Juanma, Juanjo, Rafa, Rafa Koldo, Sara and Vanesa, who allowed me to keep doing politics in the Netherlands. To Juan Luis, *por tomarnos algo de vez en cuando.* To Cami, Clara, Garazi, Jelena, Lorenzo, Said, Simona, Shane and Toto for protesting together. To Quique, Isa and Marina, for the drinks and the Catan nights. To Natalia, for the museum afternoons together. To Amadeo, for being a good friend and supportive in my first year in the Netherlands. To Robin, for the coffee conversations at Nikhef. To Marta, for the cinema sessions, and for designing the beautiful cover of this thesis. To Fede, for the football, listening, and sharing. To Ric, for always having an original conversation to start. To Angel, for the walks and reflections together. To Milagros, for constantly caring for your friends. To Yasan, for the clothes I have inherited from you, and for always being a good *habibi*. To Eva, for being there in the cinema sessions, the personal conversations, and the parties. To Joie, for bringing me happiness and hope. Dank je dat je me gelukkig maakt. To Bernat, for being the most fun possible roommate ever. To Llorenç, for being my *husband* and my best friend these four years. Also, thanks to Ana, Blair, Carmen, Ferran, Keelan, Matas, Nuria and Ramón. My people from Spain have also supported me during these years, being proud and happy of me every time I saw them. The last pieces of gratitude are for them.

Muchas gracias a mis *Hoplitas,* a mis chicos y chicas de *Mortero,* a mis viajeros de *Siguiente País en Conflicto*, a mis colegas del *Hoy se lía*, y a mis *Camineiros.* También gracias a vosotros: Ach, Benjamín, camaradas del Círculo, Deif, Fer, Fiol, Geri, Igna, Radu, soles de la Conce, y Tamara.

Finalmente, quiero dar muchas gracias a toda mi familia, por vuestro apoyo incondicional durante toda mi vida y este doctorado. Esta tesis está dedicada a vosotros: mamá, papá y Nacho. También a Raquel, por lo orgullosa que siempre estabas de mí, al igual que yo lo estoy y siempre estaré de Rubén.

Contents

Chapter 1

Introduction

Motivated by the necessity to perform calculations, humankind has developed *algorithms*: sequences of simple computation steps designed to perform (more) complex computations. Maybe the most basic calculation is the sum of two natural numbers, and the best-known example of an algorithm is the set of instructions we learn in primary school to perform this task. A slightly more involved algorithm is the one to multiply two natural numbers, but it is still taught at primary school. Considerably more difficult is the inverse process of *factoring*, consisting of writing a natural number as a product of *prime* numbers, which are the natural numbers that cannot be written as the product of two other natural numbers (apart from the product of 1 and themselves). The definition of factoring involves abstract concepts, and thus, understanding the algorithms for that problem requires elaborate mathematical theories. Despite what its description might suggest, factoring is a problem with practical relevance, as cybersecurity is based on the inability of current computers to factor big natural numbers efficiently. The capacities of these computers are determined by the behavior of the physical units they are composed of, which are governed by classical physics, and thus we refer to them as *classical computers*. By contrast, in the 80's Yuri Manin and Paul Benioff proposed an idea later popularized by Feynman: to build computers with quantum particles, described by the richer theory of quantum mechanics [Ben80, Man80]. This gave birth to the idea of *quantum computers*, whose capabilities would exceed those of classical computers. Notably, Peter Shor discovered a *fast* algorithm for factoring large numbers in a *quantum computer* [Sho97]. As of today, we are on the path towards realizing a fully functional quantum computer.

Factoring is not an isolated case, but rather one of many computational problems

with practical relevance whose analysis benefits from the use of mathematical structures. In this thesis, we will approach several of such problems, all related to quantum computing, from the perspective of *theoretical computer science*, i.e., establishing rigorous guarantees. The questions we will consider fall into two categories. The first is *query complexity*, a model of computation where the power and limitations of quantum and classical computers can be rigorously studied. In particular, the foundational quantum algorithms by Peter Shor, to factor large numbers, and by Lov Grover, to find a marked element in a list, belong to this model [Gro96, Sho97]. The second category is *learning theory*, where the meta-question is how to characterize an unknown object to which we have limited access. In contrast with recent breakthroughs in machine learning, such as ChatGPT, our findings do not focus on immediate applicability but instead provide rigorous performance guarantees. In other words, our results fit into the subfield of *computational learning theory* initiated by Leslie Valiant in 1984 [Val84].

To address these questions, we will borrow and develop tools of non-applied mathematical branches, especially from *functional analysis*. Functional analysis is the branch of mathematics that studies functions in a very general way, even when they depend on an unbounded number of variables. Relevant examples of functions are *polynomials*, which can be written as sums of *monomials*, that are products of numbers and variables. Thanks to their elementary definition, polynomials are ubiquitous and motivate deep mathematical questions. Some of these questions are related to the comparison of *norms* of polynomials, which are measures of the *size* of the polynomial. How these norms relate has motivated many celebrated advances of mathematics during the last century, such as the works of Alexander Grothendieck, Frederic Bohnenblust and Carl Einar Hille [Gro53, BH31].

Polynomials have found applications in theoretical computer science, like the *polynomial method* of query complexity, a tool to prove limitations of quantum computers [NS94, BBC$^+$01]. Maybe more implicitly, polynomials have also played a role in quantum mechanics, as, for instance, Hamiltonians, which model the interactions between quantum particles, can be regarded as polynomials [MO08].

In this thesis, as explained in more detail in Section 1.1, we make progress in the understanding of the polynomial method and propose new algorithms to learn quantum objects. In parallel, we will prove functional-analytic statements relating different norms of polynomials.

This thesis is a product of a four-year PhD program developed at the Centrum Wiskunde & Informatica (CWI) in Amsterdam.

1.1 Overview

This thesis is divided into three parts, preceded by a preliminary chapter (Chapter 2) where we introduce notation, definitions, and basic results.

In Part I, we study several questions related to quantum query complexity and polynomials. Query complexity is a model of computing where the aim is to approximate a known function f on a hidden input x. The algorithm can access x via queries that reveal units of information about x. The goal is to determine how many queries an algorithm needs to solve such a task. Depending on the nature of these queries, one can speak about classical or quantum query complexity.

In Chapter 3, we show that the polynomial method is complete. This method was introduced and successfully applied to show lower bounds on quantum query complexity [BBC$^+$01, AS04]. Recently, it was refined as a tool to prove upper bounds [ABP19], which also led to a classical optimization algorithm to determine quantum query complexity [GL19]. In this chapter, we review these advancements and complete this picture by proposing a constructive method to design quantum algorithms via polynomials.

In Chapter 4, we show two technical results. The first is that the Grothendieck constant, which appears in the celebrated Grothendieck inequality that relates two norms of polynomials of degree 2, can be characterized in terms of quantum algorithms that make 1 query. The second result shows that quantum algorithms that make 2 queries are not equivalent to a certain class of polynomials because of the failure of Grothendieck's inequality for polynomials of degree greater than 2, answering a question by Aaronson, Ambainis, Iraids, Kokainis, and Smotrovs [AAI$^+$16].

In Chapter 5, we make progress on the Aaronson and Ambainis (AA) conjecture. The AA conjecture asserts that certain low-degree polynomials have a very influential variable [AA09]. Although this is a functional-analytic conjecture, it implies that quantum query complexity can only be much lower than classical query complexity for functions where some structure about the hidden input is promised beforehand. Our contributions in this chapter are formulating a weaker conjecture that retains the implications for query complexity, and proving it in a particular case.

In Part II, we explore several questions related to learning quantum objects that can be understood as polynomials. This analogy was pointed out by Montanaro and Osborne, who showed that polynomials can be embedded into quantum operators while preserving relevant properties such as the degree [MO08].

In Chapter 6, we prove two new versions of the Bohnenblust-Hille inequality, which

compares two norms of polynomials. We apply the first of them to learn quantum query algorithms. Then, we use the second to learn quantum channels, which are the operations allowed in a quantum computer.

In Chapter 7, we prove some of the first results about learning quantum Hamiltonians. Due to the Schrödinger equation, Hamiltonians model the evolution of quantum systems. In this chapter, we design algorithms to infer properties of and, in some cases, fully characterize an unknown Hamiltonian by accessing the corresponding time evolution.

Part III is a *bonus* that just consists of Chapter 8, where we gather three new proofs, that we find elegant and concise, of known results related to the analysis of Boolean functions.

1.2 Relation to literature

The content of this thesis is based on the following papers.

- Section 3.4 is a quantum-oriented exposition of the following work, originally written for a non-applied mathematical audience:

  [Esc25] Francisco Escudero Gutiérrez. Christensen-Sinclair factorization via semidefinite programming. *Linear Algebra and its Applications*, 714:28–44, 2025.

- Chapter 4 is based on:

  [BE22] Jop Briët and Francisco Escudero Gutiérrez. On Converses to the Polynomial Method. In *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022), volume 232 of Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:10. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

  [BEG24] Jop Briët, Francisco Escudero Gutiérrez, and Sander Gribling. Grothendieck inequalities characterize converses to the polynomial method. *Quantum*, 8:1526, 2024.

- Chapter 5 is based on:

  [Esc24a] Francisco Escudero Gutiérrez. Influences of Fourier completely bounded polynomials and classical simulation of quantum algorithms. *Chicago Journal of Theoretical Computer Science*, 2024.

  Also presented at the *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*.

- Chapter 6 is based on:

  [ADEP25] Srinivasan Arunachalam, Arkopal Dutt, Francisco Escudero Gutiérrez, and Carlos Palazuelos. A cb-Bohnenblust–Hille inequality with constant one and its applications in learning theory. *Mathematische Annalen*, pages 1–30, 2025.

  A complementary version of [ADEP25] also appeared in the proceedings of ICALP'24:

  [ADEP24] Srinivasan Arunachalam, Arkopal Dutt, Francisco Escudero Gutiérrez, and Carlos Palazuelos. Learning low-degree quantum objects. In *51st International Colloquium on Automata, Languages, and Programming* (ICALP 2024), pages 13–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.

  Also presented at the *19th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2024)*.

- Chapter 7 is based on:

  [ADE25] Srinivasan Arunachalam, Arkopal Dutt, and Francisco Escudero Gutiérrez. Testing and learning structured quantum Hamiltonians. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1263–1270, New York, NY, USA, 2025. Association for Computing Machinery.

  Also presented at the *20th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2025)*. A journal version has already been accepted in Communications in Mathematical Physics.

Additionally, this thesis contains previously unpublished work. More precisely:

- Chapter 3 contains a new exposition of the (completely bounded) polynomial method. In particular, we prove that the refinement of polynomial degree proposed in [ABP19] upper bounds quantum query complexity in an almost self-contained way. For this, we avoid the functional analytic black-boxes of the original proof, and only use tools from theoretical computer science.

- Chapter 8 contains new proofs of known results, with marginal improvements, that we find elegant and concise. All of these results are related to the analysis of Boolean functions $f : \{-1, 1\}^n \to \{-1, 1\}$.

Finally, the following papers were also completed during my PhD.

- [EFFJ+23] Francisco Escudero Gutiérrez, David Fernández-Fernández, Gabriel Jaumá, Guillermo F. Peñas, and Luciano Pereira. Hardware-efficient entangled

measurements for variational quantum algorithms. *Physical Review Applied*, 20(3), 2023.

- [EM24] Francisco Escudero Gutiérrez and Garazi Muguruza. All S_p notions of quantum expansion are equivalent. *arXiv preprint* arXiv:2405.03517, 2024.

- [BE24] Jinge Bao and Francisco Escudero Gutiérrez. Learning junta distributions, quantum junta states, and QAC^0 circuits. *arXiv preprint* arXiv:2410.15822, 2024.

  Presented at the *25th Asian Quantum Information Science Conference* (AQIS'25).

- [ACE+25] Amira Abbas, Nunzia Cerrato, Francisco Escudero Gutiérrez, Dmitry Grinko, Francesco Anna Mele, and Pulkit Sinha. Nearly optimal algorithms to learn sparse quantum Hamiltonians in physically motivated distances. *arXiv preprint* arXiv:2509.09813, 2025.

- [BCE+25] Andreas Bluhm, Matthias C. Caro, Francisco Escudero Gutiérrez, Aadil Oufkir, and Cambyse Rouzé. Certifying and learning quantum Ising Hamiltonians. *arXiv preprint* arXiv:2509.10239, 2025.

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 ℓ_p norms of such vectors are

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

The ℓ_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^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 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 ℓ_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 ℓ_2 to ℓ_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 2×2 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 σ_0 to refer to I, σ_1 to X, σ_2 to Y and σ_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 ρ is an element of M_N that is positive semidefinite and has trace one. A state ρ 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 ρ 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 Φ 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 ρ, 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

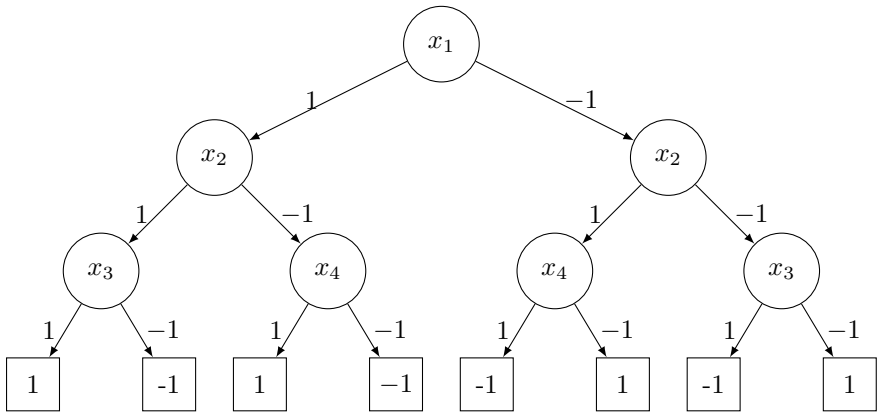

Figure 2.1: Decision tree representing a 3-query classical algorithm that computes the function $f(x_1, x_2, x_3, x_4) = (x_1 + x_2)x_3/2 + (x_1 - x_2)x_4/2$.

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 ε* 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

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

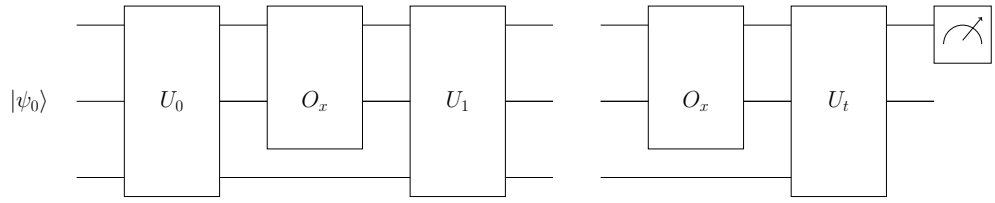

Figure 2.2: Quantum query algorithm.

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 ε* 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 ε. 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 ℓ_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 ℓ_∞ 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 ith 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 Φ 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 Φ 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 Φ 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*

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

Part I

Quantum query complexity via polynomials

Chapter 3

The quantum polynomial method is complete

3.1 Introduction

In this chapter, we will review the evolution of the polynomial method in quantum query complexity. Initially, it was proposed by Beals, Buhrman, Cleve, Mosca and de Wolf as a tool to lower bound quantum query complexity [BBC$^+$01], who were inspired by the *classical* polynomial method of Nisan and Szegedy to lower bound the randomized query complexity [NS94]. This technique has been proven useful in many problems, often providing optimal lower bounds (see e.g., [BKT20] and references therein). More than 15 years after its birth, Arunachalam, Briët and Palazuelos refined the method using completely bounded polynomials. This way, it became a tool that potentially allows one to prove upper bounds to quantum query complexity [ABP19]. In this chapter, based on unpublished joint work with Jop Briët, we show how to use completely bounded polynomials to prove several previously known upper bounds to quantum query complexity. In particular, we reprove the upper bounds by Grover, by Deutsch and Jozsa, and by Bernstein and Vazirani [DJ92, BV93, Gro96], and we show that k-fold forrelation can be computed by k quantum queries [AA15, BS21]. Following the result of Arunachalam et al., Gribling and Laurent proposed a hierarchy of semidefinite programs to compute quantum query complexity [GL19]. However, these semidefinite programs do not give any information about how optimal quantum algorithms look like. Finally, we proposed an alternative hierarchy of semidefinite

programs, also based on completely bounded polynomials, that not only compute quantum query complexity, but also output the description of optimal quantum query algorithms [Esc25]. Putting everything together, we can say that the polynomial method is *complete*, in the sense that it has all the capabilities desirable from a method to understand quantum query complexity; it can be used to show lower bounds and upper bounds, to compute quantum query complexity, and to extract optimal quantum algorithms.

A novelty of this chapter is that the exposition of all the results is elementary and almost self-contained. In particular, we follow [Esc25] and reprove the Christensen and Sinclair factorization theorem of operator spaces via semidifenite programming [CS87]. This result is the key in the refinement of the polynomial method by Arunachalam et al., but it does not belong to the usual toolbox of the theoretical computer scientist [ABP19]. Thus, this chapter offers to the computer scientist a way to fully understand the method of Arunachalam et al. without requiring a background in operator spaces.

3.2 Quantum lower bounds by polynomials

The key observation by Beals et al. that linked quantum query algorithms to polynomials is that the bias of a quantum algorithm that makes t queries is a multilinear polynomial of degree at most $2t$ [BBC$^+$01].

Theorem 3.1. *Let $\mathcal{A} : \{-1,1\}^n \to [-1,1]$ be the bias of t-query quantum algorithm. Then, $\mathcal{A}$ is a polynomial of degree at most $2t$.*

Proof. Before the measurement, on input x, the algorithm prepares a pure quantum state that can be written as

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

for some fixed unitary matrices $U_0, \dots U_t$ and some fixed pure state $|\psi_0\rangle$. Note that by definition of matrix multiplication, the coefficients of $|\psi_t(x)\rangle$ in the computational basis are multilinear polynomials of degree at most t. Hence, if $\{M_{-1}, M_1\}$ is the binary measurement performed by the algorithm, then the bias, $\langle\psi_t(x)|M_1|\psi_t(x)\rangle - \langle\psi_t(x)|M_{-1}|\psi_t(x)\rangle$, is a polynomial of degree at most $2t$. □

A direct consequence of Theorem 3.1 is that to lower bound the quantum query complexity of a Boolean function f, it suffices to show that it cannot be approximated

by polynomials of low degree. More formally, we have the following.

Definition 3.2. Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon \geq 0$. The *ε-approximate degree of f* is the minimum degree of a *bounded* polynomial $p : \{-1,1\}^n \to [-1,1]$ such that $|p(x) - f(x)| \leq \varepsilon$ for every $x \in D$. We use $\widetilde{\deg}_\varepsilon(f)$ to refer to this quantity. We also use $\widetilde{\deg}(f)$ to refer to $\widetilde{\deg}_{2/3}(f)$ and $\deg(f)$ to refer to $\widetilde{\deg}_0(f)$.

Corollary 3.3. *Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ be a Boolean function and let $\varepsilon \geq 0$. Then, $\widetilde{\deg}_\varepsilon(f)/2 \leq Q_\varepsilon(f)$.*

As an example of an application of Corollary 3.3, we will show that the quantum query complexity of the OR_n function is $\Omega(\sqrt{n})$, which implies that Grover's algorithm is optimal [Gro96]. To do that, we prove that $\widetilde{\deg}(\mathrm{OR}_n) = \Omega(\sqrt{n})$, originally shown in [NS94], and then apply Corollary 3.7. We define the OR_n function as $\mathrm{OR}_n(x) = 1$ if $x = 1^n$ and $\mathrm{OR}_n(x) = -1$ otherwise.

Proposition 3.4. $Q(\mathrm{OR}_n) = \Omega(\sqrt{n})$.

Proof. By Corollary 3.3 it suffices to show that $\widetilde{\deg}(\mathrm{OR}_n) = \Omega(\sqrt{n})$. Let $p : \{-1,1\}^n \to [-1,1]$ be a degree-t polynomial that satisfies

$$|p(x) - \mathrm{OR}_n(x)| \leq 2/3$$

for every $x \in \{-1,1\}^n$. Consider the symmetrization p' of p, given by $p'(x) := \sum_{\pi \in \mathcal{S}_n} p(\pi \circ x)/n!$. The symmetric polynomial $p' : \{-1,1\}^n \to \mathbb{R}$ also has degree t, takes values between -1 and 1 and satisfies that

$$|p'(x) - \mathrm{OR}_n(x)| \leq 2/3.$$

By the Minsky-Papert symmetrization technique, Proposition 2.13, there is a univariate polynomial q of degree t such that $q(x) = p'(\sum_i x_i/n)$ for every $x \in \{-1,1\}^n$ and $q([-1,1]) \subseteq [-1,1]$. In particular, $|q((n-2)/n) - (-1)| \leq 2/3$ and $|q(1) - 1| \leq 2/3$. Hence, $|q((n-2)/n) - q(1)| \geq 2/3$. By Markov brothers' inequality, Proposition 2.11, this implies that

$$t = \sqrt{\frac{2/3}{1 - (n-2)/n}} = \Omega(\sqrt{n}),$$

as desired. □

3.2.1 Quantum upper bounds by polynomials

The two main techniques to prove lower bounds for quantum query complexity are the polynomial and the adversary method. The latter was proposed in 2000 by Ambainis [Amb00], and it was quickly refined to also serve as a tool to prove quantum query upper bounds [HLv07]. However, since 2003 it is known that there are functions f such that $Q(f) > (\widetilde{\deg}(f))^c$ for some constant $c > 1$ [Amb03], so the polynomial method does not provide upper bounds to quantum query complexity. A natural question was whether a refinement of the polynomial method would allow it to serve as a tool to prove quantum upper bounds. An attempt of this refinement was proposed by Aaronson, Ambainis, Iraids, Kokainis, Smotrovs [AAI$^+$16]. They strengthened Theorem 3.1 by noticing that the bias of every quantum t-query algorithm is not only a multilinear polynomial of degree at most $2t$, but also the amplitudes of such algorithms are multilinear forms of degree t. This is true because if one looks at the state prepared by the quantum algorithm after t queries it has the form of

$$U_t(O_x \otimes \mathrm{Id}_d)U_{t-1} \dots U_1(O_x \otimes \mathrm{Id}_d)U_0|\psi_0\rangle.$$

In particular, if one queried different inputs $x_1, \dots, x_t$ on every query,

$$U_t(O_{x_t} \otimes \mathrm{Id}_d)U_{t-1} \dots U_1(O_{x_1} \otimes \mathrm{Id}_d)U_0|\psi_0\rangle,$$

then the amplitudes of the resulting state would be linear in every input. Hence, the polynomials representing the bias of quantum query algorithms are more structured than initially noted by Beals et al. [BBC$^+$01]. Unfortunately, as shown in the work by Aaronson et al., the corresponding notion of polynomial degree also fails to provide upper bounds to quantum query complexity. However, the idea of Aaronson et al. was in the correct direction. Shortly after, Arunachalam, Briët and Palazuelos realized that if instead of querying binary strings the algorithms queried any contractions (matrices with operator norm at most 1) $X_1 \dots, X_t$ the amplitudes of the resulting vector,

$$U_t X_t U_{t-1} \dots U_1 X_1 U_0|\psi_0\rangle,$$

would still be linear in $X_1, \dots, X_t$ and bounded by 1 in absolute value [ABP19]. Furthermore, the same is true if one takes *tensor products with identity,* meaning that for every $m \in \mathbb{N}$, every m-dimensional vector $|\phi\rangle$ and contractions $X_1, \dots, X_t$ we have

that the amplitudes of

$$(U_t \otimes \mathrm{Id}_m)X_t(U_{t-1} \otimes \mathrm{Id}_m)\dots(U_1 \otimes \mathrm{Id}_m)X_1(U_0 \otimes \mathrm{Id}_m)(|\psi_0\rangle \otimes |\phi\rangle)$$

are linear in every $X_1, \dots, X_t$ and bounded by 1. As this is true for every $m \in \mathbb{N}$, the bias of quantum query algorithms are, in some sense that we specify below, *completely bounded* polynomials. Surprisingly, Arunachalam et al. showed that the corresponding notion of degree fully characterizes quantum query complexity, enabling the polynomial method to be a potential tool to prove quantum upper bounds. In the rest of the section, we will make this idea rigorous, and give examples of quantum upper bounds by polynomials.

3.3 The completely bounded polynomial method

We start by defining a notion of completely bounded degree, and we will later prove that it characterizes quantum query complexity.

Definition 3.5. Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon \geq 0$. The *ε-approximate completely bounded degree of f* is the minimum $t \in \mathbb{N}$ such that there exists a t-linear form $T : \mathbb{R}^{2n} \times \cdots \times \mathbb{R}^{2n} \to \mathbb{R}$ such that

- $\|T\|_{\mathrm{cb}} \leq 1$,
- and $|T((x,1^n),\dots,(x,1^n)) - p(x)| \leq \varepsilon \quad \forall\, x \in D$.

We use $\widetilde{\mathrm{cbdeg}}_\varepsilon(f)$ to refer to this quantity and $\widetilde{\mathrm{cbdeg}}(f)$ to refer to $\widetilde{\mathrm{cbdeg}}_{2/3}(f)$.

As we argued at the beginning of this section, every t-query quantum algorithm determines a completely bounded form T, so we have that $Q_\varepsilon(f) \geq \widetilde{\mathrm{cbdeg}}_\varepsilon(f)/2$. This strengthens the original polynomial method, because $\|T\|_\infty \leq \|T\|_{\mathrm{cb}}$. Given that there exist separations between the infinity and the completely bounded norms, see for instance [BP19], it is expected that this refinement of the polynomial method allows one to prove stronger quantum lower bounds. Additionally, Arunachalam et al. showed that $Q_\varepsilon(f) = \widetilde{\mathrm{cbdeg}}_\varepsilon(f)/2$, turning the polynomial method into a tool to prove quantum upper bounds.

Theorem 3.6 (Quantum query algorithms are completely bounded forms [ABP19])**.** *Let $p : \{-1,1\}^n \to \mathbb{R}$. Then, the following are equivalent;*

(a) p is the bias of a t-query quantum algorithm.

(b) There exists a $2t$-linear form $T : \mathbb{R}^{2n} \times \cdots \times \mathbb{R}^{2n} \to \mathbb{R}$ such that

$$\|T\|_{\mathrm{cb}} \leq 1 \quad \text{and} \quad T((x, 1^n), \ldots, (x, 1^n)) = p(x) \quad \forall\, x \in \{-1, 1\}^n.$$

Corollary 3.7 (The completely bounded polynomial method)**.** *Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon \geq 0$. Then, $Q_\varepsilon(f) = \widetilde{\mathrm{cbdeg}}_\varepsilon(f)$*72*.*

In order to prove Theorem 3.6, Arunachalam et al. established a relation between operator spaces, where the completely bounded norm has been widely studied [Pau03], and quantum algorithms. In particular, they realized that a seminal result by Christensen and Sinclair, which asserts that multilinear forms are completely bounded if and only if they factor in a way resembling the structure of quantum algorithms, allows one to determine which polynomials can be produced by quantum query algorithms.

Theorem 3.8 (Christensen and Sinclair factorization [CS87])**.** *Let $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ be a t-linear form. Then, $\|T\|_{\mathrm{cb}} \leq 1$ if and only if there exist $d \in \mathbb{N}$, $(n+d)$-dimensional contractions $A_0, \ldots, A_t$, an $(n+d)$-dimensional unit vector v such that*

$$T(x_1, \ldots, x_t) = \langle v, A_t(\mathrm{Diag}(x_t) \otimes \mathrm{Id}_d) A_{t-1} \ldots A_1 (\mathrm{Diag}(x_1) \otimes \mathrm{Id}_d) A_0 v \rangle,$$

for every $x_1, \ldots, x_t \in \mathbb{R}^n$.

The original statement of Theorem 3.8 works for any operator space, and the one we use corresponds to the particular case of the natural operator space defined by ℓ_∞. Also, the usual formulation of Theorem 3.8 is for complex operator spaces, which was the one applied by Arunachalam et al. [ABP19]. However, Theorem 3.1 is sufficient to prove Theorem 3.6, provided that we assume, without loss of generality, that we use real numbers for quantum query algorithms (see Remark 2.5). In Section 3.4 we will give a new proof of Theorem 3.8, based on [Esc25], via semidefinite programming. Now, we are ready to prove Theorem 3.6.

Proof of Theorem 3.6. We first prove that *a)* $\implies$ *b)*. By Remark 2.5, we have that the bias of a t-query quantum algorithm can be written as

$$\begin{aligned}\mathcal{A}(x) =& \langle v, A_0^\dagger (\mathrm{Diag}(1^n, x) \otimes \mathrm{Id}_d) A_1^\dagger \ldots A_{t-1}^\dagger (\mathrm{Diag}(1^n, x) \otimes \mathrm{Id}_d) A_t^\dagger \\ & \cdot (M_1 - M_{-1}) A_t (\mathrm{Diag}(1^n, x) \otimes \mathrm{Id}_d) A_{t-1} \ldots A_1 (\mathrm{Diag}(1^n, x) \otimes \mathrm{Id}_d) A_0 v \rangle,\end{aligned}$$

where $A_0, \ldots, A_T$ are $(n+d)$-dimensional contractions, v is an $(n+d)$-dimensional unit vector and $\{M_{-1}, M_1\}$ is a $(n+d)$ POVM. If we define the $(2t)$-linear form

$T : \mathbb{R}^{2n} \times \cdots \times \mathbb{R}^{2n} \to \mathbb{R}$ given by

$$\begin{aligned} T(y_1, \ldots, y_{2t}) =& \langle v, A_0^\mathsf{T}(\mathrm{Diag}(y_{2t}) \otimes \mathrm{Id}_d) A_1^\mathsf{T} \ldots A_{t-1}^\mathsf{T}(\mathrm{Diag}(y_{t+1}) \otimes \mathrm{Id}_d) A_t^\mathsf{T} \\ & \cdot (M_1 - M_{-1}) A_t(\mathrm{Diag}(y_t) \otimes \mathrm{Id}_d) A_{t-1} \ldots A_1(\mathrm{Diag}(y_1) \otimes \mathrm{Id}_d) A_0 v \rangle, \end{aligned}$$

we have that $T((1^n, x), \ldots, (1^n, x)) = \mathcal{A}(x)$. Furthermore, as $\|M_1 - M_{-1}\|_{\mathrm{op}} \leq 1$, by Theorem 3.8 it follows that $\|T\|_{\mathrm{cb}} \leq 1$. Hence, we have showed that $a) \implies b)$.

We now prove that $b) \implies a)$. Let $p : \{-1, 1\}^n \to \mathbb{R}$ be such that there exists a $2t$-linear form $T : \mathbb{R}^{2n} \times \cdots \times \mathbb{R}^{2n} \to \mathbb{R}$ satisfying that

$$\|T\|_{\mathrm{cb}} \leq 1 \quad \text{and} \quad T((x, 1^n), \ldots, (x, 1^n)) = p(x) \quad \forall\, x \in \{-1, 1\}^n.$$

By Theorem 3.8, there exist $d \in \mathbb{N}$, $(n+d)$-dimensional contractions $A_0, \ldots, A_{2t}$ and $(n+d)$-dimensional unit vectors u, v such that

$$T(y_1, \ldots, y_{2t}) = \langle v, A_{2t}(\mathrm{Diag}(y_{2t}) \otimes \mathrm{Id}_d) A_{2t-1} \ldots A_1(\mathrm{Diag}(y_1) \otimes \mathrm{Id}_d) A_0 v \rangle,$$

for every $y_1, \ldots, y_{2t} \in \mathbb{R}^{2n}$. For every $x \in \{-1, 1\}^n$ we define

$$\begin{aligned} v_1(x) &= A_t(\mathrm{Diag}(x, 1^n) \otimes \mathrm{Id}_d) A_{t-1} \ldots A_1(\mathrm{Diag}(x, 1^n) \otimes \mathrm{Id}_d) A_0 v, \\ v_2(x) &= (\mathrm{Diag}(x, 1^n) \otimes \mathrm{Id}_d) A_{t+1}^\mathsf{T} \ldots A_{2t-1}^\mathsf{T}(\mathrm{Diag}(x, 1^n) \otimes \mathrm{Id}_d) A_{2t}^\mathsf{T} v. \end{aligned}$$

Note that $\langle v_2(x), v_1(x) \rangle = T((x, 1^n), \ldots, (x, 1^n))$. Hence, it just remains to define a t-query quantum algorithm whose bias is $\langle v_2(x), v_1(x) \rangle$. To do that, we define $2(n+d)$-dimensional contractions

$$\begin{aligned} \widetilde{A}_0 &= (X \otimes \mathrm{Id}_{n+d}) c\text{-}A_0 (X \otimes \mathrm{Id}_{n+d}) c\text{-}A_{2t}^\mathsf{T} (H \otimes \mathrm{Id}_{n+d}), \\ \widetilde{A}_i &= (X \otimes \mathrm{Id}_{n+d}) c\text{-}A_i (X \otimes \mathrm{Id}_{n+d}) c\text{-}A_{2t-i}^\mathsf{T}, \quad \text{for } i \in [t-1], \\ \widetilde{A}_t &= (H \otimes \mathrm{Id}_{n+d}) c\text{-}A_t (X \otimes \mathrm{Id}_{n+d}), \end{aligned}$$

where c-A is the controlled version of A. Then, we have that the vector prepared by the corresponding quantum query algorithm is

$$\begin{aligned} |\psi(x)\rangle &= \widetilde{A}_t(\mathrm{Id}_2 \otimes \mathrm{Diag}(x, 1^n) \otimes \mathrm{Id}_d) \widetilde{A}_{t-1} \ldots \widetilde{A}_1(\mathrm{Id}_2 \otimes \mathrm{Diag}(x, 1^n) \otimes \mathrm{Id}_d) \widetilde{A}_0(|0\rangle \otimes |v\rangle) \\ &= \frac{1}{2}(|0\rangle \otimes (|v_1(x)\rangle + |v_2(x)\rangle) + (|1\rangle \otimes (|v_1(x)\rangle - |v_2(x)\rangle). \end{aligned}$$

Finally, if we choose the measurement $\{M_{-1}, M_1\}$ to be $M_1 = |0\rangle\langle 0| \otimes \mathrm{Id}_{n+d}$ and

$M_{-1} = |1\rangle\langle 1| \otimes \mathrm{Id}_{n+d}$, then we have that the bias of the quantum algorithm is

$$\mathcal{A}(x) = \langle \psi(x)|(M_1 - M_{-1})|\psi(x)\rangle = \langle v_1(x), v_2(x)\rangle,$$

as desired. □

3.3.1 Examples of quantum upper bounds by polynomials

In this section, we will reprove several quantum upper bounds via the polynomial method. We will show that certain functions are completely bounded polynomials of degree $2t$, and we will invoke Theorem 3.6, which ensures that they are the bias of a t-query quantum algorithm.

Interestingly, for all of the examples of this section, the following non-commutative version of the Cauchy-Schwarz inequality will play a key role.

Lemma 3.9. *Let $X_1, \ldots, X_n \in M_m$ and let $Y_1, \ldots, Y_n \in M_m$. Then,*

$$\Big\| \sum_{i=1}^n X_i Y_i \Big\|_{\mathrm{op}}^2 \leq \Big\| \sum_{i=1}^n X_i X_i^\mathsf{T} \Big\|_{\mathrm{op}} \Big\| \sum_{i=1}^n Y_i^\mathsf{T} Y_i \Big\|_{\mathrm{op}}.$$

Proof. Consider the following matrices

$$X = \begin{pmatrix} X_1 & \ldots & X_n \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} Y_1 & \ldots & 0 \\ Y_2 & \ldots & 0 \\ \vdots & & \vdots \\ Y_n & \ldots & 0 \end{pmatrix}.$$

First, we have that $\|XY\|_{\mathrm{op}}^2 \leq \|XX^\mathsf{T}\|_{\mathrm{op}} \|Y^\mathsf{T} Y\|_{\mathrm{op}}$. Finally, we have that $\|XY\|_{\mathrm{op}}^2 = \|\sum_{i=1}^n X_i Y_i\|_{\mathrm{op}}^2$, $\|XX^\mathsf{T}\|_{\mathrm{op}} = \|\sum_{i=1}^n X_i X_i^\mathsf{T}\|_{\mathrm{op}}$ and $\|Y^\mathsf{T} Y\|_{\mathrm{op}} = \|\sum_{i=1}^n Y_i^\mathsf{T} Y_i\|_{\mathrm{op}}$. □

Reproving Deutsch-Jozsa

Deutsch and Jozsa gave a 1-query quantum algorithm whose bias is a Boolean function $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ whose classical query complexity is $\Omega(n)$ [DJ92]. Here,

$$D = \{x \in \{-1^n, 1^n\} : x \text{ is balanced}\} \cup \{-1^n, 1^n\},$$

where x is balanced if it has the same number of -1's and 1's, and f is given by

$$f(x) = \begin{cases} 1 & \text{if } x \in \{-1^n, 1^n\}, \\ -1 & \text{if } x \text{ is balanced,} \end{cases}$$

Here, we reprove the result by Deutsch and Jozsa showing that there exists a bilinear form $T : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{R}$ such that

$$\|T\|_{\mathrm{cb}} \leq 1 \quad \text{and} \quad T((x, 1^n), (x, 1^n)) = f(x) \quad \forall\, x \in D.$$

This bilinear form is given by

$$T((x, x'), (y, y')) = 2\mathbb{E}_{i\in[n]} x_i \mathbb{E}_{j\in[n]} y_j - \mathbb{E}_{i\in[n]} x_i y_i,$$

where $x, x', y, y' \in \{-1, 1\}^n$ and the expectation is taken with respect to the uniform distribution on $[n]$. (The form T does not depend on the variables x' and y', but we write it like that for consistency with Theorem 3.6). It is routine to check that $T((x, 1^n), (x, 1^n)) = f(x)$ if $x \in \{-1^n, 1^n\}$ or x is balanced. To show that $\|T\|_{\mathrm{cb}} \leq 1$, note that for any contractions $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ it follows from Lemma 3.9 that

$$\begin{aligned} \|\mathbb{E}_i X_i (2\mathbb{E}_j Y_j - Y_i)\|_{\mathrm{op}}^2 &\leq \|\mathbb{E}_i X_i X_i^{\mathsf{T}}\|_{\mathrm{op}} \|\mathbb{E}_i (2\mathbb{E}_j Y_j - Y_i)^{\mathsf{T}} (2\mathbb{E}_k Y_k - Y_i)\|_{\mathrm{op}} \\ &\leq \|4\mathbb{E}_{j,k} Y_j^{\mathsf{T}} Y_k - 2\mathbb{E}_{i,j} Y_j^{\mathsf{T}} Y_i - 2\mathbb{E}_{i,k} Y_i^{\mathsf{T}} Y_k + \mathbb{E}_i Y_i^{\mathsf{T}} Y_i\|_{\mathrm{op}} \\ &= \|\mathbb{E}_i Y_i^{\mathsf{T}} Y_i\|_{\mathrm{op}} \\ &\leq 1. \end{aligned}$$

Reproving k-fold forrelation

We now consider the problem where, given k-Boolean functions $f_1, \ldots, f_k : \{0,1\}^n \to \{-1, 1\}$, the goal is to compute its k-fold forrelation (standing for *Fourier correlation*)$\mathrm{forr}_{\mathrm{k}} : \{-1, 1\}^{2^n} \times \cdots \times \{-1, 1\}^{2^n} \to \mathbb{R}$, which is given by

$$\begin{aligned} \mathrm{forr}_{\mathrm{k}}(f_1, \ldots, f_k) = \frac{1}{2^{\frac{n(k-1)}{2}}} \sum_{x_1, \ldots, x_{k-1} \in \{0,1\}^n} & f_1(x_1)(-1)^{\langle x_1, x_2\rangle} f_2(x_2) \ldots \\ & \cdot (-1)^{\langle x_{k-2}, x_{k-1}\rangle} f_{k-1}(x_{k-1}) \widehat{f_k}(x_{k-1}), \end{aligned}$$

where $\langle x, y\rangle = \sum_i x_i y_i$. Here, the queries are made to the truth tables of $f_1, \ldots, f_k$. Aaronson and Ambainis introduced this problem as a candidate to witness the largest possible separation between quantum and query complexities [AA15], which was later

confirmed by Bansal and Sinha [BS21]. Here, we reprove that f can be computed as the bias of a quantum algorithm that makes k queries, one to each $f_1, \dots, f_k$. Note that this is not the model that we have considered so far, where all the queries where made to the same input. However, a simple modification of Theorem 3.6 ensures that such an algorithm exists if $\mathrm{forr}_{\mathrm{k}}$, which is a k-linear form, satisfies $\|\mathrm{forr}_{\mathrm{k}}\|_{\mathrm{cb}} \leq 1$. Thus, it suffices to check the latter. Indeed, for m-dimensional orthogonal matrices $F_1(x_1), \dots, F_k(x_k)$ we have that

$$\|(\mathrm{forr}_{\mathrm{k}})_m(F_1, \dots, F_k)\|_{\mathrm{op}}^2 = \frac{1}{2^{n(k-1)}} \| \sum_{x_1} F_1(x_1) \sum_{x_2 \dots x_n} (-1)^{\langle x_1, x_2 \rangle} F_2(x_2) \dots \widehat{F_k}(x_{k-1}) \|_{\mathrm{op}}^2,$$

where $\widehat{F}_k(x_{k-1}) = \mathbb{E}_{x_k} (-1)^{\langle x_{k-1}, x_k \rangle} F_k(x_k)$ is the matrix-valued Fourier coefficient. Next,

$$\begin{aligned}
\|(\mathrm{forr}_{\mathrm{k}})_m(F_1, \dots, F_k)\|_{\mathrm{op}}^2 &\leq \frac{1}{2^n} \| \sum_{x_1} F_1(x_1) F_1^{\mathsf{T}}(x_1) \|_{\mathrm{op}} \\
&\quad \cdot \frac{1}{2^{n(k-2)}} \| \underbrace{\sum_{x_2, \dots, x_n'} \dots F_2^{\mathsf{T}}(x_2) \underbrace{\left(\sum_{x_1} (-1)^{\langle x_1, x_2 \rangle} (-1)^{\langle x_1, x_2' \rangle} \right)}_{2^n \delta_{x_2, x_2'}} F_2(x_2') \dots}_{(*)} \|_{\mathrm{op}} \\
&\leq \frac{1}{2^{n(k-3)}} \| \sum_{x_2} \left(\sum_{x_3, \dots, x_n} (-1)^{\langle x_2, x_3 \rangle} F_3(x_3) \dots \right)^{\mathsf{T}} F_2^{\mathsf{T}}(x_2) F_2(x_2) \\
&\quad \cdot \left(\sum_{x_3, \dots, x_n} (-1)^{\langle x_2, x_3 \rangle} F_3(x_3) \dots \right) \|_{\mathrm{op}},
\end{aligned}$$

where in the first line we have applied Lemma 3.9, and in the third line that $F_1(x_1)$ are orthogonal matrices. Now, as $F_2^{\mathsf{T}}(x_2) F_2(x_2) = \mathrm{Id}_m$, we have that

$$\begin{aligned}
&\|(\mathrm{forr}_{\mathrm{k}})_m(F_1, \dots, F_k)\|_{\mathrm{op}}^2 \\
&\leq \frac{1}{2^{n(k-3)}} \| \underbrace{\sum_{x_3, x_3', \dots, x_n', x_n'} \dots F_3^{\mathsf{T}}(x_3) \left(\sum_{x_2} (-1)^{\langle x_2, x_3 \rangle} (-1)^{\langle x_2', x_3' \rangle} \right) F_3(x_3') \dots}_{(**)} \|_{\mathrm{op}}.
\end{aligned}$$

Now, $(**)$ is essentially the same as $(*)$, so iterating the argument that led us from $(*)$ to $(**)$ we arrive at

$$\|(\mathrm{forr_k})_m(F_1,\ldots,F_k)\|_{\mathrm{op}}^2 \leq \|\sum_{x_{k-1}} \widehat{F}_k^{\mathsf{T}}(x_{k-1})\widehat{F}_k(x_{k-1})\|_{\mathrm{op}} = \|\,\mathbb{E}_x\, F_k^{\mathsf{T}}(x)F_k(x)\|_{\mathrm{op}} = 1,$$

where in the first equality we have used Parseval identity and in the second that $F_k(x)$ are orthogonal. Thus, $\mathrm{forr_k}$ is completely bounded, as desired.

Other examples

One can also reprove other well-known quantum upper bounds using polynomials. Briët reproved Grover's upper bound of $O(\sqrt{n})$ quantum queries to compute the OR_n function by showing that the polynomials constructed by Nisan and Szegedy to approximate OR_n are completely bounded [Bri19, NS94]. Also, using a modification of Theorem 3.6, we could show that there exists an algorithm that with one quantum query to the truth table of a Boolean function can sample from its Fourier distribution, reproving Bernstein-Vazirani's celebrated result [BV93]. We will not prove the latter claim because it would require introducing more notation and would not add conceptual value, as we have already accomplished the purpose of this section: demonstrating that quantum upper bounds can follow from the polynomial method.

3.4 From polynomials to quantum algorithms

In this section, we will start by giving an alternative proof of the Christensen-Sinclair factorization theorem, Theorem 3.8, via semidefinite programming. Contrary to the original proof, ours is elementary, constructive and does not need to use the Hahn-Banach theorem (just a finite-dimensional separation result). We will follow [Esc25], where a more general version of Christensen and Sinclair's result is proven. After, we will use the fact that this proof is based on semidefinite programming and is constructive to give a hierarchy of semidefinite programs that computes quantum query complexity and outputs optimal quantum query algorithms.

3.4.1 Christensen-Sinclair factorization via SDPs

We will prove an equivalent version of Theorem 3.8. To state it, we should introduce the representation norm of a t-linear form $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$, which is given by

$$\begin{aligned} \|T\|_{\mathrm{rep}} = \inf \quad & w \\ \text{s.t.} \quad & T(x_1, \ldots, x_t) = \langle u, A_0(\mathrm{Diag}(x_1) \otimes \mathrm{Id}_d) A_1 \ldots A_{t-1} (\mathrm{Diag}(x_t) \otimes \mathrm{Id}_d) A_t v \rangle \,, \\ & \forall\ x_1, \ldots, x_t \in \mathbb{R}^n, \\ & d \in \mathbb{N},\ u, v \in \mathbb{R}^d,\ \|u\|_2^2 = \|v\|_2^2 = w, \\ & A_0 \in M_{d,nd},\ A_1, \ldots, A_{t-1} \in M_{nd,nd},\ A_t \in M_{nd,d} \text{ contractions.} \end{aligned} \tag{3.1}$$

Now, we can rewrite Theorem 3.10 in the following way.

Theorem 3.10 (Christensen and Sinclair factorization [CS87])**.** *Let $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ be a t-linear form. Then, $\|T\|_{\mathrm{cb}} = \|T\|_{\mathrm{rep}}$.*

We will prove the following result, which is stronger than Theorem 3.10.

Theorem 3.11. *Given a t-linear form $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$, there is a pair of semidefinite programs (P_{CS}) and (D_{CS}) such that*

(i) (P_{CS}) optimal value equals $\|T\|_{\mathrm{rep}}$,

(ii) (D_{CS}) optimal value equals $\|T\|_{\mathrm{cb}}$,

(iii) (D_{CS}) is the dual of (P_{CS}) and their optimal values are equal.

Theorem 3.11 has three important consequences. The first one is already clear from the statement, and the other two will become clear later (see Remark 3.12). These consequences are:

(a) Theorem 3.11 implies Theorem 3.10;

(b) (P_{CS}) and (D_{CS}) have $O(\mathrm{poly}(n)^t)$ variables, so the known algorithms to approximate semidefinite programs can be used to efficiently compute the completely bounded norm. This will imply that there is a hierarchy of SDPs to compute quantum query complexity.

(c) From the solution returned by these algorithms one can extract a description of the vectors and matrices appearing in a factorization as in Eq. (3.1). This will imply that optimal quantum query algorithms can be extracted from the hierarchy of SDPs mentioned in Item (b).

We divide the proof of Theorem 3.11 in 3 parts. In the first, we introduce (P_{CS}) and prove Theorem 3.11 (*i*), in the second we introduce (D_{CS}) and prove Theorem 3.11 (*ii*), and in the third we show that (P_{CS}) and (D_{CS}) are semidefinite programs and prove Theorem 3.11 (*iii*).

The primal semidefinite program

In this section, we introduce (P_{CS}) and prove Theorem 3.11 Item (i). Before doing that, we give some intuition for why $\|T\|_{\mathrm{rep}}$ can be formulated as a semidefinite program. Assume that T factors as in Eq. (3.1). Then, we consider the following block structure for the contractions A_s:

$$A_0 = \begin{pmatrix} A_0(1) & \dots & A_0(n) \end{pmatrix}, \; A_s = \begin{pmatrix} A_s(1,1) & \dots & A_s(1,n) \\ \vdots & \ddots & \vdots \\ A_s(n,1) & \dots & A_s(n,n) \end{pmatrix}, \; A_t = \begin{pmatrix} A_t(1) \\ \dots \\ A_t(n) \end{pmatrix}, \tag{3.2}$$

for $s \in [t-1]$. We also define the following vectors,

$$v_i = A_t(i)v, \text{ for } i \in [n], \tag{3.3}$$

$$v_{\mathbf{i}} = A_{t-s}((i_1, i_2)) \dots A_{t-1}((i_s, i_{s+1}))A_t(i_{s+1})v, \text{ for } \mathbf{i} \in [n]^{s+1}, \; s \in [t-1], \tag{3.4}$$

$$v'_{\mathbf{i}} = A_0(i_1)A_1((i_1, i_2)) \dots A_t(i_t)v, \text{ for } \mathbf{i} \in [n]^t. \tag{3.5}$$

We note that $T_{\mathbf{i}} = \langle u, v'_{\mathbf{i}} \rangle$. Hence, $T_{\mathbf{i}}$ is encoded in the entries of $Y = \mathrm{Gram}\{u, v_{\mathbf{i}}, v'_{\mathbf{i}}\}$ (which corresponds to (3.7) below). In addition, the fact that the A_i are contractions can be encoded in the entries of this Gram matrix (which gives rise to Eqs. (3.9)

to (3.11) below). With these intuitions, we are ready to state (P_{CS}):

$$\begin{aligned}
\inf \quad & w && (P_{\mathrm{CS}})\\
\text{s.t.} \quad & w \geq 0,\ Y, Y' \succeq 0, && (3.6)\\
& Y'_{0,\mathbf{i}} = T_{\mathbf{i}},\ \mathbf{i} \in [n]^t, && (3.7)\\
& Y'_{0,0} = w, && (3.8)\\
& \sum_{i\in[n]} Y_{i,i} \leq w, && (3.9)\\
& \sum_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s} \preceq \oplus_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}},\ s \in [t-1], && (3.10)\\
& (Y'_{\mathbf{j},\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^t} \preceq \oplus_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{t-1}}, && (3.11)
\end{aligned}$$

where $Y \in M_{n+\cdots+n^t}$ and $Y' \in M_{1+n^t}$. The rows and columns of Y are labeled by the elements of $[n] \cup \cdots \cup [n]^t$, and for Y' they are labeled by the elements of $\{0\} \cup [n]^t$. [1]

Proof of Theorem 3.11. Assume first that T factors as in Eq. (3.1) for some vectors with $\|u\|^2 = \|v\|^2 = w$. Consider the block structure for the contractions A_s given in Eq. (3.2), and define the vectors $v_{\mathbf{i}}$ and $v'_{\mathbf{i}}$ as in Eqs. (3.3) to (3.5). Then, $T_{\mathbf{i}} = \langle u, v_{\mathbf{i}} \rangle$, for every $\mathbf{i} \in [n]^t$. Consider the positive semidefinite matrices

$$Y' := \mathrm{Gram}\{u, v'_{\mathbf{i}} :\ \mathbf{i} \in [n]^t\} \quad \text{and} \quad Y := \mathrm{Gram}\{v_{\mathbf{i}} :\ \mathbf{i} \in [n] \cup \cdots \cup [n]^t\},$$

and label the rows and columns corresponding to u with 0 and the ones corresponding to $v_{\mathbf{i}}$ and $v'_{\mathbf{i}}$ with $\mathbf{i}$. First, we have that $T_{\mathbf{i}} = Y'_{0,\mathbf{i}}$, so Eq. (3.7) is satisfied. Eq. (3.8) follows from the fact that $\|u\|^2 = w$. From the fact that A_t is a contraction, Eq. (3.9) follows:

$$\sum_{i\in[n]} Y_{i,i} = \sum_{i\in[n]} \langle v_i, v_i \rangle = \left\langle v, \sum_{i\in[n]} A_t(i)^{\mathsf{T}} A_t(i) v \right\rangle = \left\langle v, A_t^{\mathsf{T}} A_t v \right\rangle \leq \langle v, v \rangle = w.$$

From the fact that A_s are contractions for $s \in [t-1]$ Eq. (3.10) follows. Indeed, let

[1] Here, given $i \in [n]$ and $\mathbf{j} \in [n]^s$, $i\mathbf{j}$ should be interpreted as the concatenation of i and $\mathbf{j}$, i.e., $i\mathbf{j} = (i, j_1, \ldots, j_s)$.

$\lambda \in \mathbb{R}^{n^s}$. Then,

$$\begin{aligned}
&\left\langle \lambda, \sum_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s} \lambda \right\rangle \\
&= \sum_{i\in[n],\mathbf{j},\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}} \langle v_{i\mathbf{j}}, v_{i\mathbf{j}'} \rangle \lambda_{\mathbf{j}'} \\
&= \sum_{i\in[n],\mathbf{j},\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}} \langle A_{t-s}(i,j_1)v_{\mathbf{j}}, A_{t-s}(i,j_1')v_{\mathbf{j}'} \rangle \lambda_{\mathbf{j}'} \\
&= \sum_{i\in[n],\mathbf{j},\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}} \langle v_{\mathbf{j}}, A_{t-s}^{\mathsf{T}}(j_1,i)A_{t-s}(i,j_1')v_{\mathbf{j}'} \rangle \lambda_{\mathbf{j}'} \\
&= \underbrace{\sum_{\mathbf{j},\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}} \langle v_{\mathbf{j}}, (A_{t-s}^{\mathsf{T}}A_{t-s})(j_1,j_1')v_{\mathbf{j}'} \rangle \lambda_{\mathbf{j}'}}_{(*)},
\end{aligned}$$

where in the second equality we have used that $v_{i\mathbf{j}} = A(i,j_1)v_{\mathbf{j}}$, and in the third line that $A_{t-s}(i,j)^{\mathsf{T}} = A_{t-s}^{\mathsf{T}}(j,i)$. Now, if we define $w_{\mathbf{j}} = (\lambda_{1\mathbf{j}}v_{1\mathbf{j}}, \ldots, \lambda_{n\mathbf{j}}v_{n\mathbf{j}})$, it follows that

$$(*) = \sum_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}} \langle w_{\mathbf{j}}, A_{t-s}^{\mathsf{T}}A_{t-s}w_{\mathbf{j}'} \rangle = \left\langle \left(\sum_{\mathbf{j}} w_{\mathbf{j}}\right), A_{t-s}^{\mathsf{T}}A_{t-s} \left(\sum_{\mathbf{j}'} w_{\mathbf{j}'}\right) \right\rangle.$$

Hence, as $A_{t-s}^{\mathsf{T}}A_{t-s} \preceq \mathrm{Id}$, it is satisfied that

$$\begin{aligned}
(*) &\leq \left\langle \left(\sum_{\mathbf{j}\in[n]^{s-1}} w_{\mathbf{j}}\right), \left(\sum_{\mathbf{j}'\in[n]^{s-1}} w_{\mathbf{j}'}\right) \right\rangle = \sum_{i\in[n],\mathbf{j},\mathbf{j}'\in[n]^{s-1}} \lambda_{i\mathbf{j}} \langle v_{i\mathbf{j}}, v_{i\mathbf{j}'} \rangle \lambda_{i\mathbf{j}'} \\
&= \langle \lambda, \oplus_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]\times[n])^{s-1}\times[n]} \lambda \rangle,
\end{aligned}$$

as desired. The fact that A_0 is a contraction implies Eq. (3.11), and this can be shown similarly to how we just showed that Eq. (3.10) holds.

Now, assume that there exist $Y, Y' \succeq 0$, satisfying equations Eqs. (3.7) to (3.11). Consider $d \in \mathbb{N}$ and vectors $\{u, v_{\mathbf{i}}, v_{\mathbf{i}}\} \in \mathbb{R}^d$ such that

$$Y = \mathrm{Gram}\{v_{\mathbf{i}}\} \quad \text{and} \quad Y' = \mathrm{Gram}\{u, v_{\mathbf{i}}'\}.$$

Eq. (3.8) implies that $\|u\|^2 = w$. We define A_t through its blocks. Let $v \in \mathbb{R}^d$ be a vector with $\|v\|^2 = w$. We define $A_t(i) \in M_d$ as the matrix that maps v to v_i and extend by 0 to the orthogonal complement of $\mathrm{span}\{v\}$. This way, A_t is a contraction,

because

$$\|A_t\|_{\mathrm{op}}^2 = \frac{\langle A_t v, A_t v\rangle}{\langle v, v\rangle} = \frac{1}{w}\sum_{i\in[n]} \langle A_t(i)v, A_t(i)v\rangle = \frac{1}{w}\sum_{i\in[n]} \langle v_i, v_i\rangle = \frac{1}{w}\sum_{i\in[n]} Y_{i,i} \leq 1,$$

where in the inequality we have used Eq. (3.9). The definition of A_{t-s} for $s \in [t-1]$ is slightly more complicated. Given $(i,j) \in [n]\times[n]$, the block $A_{t-s}(i,j)$ is defined as the linear map on $\mathrm{span}\{v_{j\mathbf{j}} : \mathbf{j} \in [n]^{s-1}\}$ by

$$A_{t-s}(i,j)v_{j\mathbf{j}} = v_{ij\mathbf{j}}$$

and extended by 0 to the orthogonal complement. First, as $\{v_{j\mathbf{j}} : \mathbf{j} \in [n]^{s-1}\}$ may not be linearly independent, we have to check that this a good definition, namely that for every $\lambda \in \mathbb{R}^{n^{s-1}}$

$$\sum_{\mathbf{j}\in[n]^{s-1}} \lambda_{j\mathbf{j}} v_{j\mathbf{j}} = 0 \implies \sum_{\mathbf{j}\in[n]^{s-1}} \lambda_{j\mathbf{j}} v_{ij\mathbf{j}} = 0. \tag{3.12}$$

Indeed, we can prove something stronger. For any $\lambda \in \mathbb{R}^{n^{s-1}}$, we define $\tilde{\lambda} \in \mathbb{R}^{n^s}$ by $\tilde{\lambda}_{j'\mathbf{j}} := \delta_{j,j'}\lambda_{\mathbf{j}}$, where j is the second index in the pair (i,j) that indexes the block $A_{t-s}(i,j)$. Then,

$$\begin{aligned}
&\left\langle \sum_{\mathbf{j}\in[n]^{s-1}} \lambda_{j\mathbf{j}} v_{ij\mathbf{j}}, \sum_{\mathbf{j}'\in[n]^{s-1}} \lambda_{j\mathbf{j}'} v_{ij\mathbf{j}'} \right\rangle \\
&= \left\langle \lambda, (Y_{(ij\mathbf{j},ij\mathbf{j}')})_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}}\lambda \right\rangle \\
&= \left\langle \tilde{\lambda}, (Y_{(i\mathbf{j},i\mathbf{j}')})_{\mathbf{j},\mathbf{j}'\in[n]^{s}}\tilde{\lambda} \right\rangle \\
&\leq \left\langle \tilde{\lambda}, \sum_{k\in[n]} (Y_{k\mathbf{j},k\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{s}}\tilde{\lambda} \right\rangle \\
&\leq \left\langle \tilde{\lambda}, \oplus_{k\in[n]} (Y_{k\mathbf{j},k\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}}\tilde{\lambda} \right\rangle \\
&= \left\langle \lambda, (Y_{j\mathbf{j},j\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}}\lambda \right\rangle \\
&= \left\langle \sum_{\mathbf{j}\in[n]^{s-1}} \lambda_{j\mathbf{j}} v_{j\mathbf{j}}, \sum_{\mathbf{j}'\in[n]^{s-1}} \lambda_{j\mathbf{j}'} v_{j\mathbf{j}'} \right\rangle,
\end{aligned}$$

where in the first inequality we have used that $(Y_{k\mathbf{j},k\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s} \succeq 0$ for every $k \in [n]$, and in the second inequality we have used (3.10). Thus, Eq. (3.12) holds. Now, we have to check that A_{t-s} is a contraction. By the definition of A_{t-s}, we just have to

check that for every $\lambda \in \mathbb{R}^{n^s}$,

$$\lambda v := \begin{pmatrix} \sum_{\mathbf{j}\in[n]^{s-1}} \lambda_{1\mathbf{j}} v_{1\mathbf{j}} \\ \vdots \\ \sum_{\mathbf{j}\in[n]^{s-1}} \lambda_{n\mathbf{j}} v_{n\mathbf{j}} \end{pmatrix}$$

is mapped to a vector with smaller or equal norm. Indeed,

$$\begin{aligned} \langle A_{t-s}\lambda v, A_{t-s}\lambda v\rangle &= \sum_{i,\mathbf{j},\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}} \langle v_{i\mathbf{j}}, v_{i\mathbf{j}'}\rangle \lambda_{\mathbf{j}'} \\ &= \left\langle \lambda, \sum_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s} \lambda \right\rangle \\ &\leq \left\langle \lambda, \oplus_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}} \lambda \right\rangle \\ &= \langle \lambda v, \lambda v\rangle , \end{aligned}$$

where in the inequality we have used Eq. (3.10). Finally, we define A_0 through its blocks. $A_0(i)$ is defined by $A_0(i)v_{i\mathbf{j}} = v'_{i\mathbf{j}}$ for $\mathbf{j} \in [n]^{t-1}$ and extended by 0 to the orthogonal complement of $\mathrm{span}\{v_{i\mathbf{j}} : \mathbf{j} \in [n]^{t-1}\}$. Using Eq. (3.11), we can check that these blocks are well-defined and that A_0 is a contraction using a similar argument to the one that we have just used to verify the same properties of A_{t-s}. It just remains to show that (u, v, A_i) defines a factorization for T as in (3.1). Eq. (3.1) holds if and only it holds for a basis of $\mathbb{R}^n$. We verify it for the canonical basis $\{e_i\}_{i\in[n]}$. On the one hand, by definition, we have that $T(e_{i_1}, \ldots, e_{i_t}) = T_{\mathbf{i}}$. On the other hand, a simple calculation shows that

$$\begin{aligned} Y'_{0,\mathbf{i}} &= \langle u, A_0(i_1)A_1((i_1, i_2)) \ldots A_{t-1}((i_{t-1}, i_t))A_t(i_t)v\rangle \\ &= \langle u, A_0(\mathrm{Diag}(e_{i_1}) \otimes \mathrm{Id}_d)A_1 \ldots A_{t-1}(\mathrm{Diag}(e_{i_t}) \otimes \mathrm{Id}_d)A_t v\rangle. \end{aligned}$$

Hence, by Eq. (3.7) follows that

$$T(e_{i_1}, \ldots, e_{i_t}) = \langle u, A_0(\mathrm{Diag}(e_{i_1}) \otimes \mathrm{Id}_d)A_1 \ldots A_{t-1}(\mathrm{Diag}(e_{i_t}) \otimes \mathrm{Id}_d)A_t v\rangle,$$

as desired. □

Remark 3.12. (P_{CS}) has $\mathrm{poly}(n)^t$ variables, so Item (b) holds. Item (c) can be inferred from the second part of the proof of Theorem 3.11 Item (i), where a recipe to extract a factorization as in Eq. (3.1) for $(Y'_{0,\mathbf{i}})_{\mathbf{i}}$ satisfying Eqs. (3.8) to (3.11) is given.

The dual semidefinite program

In this section, we introduce (D_{CS}) and prove Theorem 3.11 Item (ii). (D_{CS}) is given by:

$$\begin{aligned}
\sup \quad & \sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}} y_{0,\mathbf{i}} && (D_{\mathrm{CS}})\\
\text{s.t.} \quad & y_0, y_0' \geq 0, \ \left(y_{\mathbf{i},\mathbf{i}'}\right)_{\mathbf{i},\mathbf{i}'\in[n]^s} \succeq 0, \ \text{for } s\in[t], && (3.13)\\
& y_0 + y_0' \leq 1, && (3.14)\\
& y_0 \geq y_{i,i}, \quad \text{for } i\in[n] && (3.15)\\
& (y_{\mathbf{j}}, y_{\mathbf{j}'})_{\mathbf{j},\mathbf{j}\in[n]^s} \geq (y_{i\mathbf{j}}, y_{i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s} \quad \text{for } i\in[n],\ s\in[t-1], && (3.16)\\
& \left(\begin{array}{c|c} y_0' & \cdots \quad (y_{0,\mathbf{i}})_{\mathbf{i}\in[n]^t}/2 \quad \cdots \\ \hline \vdots & \\ \frac{(y_{0,\mathbf{i}})_{\mathbf{i}\in[n]^t}}{2} & \left(y_{\mathbf{i},\mathbf{i}'}\right)_{\mathbf{i},\mathbf{i}'\in[n]^t} \\ \vdots & \end{array}\right) \succeq 0, && (3.17)
\end{aligned}$$

Before diving into the proof, we give some intuition of why the optimal value of (D_{CS}) is $\|T\|_{\mathrm{cb}}$. One should note that Eq. (3.17) means that the variables $y_{0,\mathbf{i}}$ can be written as $\langle u, v_{\mathbf{i}}\rangle$ for some vectors u, $v_{\mathbf{i}}$. Then, roughly speaking, Eqs. (3.15) and (3.16) encode that the $v_{\mathbf{i}}$ equal $X_1(i_1)\dots X_t(i_t)v$ for some contractions $X_1(i_1),\dots,X_t(i_t)$ and a vector v, and Eq. (3.14) encodes that u and v are bounded vectors.

Proof of Theorem 3.11 Item (ii). First, we note that Eq. (3.13) means that there exist $d\in\mathbb{N}$ and vectors $\{u, v, v_{\mathbf{i}} : \mathbf{i}\in[n]^s,\ s\in[t]\}\subset\mathbb{R}^m$ such that $y_0' = \langle u,u\rangle$, $y_0 = \langle v,v\rangle$, and $y_{\mathbf{i},\mathbf{i}'} = \langle v_{\mathbf{i}}, v_{\mathbf{i}'}\rangle$ for every $\mathbf{i}\in[n]^s$ and $s\in[t]$. Then, Eq. (3.15) means that $\langle u,u\rangle + \langle v,v\rangle \leq 1$ and Eq. (3.17) means that $y_{0,\mathbf{i}} = 2\langle u, v_{\mathbf{i}}\rangle$ for every $\mathbf{i}\in[n]^t$. Thus,

we can rewrite Eq. (D_{CS}) as

$$\begin{aligned}
\sup \quad & 2 \sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}} \langle u, v_{\mathbf{i}} \rangle, && (3.18)\\
\text{s.t.} \quad & m \in \mathbb{N},\ u, v, v_{\mathbf{i}} \in \mathbb{R}^m,\ \mathbf{i} \in [n]^s,\ s \in [t], \\
& \langle u, u \rangle + \langle v, v \rangle \leq 1, \\
& \langle v, v \rangle \geq \langle v_i, v_i \rangle, \quad \text{for } i \in [n] && (3.19)\\
& (\langle v_{\mathbf{j}}, v_{\mathbf{j}'} \rangle)_{\mathbf{j},\mathbf{j}\in[n]^s} \geq (\langle v_{i\mathbf{j}}, v_{i\mathbf{j}'} \rangle)_{\mathbf{j},\mathbf{j}'\in[n]^s} \quad \text{for } i \in [n],\ s \in [t-1] && (3.20)\\
& && (3.21)
\end{aligned}$$

Next, we will show that Eqs. (3.19) and (3.20) are equivalent to the existence of contractions $X_1, \ldots, X_t \in M_m$ such that

$$v_{\mathbf{i}} = X_{t-s+1}(i_1) \ldots X_t(i_s) v, \tag{3.22}$$

for every $\mathbf{i} \in [n]^s$ and every $s \in [t]$. Indeed, assume that Eqs. (3.19) and (3.20) hold. Then, for every $i \in [n]$ and every $s \in \{0\} \cup [t]$, we define

$$X_{t-s}(i) v_{\mathbf{j}} := v_{i\mathbf{j}}$$

for every $\mathbf{j} \in [n]^s$ and extend it by 0 on the orthogonal complement of $\mathrm{span}\{v_{\mathbf{j}} : \mathbf{j} \in [n]^s\}$. We have to check that the $X_{t-s}(i)$ are well-defined as linear maps. Namely, that for every $\lambda \in \mathbb{R}^{n^s}$ we have

$$\sum_{\mathbf{j}\in[n]^s} \lambda_{\mathbf{j}} v_{\mathbf{j}} = 0 \implies \sum_{\mathbf{j}\in[n]^s} \lambda_{\mathbf{j}} v_{i\mathbf{j}} = 0.$$

In fact, we can prove that the $X_{t-s}(i)$ are well-defined and contractions at the same time. Indeed, for $\lambda \in \mathbb{R}^{n^s}$ we have that

$$\begin{aligned}
\left\langle \sum_{\mathbf{j}\in[n]^s} \lambda_{\mathbf{j}} v_{i\mathbf{j}}, \sum_{\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}'} v_{i\mathbf{j}'} \right\rangle &= \left\langle \lambda, \left(\langle v_{i\mathbf{j}}, v_{i\mathbf{j}'} \rangle \right)_{\mathbf{j},\mathbf{j}'\in[n]^s} \lambda \right\rangle \\
&\leq \left\langle \lambda, \left(\langle v_{\mathbf{j}}, v_{\mathbf{j}'} \rangle \right)_{\mathbf{j},\mathbf{j}'\in[n]^s} \lambda \right\rangle \\
&= \left\langle \sum_{\mathbf{j}\in[n]^s} \lambda_{\mathbf{j}} v_{\mathbf{j}}, \sum_{\mathbf{j}'\in[n]^s} \lambda_{\mathbf{j}'} v_{\mathbf{j}'} \right\rangle,
\end{aligned}$$

where we have used Eq. (3.20) (or Eq. (3.19) if $s = 0$).

On the other hand, if Eq. (3.22) holds, it is a routine check showing that Eqs. (3.19) and (3.20) hold. Putting everything together, we can rewrite (3.18) as

$$\begin{aligned} \sup \quad & 2 \sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}} R_{\mathbf{i}}, \\ \text{s.t.} \quad & R \in \mathbb{R}^{n^t},\ m \in \mathbb{N},\ u, v \in \mathbb{R}^m, X_s \in M_m \text{ contractions for } s \in [t], \\ & \langle u, u\rangle + \langle v, v\rangle \leq 1, \\ & R_{\mathbf{i}} = \langle u, X_1(i_1) \dots X_t(i_t) v\rangle\ , \text{ for } \mathbf{i} \in [n]^t. \end{aligned} \tag{3.23}$$

We finally claim that the above optimization problem is equivalent to

$$\begin{aligned} \sup \quad & 2 \sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}} R_{\mathbf{i}}, \\ \text{s.t.} \quad & R \in \mathbb{R}^{n^t},\ m \in \mathbb{N},\ u, v \in \mathbb{R}^m, X_s \in M_m \text{ contractions for } s \in [t], \\ & \langle u, u\rangle\ , \langle v, v\rangle \leq 1/2, \\ & R_{\mathbf{i}} = \langle u, X_1(i_1) \dots X_t(i_t) v\rangle\ , \text{ for } \mathbf{i} \in [n]^t. \end{aligned} \tag{3.24}$$

We first note that the optimum of Eq. (3.23) is greater or equal than the optimum of Eq. (3.24), because the feasible region is larger in the case of Eq. (3.23). On the other hand, if one picks a feasible instance (u, v, X) of Eq. (3.23), one can define the instance $(\tilde{u}, \tilde{v}, X)$ by

$$\tilde{u} = \frac{u\sqrt{\|u\|^2 + \|v\|^2}}{\sqrt{2}\|u\|},\ \tilde{v} = \frac{v\sqrt{\|u\|^2 + \|v\|^2}}{\sqrt{2}\|v\|},$$

which is feasible for Eq. (3.24) and attains a value greater or equal than (u, v, X), because

$$\begin{aligned} \Big| \sum T_{\mathbf{i}} \langle \tilde{u}, X_1(i_1) \dots X_t(i_t) \tilde{v}\rangle \Big| &= \frac{\|u\|^2 + \|v\|^2}{2\|u\|\|v\|} \Big| \sum T_{\mathbf{i}} \langle u, X_1(i_1) \dots X_t(i_t) v\rangle \Big| \\ &\geq \Big| \sum T_{\mathbf{i}} \langle u, X_1(i_1) \dots X_t(i_t) v\rangle \Big|. \end{aligned}$$

Now, the result follows from the fact that the optimal value of Eq. (3.24) is $\|T\|_{\mathrm{cb}}$. □

Strong duality

Finally, we prove Theorem 3.11 Item (iii).

Proof of Theorem 3.11 Item (iii). First, we show that (P_{CS}) can be expressed as in the canonical form of (P) in Eq. (2.15). To do that we introduce the slack matrix variables Z and Z' and write (P_{CS}) as

$$\begin{aligned}
&\inf \quad w && (\widetilde{P}_{\mathrm{CS}})\\
&\text{s.t.} \quad X := \begin{pmatrix} w & 0 & 0 & 0 & 0 \\ 0 & Y & 0 & 0 & 0 \\ 0 & 0 & Y' & 0 & 0 \\ 0 & 0 & 0 & Z & 0 \\ 0 & 0 & 0 & 0 & Z' \end{pmatrix} \succeq 0 \\
&\qquad Y_{0,\mathbf{i}} = T_{\mathbf{i}}, \ \mathbf{i} \in [n]^t, && (3.25)\\
&\qquad w - Y'_{0,0} = 0, && (3.26)\\
&\qquad w - \sum_{i\in[n]} Y_{i,i} = Z_{0,0}, && (3.27)\\
&\qquad \oplus_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{s-1}} - \sum_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s} = (Z_{\mathbf{j},\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^s}, \ s \in [t-1], && (3.28)\\
&\qquad \oplus_{i\in[n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^{t-1}} - (Y'_{\mathbf{j},\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[n]^t} = Z', && (3.29)
\end{aligned}$$

One can regard X as a positive semidefinite matrix with some entries set to 0, which can be imposed via linear constraints. Additionally, note that the objective function w is a linear function of the entries of X, and so are the restrictions Eqs. (3.25) to (3.29). Hence, (P_{CS}) has the form of (P) in Eq. (2.15).

Second, we show that (D_{CS}) can be expressed as in the canonical form of (D) in Eq. (2.15). We can rewrite (D_{CS}) as

$$\begin{aligned}
\sup \quad & \sum_{\mathbf{i}\in[n]^t} T_{\mathbf{i}} R_{\mathbf{i}} && (\widetilde{D}_{\mathrm{CS}})\\
\text{s.t.} \quad & y_0, y_0', R_{\mathbf{i}}, y_{\mathbf{i},\mathbf{i}'}, \mathbf{i}, \mathbf{i}' \in [n]^s,\ s \in [t] \\
& y_0 \geq 0,\ y_0' \geq 0,\ \sum_{\mathbf{j},\mathbf{j}'\in[n]^s} y_{\mathbf{j},\mathbf{j}'} E_{\mathbf{j},\mathbf{j}'} \succeq 0,\ \text{for } s \in [t], && (3.30)\\
& y_0 + y_0' \leq 1, && (3.31)\\
& y_0 \geq y_{i,i}, \quad \text{for } i \in [n] && (3.32)\\
& \sum_{\mathbf{j},\mathbf{j}'\in[n]^s} (y_{\mathbf{j},\mathbf{j}'} - y_{i\mathbf{j},i\mathbf{j}'}) E_{\mathbf{j},\mathbf{j}'} \succeq 0, \quad \text{for } i \in [n],\ s \in [t-1] && (3.33)\\
& y_0' E_{0,0} + \sum_{\mathbf{j}\in[n]^t} R_{\mathbf{j}} \frac{E_{0,\mathbf{j}} + E_{\mathbf{j},0}}{2} + \sum_{\mathbf{i},\mathbf{i}'\in[n]^t} y_{\mathbf{i},\mathbf{i}'} E_{\mathbf{i},\mathbf{i}'} \succeq 0. && (3.34)
\end{aligned}$$

Thus, we have written (D_{CS}) as an optimization problem $(\widetilde{D}_{\mathrm{CS}})$ on the variables $y_0, y_0', R_{\mathbf{i}}, y_{\mathbf{i},\mathbf{i}'}$. Moreover, the objective function is a linear combination of these variables. Also, the constraints are positive semidefinite constraints on matrices that are linear combinations of other matrices, where the coefficients of these linear combinations are $y_0, y_0', R_{\mathbf{i}}, y_{\mathbf{i},\mathbf{i}'}$. Putting everything together, it follows that (D_{CS}) is of the form of (D) in Eq. (2.15).

Third, we show that (D_{CS}) is the dual of (P_{CS}). Equivalently, we prove that $(\widetilde{D}_{\mathrm{CS}})$ is the dual of $(\widetilde{P}_{\mathrm{CS}})$. To take the dual of a primal semidefinite program such as $(\widetilde{P}_{\mathrm{CS}})$ it is convenient to assign a dual variable to every linear constraint. We assign $R_{\mathbf{i}}$ to the constraints in Eq. (3.25), y_0' to Eq. (3.26), y_0 to Eq. (3.27), and $y_{\mathbf{i},\mathbf{i}}$ to Eqs. (3.28) and (3.29). In addition, one should note that every variable in the primal corresponds to a restriction in the dual. With this in mind, from the definition of the dual given in Eq. (2.15), it follows that $(\widetilde{D}_{\mathrm{CS}})$ is the dual of $(\widetilde{P}_{\mathrm{CS}})$, and that the constraints of Eq. (3.30) correspond to variable Z in $(\widetilde{P}_{\mathrm{CS}})$, Eq. (3.31) to varible w, and Eqs. (3.32) to (3.34) to variable Y.

Finally, we show that the conditions of Theorem 2.20 are satisfied by $(\widetilde{P}_{\mathrm{CS}})$ and $(\widetilde{D}_{\mathrm{CS}})$, which implies that their values are equal. $(\widetilde{P}_{\mathrm{CS}})$ is feasible, as every T factors as in Eq. (3.1) for some u, v with sufficiently large norm (if this was not true, $\|T\|_{\mathrm{cb}}$ would not be a norm). In addition, we claim that the following parameters define a

strictly positive feasible instance for $(\widetilde{D}_{\mathrm{CS}})$

$$\begin{aligned} y_0 = y_0' &= \frac{1}{3}, \\ y_{\mathbf{i},\mathbf{j}} &= \frac{\delta_{\mathbf{i},\mathbf{j}}}{3(n+1)^s}, \text{ for } \mathbf{i},\mathbf{j} \in [n]^s,\ s \in [t], \\ R_{\mathbf{i}} &= 0, \text{ for } \mathbf{i} \in [n]^t. \end{aligned}$$

Indeed, with these parameters Eqs. (3.30) to (3.34) read as follows:

$$\begin{aligned} &\frac{1}{3} \geq 0,\ \mathrm{Id} \succ 0 \\ &\frac{1}{3} + \frac{1}{3} \leq 1, \\ &\frac{1}{3} \succ \frac{n}{3(n+1)}, \\ &\frac{1}{3(n+1)^s}\mathrm{Id}_{n^s} \succ \frac{n}{3(n+1)^{s+1}}\mathrm{Id}_{n^s}, \text{ for } s \in [t-1], \\ &\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3(n+1)^t}\mathrm{Id}_{n^t} \end{pmatrix} \succ 0, \end{aligned}$$

and these identities are true because $1 > n/(n+1)$. □

3.4.2 A hierarchy of SDPs to find quantum algorithms

To introduce the announced hierarchy of SDPs, we first note that by Theorem 3.6 it follows that the smallest error that can be achieved when approximating a function $f : D \subseteq \{-1,1\}^n \to \mathbb{R}$ with a t-query quantum algorithm is

$$\begin{aligned} \mathcal{E}(f,t) = \inf \big\{ \varepsilon \geq 0 \mid\ & \exists\ 2t\text{-linear form } T : \mathbb{R}^{2n} \times \cdots \times \times \mathbb{R}^{2n} \to \mathbb{R} \\ & |f(x) - T((x,1^n),\ldots,(x,1^n))| \leq \varepsilon \quad \forall x \in D, \\ & \|T\|_{\mathrm{cb}} \leq 1 \big\}. \end{aligned}$$

Now, an immediate corollary of Theorem 3.11 is the following formulation of $\mathcal{E}(p,t)$ as an SDP.

Corollary 3.13. *Let $f : \{-1,1\}^n \to [-1,1]$ and $t \in \mathbb{N}$. Then,*

$$
\begin{aligned}
\mathcal{E}(f,t) = \inf \quad & \varepsilon \\
\text{s.t.} \quad & \varepsilon \geq 0,\ Y, Y' \succeq 0, \\
& |p(x) - \sum_{\mathbf{i}\in[n]^{2t}} Y'_{0,\mathbf{i}} y_{i_1} \dots y_{i_{2t}}| \leq \varepsilon, \qquad y = (x, 1^n), \forall\ x \in \{-1,1\}^n, \\
& Y'_{0,0} = w, \\
& \sum_{i\in[2n]} Y_{i,i} \leq w, \\
& \sum_{i\in[2n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[2n]^s} \preceq \oplus_{i\in[2n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[2n]^{s-1}},\ s \in [2t-1], \\
& (Y'_{\mathbf{j},\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[2n]^{2t}} \preceq \oplus_{i\in[2n]} (Y_{i\mathbf{j},i\mathbf{j}'})_{\mathbf{j},\mathbf{j}'\in[2n]^{2t-1}},
\end{aligned}
$$

We observe that, as a consequence of Corollary 3.13, we have that $(\mathcal{E}(f,t))_t$ determines a hierarchy of SDPs that computes quantum query complexity. Indeed, to compute $Q_\varepsilon(f)$ one can solve $\mathcal{E}(f,1),\ \mathcal{E}(f,2), \dots$ and stop at the smallest t_0 satisfying $\mathcal{E}(f,t_0) \leq \varepsilon$. Then, we will have that $t_0 = Q_\varepsilon(f)$. Additionally, from an optimal solution to $\mathcal{E}(f,t_0)$ one can obtain an optimal quantum algorithm. This can be easily (but tediously) done following the constructions in the proofs of Theorem 3.6 and Theorem 3.11 Item (i),

Comparison with other methods

There are other formulations of $\mathcal{E}(f,t)$ as a SDP: the aforementioned work by Gribling and Laurent [GL19] and by Barnum, Saks, and Szegedy [BSS03]. We will compare these three methods with ours, and also with the adversary method, which does not compute $\mathcal{E}(f,t)$, but provides a SDP that directly computes the quantum query complexity. We remark the following:

- The method of Gribling and Laurent does not provide a description of the approximating quantum algorithm, while the others method do.
- The sizes of the SDPs differ, as shown in Table 3.1. The ones of Corollary 3.13 are considerably smaller than the ones in [BSS03] and the size of the SDP of the adversary method, but they are slightly bigger than the ones in [GL19].
- The adversary method loses constant factors in the characterization of quantum query complexity, and it does not work for exact quantum query complexity. On

	# blocks	block size	# lin. ineq.	# lin. eq.
Adversary method [HLv07]	n	$\|D\|$	0	$\|f^{-1}(1)\|\|f^{-1}(0)\|$
Barnum-Saks-Szegedy [BSS03]	$nt+2$	$\|D\|$	$\|D\|$	$\Theta(t\|D\|^2)$
Gribling-Laurent [GL19]	1	$\Theta(n^t)$	$2\|D\|+1$	$\Theta(n^{2t})$
Corollary 3.13	$4t-2$	$\Theta((2n)^{2t})$	$2\|D\|+3$	$\Theta(2t(2n)^{2t})$

Table 3.1: A comparison of the sizes of the SDPs to compute quantum query complexity. We count the number of linear equalities, inequalities, and PSD blocks, keeping track of the size of the largest block.

the other hand, the other three hierarchies of SDPs do characterize quantum query complexity, including the exact case, without losing constant factors.

Chapter 4

Grothendieck inequalities characterizes converses to the polynomial method

4.1 Introduction

For a Boolean function $f : D \to \{-1, 1\}$ defined on a set $D \subseteq \{-1, 1\}^n$, the celebrated *polynomial method* of Beals, Buhrman, Cleve, Mosca and de Wolf [BBC+01], introduced in Chapter 3, gives a lower bound on the quantum query complexity of f in terms of the approximate degree. Using this method, many well-known quantum algorithms were proved to be optimal in terms of query complexity (see e.g., [BKT20] and references therein).

Since polynomials are simpler objects than quantum query algorithms, it is of interest to know how well approximate degree approximates quantum query complexity. There are total functions f that satisfy $Q(f) \geq \widetilde{\deg}(f)^c$ for some absolute constant $c > 1$ [Amb06, ABDK16]; the second reference gives an exponent $c = 4 - o(1)$, which was shown to be optimal in [ABDK16]. For partial functions it was recently shown that this separation can even be exponential [AB23]. These separations rule out a direct converse to the polynomial method, whereby a given bounded degree-$2t$ polynomial p can be computed by a t-query quantum algorithm $\mathcal{A}$. However, since these results concern functions whose approximate degree grows with n, they leave room for the possibility that such an $\mathcal{A}$ approximates p with some error that depends on t.

For bounded polynomials of degree at most 2, a *multiplicative converse* to the polynomial method was proved in [AAI+16], showing that up to an absolute constant scaling, quadratic polynomials can indeed be computed by 1-query quantum algorithms.

Theorem 4.1 (Quadratic multiplicative converse [AAI+16])**.** *There exists an absolute constant $C \in (0,1]$ such that $\mathcal{E}(Cp,1) = 0$ for every bounded polynomial p of degree at most 2.*

This result directly implies the following *additive* version.

Corollary 4.2 (Quadratic additive converse)**.** *There exists an absolute constant $\varepsilon \in (0,1)$ such that the following holds. For every bounded polynomial p of degree at most 2, we have $\mathcal{E}(p,1) \leq \varepsilon$. In particular, one can take $\varepsilon = 1 - C$ for the constant C appearing in Theorem 4.1.*

In light of the polynomial method, Corollary 4.2 shows that one-query quantum algorithms are roughly equivalent to bounded quadratic polynomials. The authors of [AAI+16] asked whether this result generalizes to higher degrees. Two ways to interpret this question are that for any k, any bounded degree-$2k$ polynomial p satisfies:

(a) Multiplicative converse: $\mathcal{E}(Cp,k) = 0$ for some $C = C(k) > 0$, or;

(b) Additive converse: $\mathcal{E}(p,k) \leq \varepsilon$ for some $\varepsilon = \varepsilon(k) < 1$.

The dependence on the degree k in these options is necessary due to the known separations between bounded-error quantum query complexity and approximate degree. Option (a), the higher-degree version of Theorem 4.1, was ruled out in [ABP19].

Theorem 4.3. *For any $C > 0$, there exist an $n \in \mathbb{N}$ and a bounded quartic n-variable polynomial p such that no two-query quantum algorithm $\mathcal{A}$ satisfies $\mathbb{E}[\mathcal{A}(x)] = Cp(x)$ for every $x \in \{-1,1\}^n$.*

Note that Option ((a)) with C implies Option ((b)) with $1 - C$, but Theorem 4.3 does not rule out Option ((a)).

Contributions of this chapter

Our first contribution concerns an error in the original proof of Theorem 4.3, which was based on a probabilistic example. Here, we show that Theorem 4.3 holds as stated, both by considering a slightly modified probabilistic example and by giving a completely explicit example. More importantly, we prove a stronger result that subsumes Theorem 4.3: we rule out the possibility of Option ((b)).

Theorem 4.4. *There is no constant $\varepsilon \in (0,1)$ such that for every bounded polynomial p of degree at most 4, we have $\mathcal{E}(p,2) \leq \varepsilon$.*

In the context of quantum query complexity of Boolean functions, this rules out arguably the most natural way to *upper* bound $Q(f)$ in terms of $\widetilde{\deg}(f)$: First, ε-approximate f by a degree-$2t$ polynomial p, then ε'-approximate p with a t-query quantum algorithm $\mathcal{A}$, with $\varepsilon + \varepsilon' < 1$, and finally boost the success probability of $\mathcal{A}$ so that it approximates f, for instance by taking the majority of independent runs of $\mathcal{A}$. Corollary 4.2 gives the only exceptional case where this is possible in general.

Our second contribution concerns 1-query quantum algorithms. For the case of bilinear forms, Theorem 4.1 was proved using a surprising application of the famous Grothendieck theorem (see Section 2.7.1). The general form of Theorem 4.1 follows from decoupling techniques. In this chapter, we show that the additive approximation implied by Theorem 4.1 is optimal.

Theorem 4.5. *The worst-case minimum error for one-query quantum algorithms satisfies*

$$\sup_p \mathcal{E}(p,1) = 1 - \frac{1}{K_G^{\mathbb{R}}},$$

where the supremum is taken over the set of bounded bilinear forms.

This complements another well-known characterization of $K_G^{\mathbb{R}}$ in terms of the largest-possible Bell-inequality violations in two-player XOR games [Tsi80].

The main technical result of this chapter

Both Theorems 4.3 and 4.4 are in fact corollaries of our main result (Theorem 4.13 below), which gives a formula for $\mathcal{E}(p,t)$ when p is a block-multilinear form. Block-multilinear forms already played an important role in other works related to quantum query complexity [OZ15, AAI+16, BSdW22], theoretical computer science [KN07, Lov10, KM13] and in the polarization theory of functional analysis [BH31, Har72].

The formula characterizes $\mathcal{E}(p,t)$ in terms of a ratio of norms appearing naturally in Grothendieck's theorem for bilinear forms (see Section 2.7.1). The dual formulation of Grothendieck's theorem asserts that for any bilinear form $A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$\|A\|_{\infty,*} \leq K_G^{\mathbb{R}} \|A\|_{\mathrm{cb},*}.$$

Similar norms can be defined for block-multilinear forms of higher degree. Endowing the space of polynomials with the standard inner product of the coefficient vectors in

the monomial basis, our formula for $\mathcal{E}(p,t)$ is as follows.

Theorem 4.6 (Informal version of Theorem 4.13)**.** *For a block-multilinear form p of degree $2t$, we have*

$$\mathcal{E}(p,t) = \sup_q \frac{\langle p,q\rangle - \|q\|_{\mathrm{cb},*}}{\|q\|_{\infty,*}}.$$

where the supremum runs over all block-multilinear forms q of degree $2t$.

The proof of Theorem 4.6 uses a characterization of quantum query algorithms in terms of completely bounded polynomials [ABP19].

Theorems 4.4 and 4.5 follow from Theorem 4.6 by taking suprema over particular sequences of bounded degree-$2t$ block-multilinear forms. From Theorem 4.6 it follows that

$$\sup_p \mathcal{E}(p,t) = \sup_q \left[\left(\sup_p \frac{\langle p,q\rangle}{\|q\|_{\infty,*}}\right) - \frac{\|q\|_{\mathrm{cb},*}}{\|q\|_{\infty,*}}\right] = 1 - \inf_q \frac{\|q\|_{\mathrm{cb},*}}{\|q\|_{\infty,*}}. \tag{4.1}$$

Now, Theorem 4.5 follows from Eq. (4.1) and the dual version of Grothendieck's inequality (Section 4.1). Similarly, Theorem 4.4 is proven by using Eq. (4.1) and constructing a family of degree-4 polynomials $(p_n)_n$ that witnesses the failure of Grothendieck inequality. By this we mean that $(p_n)_n$ exhibit the separation

$$\frac{\|p_n\|_{\mathrm{cb}}}{\|p_n\|_\infty} \to \infty. \tag{4.2}$$

By duality this implies that there is a sequence $(r_n)_n$ with $\|r_n\|_{\mathrm{cb},*}/\|r_n\|_{\infty,*} \to 0$, which alongside Eq. (4.1) implies that $\sup_p \mathcal{E}(p,2) = 1$, as desired.

4.2 Preliminaries

Polynomials, norms and quantum query complexity

As usual we let $\mathbb{R}[x_1,\ldots,x_n]$ be the ring of n-variate polynomials with real coefficients, whose elements we write as

$$p(x) = \sum_{\alpha\in\mathbb{Z}^n_{\geq 0}} c_\alpha x^\alpha, \tag{4.3}$$

where $x^\alpha = x_1^{\alpha_1}\cdots x_n^{\alpha_n}$ and $c_\alpha \in \mathbb{R}$. We define the support of p by

$$\mathrm{supp}(p) = \{\alpha \in \mathbb{Z}^n_{\geq 0} \mid c_\alpha \neq 0\}.$$

For $\alpha \in \mathbb{Z}^n_{\geq 0}$, write $|\alpha| = \alpha_1 + \cdots + \alpha_n$, which is the degree of the monomial x^α. A form of degree d is a homogeneous polynomial of degree d, i.e., a polynomial whose support consists of α for which $|\alpha| = d$. Denote by $\mathbb{R}[x_1, \ldots, x_n]_{=d}$ the space of forms of degree d. For p as in Eq. (4.3), define its homogeneous degree-d part by

$$p_{=d}(x) = \sum_{|\alpha|=d} c_\alpha x^\alpha.$$

We endow $\mathbb{R}[x_1, \ldots, x_n]$ with the inner product given by

$$\langle p, q \rangle = \sum_{\alpha \in \mathbb{Z}^n_{\geq 0}} c_\alpha c'_\alpha,$$

where c_α and c'_α are the coefficients of p and q, respectively.

We recall the definition of $\|\cdot\|_1$ and $\|\cdot\|_\infty$, which are seminorms of polynomials in $\mathbb{R}[x_1, \ldots, x_n]$, and norms on the space of multilinear polynomials.

$$\|p\|_\infty := \sup_{x \in \{-1,1\}^n} |p(x)|,$$
$$\|p\|_1 := \mathbb{E}_{x \in \{-1,1\}^n} |p(x)|,$$

where the expectation is taken with respect to the uniform probability measure.

We will work with a reformulation of the completely bounded polynomial method, Theorem 3.6. To state it, we define the completely bounded norm of a form p.

Definition 4.7. Let $p \in \mathbb{R}[x_1, \ldots, x_n]_{=t}$. Then, its completely bounded norm is defined by

$$\|p\|_{\mathrm{cb}} = \inf \Big\{ \|T\|_{\mathrm{cb}} \mid p(x) = T(x, \ldots, x) \ \forall x \in \mathbb{R}^n \Big\},$$

where the infimum runs over all t-linear forms $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$.

Note that we are slightly abusing notation because we have introduced two notions of completely bounded norm for t-linear forms $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$. The first one in Definition 2.17, where we regard T as a multilinear form. Furthermore, such T can also be regarded as a homogeneous polynomial in n^t variables, so we have defined a second notion of completely bounded norm for it in Definition 4.7. For the rest of the chapter, we will use the definition of Definition 4.7. However, to prove Theorem 4.5 we should show that for bilinear forms both norms are equal (see Proposition 4.25 below).

Now, we can restate Theorem 3.6.[1]

Theorem 4.8 (Completely bounded polynomial method)**.** *Let* $p : \{-1,1\}^n \to [-1,1]$ *and let* $t \in \mathbb{N}$*. Then,*

$$\begin{aligned}\mathcal{E}(p,t) = \inf \quad & \|p-q\|_\infty \\ \text{s.t.} \quad & h \in \mathbb{R}[x_1,\dots,x_{n+1}]_{=2t} \text{ with } \|h\|_{\text{cb}} \le 1 \\ & q : \{-1,1\}^n \to \mathbb{R}, \text{ with } q(x) = h(x,1) \ \forall\ x \in \{-1,1\}^n.\end{aligned}$$

Block-multilinear forms

Theorem 4.13 is stated for a special kind of polynomials, which are the block-multilinear forms.

Definition 4.9. Let $\mathcal{P} = \{I_1,\dots,I_t\}$ be a partition of $[n]$ into t (pairwise disjoint) non-empty subsets. Define the set of *block-multilinear polynomials with respect to* $\mathcal{P}$ to be the linear subspace

$$V_{\mathcal{P}} = \text{Span}\left\{x_{i_1}\cdots x_{i_t} \mid i_1 \in I_1,\dots,i_t \in I_t\right\}.$$

We also work with the larger space of polynomials spanned by monomials where in the above we replace linearity by odd degree.

Definition 4.10. For a family $\mathcal{Q} \subseteq 2^{[m]}$ of pairwise disjoint subsets, let $W_{\mathcal{Q}} \subseteq \mathbb{R}[x_1,\dots,x_m]$ be the subspace of polynomials spanned by monomials x^α with $\alpha \in \mathbb{Z}^m_{\ge 0}$ satisfying

$$\sum_{i\in I} \alpha_i \equiv 1 \bmod 2 \ \forall I \in \mathcal{Q}. \tag{4.4}$$

We use $\Pi_{\mathcal{Q}} : \mathbb{R}[x_1,\dots,x_m] \to W_{\mathcal{Q}}$ to refer to the projector onto $W_{\mathcal{Q}}$.

Remark 4.11. Given a partition $\mathcal{P}$ of $[n]$, we have $V_{\mathcal{P}} \subset W_{\mathcal{P}}$. In particular, $V_{\mathcal{P}}$ consists of precisely the multilinear polynomials in $W_{\mathcal{P}}$.

Although the projector $\Pi_{\mathcal{Q}}$ onto $W_{\mathcal{Q}}$ is properly defined on the space of polynomials of n variables, we will slightly abuse notation and let it act on a t-tensor $T \in \mathbb{R}^{n\times\cdots\times n}$ as follows. Define $\mathcal{I}_{\mathcal{Q}} \subseteq [n]^t$ to be the set of t-tuples that contain an odd number of

[1] A direct reformulation of Theorem 3.6 would be with the polynomial h below belonging to $\mathbb{R}[x_1,\dots,x_{2n}]_{=2t}$, instead of $\mathbb{R}[x_1,\dots,x_{n+1}]_{=2t}$, However, in [GL19] it was observed that only *one extra variable* is needed.

elements from each set in $\mathcal{Q}$. Then, we let $\Pi_{\mathcal{Q}} T$ be the tensor given by

$$(\Pi_{\mathcal{Q}} T)_{\mathbf{i}} := \begin{cases} T_{\mathbf{i}} & \text{if } \mathbf{i} \in \mathcal{I}_{\mathcal{Q}}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

It is not hard to see that if p is a polynomial satisfying $T(x, \dots, x) = p(x)$ for every $x \in \{-1, 1\}^n$, then $\Pi_{\mathcal{Q}} T(x, \dots, x) = \Pi_{\mathcal{Q}} p(x)$ for every $x \in \{-1, 1\}^n$.

We note that all the norms and seminorms we have mentioned are norms on the space $V_{\mathcal{P}}$ for any partition $\mathcal{P}$ of $[n]$. Hence, we can take the dual of these norms with respect to this subspace, so from now on $\|p\|_{\infty,*}$ and $\|p\|_{\mathrm{cb},*}$ will be the dual of $\|p\|_{\infty}$ and $\|p\|_{\mathrm{cb}}$ of $V_{\mathcal{P}}$, respectively. By contrast, when we write $\|R\|_{\mathrm{cb},*}$ for some t-tensor $\mathbb{R}^{n \times \cdots \times n}$ we refer to the dual norm of the completely bounded norm of R with respect to the whole space of t-tensors.

We stress that $\|\cdot\|_{\infty,*}$ need not be equal to $\|\cdot\|_1$. This is because we are taking the dual norms with respect to $V_{\mathcal{P}}$ and not with respect to the space of all multilinear maps, in which case the dual norm would be $\|p\|_1$. The following example shows that $\|p\|_{\infty,*} \neq \|p\|_1$ in general.

Example 4.12. Consider $n = 3$, $t = 1$ and $p = (x_1 + x_2 + x_3)/3$. Then, $\|p\|_1 > 1/3$, but $\|p\|_{\infty,*} \leq 1/3$. Indeed, as $|p(x)| \geq 1/3$ for every $x \in \{-1, 1\}^3$ and $|p(x)| > 1/3$ for some $x \in \{-1, 1\}^3$, we have that $\|p\|_1 > 1/3$. On the other hand, in this case $\mathcal{P} = \{[3]\}$ so $V_{\mathcal{P}}$ is the set of linear polynomials. Note that if q is linear, then $\|\hat{q}\|_1 = \|q\|_{\infty}$, where $\hat{q}$ is the Fourier transform of q. Hence

$$\|p\|_{\infty,*} = \sup_{q \in V_{\mathcal{P}}, \|q\|_{\infty} \leq 1} \langle p, q \rangle = \sup_{q \in V_{\mathcal{P}}, \|\hat{q}\|_1 \leq 1} \langle \hat{p}, \hat{q} \rangle \leq \sup_{\|\hat{q}\|_1 \leq 1} \|\hat{p}\|_{\infty} \|\hat{q}\|_1 = \frac{1}{3},$$

where in second equality we used Parseval's identity.

4.3 $\mathcal{E}(p, t)$ for block-multilinear forms

In this section we formally state and prove our main result:

Theorem 4.13. *Let $\mathcal{P}$ be a partition of $[n]$ in $2t$ subsets and $p \in V_{\mathcal{P}}$. Then,*

$$\mathcal{E}(p, t) = \sup \left\{ \langle p, r \rangle - \|r\|_{\mathrm{cb},*} \mid r \in V_{\mathcal{P}}, \ \|r\|_{\infty,*} \leq 1 \right\}.$$

For the proof, we use more convenient expressions for the completely bounded norms and the fact that the projector Π_Q is contractive under several norms.

Contractivity of the projector Π_Q.

A key element of the proof of Theorem 4.13 is that can restrict the infimum in Theorem 4.8 to the space of polynomials $W_{\mathcal{Q}}$ given in Definition 4.10. To do that, we prove that the orthogonal projector onto this space, $\Pi_{\mathcal{Q}}$ is contractive in several norms. This will follow from the fact that $\Pi_{\mathcal{Q}}$ has a particularly nice structure in the form of an averaging operator. Let $\mathcal{Q}$ be a family of disjoint subsets of $[n]$. For each $I \in \mathcal{Q}$ let z_I be a random variable that takes the values -1 and 1 with probability $1/2$ and let $z = (z_I)_{I\in\mathcal{Q}}$. For a bit string $x \in \{-1,1\}^n$, we define the random variable $x \cdot z \in \{-1,1\}^n$ as

$$(x \cdot z)(i) := \begin{cases} x_i z_I & \text{if } i \in I \text{ for some } I \in \mathcal{Q}, \\ x_i & \text{otherwise.} \end{cases}$$

For a matrix-valued map $A : [n] \to M(d)$ we define the random variable $A \cdot z$ in an analogous way.

Proposition 4.14. *For any $p \in \mathbb{R}[x_1, \ldots, x_n]$ and $x \in \mathbb{R}^n$, we have that*

$$\Pi_{\mathcal{Q}} p(x) = \mathbb{E}_z \Big[p(x \cdot z) \prod_{I\in\mathcal{Q}} z_I \Big].$$

Similarly, for any t-tensor $T \in \mathbb{R}^{n\times\cdots\times n}$, positive integer d and matrix-valued map $A\colon [n] \to M(d)$, we have that

$$\Pi_{\mathcal{Q}} T(A) = \mathbb{E}_z \left[T(A \cdot z) \prod_{I\in\mathcal{Q}} z_I \right].$$

Proof. By linearity, it suffices to prove the equality for monomials. Let $\alpha \in \mathbb{Z}^n_{\geq 0}$. Then we have

$$(x \cdot z)^\alpha \prod_{I\in\mathcal{Q}} z_I = x^\alpha \prod_{I\in\mathcal{Q}} z_I^{1+\sum_{i\in I}\alpha_i}.$$

It follows that

$$\mathbb{E}_z \Big[(x \cdot z)^\alpha \prod_{I\in\mathcal{Q}} z_I \Big] = \begin{cases} x^\alpha & \text{if } 1 + \sum_{i\in I} \alpha_i = 0 \bmod 2 \ \forall I \in \mathcal{Q}, \\ 0 & \text{otherwise.} \end{cases}$$

It remains to observe that this is precisely the projection of x^α on $W_{\mathcal{Q}}$. The statement for tensors follows analogously. □

Finally, we prove that $\Pi_{\mathcal{Q}}$ is contractive with respect to the relevant norms.

Lemma 4.15. *Let $\mathcal{Q}$ be a family of disjoint subsets of $[n]$ and $p \in \mathbb{R}[x_1, \ldots, x_n]$ and let* norm $\in \{\mathrm{cb}, \infty, 1\}$ *where for the* cb*-norm we moreover require p to be homogeneous. Then*

$$\|\Pi_{\mathcal{Q}} p\|_{\mathrm{norm}} \leq \|p\|_{\mathrm{norm}}.$$

Proof. First, we consider the $\|\cdot\|_\infty$ norm. For every $x \in \{-1,1\}^n$, we have that $x \cdot z \in \{-1,1\}^n$, so

$$|\Pi_{\mathcal{Q}} p(x)| \leq \mathbb{E}_z |p(x \cdot z) \prod_{I \in \mathcal{Q}} z_I| = \mathbb{E}_z |p(x \cdot z)| \leq \mathbb{E}_z \|p\|_\infty = \|p\|_\infty,$$

where in the first inequality we used Proposition 4.14 and the triangle inequality.

Second, we consider $\|\cdot\|_{\mathrm{cb}}$. Arguing as in the $\|\cdot\|_\infty$ case and using Definition 4.7, it follows that for any t-tensor $T \in \mathbb{R}^{n \times \cdots \times n}$ we have that $\|\Pi_{\mathcal{Q}} T\|_{\mathrm{cb}} \leq \|T\|_{\mathrm{cb}}$. Given that $\Pi_{\mathcal{Q}} p(x) = \Pi_{\mathcal{Q}} T(x)$ if $p(x) = T(x)$, it follows that

$$\|\Pi_{\mathcal{Q}} p\|_{\mathrm{cb}} \leq \|\Pi_{\mathcal{Q}} T\|_{\mathrm{cb}} \leq \|T\|_{\mathrm{cb}}$$

for every t-tensor $T \in \mathbb{R}^{n \times \cdots \times n}$ such that $T(x) = p(x)$. Taking the infimum over all those T we arrive at $\|\Pi_{\mathcal{Q}} p\|_{\mathrm{cb}} \leq \|p\|_{\mathrm{cb}}$.

Finally, for $\|\cdot\|_1$ we have

$$\|\Pi_{\mathcal{Q}} p\|_1 = \mathbb{E}_x |\mathbb{E}_z p(x \cdot z) \prod_{I \in \mathcal{Q}} z_I| \leq \mathbb{E}_x \mathbb{E}_z |p(x \cdot z)| = \mathbb{E}_z \mathbb{E}_x |p(x)| = \|p\|_1,$$

where in the first equality we have used Proposition 4.14 and in the third we have used the fact that the uniform measure is invariant under multiplication by $z \in \{-1,1\}^n$. □

Putting everything together

We are now ready to prove Theorem 4.13. To this end, we start from the expression given in Theorem 4.8 for $\mathcal{E}(p,t)$ and let $h \in \mathbb{R}[x_1, \ldots, x_{n+1}]_{=2t}$ with $\|h\|_{\mathrm{cb}} \leq 1$ and let $q : \{-1,1\}^n \to \mathbb{R}$ be defined by $q(x) = h(x,1)$ for every $x \in \{-1,1\}^n$.

We first show that we can project q (and h) onto $W_{\mathcal{P}}$ and obtain a feasible solution whose objective value is at least as good as q. Since $\mathcal{P}$ is a partition of $[n]$, it defines a family of disjoint subsets of $[n+1]$, so by Lemma 4.15, we have $\|\Pi_{\mathcal{P}} h\|_{\mathrm{cb}} \leq \|h\|_{\mathrm{cb}} \leq 1$. Since the degree of h is at most $2t$, the polynomial $\Pi_{\mathcal{P}} h$ has degree at most $2t$.

This shows that each monomial in its support contains exactly one variable from each of the $2t$ sets in $\mathcal{P}$. We can therefore observe that $\Pi_{\mathcal{P}}h$ does not depend on x_{n+1}. Since $h(x,1) = q(x)$, we have $\Pi_{\mathcal{P}}h(x,1) = \Pi_{\mathcal{P}}q(x)$ and therefore $\Pi_{\mathcal{P}}q \in V_{\mathcal{P}}$. From Definition 4.7 follows that $\|\Pi_{\mathcal{P}}q\|_{\mathrm{cb}} \leq 1$. Indeed, applying $\Pi_{\mathcal{P}}$ to a $2t$-tensor $T \in \mathbb{R}^{(n+1)\times\cdots(n+1)}$ that certifies $\|h\|_{\mathrm{cb}} \leq 1$ results in a tensor $\Pi_{\mathcal{P}}T$ that satisfies $\Pi_{\mathcal{P}}T(\mathbf{i}) = 0$ whenever $\mathbf{i}$ contains an index equal to $n+1$. So, $\Pi_{\mathcal{P}}T(x,1) = \Pi_{\mathcal{P}}q(x)$ for every $x \in \{-1,1\}^n$ and thus $\Pi_{\mathcal{P}}T$, viewed as a $2t$-tensor in $\mathbb{R}^{n\times\cdots\times n}$, certifies $\|\Pi_{\mathcal{P}}q\|_{\mathrm{cb}} \leq 1$. For the objective value of $\Pi_{\mathcal{P}}q$ we finally observe that

$$\|p - \Pi_{\mathcal{P}}q\|_\infty = \|\Pi_{\mathcal{P}}(p-q)\|_\infty \leq \|p-q\|_\infty,$$

where we used that $p \in V_{\mathcal{P}}$ in the equality and Lemma 4.15 in the inequality. This shows that

$$\mathcal{E}(p,t) \geq \inf\{\|p-q\|_\infty \mid q \in V_{\mathcal{P}} \text{ with } \|q\|_{\mathrm{cb}} \leq 1\}.$$

To show that the above inequality is in fact an equality it suffices to observe that given a polynomial $q \in V_{\mathcal{P}}$, we can define $h \in \mathbb{R}[x_1,\ldots,x_{n+1}]$ as $h(x,x_{n+1}) = q(x)$ and then we have $\|h\|_{\mathrm{cb}} \leq \|q\|_{\mathrm{cb}}$.

Finally, in the above reformulation of $\mathcal{E}(p,t)$, we can express $\|p-q\|_\infty$ in terms of its dual norm and obtain

$$\begin{aligned}\mathcal{E}(p,t) = \inf_q \sup_r \quad & \langle p-q, r\rangle \\ \text{s.t.} \quad & q \in V_{\mathcal{P}} \text{ with } \|q\|_{\mathrm{cb}} \leq 1, \\ & r \in V_{\mathcal{P}} \text{ with } \|r\|_{\infty,*} \leq 1.\end{aligned}$$

Finally, we need the von Neumann's minimax theorem (see [Nik54] for a proof).

Theorem 4.16 (Minimax)**.** *Let X and Y convex compact sets. Let $f : X \times Y \to \mathbb{R}$ such that f is concave in the first variable and convex in the second. Then,*

$$\sup_{x\in X}\inf_{y\in Y} f(x,y) = \inf_{y\in Y}\sup_{x\in X} f(x,y).$$

The desired result then follows by exchanging the infimum and supremum, which we are allowed to do by Theorem 4.16.

4.4 Separations between infinity and completely bounded norms

In this section we show that the completely bounded norm of a degree 4 bounded polynomial can be unbounded. In other words, we prove the following Theorem.

Theorem 4.17. *There is a sequence $p_n \in \mathbb{R}[x_1, \dots, x_n]_{=4}$ such that*

$$\frac{\|p_n\|_{\mathrm{cb}}}{\|p_n\|_\infty} \to \infty.$$

To prove Theorem 4.17 we first provide a framework to lower bound the completely bounded norm inspired on a technique due to Varopoulos [Var74].[2] Second, we construct two sequences of bounded polynomials, one random and one explicit, that fit in that framework and have unbounded completely bounded norm.

Lower bounding the completely bounded norm

We will first talk about general cubic forms, that is polynomials given by:

$$p(x) = \sum_{S \in \binom{[n]}{3}} c_S \prod_{i \in S} x_i, \tag{4.6}$$

where the c_S are some real coefficients. We will lower bound its completely bounded norm. Then, we will extent this lower bound to an associated quartic form, given by $x_0 p(x)$. For $i \in [n]$, define the ith *slice* of p to be the symmetric matrix $M_i \in \mathbb{R}^{n \times n}$ with (j,k)-coefficient equal to $c_{\{i,j,k\}}$ if i, j, k are pairwise distinct and 0 otherwise. Then, define

$$\Delta(p) = \max_{i \in [n]} \|M_i\|_{\mathrm{op}}.$$

Lemma 4.18 (tri-linear Varopoulos decomposition)**.** *Let p be an n-variate multilinear cubic form as in* (4.6)*. Then, for some $d \in \mathbb{N}$, there exist contractions $A(1), \dots, A(n) \in$*

[2] We use the same construction as the one proposed by Varopoulos, but we apply it to multilinear polynomials, which gives it the extra property displayed in Eq. (4.7)

M_d and orthogonal vectors $u, v \in S^{d-1}$ such that $[A(j), A(i)] = 0$, and

$$A(i)^2 = 0 \tag{4.7}$$

$$\langle u, A(i)v \rangle = 0 \tag{4.8}$$

$$\langle u, A(i)A(j)v \rangle = 0 \tag{4.9}$$

$$\langle u, A(i)A(j)A(k)v \rangle = \frac{c_{\{i,j,k\}}}{\Delta(p)} \tag{4.10}$$

for all pairwise distinct $i, j, k \in [n]$.

Proof. For each $i \in [n]$, define M_i as above. Define $W_i = \Delta(p)^{-1} M_i$ and note that W_i has operator norm at most 1. For each $i \in [n]$, define the $(2n+2) \times (2n+2)$ block matrix

$$A(i) = \left[\begin{array}{c|c|c|c} & & & \\ \hline e_i & & & \\ \hline & W_i^{\mathsf{T}} & & \\ \hline & & e_i^{\mathsf{T}} & \end{array}\right],$$

where the first and last rows and columns have size 1, the second and third have size n and where the empty blocks are filled with zeros. Define $u = e_{2n+1}$ and $v = e_1$. The rest of the proof is identical to the proof of [BP19, Lemma 2.11], except for the property that $A(i)^2 = 0$. This follows from the fact that

$$A(i)^2 = \left[\begin{array}{c|c|c|c} & & & \\ \hline & & & \\ \hline W_i^{\mathsf{T}} e_i & & & \\ \hline & e_i^{\mathsf{T}} W_i^{\mathsf{T}} & & \end{array}\right]$$

and that the ith row and ith column of M_i (and hence W_i) are zero. □

Corollary 4.19. *Let p be an n-variate multilinear cubic form as in (4.6). Suppose that an $(n+2)$-variate quartic form $h \in \mathbb{R}[x_0, x_1, \ldots, x_n, z]$ satisfies $h(x, 1) = x_0 p(x_1, \ldots, x_n)$ for every $x \in \{-1, 1\}^{n+1}$. Then,*

$$\|h\|_{\mathrm{cb}} \geq \frac{\|p\|_2^2}{\Delta(p)}.$$

Proof. From the orthonormality of the characters, it follows that h and $x_0 p$ have equal coefficients for each quartic multilinear monomial in the variables $x_0, \ldots, x_n$, which are c_S for $x_0 \chi_S$ with $S \in \binom{[n]}{3}$ and 0 otherwise. Let $A(1), \ldots, A(n) \in B_{M_d}$ and $u, v \in S^d$ be as in Lemma 4.18, and extend A by $A(0) = I, A(n+1) = 0$.

Commutativity and properties (4.7)–(4.9) imply that if a quartic monomial expression $A((i,j,k,l))$ with $i,j,k,l \in \{0,\dots,n+1\}$ has repeated indices or an index equal to $n+1$, then $\langle u, A((i,j,k,l))v\rangle = 0$. With this, it follows that, for every T_h such that $T_h(x,\dots,x) = h(x)$, we have

$$\|T_h\|_{\mathrm{cb}} \geq \sum_{\mathbf{i}\in(\{0\}\cup[n+1])^4} T_{\mathbf{i}}\Big\langle u, A(\mathbf{i})v\Big\rangle = \sum_{S\in\binom{[n]}{3}} c_S\Big\langle u,\, A(0)\prod_{i\in S} A(i)v\Big\rangle. \tag{4.11}$$

Finally, if we use that $A(0) = \mathrm{Id}$, property (4.10) and Parseval's identity, we obtain the desired result:

$$\|h\|_{\mathrm{cb}} = \inf\|T_h\|_{\mathrm{cb}} \geq \sum_{S\in\binom{[n]}{3}} c_S\langle u, \prod_{i\in S} A(i)v\rangle = \Delta(p)^{-1}\sum_{S\in\binom{[n]}{3}} c_S^2 = \frac{\|p\|_2^2}{\Delta(p)}.$$

□

A separation based on a random example

We begin by defining a random cubic form as in (4.6) where the coefficients c_S are chosen to be independent uniformly distributed random signs. Parseval's identity then gives $\|p\|_2^2 = \binom{n}{3}$. We now use a standard random-matrix inequality to upper bound $\Delta(p)$ (see [Tao12, Corollary 2.3.6] for a proof).

Lemma 4.20. *There exist absolute constants $C, c \in (0,\infty)$ such that the following holds. Let n be a positive integer and let M be a random $n\times n$ symmetric random matrix such that for $j \geq i$, the entries M_{ij} are independent random variables with mean zero and absolute value at most 1. Then, for any $\tau \geq C$, we have*

$$\Pr\big[\|M\|_{\mathrm{op}} > \tau\sqrt{n}\big] \leq Ce^{-c\tau n}.$$

Applying Lemma 4.20 to the slices M_i and the union bound then imply that $\Delta(p) \leq C\sqrt{n}$ with probability $1-\exp(-Cn)$. By Hoeffding's inequality [BLM13, Theorem 2.8] and the union bound, we have that $\|p\|_\infty \leq Cn^2$ with probability $1-\exp(-Cn)$. Rescaling p then gives that there exists a bounded multilinear cubic form such that $\|p\|_2^2/\Delta(p) \geq C\sqrt{n}$. Now Theorem 4.17 follows from Corollary 4.19.

A construction based on an explicit example

We also give an explicit construction using techniques from [BP19], which were used there to disprove a conjecture on a tri-linear version of Grothendieck's theorem. We do not exactly use the construction from that paper because it involves complex functions. Instead, we will use the Möbius function (defined below), which is real valued and has the desired properties.

The construction uses some notions from additive combinatorics. For a function $f : \mathbb{Z}_n \to [-1, 1]$ (on the cyclic group of order n), define the 3-linear form

$$p(x_1, x_2, x_3) = \sum_{a,b \in \mathbb{Z}_n} x_{1,a} x_{2,a+b} x_{3,a+2b} f(a + 3b).$$

where $x_1, x_2, x_3 \in \{-1, 1\}^n$ and the sums of a and b are done in $\mathbb{Z}_n$.

We begin by upper bounding $\Delta(p)$. The polynomial p has $3n$ slices, $M_{i,a} \in \mathbb{R}^{[3] \times \mathbb{Z}_n}$ for each $i \in [3]$ and $a \in \mathbb{Z}_n$, which we view as 3×3 block-matrices with blocks indexed by $\mathbb{Z}_n$. The slice $M_{1,a}$ is supported only on the $(2, 3)$ and $(3, 2)$ blocks, which are each others' transposes. On its $(2, 3)$ block it has value $f(a+3b)$ on coordinate $(a+b, a+2b)$ for each b. In particular, this matrix has at most one nonzero entry in each row and column. It follows that a relabeling of the rows turns $M_{1,a}$ into a diagonal matrix with diagonal entries in $[-1, 1]$, and therefore $\|M_{1,a}\|_{\mathrm{op}} \leq 1$. Similarly, we get that $\|M_{i,a}\|_{\mathrm{op}} \leq 1$ for $i = 2, 3$. Hence,

$$\Delta(p) \leq 1. \tag{4.12}$$

for any choice of f.

Now we will choose a specific f for which we will be able to upper bound $\|p\|_\infty$ and lower bound $\|p\|_2^2$. Identify $\mathbb{Z}_n$ with $\{0, 1 \ldots, n-1\}$ in the standard way. We choose f to be the Möbius function restricted to this interval. That is, set $f(0) = 0$ and for $a > 0$, set

$$f(a) = \begin{cases} 1 & \text{if } a \text{ is square-free with an even number of prime factors} \\ -1 & \text{if } a \text{ is square-free with an odd number of prime factors} \\ 0 & \text{otherwise.} \end{cases}$$

The infinity norm of p can be upper bounded in terms of the Gowers 3-uniformity

norm of f. This norm plays a central role in additive combinatorics and is defined by

$$\|f\|_{U^3} = \Big(\mathbb{E}_{a,b_1,b_2,b_3 \in \mathbb{Z}_n} \prod_{c \in \{0,1\}^3} f(a + c_1 b_1 + c_2 b_2 + c_3 b_3) \Big)^{\frac{1}{8}}.$$

The proof of the announced bound can be found in [Gre07, Proposition 1.11].

Lemma 4.21 (generalized von Neumann inequality)**.** *Suppose that n is coprime to 6. Then, for any $f : \mathbb{Z}_n \to [-1, 1]$, we have that*

$$\|p\|_\infty \leq n^2 \|f\|_{U^3}.$$

A recent result by Tao and Teräväinen [TT23] given an upper bound to the Gowers 3-uniformity norm of the Möbius function.

Theorem 4.22. *Let $f : \mathbb{Z}_n \to \mathbb{R}$ be the Möbius function. Then,*

$$\|f\|_{U^3} \leq \frac{1}{(\log \log n)^C}.$$

for some constant $C > 0$.

Combining Lemma 4.21 and Theorem 4.22 it follows that

$$\|p\|_\infty \leq \frac{n^2}{(\log \log n)^C} \tag{4.13}$$

for some constant $C > 0$.

To lower bound $\|p\|_2^2$ we begin using Parseval's identity, which implies that

$$\|p\|_2^2 = n \sum_{a \in \mathbb{Z}_n} f(a)^2. \tag{4.14}$$

Given that $|f(a)|^2$ is 1 if a is square-free and 0 otherwise, we can use a classical result of number theory to lower bound $\|p\|_2^2$ (see [HW+79, page 269] for a proof).

Proposition 4.23. *There are $\frac{6}{\pi^2} n - O(\sqrt{n})$ natural numbers between 1 and n that are square-free.*

From Eq. (4.14) and Proposition 4.23 follows that

$$\|p\|_2^2 = \frac{6}{\pi^2} n^2 - O(\sqrt{n^3}). \tag{4.15}$$

Finally, we substitute p by $p/(n^2/(\log\log n)^C)$, and it follows from Eqs. (4.12), (4.13) and (4.15) that p is bounded and

$$\frac{\|p\|_2^2}{\Delta(p)} \geq \frac{6}{\pi^2}(\log\log n)^C - o(1).$$

Again, Theorem 4.17 now follows from Corollary 4.19.

Remark 4.24. The *jointly completely bounded norm* of p is given by

$$\|p\|_{\text{jcb}} = \sup_{d\in\mathbb{N}} \| \sum_{a,b\in\mathbb{Z}_n} A(1,a)A(2,a+b)A(3,a+2b)f(a+3b)\|,$$

where the supremum is taken over maps $A : [3]\times[n] \to \mathbb{C}^{d\times d}$ such that $\|A(i,a)\|_{\text{op}} \leq 1$ and $[A(i,a), A(j,b)] = [A(i,a), A(j,b)^\dagger] = 0$ for all $i \neq j$ and $a, b \in \mathbb{Z}_n$. This norm can also be stated in terms of tensor products and the supremum is attained by observable-valued maps. As such, this norm appears naturally in the context of non-local games. It was shown in [BBB+19] that Proposition 4.21 also holds for the jointly completely bounded norm, that is $\|p\|_{\text{jcb}} \leq n^2\|f\|_{U^3}$. The proof of Corollary 4.19 easily implies that $\|p\|_{\text{cb}} \geq \|p\|_2^2/\Delta(p)$. This was used in [BP19] to prove that the jcb and cb norms are inequivalent.

4.5 Grothendieck inequalities characterize converses to the polynomial method

In this section, we show, as a corollary of Theorem 4.13, that Grothendieck inequalities characterize converses to the polynomial method. By this we mean that: i) for 1-query algorithms an additive converse is possible and moreover this converse characterizes $K_G^{\mathbb{R}}$; and ii) for 2-query algorithms no additive converse is possible, because Grothendieck's inequality fails for 3-linear forms.

4.5.1 Characterizing $K_G^{\mathbb{R}}$ with 1-query quantum algorithms

Here we prove Theorem 4.5. Before doing that, we should prove Definition 2.17 and Definition 4.7 coincide for bilinear forms, so we can apply Grothendieck's Theorem, which uses Definition 2.17, into Theorem 4.13, which uses Definition 4.7.

Proposition 4.25. *For bilinear forms Definitons 2.17 and 4.7 coincide.*

Proof. Let $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear form. In this proof we will use $\|T\|_{\mathrm{cb}}$ to refer to the quantity defined in Definition 2.17, and we will write the quantity of Definition 4.7 as

$$\|T\|_{\tilde{\mathrm{cb}}} = \inf\Big\{\|R\|_{\mathrm{cb}} \mid T(x) = R(x,x)\ \forall\ x \in \mathbb{R}^n \times \mathbb{R}^n\Big\},$$

where the infimum runs over all bilinear forms $R : (\mathbb{R}^n \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R}$.

We first prove that $\|T\|_{\tilde{\mathrm{cb}}} = \|T_{\mathrm{sym}}\|_{\mathrm{cb}}$, where $T_{\mathrm{sym}} : (\mathbb{R}^n \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R}$ is the only symmetric bilinear form such that $T(x) = T_{\mathrm{sym}}(x,x)$ for every $x \in \mathbb{R}^n \times \mathbb{R}^n$. On the one hand, by definition, it follows that $\|T\|_{\tilde{\mathrm{cb}}} \leq \|T_{\mathrm{sym}}\|_{\mathrm{cb}}$. On the other hand, consider a bilinear form $R : (\mathbb{R}^n \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R}$ such that $T(x) = R(x,x)$ for every $x \in \mathbb{R}^n \times \mathbb{R}^n$. We define $R^{\mathsf{T}} : (\mathbb{R}^n \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R}$ as the bilinear form obtained by transposing the matrix associated to R as in Definition 2.14. We have that $T_{\mathrm{sym}} = (R + R^{\mathsf{T}})/2$ and that $T(x) = R^{\mathsf{T}}(x,x)$ for every $x \in \mathbb{R}^n \times \mathbb{R}^n$. Furthermore, it is satisfied that

$$\begin{aligned}
\|R^{\mathsf{T}}\|_{\mathrm{cb}} &= \sup\Big\{\|\sum_{i,j} R_{j,i} A(i) B(j)\| \mid A(i), B(j) \in B_{M_d}\Big\} \\
&= \sup\Big\{\|\sum_{i,j} R_{j,i} B(j)^{\mathsf{T}} A(i)^{\mathsf{T}}\| \mid A(i), B(j) \in B_{M_d}\Big\} \\
&= \|R\|_{\mathrm{cb}},
\end{aligned} \tag{4.16}$$

where we use (twice) that for any matrix M we have $\|M\| = \|M^{\mathsf{T}}\|$. Thus, we have that $\|R\|_{\mathrm{cb}} \geq \|T_{\mathrm{sym}}\|_{\mathrm{cb}}$, so $\|T\|_{\tilde{\mathrm{cb}}} \geq \|T_{\mathrm{sym}}\|_{\mathrm{cb}}$.

Second, we prove that $\|T\|_{\mathrm{cb}} = \|T_{\mathrm{sym}}\|_{\mathrm{cb}}$. We observe that $T_{\mathrm{sym}} = \frac{1}{2}\begin{pmatrix} 0 & T \\ T^{\mathsf{T}} & 0 \end{pmatrix}$. Thus, we immediately have that $\|T\|_{\mathrm{cb}} \leq \|T_{\mathrm{sym}}\|_{\mathrm{cb}}$. Also, it is satisfied that

$$\begin{aligned}
\|T_{\mathrm{sym}}\|_{\mathrm{cb}} &\leq \frac{1}{2}\left(\left\|\begin{pmatrix} 0 & T \\ 0 & 0 \end{pmatrix}\right\|_{\mathrm{cb}} + \left\|\begin{pmatrix} 0 & 0 \\ T^{\mathsf{T}} & 0 \end{pmatrix}\right\|_{\mathrm{cb}}\right) \\
&\leq \frac{1}{2}\left(\|T\|_{\mathrm{cb}} + \|T^{\mathsf{T}}\|_{\mathrm{cb}}\right) \\
&= \|T\|_{\mathrm{cb}},
\end{aligned}$$

where the last equality uses (4.16). □

We recall that it was shown in [AAI+16] that for every bilinear form there exists a 1-query quantum algorithm that makes additive error at most $1 - 1/K_G^{\mathbb{R}}$. It thus

remains to show the lower bound.

Theorem 4.5. *The worst-case minimum error for one-query quantum algorithms satisfies*

$$\sup_p \mathcal{E}(p,1) = 1 - \frac{1}{K_G^{\mathbb{R}}},$$

where the supremum is taken over the set of bounded bilinear forms.

Proof. Theorem 4.13 shows the following:

$$\begin{aligned} \sup_{p\in\mathcal{BB}} \mathcal{E}(p,1) &= \sup_{\|p\|_\infty\le 1} \sup_{\|r\|_{\infty,*}\le 1} \langle p, r\rangle - \|r\|_{\mathrm{cb},*} && (4.17)\\ &= \sup_{\|r\|_{\infty,*}\le 1} \|r\|_{\infty,*} - \|r\|_{\mathrm{cb},*} && (4.18)\\ &= \sup_{\|r\|_{\infty,*}=1} 1 - \|r\|_{\mathrm{cb},*}. \end{aligned}$$

It thus remains to show that for bilinear forms $\|r\|_{\infty,*} \le K_G^{\mathbb{R}}\|r\|_{\mathrm{cb},*}$. We do so starting from Grothendieck's theorem for matrices. It states that for $A \in \mathbb{R}^{n\times n}$ we have $\|A\|_{\mathrm{cb}} \le K_G^{\mathbb{R}}\|A\|_\infty$. Each bilinear form $q : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ uniquely corresponds to a matrix $A \in \mathbb{R}^{n\times n}$ such that $q(x,y) = x^{\mathsf{T}}Ay$. Moreover, for such q and A one has $\|q\|_\infty = \|A\|_\infty$ (immediate) and in Proposition 4.25 we showed $\|q\|_{\mathrm{cb}} = \|A\|_{\mathrm{cb}}$, so $\|q\|_{\mathrm{cb}} \le K_G^{\mathbb{R}}\|q\|_\infty$. A duality argument then concludes the proof:

$$\|r\|_{\infty,*} = \sup_{\|q\|_\infty\le 1} \langle r, q\rangle \le \sup_{\|q\|_{\mathrm{cb}}\le K_G^{\mathbb{R}}} \langle r, q\rangle = K_G^{\mathbb{R}}\|r\|_{\mathrm{cb},*}.$$

□

Remark 4.26. If in Theorem 4.5 we restrict the supremum to bilinear forms on $n+n$ variables, for a fixed n, then we obtain a characterization of $K_G^{\mathbb{R}}(n)$ instead of $K_G^{\mathbb{R}}$. Here, $K_G^{\mathbb{R}}(n) = \sup \|A\|_{\mathrm{cb}}/\|A\|_\infty$, where the supremum is taken over all non-zero $n\times n$ real matrices.

4.5.2 No converse for the polynomial method

In this section we show that there is no additive nor multiplicative converse for polynomials of degree 4 and 2-query algorithms. In other words, we will prove Theorems 4.3 and 4.4. Before doing that, we explain what was the error in the proof of Theorem 4.3 given in [ABP19].

Their proof arrives at the equation

$$\sum_{\alpha,\beta\in\{0,1,2,3,4\}^n:|\alpha|+|\beta|=4} d'_{\alpha,\beta}x^\alpha = C \sum_{\alpha\in\{0,1\}^n:|\alpha|=4} d_\alpha x^\alpha \quad \forall\, x\in\{-1,1\}^n, \tag{4.19}$$

where $d'_{\alpha,\beta}$, d_α and C are some real numbers, x^α stands for $\prod_{i=1}^n x_i^{\alpha_i}$ and $|\alpha|$ for $\sum_{i=1}^n \alpha_i$. It follows from the orthogonality of the characters that $d'_{\alpha,0} = Cd_\alpha$ for all $\alpha\in\{0,1\}^n$ such that $|\alpha|=4$. What is used, however, is that $d'_{\alpha,0} = Cd_\alpha$ for all $\alpha\in\{0,1,2,3,4\}^n$ such that $|\alpha|=4$, which is not true in general. For instance if $n=1$, $C=1$ and $d'_{(2,0),(0,2)} = -d'_{(0,0),(4,0)} = 1$ and the rest of the coefficients set to 0, then (4.19) becomes $x^2-1=0,\ \forall\, x\in\{-1,1\}$.

We now prove that there is no additive converse, from which the non-multiplicative converse result quickly follows.

Theorem 4.4. *There is no constant $\varepsilon\in(0,1)$ such that for every bounded polynomial p of degree at most 4, we have $\mathcal{E}(p,2)\le\varepsilon$.*

Proof. For any partition $\mathcal{P}$ of $\{0\}\cup[3n]$ in $2t$ subsets, Theorem 4.13 shows that

$$\begin{aligned}\sup_{p\in V_\mathcal{P},\|p\|_\infty\le 1}\mathcal{E}(p,t) &= \sup_{p\in V_\mathcal{P},\|p\|_\infty\le 1}\sup_{r\in V_\mathcal{P},\|r\|_{\infty,*}\le 1}\langle p,r\rangle - \|r\|_{\mathrm{cb},*}\\ &= \sup_{r\in V_\mathcal{P},\|r\|_{\infty,*}\le 1}\|r\|_{\infty,*}-\|r\|_{\mathrm{cb},*}\\ &= \sup_{r\in V_\mathcal{P},\|r\|_{\infty,*}=1} 1-\|r\|_{\mathrm{cb},*}.\end{aligned}$$

Consider now the case $t=2$ and the partition $\mathcal{P}_n = \{\{0\},\{1,\dots,n\},\{n+1,\dots,2n\},\{2n+1,\dots,3n\}\}$ of $\{0\}\cup[3n]$. In Theorem 4.17 a sequence of forms $p_n\in V_{\mathcal{P}_n}$ was constructed with the property that

$$\frac{\|p_n\|_{\mathrm{cb}}}{\|p_n\|_\infty}\to\infty. \tag{4.20}$$

Hence, by a duality argument we get that there is a sequence $r_n\in V_{\mathcal{P}_n}$ such that $\|r_n\|_{\mathrm{cb},*}/\|r_n\|_{\infty,*}\to 0$. Indeed, suppose towards a contradiction that there is a $K>0$ such that for every $n\in\mathbb{N}$ and every $r\in V_{\mathcal{P}_n}$ we have that $\|r\|_{\mathrm{cb},*}\ge K\|r\|_{\infty,*}$. Then,

$$\|p\|_{\mathrm{cb}} = \sup_{\|r\|_{\mathrm{cb},*}\le 1}\langle r,p\rangle \le \frac{1}{K}\sup_{\|r\|_{\infty,*}\le 1}\langle r,p\rangle = \frac{1}{K}\|p\|_\infty,$$

which contradicts Eq. (4.20). The sequence r_n shows that

$$\sup_{p\in V_{\mathcal{P}_n}, \|p\|_\infty \leq 1, n\in\mathbb{N}} \mathcal{E}(p,2) = 1,$$

which implies the stated result. □

Theorem 4.3. *For any $C > 0$, there exist an $n \in \mathbb{N}$ and a bounded quartic n-variable polynomial p such that no two-query quantum algorithm $\mathcal{A}$ satisfies $\mathbb{E}[\mathcal{A}(x)] = Cp(x)$ for every $x \in \{-1,1\}^n$.*

Proof. First note that we can assume $C \leq 1$, because $|\mathbb{E}[\mathcal{A}(x)]| \leq 1$ for any algorithm $\mathcal{A}$ and any $x \in \{-1,1\}^n$. Assume that there exists $0 < C \leq 1$ such that for every bounded p of degree 4 there is a 2-query algorithm $\mathcal{A}$ with $\mathbb{E}[\mathcal{A}(x)] = p(x)$ for every $x \in \{-1,1\}^n$. We claim that that $\mathcal{A}$ approximates p up to an additive error $1 - 1/C$, which contradicts Theorem 4.4. Indeed,

$$|p(x) - \mathbb{E}[\mathcal{A}(x)]| = |p(x)(1-C)| \leq 1 - C.$$

□

Chapter 5

Towards Aaronson and Ambainis conjecture via Fourier completely bounded polynomials

5.1 Introduction

Understanding the quantum query complexity of Boolean functions $f : D \to \{-1, 1\}$, where D is a subset of $\{-1, 1\}^n$, has been a crucial task of quantum information science [Amb18]. Many celebrated quantum algorithms show an advantage in terms of query complexity, for example in unstructured search [Gro96], period finding [Sho97], Simon's problem [Sim97], NAND-tree evaluation [FGG07] and element distinctness [Amb07]. However, these advantages are limited to be polynomial in the case of total functions (those with $D = \{-1, 1\}^n$), while they can be exponential for highly structured problems (informally, this means that $|D| = o(2^n)$), such as for Simon's problem [Sim97], period finding [Sho97] or k-fold forrelation [AA15, Tal20, BS21, SSW21]. It is widely believed that a lot of structure is necessary for superpolynomial speedups[1]. The following folklore conjecture, which has circulated since the late 90s,

[1]Recently, Yamakawa and Zhandry showed that superpolynomial speedups can be attained in unstructured search problems. That does not contradict that structure is needed to achieve superpolynomial speedups in decision problems, which are those modeled by Boolean functions [YZ22].

but was first formally posed by Aaronson and Ambainis [AA09], formalizes this idea.

Conjecture 5.1 (Folklore)**.** The biases of t-query quantum algorithms can be simulated with error at most ε on at least a $(1-\delta)$-fraction of the inputs using $\mathrm{poly}(t, 1/\varepsilon, 1/\delta)$ classical queries.

In other words, it is believed that quantum query algorithms can be approximated almost everywhere by classical query algorithms with only a polynomial overhead.

A route towards proving Conjecture 5.1 was designed by Aaronson and Ambainis using that the bias of quantum query algorithms are polynomials. Indeed, Beals et al. [BBC+01], proved that the bias of a t-query quantum algorithm is a bounded polynomial $p : \{-1,1\}^n \to \mathbb{R}$ of degree at most $2t$. Based on this observation, Aaronson and Ambainis conjectured in [AA09] that every bounded polynomial of bounded degree has an influential variable.

Conjecture 5.2 (Aaronson-Ambainis (AA))**.** Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most t with $\|p\|_\infty \leq 1$. Then, p has a variable with influence at least $\mathrm{poly}(\mathrm{Var}[p], 1/t)$.

The argument of [AA09, Theorem 7] to show that Conjecture 5.2 would imply Conjecture 5.1 works as follows. Let p the bounded polynomial of degree at most $2t$ that represents the bias of t query quantum algorithm. Say that we want to approximate $p(y)$ for some $y \in \{-1,1\}^n$. First, query an influential variable i of y. Then, the restricted polynomial $p|_{x(i)=y(i)}$ would also be a bounded polynomial of degree at most $2t$, so we can query again an influential variable. Given that the influences of these variables are big, after a *small* number of queries the remaining polynomial would have a low variance, so if we output its expectation it would be close to $p(y)$ with high probability.

A few reductions to other conjectures have been made. The first one is that is sufficient to prove the conjecture for one-block decoupled polynomials [OZ15]. Very recently, Lovett and Zhang stated two conjectures related to fractional certificate complexity that, if true, would imply the AA conjecture [LZ22]. Also recently, Austrin et al. showed a connection of the AA conjecture with cryptography: they proved that if the AA conjecture is false, then there is a secure key agreement in the quantum random oracle model that cannot be broken classically [ACC+22]. The most recent work in this line of research is the one by Bhattacharya, who showed that the conjecture is true for random restrictions of the polynomial [Bha25]. Regarding particular cases, it is only known to be true in a few scenarios: Boolean functions

$f : \{-1,1\}^n \to \{-1,1\}$ [Mid05, OSSS05, JZ11], symmetric polynomials [Iva19], multilinear forms whose Fourier coefficients are all equal in absolute value [Mon12] and block-multilinear completely bounded polynomials [BSdW22].

The last result is relevant in this context because Arunachalam, Briët and Palazuelos showed that the biases of quantum query algorithms are polynomials that are not only bounded, but also completely bounded [ABP19] (see Theorem 3.6). This is a more restricted normalization condition, which can be informally understood as the polynomial taking bounded values when evaluated not only on bounded scalars, but also on bounded matrix inputs. This way, one could try to use this extra condition to prove results about quantum query algorithms.

This idea was first put in practice by Bansal, Sinha and de Wolf [BSdW22]. They showed that the AA conjecture holds for completely bounded block-multilinear forms, which implies an almost everywhere classical simulation result, similar to Conjecture 5.1, for the amplitudes of certain quantum query algorithms. These algorithms query different (non-controlled) bit strings on every query, while Conjecture 5.1 concerns algorithms that query the same controlled bit string on every query.

Results of this chapter

We follow that line of work and use the characterization of [ABP19] to design a route towards Conjecture 5.1. Our first result is a new presentation of that characterization that is more convenient for our purposes. To do this we introduce the Fourier completely bounded t-norms ($\|\cdot\|_{\mathrm{fcb},t}$), which are relaxations of the supremum norm. In these norms we not only take the supremum of the values that the polynomial takes over Boolean strings as in Eq. (2.12), but also on matrix inputs that behave like Boolean strings. We will not include formal definitions in the introduction, but we illustrate the concept of having *Boolean behavior of degree t* with an example. For $m \in \mathbb{N}$, we denote the $m \times m$ real matrices by M_m. Say that $t = 4$ and $n = 6$, then if a pair of vectors $u, v \in \mathbb{R}^m$ and a string of matrices $A \in (M_m)^6$ have Boolean behaviour of degree 4, they satisfy, for instance,

$$\langle u, A(1)A(1)A(2)A(3)v\rangle = \langle u, A(5)A(2)A(3)A(5)v\rangle,$$

because they should simulate the relation $x(1)x(1)x(2)x(3) = x(5)x(2)x(3)x(5)$ satisfied by any Boolean string $\{-1,1\}^6$. As the reader might guess, (u, v, A) will have Boolean behavior of degree t if it simulates the relations of $\mathbb{F}_2^n$ that involve product of t of the canonical generators.

5.1. Introduction

Using the Fourier expansion of polynomials defined on the Boolean hypercube we will introduce a natural way of evaluating polynomial in matrix inputs that have Boolean behavior, which allows us to introduce the Fourier completely bounded t-norm.

Definition 5.3. (Informal version of Definition 5.10) Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most t. Its *Fourier completely bounded t-norm* is given by

$$\|p\|_{\mathrm{fcb},t} := \sup |p(u,v,A)| \tag{5.1}$$

where the supremum is taken over all (u,v,A) that have Boolean behavior of degree t.

After a reinterpretation of the semidefinite programs proposed in [GL19] to characterize quantum query complexity, based on [ABP19], we show that the Fourier completely bounded t-norms are those that characterize quantum query algorithms.

Theorem 5.4. *Let $p : \{-1,1\}^n \to \mathbb{R}$. Then, p is the bias of a t-query quantum algorithm if and only if its degree is at most $2t$ and $\|p\|_{\mathrm{fcb},2t} \leq 1$.*

This new presentation of the main result of [ABP19] is more compact than the original one. It is presented directly in terms of polynomials of the Boolean hypercube, does not involve a minimization over possible completely bounded extensions of p as in Definition 4.7, and eludes the use of tensors/multilinear forms.

Given that the Fourier completely bounded t-norms are at least the supremum norm[2], Theorem 5.4 suggests that Conjecture 5.2 may be more general than necessary. Hence, we propose the following weaker conjecture, that would also imply Conjecture 5.1.

Conjecture 5.5. Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most t with $\|p\|_{\mathrm{fcb},t} \leq 1$. Then, p has a variable with influence at least $\mathrm{poly}(\mathrm{Var}[p], 1/t)$.

Using a generalization through *creation* and *annihilation* operators of the construction used by Varopoulos to rule out a von Neumann's inequality for degree 3 polynomials [Var74], we can prove a particular case of Conjecture 5.5.

Theorem 5.6. *Let $t \in \mathbb{N}$. Let $p : \{-1,1\}^n \to \mathbb{R}$ be a homogeneous polynomial of degree t and with $\|p\|_{\mathrm{fcb},t} \leq 1$. Then, the maximum influence of p is at least $\mathrm{Var}[p]^2$.*

[2]From the results of [BP19] it can be inferred that there is a sequence of polynomials p_n of degree 3 such that $\|p_n\|_{\mathrm{fcb},3}/\|p_n\|_\infty \to_n \infty$.

The proof of the homogeneous case does not straightforwardly generalize (see Remark 5.20), but it suggests a way to solve the general case (see Remark 5.21). In particular, we propose Question 5.22 (that reminds of tensor networks and almost-quantum correlations), which if answered affirmatively would imply Conjecture 5.5.

Theorem 5.6 is the first result concerning the AA conjecture whose constant has no dependence on the degree (to prove Conjecture 5.1 we could afford a polynomial dependence on the degree). Also, it requires considerably fewer algebraic constraints than the other particular cases for which we know AA conjecture to hold. In addition, thanks to Theorem 5.4, it can be interpreted directly in terms of quantum query algorithms.

Corollary 5.7. *Let $t \in \mathbb{N}$. Let $\mathcal{A}$ be a t-query quantum algorithm whose bias is a homogeneous polynomial $p : \{-1,1\}^n \to \mathbb{R}$ of degree $2t$. Then, the maximum influence of p is at least* $\mathrm{Var}[p]^2$.

With a similar construction as the one we used for Theorem 5.6, we can reprove the results of [BSdW22] regarding the influence of block-multilinear completely bounded polynomials. These polynomials have a particular algebraic structure and also a normalization condition when evaluated on matrix inputs (see Section 5.4.1 below).

Theorem 5.8. *Let $t \in \mathbb{N}$. Let $p : \{-1,1\}^{n\times t} \to \mathbb{R}$ be a block-multilinear degree t polynomial with $\|p\|_{\mathrm{cb}} \leq 1$. Then, p has a variable of influence at least* $(\mathrm{Var}[p]/t)^2$. *What is more, if p is homogeneous of degree t, then it has a variable of influence at least* $\mathrm{Var}[p]^2$.

Theorem 5.8 corresponds to [BSdW22, Theorem 1.4], where Bansal et al. proved the same result but with influences at least $\mathrm{Var}[p]^2/[e(t+1)^4]$ in the general case and with $\mathrm{Var}[p]^2/(t+1)^2$ in the homogeneous degree t case. Their proofs involve evaluating p in *random infinite dimensional matrix inputs*, which they can control using ideas of free probability. However, our proof evaluates p in explicit finite dimensional matrix inputs, is shorter and obtains better constants. In particular, our constant for the homogeneous case is optimal.

5.2 The Fourier completely bounded t-norms

There is a vast theory concerning the properties of multilinear maps $T : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ that are completely bounded, i.e., bounded when they are extended to matrix domains [Pau03]. However, to the best of our knowledge, there is no notion of being

completely bounded for polynomials $p : \{-1,1\}^n \to \mathbb{R}$ defined on the Boolean hypercube. Here, we propose a matrix notion of behaving like a Boolean string. Then, using the Fourier expansion of these polynomials we define the evaluation of the polynomials on these matrix inputs that behave like Boolean strings. Finally, we introduce the Fourier completely bounded t-norms and prove a few of their properties.

We recall that every $p : \{-1,1\}^n \to \mathbb{R}$ can be written as

$$p(x) = \sum_{S \subseteq [n]} \hat{p}(S) \prod_{i \in S} x(i), \tag{5.2}$$

where $\hat{p}(S)$ are the Fourier coefficients of p. We say that p has degree at most t if $\hat{p}(S) = 0$ for every $|S| > t$, where $|S|$ denotes the cardinality of S.

We will be interested on simulating the behavior of bit strings $x \in \{-1,1\}^n \times \{1\}$ with one extra frozen variable[3]. Given $t \in \mathbb{N}$ and $\mathbf{i}, \mathbf{j} \in [n+1]^t$ we say that $\mathbf{i} \sim \mathbf{j}$, if

$$x(i_1) \dots x(i_d) = x(j_1) \dots x(j_d) \text{ for every } x \in \{-1,1\}^n \times \{1\}. \tag{5.3}$$

In other words, if we define

$$S_{\mathbf{i}} := \{k \in [n] : \ k \text{ occurs an odd number of times in } \mathbf{i}\},$$

then $\mathbf{i} \sim \mathbf{j}$ if and only if $S_{\mathbf{i}} = S_{\mathbf{j}}$. Note that $n+1$ does not belong to these sets $S_{\mathbf{i}}$. Given $S \subseteq [n]$ with $|S| \leq t$, we write $[\mathbf{i}^S]$ to denote the equivalence class of indices $\mathbf{i}$ such that $S_{\mathbf{i}} = S$.

Definition 5.9. Let $n, t, m \in \mathbb{N}$. Let $u, v \in S^{m-1}$ and let $A \in (B_m)^n$. We say that (u, v, A) has *Boolean behavior of degree* t if

$$\langle u, A(i_1) \dots A(i_d) v \rangle = \langle u, A(j_1) \dots A(j_d) v \rangle$$

for all $\mathbf{i}, \mathbf{j} \in [n+1]^t$ such that $\mathbf{i} \sim \mathbf{j}$. We call $\mathscr{BB}^t$ to the set of (u, v, A) with Boolean behavior of degree t.

Informally, having Boolean behavior of degree t means that the relations of Eq. (5.3) and some normalization conditions are satisfied. In particular, for any bit string $x \in \{-1,1\}^n \times \{1\}$ and any $t \in \mathbb{N}$, we have that $(1, 1, x)$ has Boolean behavior of degree t.

[3]The extra variable set to 1 is there because quantum query algorithms query a controlled bit string. A non-controlled version, which would not require that extra variable.

Also note that given $t \in \mathbb{N}$, for every $S \subseteq [n]$ with $|S| \leq t$ there is at least one $\mathbf{i} \in [n+1]^t$ such that $S_{\mathbf{i}} = S$. Thus, given (u, v, A) with Boolean behavior of degree t, for every $|S| \leq t$ the product $\prod_{i \in S} x(i)$ can be simulated (in a unique manner) by $\langle u, A(i_1^S) \dots A(i_d^S) v \rangle$. In particular, this means that for a polynomial p of degree at most t, we can define through Eq. (5.2) an evaluation of p on every (u, v, A) that has Boolean behavior of degree t, which leads to the definition Fourier completely bouded t-norm.

Definition 5.10. Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most t. Then, its *Fourier completely bounded t-norm* is defined by

$$\|p\|_{\mathrm{fcb},t} = \sup_{(u,v,A) \in \mathscr{B}\mathscr{B}^t} \sum_{S \subseteq [n], |S| \leq t} \hat{p}(S) \langle u, A(i_1^S) \dots A(i_d^S) v \rangle.$$

The rest of the section is devoted to prove a few results concerning the Fourier completely bounded t-norms. First of all we show that, indeed, they are norms.

Proposition 5.11. *Let $t \in \mathbb{N}$. Then, $\| \cdot \|_{\mathrm{fcb},t}$ is a norm in the space of polynomials $p : \{-1,1\}^n \to \mathbb{R}$ of degree at most t.*

Proof. It clearly satisfies the triangle inequality and is homogeneous. Also, if $p = 0$ then $\|p\|_{\mathrm{fcb},t} = 0$, and vice versa, because $\|p\|_\infty \leq \|p\|_{\mathrm{fcb},t}$. □

One nice property of these norms is that they can be computed as semidefinite programs.

Proposition 5.12. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most t. Then, its Fourier completely bounded t-norm can be written as the following SDP*

$$\|p\|_{\mathrm{fcb},t} = \sup \sum_{S \in [n], |S| \leq t} \hat{p}(S) \langle u, v_{\mathbf{i}^S} \rangle, \tag{5.4}$$

$$u, v, v_{\mathbf{i}} \in \mathbb{R}^m,\ m \in \mathbb{N},\ \mathbf{i} \in [n+1]^s,\ s \in [t],$$

$$\langle u, v_{\mathbf{i}} \rangle = \langle u, v_{\mathbf{j}} \rangle,\ \text{if } \mathbf{i} \sim \mathbf{j},\ \mathbf{i}, \mathbf{j} \in [n+1]^t, \tag{5.5}$$

$$\langle u, u \rangle = \langle v, v \rangle = 1, \tag{5.6}$$

$$\mathrm{Gram}_{\substack{\mathbf{j} \in [n+1]^s, \\ s \in [t-1]_0}} \{v_{i\mathbf{j}}\} \preccurlyeq \mathrm{Gram}_{\substack{\mathbf{j} \in [n+1]^s, \\ s \in [t-1]_0}} \{v_{\mathbf{j}}\},\ \text{for } i \in [n+1], \tag{5.7}$$

where we by $v_{\mathbf{j}}$ with $\mathbf{j} \in [n+1]^0$ we mean v, Gram *denotes the gram matrix and the symbol '$\preccurlyeq$' the usual matrix inequality.*

Proof. Let $|||p|||$ be the expression on the right-hand side of Eq. (5.4). Note that Eq. (5.5) represents the relations of bit strings of Eq. (5.3), while Eqs. (5.6) and (5.7) encode normalization conditions.

On the one hand, every $(u, v, A) \in \mathscr{BB}^t$ defines a feasible instance for $|||p|||$ through

$$v_{\mathbf{i}} := A(i_1) \dots A(i_s)v$$

for every $\mathbf{i} \in [n+1]^s$ and every $s \in [t]$. Given that the value of this instance is

$$\sum_{S \subseteq [n], |S| \leq t} \widehat{p}(S) \langle u, A(i_1^S) \dots A(i_d^S)v \rangle$$

we have that $|||p||| \geq \|p\|_{\mathrm{fcb},t}$.

On the other hand, let $u, v, v_{\mathbf{i}} \in \mathbb{R}^m$ be a feasible instance of $|||p|||$. For $i \in [n+1]$ define $A(i) \in M_m$ as the linear map from $\mathbb{R}^m$ to $\mathbb{R}^m$ that takes $v_{\mathbf{j}}$ to $v_{i\mathbf{j}}$ for every $\mathbf{j} \in [n+1]^s$ and every $s \in [t-1]_0$, and it is extended to the orthogonal complement as 0. First of all, we should check that this is a correct definition, meaning that for every $\lambda \in \mathbb{R}^m$, with $m = (n+1)^{t-1} + \dots + (n+1)^0$, we have that

$$\sum_{\mathbf{j}} \lambda_{\mathbf{j}} v_{\mathbf{j}} = 0 \implies \sum_{\mathbf{j}} \lambda_{\mathbf{j}} v_{i\mathbf{j}} = 0.$$

Indeed, we can prove something stronger:

$$\begin{aligned}(\sum_{\mathbf{j}} \lambda_{\mathbf{j}} v_{i\mathbf{j}})^{\mathsf{T}} \sum_{\mathbf{j}'} \lambda_{\mathbf{j}'} v_{i\mathbf{j}'} &= \lambda^{\mathsf{T}} \mathrm{Gram}_{\substack{\mathbf{j} \in [n+1]^s, \\ s \in [t-1]_0}} \{v_{i\mathbf{j}}\} \lambda \leq \lambda^{\mathsf{T}} \mathrm{Gram}_{\substack{\mathbf{j} \in [n+1]^s, \\ s \in [t-1]_0}} \{v_{\mathbf{j}}\} \lambda \\ &= (\sum_{\mathbf{j}} \lambda_{\mathbf{j}} v_{\mathbf{j}})^{\mathsf{T}} \sum_{\mathbf{j}'} \lambda_{\mathbf{j}'} v_{\mathbf{j}'}.\end{aligned}$$

The above calculation also proves that the $A(i)$'s are contractions, and thanks to Eq. (5.5) it follows that (u, v, A) has Boolean behavior of degree t. Finally, note that the value of this (u, v, A) for $\|p\|_{\mathrm{fcb},t}$ is the same as the value of $(u, v, v_{\mathbf{j}})$ for $|||p|||$, so $\|p\|_{\mathrm{fcb},t} \geq |||p|||$. □

Given $t, t' \in \mathbb{N}$ with $t' > t$ and a polynomial $p : \{-1, 1\}^n \to \mathbb{R}$ of degree at most t, $\|p\|_{\mathrm{fcb},t}$ and $\|p\|_{\mathrm{fcb},t'}$ have different definitions, but they are comparable. In particular, we prove that the Fourier completely bounded t-norms are not increasing.

Proposition 5.13. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most t. Then,*

$$\|p\|_{\mathrm{fcb},t+1} \leq \|p\|_{\mathrm{fcb},t}.$$

Remark 5.14. Proposition 5.13 is coherent with Theorem 5.4 (proved below), because allowing more queries to quantum algorithms only increases their power. Theorem 5.4 also suggests that $\|p\|_{\mathrm{fcb},n} = \|p\|_\infty$ should hold, because n quantum queries should be enough to output any bounded polynomial. If true, alongside Propositions 5.12 and 5.13, it would mean that $(\|p\|_{\mathrm{fcb},t})_{t\in[n]}$ is a decreasing hierarchy of SDPs that tend to $\|p\|_\infty$.

Proof of Proposition 5.13. Let (u, v, A) have Boolean behavior of degree $t+1$. Then,

$$(\widetilde{u}, \widetilde{v}, \widetilde{A}) = (u, \frac{A(n+1)v}{\|A(n+1)v\|}, A) \tag{5.8}$$

has Boolean behavior of degree t. Also, given that $t+1 > t$, we have that for every $S \subseteq [n]$ with $|S| \leq t$, there exists $\mathbf{i} \in [n+1]^{t+1}$ such that $S_{\mathbf{i}} = S$, $i_{t+1} = n+1$, and

$$\langle u, A(i_1)\dots A(i_{t+1})v\rangle = \|A(n+1)v\|\langle \widetilde{u}, \widetilde{A}(i_1)\dots\widetilde{A}(i_d)\widetilde{v}\rangle. \tag{5.9}$$

This way,

$$\begin{aligned}
\|p\|_{\mathrm{fcb},t+1} &= \sup_{(u,v,A)\in\mathscr{B}\mathscr{B}^{t+1}} \sum_{S\subseteq[n],|S|\leq t} \widehat{p}(S)\langle u, A(i_1^S)\dots A(i_{t+1}^S)v\rangle \\
&= \sup_{(u,v,A)\in\mathscr{B}\mathscr{B}^{t+1}} \|A(n+1)v\| \sum_{S\subseteq[n],|S|\leq t} \widehat{p}(S)\langle \widetilde{u}, \widetilde{A}(i_1^S)\dots\widetilde{A}(i_d^S)\widetilde{v}\rangle \\
&\leq \sup_{(u,v,A)\in\mathscr{B}\mathscr{B}^{t+1}} \sum_{S\subseteq[n],|S|\leq t} \widehat{p}(S)\langle \widetilde{u}, \widetilde{A}(i_1^S)\dots\widetilde{A}(i_d^S)\widetilde{v}\rangle \\
&\leq \sup_{(u',v',A')\in\mathscr{B}\mathscr{B}^{t}} \sum_{S\subseteq[n],|S|\leq t} \widehat{p}(S)\langle u', A'(i_1^S)\dots A'(i_d^S)v'\rangle \\
&= \|p\|_{\mathrm{fcb},t},
\end{aligned}$$

where in the second line we have used Eq. (5.9), and in the third line that $\|A(n+1)v\| \leq 1$, and in the fourth that $(\widetilde{u}, \widetilde{v}, \widetilde{A})$ has Boolean behavior of degree t. □

The next proposition states that $\|\cdot\|_{\mathrm{fcb},t}$ does not increase after restrictions, which is a relevant feature to ensure that Conjecture 5.5 implies Conjecture 5.1. Given a polynomial $p : \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$, the *restriction of p to the i-th variable*

being set to $y \in \{-1,1\}$ is the polynomial $q : \{-1,1\}^{n-1} \to \mathbb{R}$ (whose variables we index with $x(1),\dots,x(i-1),x(i+1),\dots,x(n)$ for convenience) defined by $q(x) := p(x(1),\dots,x(i-1),y,x(i+1),\dots,x(n))$.

Proposition 5.15. *Let* $p : \{-1,1\}^n \to \mathbb{R}$ *be a polynomial of degree at most* t *and let* $i \in [n]$. *Let* $q : \{-1,1\}^{n-1} \to \mathbb{R}$ *be the restriction of* p *to the* i*-th variable being set to* $y \in \{-1,1\}$. *Then,*

$$\|q\|_{\mathrm{fcb},t} \leq \|p\|_{\mathrm{fcb},t}.$$

Proof. Consider a pair of vectors and a string of matrices $(u, v, A(1), \dots, A(i-1), A(i+1), \dots, A(n+1))$ with Boolean behavior of degree t. Define $\widetilde{u} := u$, $\widetilde{v} := v$ and $\widetilde{A}(j)$ for $j \in [n+1]$ as

$$\widetilde{A}(j) = \begin{cases} A(j) & \text{if } j \neq i, \\ yA(n+1) & \text{if } j = i. \end{cases}$$

It can be verified that $(\widetilde{u}, \widetilde{v}, \widetilde{A}(1), \dots, \widetilde{A}(n+1))$ has Boolean behavior of degree t. Now note that for every $S \subseteq [n] - \{i\}$, it is satisfied that

$$\widehat{q}(S) = \widehat{p}(S) + y\widehat{p}(S \cup \{i\}). \tag{5.10}$$

Also, for every $S \subseteq [n] - \{i\}$ with $|S| \leq t-1$, it is satisfied that

$$\langle \widetilde{u}, \widetilde{A}(j_1^S) \dots \widetilde{A}(j_d^S) \widetilde{v} \rangle = y \langle \widetilde{u}, \widetilde{A}(j_1^{S \cup \{i\}}) \dots \widetilde{A}(j_d^{S \cup \{i\}}) \widetilde{v} \rangle. \tag{5.11}$$

Thus,

$$\begin{aligned}
\|q\|_{\mathrm{fcb},t} &= \sup_{\substack{(u,v,A(j))\in\mathscr{B}\mathscr{B}^t \\ j\in[n+1]-\{i\}}} \sum_{S\subseteq[n]-\{i\},|S|\le t} \widehat{q}(S)\langle u, A(j_1^S)\dots A(j_t^S)v\rangle \\
&= \sup_{\substack{(u,v,A(j))\in\mathscr{B}\mathscr{B}^t \\ j\in[n+1]-\{i\}}} \sum_{S\subseteq[n]-\{i\},|S|\le t} \widehat{p}(S)\langle u, A(j_1^S)\dots A(j_d^S)v\rangle \\
&+ \sum_{S\subseteq[n]-\{i\},|S|\le t-1} y\widehat{p}(S\cup\{i\})\langle u, A(j_1^S)\dots A(j_d^S)v\rangle \\
&= \sup_{\substack{(u,v,A(j))\in\mathscr{B}\mathscr{B}^t \\ j\in[n+1]-\{i\}}} \sum_{S\subseteq[n]-\{i\},|S|\le t} \widehat{p}(S)\langle \widetilde{u}, \widetilde{A}(j_1^S)\dots \widetilde{A}(j_d^S)\widetilde{v}\rangle \\
&+ \sum_{S\subseteq[n]-\{i\},|S|\le t-1} \widehat{p}(S\cup\{i\})\langle \widetilde{u}, \widetilde{A}(j_1^{S\cup\{i\}})\dots \widetilde{A}(j_d^{S\cup\{i\}})\widetilde{v}\rangle \\
&\le \sup_{\substack{(u',v',A'(j))\in\mathscr{B}\mathscr{B}^t \\ j\in[n+1]}} \sum_{S\subseteq[n],|S|\le t} \widehat{p}(S)\langle u', A'(j_1^S)\dots A'(j_d^S)v'\rangle \\
&= \|p\|_{\mathrm{fcb},t},
\end{aligned}$$

where in the second line we have used Eq. (5.10), in the fourth line Eq. (5.11), and in the sixth line that $(\tilde{u},\tilde{v},\tilde{A})$ has Boolean behavior. □

5.3 Quantum query algorithms are Fourier completely bounded polynomials

Now we are ready to prove Theorem 5.4, that fully characterizes quantum query algorithms in terms of the Fourier completely bounded t-norms.

Theorem 5.4. *Let $p : \{-1,1\}^n \to \mathbb{R}$. Then, p is the bias of a t-query quantum algorithm if and only if its degree is at most $2t$ and $\|p\|_{\mathrm{fcb},2t} \le 1$.*

To prove Theorem 5.4 we reinterpret the semidefinite programs of [GL19], based on [ABP19].

Theorem 5.16 (Gribling-Laurent)**.** *Let $p : \{-1,1\}^n \to \mathbb{R}$. Then, p is the bias of t-query quantum algorithm if and only if its degree is at most $2t$ and the value of the*

following semidefinite program is at most 0,

$$\begin{aligned} \max\ & -w + \sum_{x\in\{-1,1\}^n} \frac{p(x)\phi(x)}{2^n} \\ \text{s.t. } & w \geq 0,\ m \in \mathbb{N},\ A_s \in (B_m)^{n+1},\ u, v \in \mathbb{R}^m,\ s \in [2t], \\ & \|\phi\|_1 = 1,\ \|u\|^2 = \|v\|^2 = w, \\ & \hat{\phi}(S_{\mathbf{i}}) = \langle u, A_1(i_1)\dots A_{2t}(i_{2t})v\rangle,\ \mathbf{i} \in [n+1]^{2t}, \end{aligned} \tag{5.12}$$

where $\|\phi\|_1 = \sum_{x\in\{-1,1\}^n} \frac{|\phi(x)|}{2^n}$.

Remark 5.17. Theorem 5.16 corresponds to [GL19, Equation (24)]. There, the authors not only ask for the $A_s(i)$ to be contractions, but also unitaries. However, that extra restriction does not change the value of the semidefinite program because we can always block-encode a contraction in the top left corner of an unitary (see for instance [AAI+16, Lemma 7]). We also want to remark that $A_s(i)$ can be taken to be equal to $A_{s'}(i)$ for every $s, s' \in [2t]$ and every $i \in [n+1]$, as this extra restriction does not change value of the semidefinite program. Indeed, let (u, v, A_s, w, ϕ) be part of feasible instance of Eq. (5.12). Define now

$$\begin{aligned} \tilde{u} &:= u \otimes e_1, \\ \tilde{v} &:= v \otimes e_{2t+1}, \\ A(i) &:= \sum_{s\in[2t]} A_s(i) \otimes e_s e_s^{\mathsf{T}}, \end{aligned}$$

where $\{e_s\}_{s\in[2t+1]}$ is an orthonormal basis of $\mathbb{R}^{2t+1}$. Then,

$$\langle u, A_1(i_1)\dots A_d(i_{2t})v\rangle = \langle \tilde{u}, \tilde{A}(i_1)\dots\tilde{A}(i_{2t})\tilde{v}\rangle,$$

for every $\mathbf{i} \in [n+1]^{2t}$. Hence, $(\tilde{u}, \tilde{v}, \tilde{A}, w, \phi)$ is a feasible instance for Eq. (5.12) that attains the same value as (u, v, A_s, w, ϕ).

Proof of 5.4. Thanks to Theorem 5.16 and Remark 5.17, we know that p is the output of t-query quantum algorithm if and only if its degree is at most $2t$ and the following

constraint is satisfied

$$\sum_{x\in\{-1,1\}^n} \frac{p(x)\phi(x)}{2^n} \leq w \tag{5.13}$$

$$\begin{aligned}
\text{s.t. } & w \geq 0,\ m \in \mathbb{N},\ A \in (B_m)^{n+1},\ u, v \in \mathbb{R}^m, \\
& \|\phi\|_1 = 1, \qquad (5.14)\\
& \|u\|^2 = \|v\|^2 = w, \\
& \widehat{\phi}(S_{\mathbf{i}}) = \langle u, A(i_1)\dots A(i_{2t})v\rangle,\ \mathbf{i} \in [n+1]^{2t}.
\end{aligned}$$

Now, note that if (u, v, A, ϕ, w) satisfies all conditions of Eq. (5.13) except for Eq. (5.14), then $(u/\sqrt{\|\phi\|_1},\ v/\sqrt{\|\phi\|_1}, A, \phi/\|\phi\|_1, w/\|\phi\|_1)$ would be a feasible instance. Furthermore, given that

$$\sum_{x\in\{-1,1\}^n} \frac{p(x)\phi(x)}{2^n} \leq w \iff \frac{1}{\|\phi\|_1} \sum_{x\in\{-1,1\}^n} \frac{p(x)\phi(x)}{2^n} \leq \frac{w}{\|\phi\|_1},$$

we can write Eq. (5.13) forgetting about the normalization condition of Eq. (5.14). In other words, Eq. (5.13) is equivalent to

$$\sum_{x\in\{-1,1\}^n} \frac{p(x)\phi(x)}{2^n} \leq w \tag{5.15}$$

$$\begin{aligned}
\text{s.t. } & w \geq 0,\ m \in \mathbb{N},\ A \in (B_m)^{n+1},\ u, v \in \mathbb{R}^m, \\
& \|u\|^2 = \|v\|^2 = w, \qquad (5.16)\\
& \widehat{\phi}(S_{\mathbf{i}}) = \langle u, A(i_1)\dots A(i_{2t})v\rangle,\ \mathbf{i} \in [n+1]^{2t}.
\end{aligned}$$

In addition, by homogeneity we can assume $w = 1$, as if (u, v, A, ϕ, w) is a feasible instance, then $(u/\sqrt{w}, v/\sqrt{w}, A, \phi/w, 1)$ also is, and Eq. (5.15) is satisfied for the first instance if and only if is satisfied for the second instance. Also note, that if (u, v, A) are part of a feasible instance of Eq. (5.15), then it automatically has Boolean behavior of degree $2t$, and any (u, v, A) defines a feasible instance for Eq. (5.15). Finally, by Parseval's identity we can rewrite $\sum_{x\in\{-1,1\}^n} \frac{p(x)\phi(x)}{2^n}$ as $\sum_{S\subseteq[n]} \widehat{p}(S)\widehat{\phi}(S)$. Putting altogether we get that p is the output of t-query quantum algorithm if and only if its

degree is at most $2t$ and

$$\sum_{S\subseteq[n],|S|\leq 2t} \widehat{p}(S)\langle u, A(i_1^S)\dots A(i_{2t}^S)v\rangle \leq 1$$
$$\text{s.t. } (u,v,A) \text{ has Boolean behavior of degree } 2t,$$

which is the same as saying that $\|p\|_{\mathrm{fcb},2t} \leq 1$. □

5.4 Aaronson and Ambainis conjecture for (Fourier) completely bounded polynomials

In this section we prove Theorem 5.6 and Theorem 5.8. Both are based on the construction used by Varopoulos to disprove a degree 3 von Neumann's inequality [Var74].

5.4.1 AA conjecture for block-multilinear completely bounded polynomials

Before proving Theorem 5.8, we shall specify what is a block-multilinear completely bounded polynomial. A *block-multilinear polynomial* of degree t is a polynomial $p : \{-1,1\}^{n\times t} \to \mathbb{R}$ such that if we divide the variables $x \in \{-1,1\}^{n\times t}$ in t blocks of n coordinates each, then the every of the monomials of p has at most one coordinate of each of the blocks. In other words, the block-multilinear polynomials of degree t are those that can be written as

$$p(x_1,\dots,x_d) = \hat{p}(\emptyset) + \sum_{s\in[t]} \sum_{\substack{\mathbf{b}\in[t]^s \\ b_1<\cdots<b_s}} \sum_{\mathbf{i}\in[n]^s} \hat{p}(\{(b_1,i_1),\dots,(b_s,i_s)\})x_{b_1}(i_1)\dots x_{b_s}(i_s), \tag{5.17}$$

for every $(x_1,\dots,x_d) \in (\{-1,1\}^n)^t$. For this kind of polynomials, there is a very natural way of evaluating them in matrix inputs,

$$p(A_1,\dots,A_d) = \hat{p}(\emptyset)\mathrm{Id}_m + \sum_{s\in[t]} \sum_{\substack{\mathbf{b}\in[t]^s \\ b_1<\cdots<b_s}} \sum_{\mathbf{i}\in[n]^s} \hat{p}(\{(b_1,i_1),\dots,(b_s,i_s)\})A_{b_1}(i_1)\dots A_{b_s}(i_s), \tag{5.18}$$

for every $A_s \in (M_m)^n$, $s \in [t]$ and $m \in \mathbb{N}$. The *completely bounded norm* of a block-multilinear polynomial is defined as[4]

$$\|p\|_{\mathrm{cb}} := \sup\{\|p(A_1, \ldots, A_d)\| : \ m \in \mathbb{N}, \ A_s \in (B_m)^n, \ s \in [t]\}. \tag{5.19}$$

Concerning these polynomials, we can show the following.

Theorem 5.8. *Let $t \in \mathbb{N}$. Let $p : \{-1,1\}^{n\times t} \to \mathbb{R}$ be a block-multilinear degree t polynomial with $\|p\|_{\mathrm{cb}} \leq 1$. Then, p has a variable of influence at least $(\mathrm{Var}[p]/t)^2$. What is more, if p is homogeneous of degree t, then it has a variable of influence at least $\mathrm{Var}[p]^2$.*

Remark 5.18. With our proof of the homogeneous case of Theorem 5.8 we can show that for the case of $p : \{-1,1\}^{n\times t} \to \mathbb{R}$ being a homogeneous degree t block-multinear polynomial we have the following non-commutative root influence inequality

$$\|p\|_{\mathrm{cb}} \geq \sum_{i\in[n]} \sqrt{\mathrm{Inf}_{s,i}[p]}, \tag{5.20}$$

for any $s \in [t]$. This improves [BSdW22, Theorem 1.4] in two ways. First, we can allow s to be any number in $[t]$, while they only prove the result of $s \in \{1, t\}$. Second, they prove a weaker statement that depends on t, namely,

$$\|p\|_{\mathrm{cb}} \geq \sum_{i\in[n]} \frac{\sqrt{\mathrm{Inf}_{s,i}[p]}}{\sqrt{e(t+1)}},$$

for $s \in \{1, t\}$.

Remark 5.19. Given that $p(x_1, \ldots, x_d) = x_1(1) \ldots x_d(1)$ is a homogeneous degree t block-multilinear completely bounded polynomial with $\mathrm{Var}[p]^2 = \mathrm{MaxInf}[p] = 1$, we have that the homogeneous case of Theorem 5.8 is optimal.

Proof of the homogeneous degree t case of Theorem 5.8. Let p be a homogeneous degree t block-multilinear polynomial. Let $s \in [t]$. We label the coordinates by (r, i), where $r \in [t]$ indicates the block, and $i \in [n]$. Our goal is defining $A \in (B_m)^n$ and $f_\emptyset, e_\emptyset \in S^{m-1}$ such that

$$\langle f_\emptyset, A(i_1) \ldots A(i_d) e_\emptyset \rangle = \frac{\widehat{p}(\{(1, i_1), \ldots, (t, i_d)\})}{\sqrt{\mathrm{Inf}_{s,i_s}[p]}}. \tag{5.21}$$

[4]We abuse notation here, as, for the case of homogeneous block-multilinear polynomials, this definition conflicts with the one given in Definition 4.7. For the rest of the chapter, we will use the one in Eq. (5.18).

Once we are there, we can prove the announced root-influence inequality Eq. (5.20). Indeed,

$$\begin{aligned}
\|p\|_{\mathrm{cb}} &\geq \sum_{i_1,\ldots,i_d\in[n]} \widehat{p}(\{(1,i_1),\ldots,(t,i_d)\})\langle f_\emptyset, A(i_1)\ldots A(i_d)e_\emptyset\rangle \\
&= \sum_{i_1,\ldots,i_d\in[n]} \widehat{p}(\{(1,i_1),\ldots,(t,i_d)\})\frac{\widehat{p}(\{(1,i_1),\ldots,(t,i_d)\})}{\sqrt{\mathrm{Inf}_{s,i_s}[p]}} \\
&= \sum_{i_s\in[n]} \frac{1}{\sqrt{\mathrm{Inf}_{s,i_s}[p]}} \underbrace{\sum_{i_1,\ldots,i_{s-1},i_{s+1},i_d\in[n]} \widehat{p}(\{(1,i_1),\ldots,(t,i_d)\})^2}_{\mathrm{Inf}_{s,i}[p]} \\
&= \sum_{i\in[n]} \sqrt{\mathrm{Inf}_{s,i}[p]}.
\end{aligned}$$

Finally, the statement about the maximal influence quickly follows from the root-influence inequality

$$\|p\|_{\mathrm{cb}} \geq \sum_{i\in[n]} \sqrt{\mathrm{Inf}_{s,i}[p]} \geq \sum_{i\in[n]} \frac{\mathrm{Inf}_{s,i}[p]}{\sqrt{\mathrm{MaxInf}[p]}} = \frac{\mathrm{Var}[p]}{\sqrt{\mathrm{MaxInf}[p]}},$$

which after rearranging yields

$$\mathrm{MaxInf}[p] \geq \left(\frac{\mathrm{Var}[p]}{\|p\|_{\mathrm{cb}}}\right)^2.$$

Hence, it suffices to design $(f_\emptyset, e_\emptyset, A) \in S^{m-1}\times S^{m-1}\times (B_m)^n$ satisfying Eq. (5.21). Let $\mathcal{S} := \{\{(r,i_r),\ldots,(t,i_t)\} : i_r,\ldots,i_t \in [n],\ s{+}1 \leq r \leq t\}$ and $\mathcal{S}' := \{\{(1,i_1),\ldots,(r,i_r)\} : i_1,\ldots,i_r \in [n],\ r \leq s{-}1\}$. Let $m := 2{+}|\mathcal{S}|{+}|\mathcal{S}'|$. Let $\{e_\emptyset, e_S, f_\emptyset, f_{S'} :\ S\in\mathcal{S},\ S'\in\mathcal{S}'\}$ be an orthonormal basis of $\mathbb{R}^m$, and define $A(i) \in M_m$ by

$$\begin{aligned}
A(i)e_S &:= e_{S\cup\{(t-|S|,i)\}},\ \text{for } 0\leq|S|\leq t-s-1,\ S\in\mathcal{S}, \\
A(i)e_S &:= \sum_{\substack{S'\in\mathcal{S}'\\|S'|=s-1}} \frac{\widehat{p}(S'\cup S\cup\{(s,i)\})}{\sqrt{\mathrm{Inf}_{s,i}(p)}} f_{S'},\ \text{for } |S| = t-s,\ S\in\mathcal{S}, \\
A(i)f_{S'} &:= \delta_{(|S'|,i)\in\mathcal{S}'} f_{S'-\{(|S'|,i)\}},\ S'\in\mathcal{S}'.
\end{aligned}$$

We claim that $(f_\emptyset, e_\emptyset, A(i))$ satisfies Eq. (5.21). This is because the first applications of the $A(i)$'s act like a *creation* operator and the last as *annihilation* operators. The first $t-s-1$ of the matrices on $e_\emptyset$ create a vector that stores the indices of these

first $t-s-1$ applications, namely

$$A(s+1)\dots A(t)e_{\emptyset} = e_{\{((s+1,i_{s+1}),\dots,(t,i_d))\}}.$$

The $t-s$ application has a unique behavior, as it maps the previous vector to a superposition of $f_{\cdot}$ vectors, namely

$$A(i_s)e_{\{((s+1,i_{s+1}),\dots,(t,i_d))\}} = \sum_{\substack{S'\in\mathcal{S}'\\|S'|=s-1}} \frac{\widehat{p}(S'\cup((s,i_s),\dots,(t,i_d)))}{\sqrt{\mathrm{Inf}_{s,i}(p)}} f_{S'}.$$

Finally, the last $s-1$ applications of the matrices act like annihilation operators, meaning that

$$A(i_1)\dots A(i_{s-1})f_{S'} = \delta_{S',((1,i_1),\dots,(s-1,i_{s-1}))}f_{\emptyset}.$$

Putting everything together we conclude that indeed Eq. (5.21) is satisfied.

Finally, we claim that $A(i)$ are contractions. Given that $\{e_S:\ 0\le|S|\le t-s-1,\ S\in\mathcal{S}\}$, $\{e_S:\ |S|=t-s,\ S\in\mathcal{S}\}$ and $\{f_{S'}:\ S'\in\mathcal{S}'\}$ are mapped to orthogonal spaces, we just have to check than when $A(i)$ is a contraction when it is restricted to the span of each of these 3 sets. For the first and third sets of vectors that is clear. For the second is true because for any $\lambda\in[n]^{t-s}$

$$\begin{aligned}
\|A(i)\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\lambda_S e_S\| &= \|\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\sum_{\substack{S'\in\mathcal{S}'\\|S'|=s-1}}\frac{\widehat{p}(S'\cup S\cup\{(s,i)\})}{\sqrt{\mathrm{Inf}_{s,i}[p]}}\lambda_S f_{S'}\| \\
&= \sqrt{\frac{\sum_{\substack{S'\in\mathcal{S}'\\|S'|=s-1}}\left(\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\widehat{p}(S'\cup S\cup\{(s,i)\})\lambda_S\right)^2}{\mathrm{Inf}_{s,i}[p]}} \\
&\le \sqrt{\frac{\sum_{\substack{S'\in\mathcal{S}'\\|S'|=s-1}}\left(\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\widehat{p}(S'\cup S\cup\{(s,i)\})^2\right)\left(\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\lambda_S^2\right)}{\mathrm{Inf}_{s,i}[p]}} \\
&= \sqrt{\frac{\mathrm{Inf}_{s,i}[p]}{\mathrm{Inf}_{s,i}[p]}}\sqrt{\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\lambda_S^2} \\
&= \|\sum_{\substack{S\in\mathcal{S}\\|S|=t-s}}\lambda_S e_S\|,
\end{aligned}$$

where in the inequality we have used Cauchy-Schwarz. □

Proof of the general case of Theorem 5.8. Let $p : \{-1,1\}^{n\times t} \to \mathbb{R}$ be a block-multilinear degree t polynomial. For every $s \in [t]$, let $p_{=s}$ be its degree s part. Let $D \in [t]$ be such that $\mathrm{Var}[p_{=D}] \geq \mathrm{Var}[p]/t$, which exists because $\mathrm{Var}[p] = \sum_{s\in[t]} \mathrm{Var}[p_{=s}]$. We will now divide the proof in two parts. One is showing that

$$\|p_{=D}\|_{\mathrm{cb}} \leq \|p\|_{\mathrm{cb}}, \tag{5.22}$$

and the other is proving that

$$\mathrm{MaxInf}(p_{=D}) \geq \left(\frac{\mathrm{Var}[p_{=D}]}{\|p_{=D}\|_{\mathrm{cb}}}\right)^2. \tag{5.23}$$

Once we had done that, the result will easily follow:

$$\mathrm{MaxInf}(p) \geq \mathrm{MaxInf}(p_{=D}) \geq \left(\frac{\mathrm{Var}[p_{=D}]}{\|p_{=D}\|_{\mathrm{cb}}}\right)^2 \geq \left(\frac{\mathrm{Var}[p]}{t\|p\|_{\mathrm{cb}}}\right)^2,$$

where in the second inequality we have used Eq. (5.23), and in the third we have used Eq. (5.22) and that $\mathrm{Var}[p_{=D}] \geq \mathrm{Var}[p]/t$.

First, we prove Eq. (5.22). Let $B \in B_{t+1}$ be defined by $B := \sum_{s\in[D]} e_s e_{s+1}^{\mathsf{T}}$, where $\{e_s\}_{s\in[D+1]}$ is an orthonormal basis of $\mathbb{R}^{D+1}$. Note that $\langle e_1, B^s e_{D+1}\rangle = \delta_{s,D}$ for all $s \in [t]_0$. Hence,

$$\begin{aligned}
\|p_{=D}\|_{\mathrm{cb}} &= \sup_{\substack{u,v\in S^{m-1},\ A\in(B_m)^n \\ m\in\mathbb{N}}} \sum_{\substack{\mathbf{b}\in[t]^D \\ b_1<\cdots<b_D}} \sum_{\mathbf{i}\in[n]^D} \widehat{p}_{=D}(\{(b_1,i_1),\ldots,(b_D,i_D)\}) \\
&\qquad \cdot \langle u, A_{b_1}(i_1)\ldots A_{b_D}(i_D)v\rangle \\
&= \sup_{\substack{u,v\in S^{m-1},\ A\in(B_m)^n \\ m\in\mathbb{N}}} \sum_{s\in[t]} \sum_{\substack{\mathbf{b}\in[t]^s \\ b_1<\cdots<b_s}} \sum_{\mathbf{i}\in[n]^s} \widehat{p}(\{(b_1,i_1),\ldots,(b_s,i_s)\}) \\
&\qquad \cdot \langle u\otimes e_1, (A_{b_1}(i_1)\otimes B)\ldots(A_{b_s}(i_s)\otimes B)v\otimes e_{D+1}\rangle \\
&\leq \|p\|_{\mathrm{cb}}.
\end{aligned}$$

Second, we prove Eq. (5.23). Let $\mathcal{S} := \{\{(b_1,i_1),\ldots,(b_{D-1},i_{D-1})\} : \ b_s \in [t],\ b_1 < \cdots < b_{D-1},\ i_s \in [n],\ s \in [D-1]\}$. Let $m := 2 + |S|$. Let $\{v, f_\emptyset, f_S : S \in \mathcal{S}\}$ be an

orthonormal basis of $\mathbb{R}^m$. For $b \in [t],\ i \in [n]$, define $A_b(i) \in M_m$ by

$$A_b(i)v := \sum_{\substack{S \in \mathcal{S} \\ |S|=D-1}} \frac{\widehat{p}_{=D}(S \cup \{(b,i)\})}{\sqrt{\mathrm{MaxInf}[p_{=D}]}} f_S,$$
$$A_b(i)f_S := \delta_{(b,i)\in S} f_{S-\{(b,i)\}},\ \text{for } S \in \mathcal{S} \cup \emptyset.$$

$A_b(i)$ are contractions because they map the vectors of an orthonormal basis to orthogonal vectors without increasing their norms. Note that for $b_1 < \cdots < b_D$ and $\mathbf{i} \in [n]^D$ we have that

$$\langle f_\emptyset, A_{b_1}(i_1) \dots A_{b_D}(i_D)v \rangle = \frac{\widehat{p}_{=D}(\{(b_1,i_1),\dots,(b_D,i_D)\})}{\sqrt{\mathrm{MaxInf}[p_{=D}]}}.$$

Thus,

$$\begin{aligned}\|p_{=D}\|_{\mathrm{cb}} &\geq \sum_{\substack{\mathbf{b}\in[t]^D \\ b_1<\cdots<b_D}} \sum_{\mathbf{i}\in[n]^D} \widehat{p}_{=D}(\{(b_1,i_1),\dots,(b_D,i_D)\}) \langle f_\emptyset, p(A_1,\dots,A_d)v\rangle \\ &= \frac{\mathrm{Var}[p_{=D}]}{\sqrt{\mathrm{MaxInf}[p_{=D}]}},\end{aligned}$$

which after rearranging yields Eq. (5.23). □

5.4.2 AA conjecture for homogeneous Fourier completely bounded polynomials

Finally, we prove a new case of the AA conjecture.

Theorem 5.6. *Let $t \in \mathbb{N}$. Let $p : \{-1,1\}^n \to \mathbb{R}$ be a homogeneous polynomial of degree t and with $\|p\|_{\mathrm{fcb},t} \leq 1$. Then, the maximum influence of p is at least* $\mathrm{Var}[p]^2$.

Proof. Let $m := 1 + \binom{n}{0} + \cdots + \binom{n}{t-1}$. Let $\{v, f_\emptyset, f_S : S \subseteq [n],\ 1 \leq |S| \leq t-1\}$ be an orthonormal basis of $\mathbb{R}^m$. Define the matrices $A(i) \in M_m$ as

$$A(i)v := \sum_{\substack{S \ni i \\ |S|=t}} \frac{\widehat{p}(S)}{\sqrt{\mathrm{MaxInf}[p]}} f_{S-\{i\}},$$
$$A(i)f_S := \delta_{S\ni i} f_{S-\{i\}},\ \text{for } S \subseteq [n],\ 0 \leq |S| \leq t-1,$$

for $i \in [n]$ and $A(n+1) := 0$. We claim that $(f_\emptyset, v, A(i))$ has Boolean behavior of degree t. $A(n+1)$ is clearly a contraction. For $i \in [n]$, $A(i)$ is a contraction, as it maps vectors of the orthonormal basis to orthogonal vectors without increasing the norm, because

$$\|A(i)v\|^2 = \sum_{S \ni i} \frac{\widehat{p}(S)^2}{\mathrm{MaxInf}[p]} = \frac{\mathrm{Inf}_i[p]}{\mathrm{MaxInf}[p]} \leq 1.$$

On the other hand, if $S \subseteq [n]$ satisfies $|S| \leq t-1$, then any $\mathbf{i} \in [n+1]^t$ with $S_\mathbf{i} = S$ either has a repeated element of $[n]$ or has an appearance of the index $n+1$, which implies that $\langle f_\emptyset, A(i_1) \dots A(i_d) \rangle = 0 = \widehat{p}(S)$. If $|S| = t$, then any $\mathbf{i} \in [n+1]^t$ with $S_\mathbf{i} = S$ has t different indices in $[n]$ (corresponding to the elements of S), so in that case

$$\langle f_\emptyset, A(i_1) \dots A(i_d)v \rangle = \frac{\widehat{p}(S)}{\sqrt{\mathrm{MaxInf}[p]}}. \tag{5.24}$$

Putting everything together we conclude that $(f_\emptyset, v, A(i))$ has Boolean behavior of degree t, so

$$\begin{aligned} \|p\|_{\mathrm{fcb},t} &\geq \sum_{S \subseteq [n]} \widehat{p}(S) \langle f_\emptyset, A(i_1) \dots A(i_d)v \rangle = \sum_{S \subseteq [n]} \frac{\widehat{p}(S)^2}{\sqrt{\mathrm{MaxInf}[p]}} \\ &= \frac{\mathrm{Var}[p]}{\sqrt{\mathrm{MaxInf}[p]}}, \end{aligned}$$

where in the first equality we have used Eq. (5.24). After rearranging, the above expression yields

$$\mathrm{MaxInf}[p] \geq \left(\frac{\mathrm{Var}[p]}{\|p\|_{\mathrm{fcb},t}} \right)^2.$$

□

Remark 5.20. Sadly, we could not extend the proof of Theorem 5.6 to the general case. Now, we aim to illustrate what would go wrong with our technique.

For example, consider a polynomial $p : \{-1,1\}^3 \to \mathbb{R}$ with $\deg(p) = 1$ and $\|p\|_{\mathrm{fcb},3} \leq 1$. Ideally, we would want to define unit vectors u and v and contractions $A(i)$ such that for every $S \subseteq [3]$ and every $\mathbf{i} \in [\mathbf{i}^S]$ they satisfied

$$\langle u, A(i_1)A(i_2)A(i_3)v \rangle = \frac{\widehat{p}(S)}{\sqrt{\mathrm{MaxInf}[p]}}. \tag{5.25}$$

If we emulated the strategy of the proof of Theorem 5.6, then $A(1)v$ should be a *normalized* superposition of orthogonal vectors whose amplitudes are all possible $\widehat{p}(S_\mathbf{i})$

that have $i_3 = 1$. In particular, all $\widehat{p}(S)$ with $|S| = 1$ must be included among these amplitudes, because if $S = \{i\}$, then $S = S_{(i,1,1)}$. Hence, the *normalizing* factor of $A(1)v$ should be $\sqrt{\operatorname{Var} p}$, instead of $\sqrt{\operatorname{MaxInf}(p)}$. Note that this extra normalization comes from the fact that given that given (i_1, i_2, i_3), it may happen that $i_3 \notin S_{(i_1,i_2,i_3)}$ and $\widehat{p}(S_{(i_1,i_2,i_3)}) \neq 0$, because p is not homogeneous of degree-3. If we mimic the rest of the proof after this first step that we were forced to modify, we would reach

$$\langle u, A(i_1) \dots A(i_3) v \rangle = \frac{\widehat{p}(S)}{\sqrt{\operatorname{Var}[p]}}$$

instead of Eq. (5.25), which would lead to $\|p\|_{\mathrm{fcb},3} \geq \sqrt{\operatorname{Var} p}$, that is trivially true, because $\|p\|_{\mathrm{fcb},3} \geq \|p\|_\infty$ and $\|p\|_\infty \geq \sqrt{\operatorname{Var} p}$.

Remark 5.21. However, there might be a different way of, given a polynomial p of degree at most t, choosing (u, v, A) with Boolean behavior of degree t such that

$$\langle u, A(i_1) \dots A(i_d) v \rangle = \frac{\widehat{p}(S_{\mathbf{i}})}{\operatorname{poly}(t, \operatorname{MaxInf}[p])},$$

for any $\mathbf{i} \in [n+1]^t$. If that was true, one could copy and paste the proof of Theorem 5.6 and conclude Conjecture 5.5.

This reduces Conjecture 5.5 to a question with flavor of tensor networks (see [CPGSV21] for an introduction to the topic). In particular, the central questions in matrix product states theory is, given a t-tensor $T \in \mathbb{C}^{n \times \cdots \times n}$, to find matrices $A_1, \dots, A_t$ of low dimension such that $T_{\mathbf{i}} = \operatorname{Tr}[A(i_1) \dots A(i_t)]$ for every $\mathbf{i} \in [n]^t$. Thus, we are asking the same question, but with a different goal: to minimize the operator norm of the matrices, instead of their dimensions.

It also has the flavor of almost-quantum correlations [NGHA15]. Almost-quantum correlations are a model for multiparty quantum mechanics that eludes tensor products and commutativity of the observables: it only imposes the commutativity on the correlations. For example, in a bipartite scenario, valid correlations would be those determined by observables $\{A_x\}_{x \in \mathcal{X}}$ and $\{B_y\}_{y \in \mathcal{Y}}$ and a state $|\psi\rangle$ such that

$$\langle \psi | A_x B_y | \psi \rangle = \langle \psi | B_y A_x | \psi \rangle, \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}.$$

In other words, almost-quatum correlations impose the commutativity conditions with respect to the *sandwiches* with $|\psi\rangle$, instead of directly imposing them to the observables, Similarly, we would like to find matrices that satisfy certain Boolean relations with respect to the product with two vectors u and v.

5.4. Aaronson and Ambainis conjecture for (Fourier) completely bounded polynomials

Question 5.22. *Given a polynomial p of degree at most t, is there $(u, v, A) \in \mathscr{B}\mathscr{B}^t$ such that*

$$\langle u, A(i_1) \dots A(i_d) v \rangle = \frac{\widehat{p}(S_{\mathbf{i}})}{\mathrm{poly}(t, \mathrm{MaxInf}[p])},$$

for any $\mathbf{i} \in [n+1]^t$?

Part II

Quantum learning theory

Chapter 6

Bohnenblust-Hille inequalities and their applications to learning theory

6.1 Introduction

The Bohnenblust-Hille inequality states that for any $d \in \mathbb{N}$ there exists a constant C_d such that every d-homogeneous polynomial $P : \mathbb{C}^n \to \mathbb{C}$, defined as $P(z) = \sum_{|\alpha|=d} a_\alpha z^\alpha$, satisfies the following inequality:

$$\|\widehat{P}\|_{\frac{2d}{d+1}} \leq C_d \|P\|_\infty, \tag{6.1}$$

where $\|\widehat{P}\|_{\frac{2d}{d+1}}$ denotes the $\ell_{\frac{2d}{d+1}}$ sum of the coefficients $(a_\alpha)_\alpha$ and $\|P\|_\infty = \sup_{z\in\mathbb{D}^n} |P(z)|$ is the infinity norm of P [BH31].

This inequality, which generalizes the well-known Littlewood's 4/3-Inequality [Lit30], has proven to be extremely useful in the study of the convergence of Dirichlet series and was crucial in determining the asymptotic behaviors of Bohr's radius obtained in [DFOC$^+$11]. In this regard, the authors demonstrated that the constant C_d can be taken equal to C^d, for a certain constant C. The work in [DFOC$^+$11] motivated numerous subsequent studies, where the search focused on the upper and lower bounds for C_d. The best known upper bound was given in [BPSS14], where it was proved that C_d can be actually taken to be $C^{\sqrt{d \log d}}$.

Interestingly, in recent years, the Bohnenblust-Hille inequality has proven to be useful in learning theory. This is perfectly illustrated in the striking work [EI22], where the authors used a version of the Bohnenblust-Hille inequality for functions defined on the hypercube $f : \{-1, 1\}^n \to \mathbb{R}$, as established in [DMP19], to enhance the seminal Low-Degree Algorithm of Linial, Mansour, and Nisan [LMN93]. After that, the applications of the Bohnenblust-Hille inequality to learning theory have reached the quantum computing community, motivating the study of this inequality in the non-commutative realm ([HCP23a, RWZ24, VZ23]). In particular, in the work [VZ23], a version of the Bohnenblust-Hille inequality is proven for $N \times N$-dimensional matrices, which can be understood as a generalization of quantum Boolean functions [MO08].

In this chapter, we explore Bohnenblust-Hille inequalities from three different angles: considering the completely bounded norm instead of the infinity norm, extending the non-commutative variant proved in [VZ23], and determining the exact constants for the case of Boolean functions.

The completely bounded Bohnenblust-Hille inequality

The completely bounded norm of a d-homogeneous polynomial P as above is defined as

$$\|P\|_{\mathrm{cb}} = \sup \Big\| \sum_{|\alpha|=d} a_\alpha Z_1^{\alpha_1} \cdots Z_n^{\alpha_n} \Big\|_{\mathrm{op}},$$

where this supremum runs over all $m \in \mathbb{N}$ and all contractions $Z_1, \ldots, Z_n$ in $M_m(\mathbb{C})$. This norm can be understood as a non-commutative version of the infinity norm and it clearly provides an upper bound for it. Thus, one might expect that the corresponding Bohnenblust-Hille inequality involves a better constant than in the classical case. On the other hand, note that by the triangle inequality, we have that $\|P\|_{\mathrm{cb}} \leq \sum_\alpha |a_\alpha|$. Simultaneously, the completely bounded norm has proven particularly suitable in the study of quantum algorithms, providing a notion of polynomial degree that gives a tight characterization of quantum query complexity (see Chapter 5) [ABP19]. Hence, a Bohnenblust-Hille inequality for the completely bounded norm is also motivated by its potential applications in quantum learning theory. The results of this chapter rigorously fulfill these expectations. Indeed, the main result of this chapter is that the Bohnenblust-Hille inequality holds with the optimal constant $C = 1$ when the infinity norm is replaced by the completely bounded norm. Additionally, we demonstrate that the exponent $2d/(d+1)$ is also optimal in the new scenario considered here, meaning that for $p < 2d/(d+1)$ there is no quantity C_d independent of n such that

$\|\widehat{P}\|_{2d/(d+1)} \leq C_d\|P\|_{\mathrm{cb}}$ for every d-homogeneous polynomial P, as it happens with the original Bohnenblust-Hille inequality Eq. (6.1).

Theorem 6.1. *For every d-homogeneous polynomial $P : \mathbb{C}^n \to \mathbb{C}$, defined as $P(z) = \sum_{|\alpha|=d} a_\alpha z^\alpha$, the following inequality is satisfied:*

$$\|\widehat{P}\|_{\frac{2d}{d+1}} \leq \|P\|_{\mathrm{cb}}.$$

Moreover, both the constant 1 in the inequalities and the exponent $\frac{2d}{d+1}$ are optimal.

In particular, our main result holds for multilinear forms. Moreover, we will also show that in the case of general (non-necessarily homogeneous) polynomials of degree d we have

$$\|\widehat{P}\|_{\frac{2d}{d+1}} \leq \sqrt{d+1}\|P\|_{\mathrm{cb}}.$$

The optimality of Theorem 6.1 shows that the completely bounded norm fits perfectly into the study of the Bohnenblust-Hille inequality. In fact, Theorem 6.1 motivates the study of the optimality of the Bohnenblust-Hille inequality from an angle not explored to date. Rather than focusing on determining the optimal constant that satisfies the inequality (6.1), it is possible to examine the norms that satisfy the associated Bohnenblust-Hille inequality with a constant value of one. It is plausible that the second problem sheds light on the first; particularly, in the problem of finding new lower bounds for the constant C_d. Indeed, in order to find good lower bounds for the classical BH inequality, we must consider polynomials for which the infinity norm is very different from any norm for which a BH inequality with constant 1 can be proven.

Theorem 6.1 entails interesting consequences in learning theory. In particular, it allows us to improve the estimates in [EI22] when we restrict ourselves to certain functions arising in quantum computing. Indeed, in that work, it is proven that it is possible to learn any bounded function $f : \{-1,1\}^n \to [-1,1]$ of degree at most d with L_2-accuracy ε and confidence $1-\delta$ by using $O(\varepsilon^{-2(d+1)}C^{d^{3/2}\sqrt{\log d}}\log(n/\delta))$ uniformly random samples on the function. A particularly interesting type of these functions are those that arise from a quantum algorithm with d queries. More precisely, we consider here quantum query algorithms that prepare a state

$$|\psi_x\rangle = U_d(O_{x_d} \otimes \mathrm{Id}_m)U_{d-1}\cdots U_1(O_{x_1} \otimes \mathrm{Id}_m)U_0|\psi_0\rangle, \tag{6.2}$$

where m is an integer, x stands for $(x_1,\ldots,x_d)$, O_y is the n-dimensional matrix that maps $|i\rangle$ to $y_i|i\rangle$, $U_1,\ldots,U_d$ are $(n+m)$-dimensional unitaries and $|\psi_0\rangle$ is an

$(n+m)$-dimensional unit vector. The algorithm succeeds according to a projective measurement that measures the projection of the final state onto some fixed $(n+m)$-dimensional unit vector $|v\rangle$. Hence, the amplitude of $|v\rangle$ is given by $T(x) = \langle v|\psi_x\rangle$, so that $|T(x)|^2$ is the acceptance probability of the algorithm. These quantum algorithms have been considered in the quantum computing literature; for example, k-fold forrelation, that witnesses the biggest possible quantum-classical separation, has this structure [AA15]. As we will explain in Section 6.4, the argument in [EI22] alongside the Bohnenblust-Hille inequality for (bounded) multilinear forms [BPSS14] imply that the amplitudes T can be learned from $O(\varepsilon^{-2(d+1)}\mathrm{poly}(d)^d \log(n/\delta))$ samples. Furthermore, using Theorem 6.1 instead of [BPSS14] allows us to obtain the following result for learning d-query quantum algorithms which, in particular, requires a number of samples that is polynomial in n when ε and δ are constants and $d = \log(n)$.

Corollary 6.2. *Consider a quantum algorithm that makes d queries as explained above. Then, its amplitudes can be learned with L_2^2-accuracy ε and confidence $1-\delta$ from $O(\varepsilon^{-2(d+1)} d^2 \log(n/\delta))$ uniform random samples.*

Extending the non-commumative Bohnenblust-Hille inequality.

Motivated by the applications to learning quantum channels, we extend the non-commutative version of the BH inequality proved in [VZ23]. This generalization concerns the Pauli coefficients of linear maps $\Phi : M_N \to M_N$, where let $N = 2^n$ and n is a natural number. These maps can be expressed as

$$\Phi(\rho) = \sum_{x,y\in\{0,1,2,3\}^n} \widehat{\Phi}(x,y)\cdot \sigma_x \rho \sigma_y, \tag{6.3}$$

where $\sigma_x = \otimes_{i\in[n]} \sigma_{x_i}$ and σ_i for $i \in \{0,1,2,3\}$ are the Pauli matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix};$$

and $\widehat{\Phi}(x,y)$ are the Pauli coefficients of the map. Given $x \in \{0,1,2,3\}^n$, $|x|$ is the number of non-zero entries of x. The degree of Φ is the minimum integer d such that $\widehat{\Phi}(x,y) = 0$ if $|x| + |y| > d$.

We show that we can upper bound the $\ell_{2d/(d+1)}$-sum of the Pauli coefficients of $\widehat{\Phi}$ in a Bohnenblust-Hille way.

Theorem 6.3. *Let $\Phi : M_N \to M_N$ be a linear map of degree d. Then,*

$$\|\widehat{\Phi}\|_{\frac{2d}{d+1}} \leq C^d \|\Phi : S_1^N \to S_\infty^N\|,$$

where C is a universal constant and S_1^N and S_∞^N denote the spaces of one and infinity Schatten classes respectively.

The proof of Theorem 6.3 follows a similar approach to the one in [VZ23] and, in fact, extends their result. Indeed, if one considers a matrix $A \in M_N$, the main result in [VZ23] follows from the application of Theorem 6.3 to the linear map $\Phi(X) = XA$.

We use our extension to improve the current results on learning quantum channels. From a physical perspective, quantum channels describe the transformations between quantum systems. Since quantum systems are represented by quantum states, which correspond to non-commutative probability distributions, specifically positive semidefinite matrices with trace 1, quantum channels map one set of non-commutative probabilities to another. Mathematically, quantum channels on n-qubits are maps $\Phi : M_N \to M_N$ that are completely positive and trace-preserving. In particular, they satisfy $\|\Phi : S_1^N \to S_\infty^N\| \leq 1$ and Theorem 6.3 applies to them. Learning an n-qubit quantum channel is in general challenging and is known to require $\Theta(4^n)$ applications (queries) of the channel [GJ14]. This exponential complexity can be drastically improved when prior information on the structure of the channel is available. For example, a recent work of Bao and Yao [BY23] considered k-junta quantum channels, i.e., n-qubit channels that act non-trivially only on at most k of the n (unknown) qubits leaving the rest of qubits unchanged. These channels were shown to be learnable using $\widetilde{\Theta}(4^k)$ queries to the channel [BY23].

Using the same learning model as the recent work of Bao and Yao (see Section 6.4 for details) we prove the following result for learning low-degree channels, which contrary to the other applications of BH inequality in quantum learning theory, it has a query complexity independent of n [HCP23a, SVZ23a, SVZ23b, VZ23].

Theorem 6.4. *Let Φ be a n-qubit degree-d quantum channel. Then it can be learned in L_2-accuracy ε and confidence $\geq 1 - \delta$ by making $\exp\left(\widetilde{O}(d^2 + d\log(1/\varepsilon))\right) \cdot \log(1/\delta)$ queries to Φ. Here, we use the notation $\widetilde{O}$ to hide logarithmic factors in d, $1/\varepsilon$, and $1/\delta$.*

Boolean functions

Since boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ are particularly important in many contexts, we also analyze this case. Remember that the classical Fourier expansion in the hypercube allows one to write any function as

$$f = \sum_{s\in\{0,1\}^n} \widehat{f}(s)\chi_s, \tag{6.4}$$

where $\chi_s(x) = \prod_{i\in\mathrm{supp}(s)} x_i$ for $s \in \{0,1\}^n$ and $(\widehat{f}(s))_s$ are the Fourier coefficients of f. Then, the degree of f is the minimum d such that $\widehat{f}(s) = 0$ if $|s| > d$.

In this chapter, we show how the granularity property of these functions allows us to prove the corresponding optimal Bohnenblust-Hille inequality.

Proposition 6.5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a function of degree at most d. Then,*

$$\Big(\sum_{s\in\{0,1\}^n} |\widehat{f}(s)|^{\frac{2d}{d+1}}\Big)^{\frac{d+1}{2d}} \leq 2^{\frac{d-1}{d}}.$$

The equality is witnessed by the address function.

The previous proposition might be of interest in functional analysis for two reasons. First, it is conjectured that the value of the BH constant for real d-linear forms is $2^{\frac{d-1}{d}}$ [PT18], so this fact proves the conjecture for the particular case of d-linear Boolean forms. Second, the address function, that saturates the inequality, is a d-linear form that gives a lower bound for the BH constant for multilinear forms of $2^{\frac{d-1}{d}}$, which matches the best lower bound known so far for the BH inequality for real multilinear forms [DMFPSS14]. Together with Proposition 6.5 about Boolean functions, in this chapter we also study the complexity of these functions from the learning theoretical point if view and improve previous estimates in [NPVY23, Corollary 34] and [EIS22, Corollary 4] (see Section 6.3.1 for details).

6.2 Bohnenblust-Hille Inequality for the completely bounded norm

In this section we will prove those results concerning the Bohnenblust-Hille Inequality for the completely bounded norm. We will first prove a general result for tensors, from where Theorem 6.1, as well as some other results will follow straightforwardly.

cb-BH inequality for d-tensors

We consider $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and a d-tensor $T = (T_{\mathbf{i}})_{\mathbf{i}\in[n]^d} \in \mathbb{K}^n \times \cdots \times \mathbb{K}^n$. Equivalently, T can be regarded as the d-linear form $T : \mathbb{K}^n \times \cdots \times \mathbb{K}^n \to \mathbb{K}$ given by

$$T(z_1, \ldots, z_d) = \sum_{\mathbf{i}\in[n]^d} T_{\mathbf{i}} z_1(i_1) \ldots z_d(i_d).$$

For one such tensor, we denote

$$\|\widehat{T}\|_{\frac{2d}{d+1}} := \left(\sum_{\mathbf{i}\in[n]^d} |T_{\mathbf{i}}|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}}, \tag{6.5}$$

The main result of the section is the following cb-BH inequality for d-tensors.

Theorem 6.6. *Let $T \in \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ be a d-tensor. Then,*

$$\|\widehat{T}\|_{\frac{2d}{d+1}} \leq \|T\|_{\mathrm{cb}}.$$

We will make use of the following lemma, originally due to Blei [Ble79]. A simple proof can be found in [BPSS14, Theorem 2.1].

Lemma 6.7 (Blei's inequality)**.** *Given a d-tensor $T \in \mathbb{K}^n \times \cdots \times \mathbb{K}^n$, we have*

$$\|\widehat{T}\|_{\frac{2d}{d+1}} \leq \left(\prod_{s\in[d]} \sum_{i_s\in[n]} \sqrt{\sum_{i_1,\ldots,i_{s-1},i_{s+1},\ldots,i_d\in[n]} |T_{\mathbf{i}}|^2} \right)^{\frac{1}{d}}.$$

Now, we prove the key technical lemma, from where Theorem 6.6 will follow.

Lemma 6.8. *Let $T \in \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ be a d-tensor and $s \in [d]$. Then,*

$$\sum_{i_s\in[n]} \sqrt{\sum_{i_1,\ldots,i_{s-1},i_{s+1},\ldots,i_d\in[n]} |T_{i_1,\ldots,i_{s-1},i_s,i_{s+1},\ldots,i_d}|^2} \leq \|T\|_{\mathrm{cb}}.$$

Proof. We fix $s \in [d]$. The proof consists of evaluating T on an explicit set of contractions. In order to define these contractions, we denote $m = \sum_{r=0}^{d-s} n^r + \sum_{r=0}^{s-1} n^r$ and let $\{e_{\mathbf{i}}, f_{\mathbf{j}} : \mathbf{i} \in [n]^r, r \in \{0\} \cup [d-s], \mathbf{j} \in [n]^t, t \in \{0\} \cup [s-1]\}$ be an orthonormal basis of $\ell_2^m(\mathbb{K})$, where we identify $[n]^0$ with $\emptyset$. For every $i \in [n]$ we define the matrix

$Z_i \in M_m$ as:

$$\begin{aligned}
Z_i e_{\mathbf{j}} &= e_{(i,\mathbf{j})}, \text{ if } \mathbf{j} \in [n]^r,\ r \in \{0\} \cup [d-s-1], \\
Z_i e_{\mathbf{j}} &= \frac{\sum_{\mathbf{k}\in[n]^{s-1}} T^*_{(\mathbf{k},i,\mathbf{j})} f_{\mathbf{k}}}{\sqrt{\sum_{k_1,\ldots,k_{s-1},k_{s+1},\ldots,k_d\in[n]} |T_{(k_1,\ldots,k_{s-1},i,k_{s+1},\ldots,k_d)}|^2}}, \text{ if } \mathbf{j} \in [n]^{d-s}, \\
Z_i f_{\mathbf{j}} &= \delta_{i,j_t} f_{(j_1,\ldots,j_{t-1})}, \text{ if } \mathbf{j} \in [n]^t,\ t \in \{0\} \cup [s-1], \\
Z_i f_{\emptyset} &= 0.
\end{aligned}$$

Assume for the moment that Z_i are contractions. One can easily check that

$$\langle f_\emptyset, Z_{i_1} \ldots Z_{i_d} e_\emptyset \rangle = \frac{T^*_{i_1,\cdots,i_d}}{\sqrt{\sum_{k_1,\ldots,k_{s-1},k_{s+1},\ldots,k_d\in[n]} |T_{k_1,\ldots,k_{s-1},i_s,k_{s+1},\ldots,k_d}|^2}}.$$

Hence, by assuming that Z_i are contractions, we can conclude

$$\begin{aligned}
\|T\|_{\mathrm{cb}} &\geq \| \sum_{\mathbf{i}\in[n]^d} T_{\mathbf{i}} Z_{i_1} \ldots Z_{i_d} \|_{B(\ell_2^m(\mathbb{K}))} \geq \sum_{\mathbf{i}\in[n]^d} T_{\mathbf{i}} \langle f_\emptyset | Z_{i_1} \ldots Z_{i_d} | e_\emptyset \rangle \\
&\geq \sum_{\mathbf{i}\in[n]^d} T_{\mathbf{i}} \frac{T^*_{\mathbf{i}}}{\sqrt{\sum_{k_1,\ldots,k_{s-1},k_{s+1},\ldots,k_d\in[n]} |T_{k_1,\ldots,k_{s-1},i_s,k_{s+1},\ldots,k_d}|^2}} \\
&= \sum_{i_s\in[n]} \sqrt{\sum_{i_1,\ldots,i_{s-1},i_{s+1},\ldots,i_d\in[n]} |T_{i_1,\ldots,i_{s-1},i_s,i_{s+1},\ldots,i_d}|^2},
\end{aligned}$$

as desired.

Thus, it remains to prove that the matrices Z_i are contractions. Given that Z_i maps the sets $\{e_{\mathbf{i}} : \mathbf{i} \in [n]^r,\ r \in \{0\} \cup [d-s-1]\}$, $\{e_{\mathbf{i}} : \mathbf{i} \in [n]^{d-s}\}$ and $\{f_{\mathbf{i}} : \mathbf{i} \in [n]^t,\ t \in \{0\} \cup [s-1]\}$ to orthogonal subspaces, it suffices to show that the Z_i are contractions when restricted to those subspaces. For the first and third sets that is clear since Z_i maps each basis vector of those sets either to a different basis vector or

to 0. For the second set, just note that for every $\lambda \in \mathbb{K}^{n^{d-s}}$ we have

$$\begin{aligned}\|Z_i \sum_{\mathbf{j}\in[n]^{d-s}} \lambda_{\mathbf{j}} e_{\mathbf{j}}\|_2^2 &= \|\frac{\sum_{\mathbf{k}\in[n]^{s-1}} \Big(\sum_{\mathbf{j}\in[n]^{d-s}} \lambda_{\mathbf{j}} T^*_{\mathbf{k}i\mathbf{j}}\Big) f_{\mathbf{k}}}{\sqrt{\sum_{k_1,\ldots,k_{s-1},k_{s+1},\ldots,k_d\in[n]} |T_{k_1,\ldots,k_{s-1},i,k_{s+1},\ldots,k_d}|^2}}\|_2^2 \\ &= \frac{\sum_{\mathbf{k}\in[n]^{s-1}} |\sum_{\mathbf{j}\in[n]^{d-s}} \lambda_{\mathbf{j}} T^*_{\mathbf{k}i\mathbf{j}}|^2}{\sum_{k_1,\ldots,k_{s-1},k_{s+1},\ldots,k_d\in[n]} |T_{k_1,\ldots,k_{s-1},i,k_{s+1},\ldots,k_d}|^2} \\ &\leq \frac{\Big(\sum_{\mathbf{k}\in[n]^{s-1}} \sum_{\mathbf{j}\in[n]^{d-s}} |T_{\mathbf{k}i\mathbf{j}}|^2\Big)\Big(\sum_{\mathbf{j}\in[n]^{d-s}} |\lambda_{\mathbf{j}}|^2\Big)}{\sum_{k_1,\ldots,k_{s-1},k_{s+1},\ldots,k_d\in[n]} |T_{k_1,\ldots,k_{s-1},i,k_{s+1},\ldots,k_d}|^2} \\ &= \sum_{\mathbf{j}\in[n]^{d-s}} |\lambda_{\mathbf{j}}|^2 = \|\lambda\|_2^2,\end{aligned}$$

where we have used Cauchy-Schwarz for the sum over $\mathbf{j}$. □

Proof of Theorem 6.6. According to Lemma 6.7 and Lemma 6.8 we have

$$\|\widehat{T}\|_{\frac{2d}{d+1}} \leq \left(\prod_{s\in[d]} \sum_{i_s\in[n]} \sqrt{\sum_{i_1,\ldots,i_{s-1},i_{s+1},\ldots,i_d\in[n]} |T_{\mathbf{i}}|^2}\right)^{\frac{1}{d}} \leq \left(\prod_{s\in[d]} \|T\|_{\mathrm{cb}}\right)^{\frac{1}{d}} = \|T\|_{\mathrm{cb}}.$$

□

cb-BH inequality for polynomials

Now we consider the case of polynomials. Given any (not necessarily homogeneous) polynomial of degree d in n variables $P : \mathbb{K}^n \to \mathbb{K}$, we can write it as

$$P = \sum_{s\in\{0\}\cup[d]} P_s, \tag{6.6}$$

where $P_s : \mathbb{K}^n \to \mathbb{K}$ is a s-homogeneous polynomial. We denote, given $s \in [d]$,

$$\mathcal{J}(s,n) = \{(j_1,\ldots,j_s) \in [n]^s : j_1 \leq \cdots \leq j_s\}.$$

Then, P_s can be written uniquely as

$$P_s(x) = \sum_{\mathbf{j}\in\mathcal{J}(s,n)} a_{\mathbf{j}} x_{\mathbf{j}}, \tag{6.7}$$

where we denote $x_{\mathbf{j}} = x_{j_1} \cdots x_{j_s}$. Hence, we can define the completely bounded norm of P as

$$\|P\|_{\mathrm{cb}} = \sup \Big\| \sum_{s \in \{0\} \cap [d]} \sum_{\mathbf{j} \in \mathcal{J}(s,n)} a_{\mathbf{j}} Z_{j_1} \cdots Z_{j_s} \Big\|_{\mathrm{op}},$$

where the supremum runs over all (real/complex) contractions of M_m and $m \in \mathbb{N}$.

Theorem 6.1, which refers to the d-homogeneous case, follows easily from Theorem 6.6.

Proof of Theorem 6.1. Given a d-homogeneous polynomial $P : \mathbb{K}^n \to \mathbb{K}$ as above, we want to prove that

$$\Big(\sum_{\mathbf{j} \in \mathcal{J}(d,n)} |a_{\mathbf{j}}|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} \leq \|P\|_{\mathrm{cb}}.$$

To do that, we reduce it to the case of tensors. We define $T_{\mathbf{j}} = a_{\mathbf{j}}$ for every $\mathbf{j} \in \mathcal{J}(d,n)$ and $T_{\mathbf{j}} = 0$ for ever $\mathbf{j} \in [n]^d \setminus \mathcal{J}(d,n)$. By Proposition 2.18, the tensor T satisfies

$$\Big(\sum_{\mathbf{j} \in \mathcal{J}(d,n)} |a_{\mathbf{j}}|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} = \Big(\sum_{\mathbf{j} \in [n]^d} |T_{\mathbf{j}}|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} \quad \text{and} \quad \|T\|_{\mathrm{cb}} = \|P\|_{\mathrm{cb}}.$$

Hence, the result follows from Theorem 6.6. □

We will now turn our attention to the case of general polynomials. To this end, we first prove the following result:

Lemma 6.9. *Let $P : \mathbb{K}^n \to \mathbb{K}$ be a polynomial of degree d. Then,*

$$\|P\|_{\mathrm{cb}} \geq \frac{1}{\sqrt{d+1}} \sup \sum_{s \in [d] \cup \{0\}} \Big| \Big\langle u, \sum_{\substack{\alpha \in \mathbb{N}_0^n \\ \sum_i \alpha_i = s}} a_\alpha Z_1^{\alpha_1} \dots Z_n^{\alpha_n} v_s \Big\rangle \Big|,$$

where the supremum runs over all (real/complex) contractions $Z_1, \dots, Z_n$ in M_m, all m-dimensional vectors u, v_s with norm less than or equal one, and all $m \in \mathbb{N}$.

Proof. Let $m \in \mathbb{N}$, $Z_1, \dots, Z_n \in M_m$ be contractions and u, v_s be m-dimensional vectors with norm less than or equal one. For $s \in \{0\} \cup [d]$, let $b_s \in \mathbb{K}$ be such that $|b_s| = 1$ and

$$\Big| \Big\langle u, \sum_{\substack{\alpha \in \mathbb{N}_0^n \\ \sum_i \alpha_i = s}} a_\alpha Z_1^{\alpha_1} \dots Z_n^{\alpha_n} v_s \Big\rangle \Big| = b_s \Big\langle u, \sum_{\substack{\alpha \in \mathbb{N}_0^n \\ \sum_i \alpha_i = s}} a_\alpha Z_1^{\alpha_1} \dots Z_n^{\alpha_n} v_s \Big\rangle.$$

Let $\{e_0, \dots, e_d\}$ be the canonical basis of $\mathbb{K}^{d+1}$. We define the unitary operator $B : \mathbb{K}^{d+1} \to \mathbb{K}^{d+1}$ such that $B(e_{p+1}) = e_p$ for every $p \in [d]$ and $B(e_0) = e_d$. We also define the unit vectors

$$\xi = \frac{1}{\sqrt{d+1}} \sum_{q \in [d] \cup \{0\}} b_q v_q \otimes e_q \in K^m \otimes \mathbb{K}^{d+1} \quad \text{and} \quad \eta = u \otimes e_0 \in \mathbb{K}^m \otimes \mathbb{K}^{d+1}.$$

Finally, we consider the new contractions $\widetilde{Z}_i = Z_i \otimes B \in M_{m(d+1)}$ for $i = 1, \cdots, n$. Then, one can easily check that

$$\Big\langle \widetilde{u}, \sum_{s \in [d] \cup \{0\}} \sum_{\substack{\alpha \in \mathbb{N}_0^n \\ \sum_i \alpha_i = s}} a_\alpha \widetilde{Z}_1^{\alpha_1} \dots \widetilde{Z}_n^{\alpha_n} \xi \Big\rangle = \frac{1}{\sqrt{d+1}} \Big| \Big\langle u, \sum_{\substack{\alpha \in \mathbb{N}_0^n \\ \sum_i \alpha_i = s}} a_\alpha Z_1^{\alpha_1} \dots Z_n^{\alpha_n} v_s \Big\rangle \Big|,$$

from where the statement follows.

□

We can now prove a cb-BH inequality for general polynomials of degree d.

Corollary 6.10. *Let $P : \mathbb{K}^n \to \mathbb{K}$ be a polynomial of degree d. Then,*

$$\|\widehat{P}\|_{\frac{2d}{d+1}} \leq \sqrt{d+1} \|P\|_{\mathrm{cb}}.$$

Proof. Let $Q : \mathbb{K}^{n+1} \to \mathbb{K}$ be the s-homogeneous polynomial defined by

$$Q(x, x_{n+1}) := \sum_{s \in \{0\} \cup [d]} P_s(x) x_{n+1}^{d-s},$$

where $x = (x_1, \cdots, x_n)$ and P_s is the d-homogeneous part of P.

It is clear that $\|\widehat{Q}\|_{\frac{2d}{d+1}} = \|\widehat{P}\|_{\frac{2d}{d+1}}$. On the other hand, we have

$$\|Q\|_{\mathrm{cb}} = \sup \Big\langle u \Big| \sum_{s \in \{0\} \cap [d]} \sum_{\substack{\alpha \in \mathbb{N}_0^{n+1} \\ \sum_i \alpha_i = s}} a_\alpha Z_1^{\alpha_1} \dots Z_n^{\alpha_n} Z_{n+1}^{d-s} \Big| v \Big\rangle,$$

where the sup is taken over all (real/complex) contractions $Z_1, \dots, Z_{n+1} \in M_m$, all m-dimensional unit vectors u and v and all $m \in \mathbb{N}$. Then, by defining $v_s = Z_{n+1}^{d-s} |v\rangle$ we can use Lemma 6.9 to deduce

$$\|Q\|_{\mathrm{cb}} \leq \sqrt{d+1} \|P\|_{\mathrm{cb}}.$$

Hence, applying Theorem 6.1 to Q concludes the proof. □

Optimality of the cb-BH inequality

We conclude this section by proving the optimality of Theorem 6.1. We will actually prove the optimality of Theorem 6.6, from where the optimality in the exponent for the corresponding cb-BH inequality for d-multilinear forms and d-homogeneous polynomials follows.

First of all, note that constant one is the best possible in the inequality since the d-linear form $T(x_1, \cdots, x_n) = x_1$ satisfies $\|\widehat{T}\|_{2d/(d+1)} = \|T\|_{\mathrm{cb}} = 1$. Regarding the optimality in the exponent, it follows from the next statement.

Theorem 6.11. *Let $d \in \mathbb{N}$, let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $q \geq 1$. For infinitely many $n \in \mathbb{N}$, there exists a d-tensor $T \in \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ such that $\|T\|_q = n^{\frac{d}{q}}$ and $\|T\|_{\mathrm{cb}} \leq n^{\frac{d+1}{2}}$.*

The optimality in the exponent of Theorem 6.6 follows easily from the previous statement. Indeed, suppose that there is a constant $C_d > 0$ such that

$$\|T\|_q \leq C_d \|T\|_{\mathrm{cb}}.$$

Then, it follows that

$$n^{\frac{d}{q}} \leq C_d n^{\frac{d+1}{2}}$$

for every $n \in \mathbb{N}$. Therefore, $q \geq 2d/(d+1)$. In order to see that this last estimate implies the optimality for the BH inequality for d-homogeneous polynomials (Theorem 6.1) just note that for any d-linear form $T : \mathbb{K}^n \times \cdots \mathbb{K}^n \to \mathbb{K}$, we can define a d-homogeneous polynomial in $d \times n$ variables $P : (\mathbb{K}^n)^d \to \mathbb{K}$ defined as

$$P\big((x_1(i_1))_{i_1}, \cdots, (x_d(i_d))_{i_d}\big) = \sum_{i_1, \cdots, i_d = 1}^{n} T_{i_1, \cdots, i_d} x_1(i_1) \cdots x_d(i_d).$$

The optimality of Theorem 6.1 follows because if we consider the lexicographical order in $[n]^d$, then $\|P\|_{2d/(d+1)} = \|T\|_{2d/(d+1)}$ and $\|P\|_{\mathrm{cb}} = \|T\|_{\mathrm{cb}}$.

Our proof is based on the proof of the optimality of the exponent in the classical BH inequality (see [DGMP19, Chapter 4]).

Proof of Theorem 6.11. Let $n \in \mathbb{N}$ and let $N = 2^n$. We will identify $[N]$ with $\mathcal{P}(n)$ (the family of subsets of n elements) and $\{-1, 1\}^n$ in an arbitrary bijective way. In

this sense, we define the matrix $a \in \mathbb{R}^{N\times N}$ via

$$a_{(x,S)} = \prod_{i\in S} x_i,$$

for every $x \in \{-1,1\}^n$ and $S \subseteq [n]$. This matrix satisfies that

$$|a_{(x,S)}| = 1, \tag{6.8}$$

$$\sum_{x\in\{-1,1\}^n} a_{(x,S)}a_{(x,S')} = N\delta_{S,S'}. \tag{6.9}$$

We define the d-tensor $T \in \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ by

$$T = \sum_{\mathbf{i}\in[N]^d} a_{(i_1,i_2)} \dots a_{(i_{d-1},i_d)}.$$

According to Eq. (6.8) we immediately deduce that $\|T\|_q = N^{\frac{d}{q}}$.

In order to prove the upper bound for $\|T\|_{\mathrm{cb}}$ we can restrict to unitary/orthogonal matrices, thanks to Remark 2.16. Now, given arbitrary unitary matrices $U^1_{i_1}, \dots, U^d_{i_d}$, if we denote

$$R_{i_1} = \sum_{\mathbf{j}\in[N]^{d-1}} a_{(i_1,j_2)} \dots a_{(j_{d-1},j_d)} U^2_{j_2} \dots U^d_{j_d},$$

we can apply Lemma 3.9 to write

$$\begin{aligned}\Big\| \sum_{\mathbf{i}\in[N]^d} a_{(i_1,i_2)} \dots a_{(i_{d-1},i_d)} U^1_{j_1} \dots U^d_{j_d} \Big\| &\le \Big\| \sum_{i_1\in[N]} U^1_{i_1}(U^1_{i_1})^\dagger \Big\|^{\frac{1}{2}} \Big\| \sum_{i_1\in[N]} R^\dagger_{i_1} R_{i_1} \Big\|^{\frac{1}{2}} \\ &= N^{\frac{1}{2}} \Big\| \sum_{i_1\in[N]} R^\dagger_{i_1} R_{i_1} \Big\|^{\frac{1}{2}}.\end{aligned}$$

Now, we note that $\sum_{i_1\in[N]} R^\dagger_{i_1} R_{i_1}$ can be written as

$$\sum_{\mathbf{j},\mathbf{k}\in[N]^{d-1}} \Big(\sum_{i_1\in[N]} a_{(i_1,j_2)} a_{(i_1,k_2)} \Big) a_{(j_2,j_3)} \cdots a_{(j_{d-1},j_d)} a_{(k_2,k_3)} \cdots a_{(k_{d-1},k_d)} \cdot (U^d_{j_d})^\dagger \cdots (U^2_{j_2})^\dagger U^2_{k_2} \cdots U^d_{k_d}.$$

By using Eq. (6.9) and that $U_i^2 = \mathrm{Id}$, for $i \in [N]$, the previous expression equals

$$\begin{aligned} & N \sum_{i_2 \in [N]} \sum_{\mathbf{j},\mathbf{k} \in [N]^{d-2}} a_{(i_2,j_3)} \cdots a_{(j_{d-1},j_d)} a_{(i_2,k_3)} \cdots a_{(k_{d-1},k_d)} (U_{j_d}^d)^\dagger \cdots (U_{j_3}^3)^\dagger U_{k_3}^3 \cdots U_{k_d}^d \\ & = N \sum_{\mathbf{j},\mathbf{k} \in [N]^{d-2}} \Big(\sum_{i_2 \in [N]} a_{(i_2,j_3)} a_{(i_2,k_3)} \Big) a_{(j_3,j_4)} \cdots a_{(j_{d-1},j_d)} \cdots a_{(k_3,k_4)} a_{(k_{d-1},k_d)} \\ & \qquad \cdot (U_{j_d}^d)^\dagger \cdots (U_{j_3}^3)^\dagger U_{k_3}^3 \cdots U_{k_d}^d. \end{aligned}$$

We see that we can iterate this process to obtain

$$\Big\| \sum_{i_1 \in [N]} R_{i_1}^\dagger R_{i_1} \Big\| \leq N^{d-1} \Big\| \sum_{i_d \in [N]} (U_{i_d}^d)^\dagger U_{i_d}^d \Big\| = N^d.$$

Therefore, we conclude that $\Big\| \sum_{\mathbf{i} \in [N]^d} a_{(i_1,i_2)} \cdots a_{(i_{d-1},i_d)} U_{i_1}^1 \ldots U_{i_d}^d \Big\| \leq N^{\frac{d+1}{2}}$.

□

Remark 6.12. The d-linear form used in the proof of Theorem 6.11 also plays a central role in quantum query complexity. Indeed, it is the linear form determined by the d-forrelation problem, that optimally separates quantum and classical query complexity and we already introduced in Section 3.2.1 [AA15, BS21]. We recall that, given d Boolean functions $f_1, \ldots, f_d : \{0,1\}^n \to \{-1,1\}$, its d-forrelation is defined as

$$\mathrm{forr_d}(f_1, \ldots, f_d) = \frac{1}{2^{n\frac{d+1}{2}}} \sum_{x_1,\ldots,x_d \in \{0,1\}^n} f(x_1)(-1)^{\langle x_1,x_2 \rangle} \ldots f(x_{d-1})(-1)^{\langle x_{d-1},x_d \rangle} f(x_d).$$

Thus, if we consider the d-linear form T defined in the proof of Theorem 6.11 and we identify the d functions $f_1, \ldots, f_d$ with the elements of $\{-1,1\}^{2^n}$ determined by their truth table, we have

$$T(f_1, \ldots, f_d) = 2^{n\frac{d+1}{2}} \mathrm{forr_d}(f_1, \ldots, f_d).$$

6.3 Bohnenblust-Hille inequality in other contexts

6.3.1 Boolean functions

We determine the exact value of the BH constant for Boolean functions. This result follows from the well-known fact that the Fourier coefficients of Boolean functions are multiples of $2^{1-d}\mathbb{Z}$. This property is usually referred to as the granularity of

Boolean functions [O'D09, Exercise 1.11]. We sketch the proof below for the sake of completeness.

Lemma 6.13. *Let $f : \{-1,1\}^n \to \{-1,1\}$ with degree at most d. Then, $\widehat{f}(s) \in 2^{1-d}\mathbb{Z}$ for every $s \in \{0,1\}^n$.*

Proof. Recall that $f : \{-1,1\}^n \to \{-1,1\}$. We define $g : \{0,1\}^n \to \{0,1\}$ by

$$g(z) = \frac{1}{2}\Big(1 - f((1-2z_1), \ldots, (1-2z_n))\Big).$$

It is not difficult to see that g can be written in a unique way as

$$g(z) = \sum_{s \in \{0,1\}^n} c_s \prod_{i:s_i=1} z_i$$

for some coefficients $c_s \in \mathbb{R}$ such that $c_s = 0$ for every s with $|s| > d$. By applying induction on $|s|$, one can actually prove that $c_s \in \mathbb{Z}$ for every s. Indeed, we first note that for $s = \emptyset$, one has $c_{0^n} = g(0^n) \in \{0,1\}$. For s with $|s| = t+1 > 0$, assuming that $c_s \in \mathbb{Z}$ for every s with $|s| \leq t$, we have

$$c_s = g(s) - \sum_{|s'|<|s|, s'_i \leq s_i} c_{s'},$$

so c_s belongs to $\mathbb{Z}$. Finally, the statement for f can be obtained by just noticing that

$$f(x) = 1 - 2g\left(\frac{1-x_1}{2}, \ldots, \frac{1-x_n}{2}\right) = 1 - 2\sum_{|s|\leq d} c_s \prod_{i:s_i=1} \frac{1-x_i}{2}.$$

□

Proposition 6.14. *Let $f : \{-1,1\}^n \to \{-1,1\}$ of degree at most d. Then,*

$$\Big(\sum_{s\in\{0,1\}^n} |\widehat{f}(s)|^{\frac{2d}{d+1}}\Big)^{\frac{d+1}{2d}} \leq 2^{\frac{d-1}{d}}.$$

The equality is witnessed by the address function.

Proof. Since for Boolean functions one has $\|f\|_2 = 1$, Lemma 6.13 and Parseval's identity imply that f has at most $2^{2(d-1)}$ non-zero Fourier coefficients. Indeed, this immediately follows from the identity $\sum_s |\widehat{f}(s)|^2 = 1$ and the fact that $|\widehat{f}(s)| \geq 2^{1-d}$

for every non-zero coefficient. Hence, Hölder's inequality implies that, for $p \in [1,2)$,

$$\sum_{s:\widehat{f}(s)\neq 0} |\widehat{f}|^p \cdot 1 \leq \Big(\sum_{s\in\{0,1\}^n} \widehat{f}^2(s) \Big)^{\frac{p}{2}} \left(2^{2(d-1)}\right)^{\frac{2-p}{2}} = 2^{(d-1)(2-p)}.$$

Taking $p = 2d/(d+1)$ the claimed inequality follows.

The equality is witnessed by the *address function* $f : (\{-1,1\}^n)^d \to \{-1,1\}$ of degree d and $n = 2^{d-1}$, which is defined as

$$f(x) = \sum_{a\in\{-1,1\}^{d-1}} \underbrace{\frac{x_1(1) - a_1 x_1(2)}{2} \cdots \frac{x_{d-1}(1) - a_{d-1}x_{d-1}(2)}{2}}_{g_a(x_1,\ldots,x_{d-1})} x_d(a), \tag{6.10}$$

where we identify $\{-1,1\}^{d-1}$ with $[2^{d-1}]$ in the canonical way. The address function is Boolean because for every $(x_1,\ldots,x_{d-1}) \in (\{-1,1\}^n)^{d-1}$ there is only one $a \in \{-1,1\}^{d-1}$ such that $g_a(x_1,\ldots,x_{d-1})$ is not 0, in which case it takes the value ± 1. Given that it has $2^{2(d-1)}$ Fourier coefficients and all of them equal 2^{1-d}, we have that

$$\Big(\sum_{s\in\{0,1\}^n} |\widehat{f}(s)|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} = 2^{1-d} 2^{2(d-1)\cdot\frac{d+1}{2d}} = 2^{\frac{d-1}{d}},$$

as promised. □

6.3.2 A non-commutative BH inequality

In this section, we prove a Bohnenblust-Hille inequality for linear maps that are bounded in the S_1 to S_∞ norm, such as quantum channels. Recall from Eq. (6.4) that any such a function can be written as

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

and it has degree d if this is the minimal number for which $\widehat{f}(s) = 0$ if $|s| > d$. The following result was proved originally in [Ble01], and with a better constant in [DMP19].

Theorem 6.15. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function of degree at most d. Then,*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \leq C^{\sqrt{d\log d}} \|f\|_\infty,$$

where $C > 0$ is a constant.

In order to prove Theorem 6.3 we follow a similar argument to the one used in [VZ23]. However, we need to modify their argument in order to consider maps from S_1 to S_∞ and not just matrices in M_N. In fact, as we explain in Remark 6.18, Theorem 6.3 generalizes the non-commutative BH inequality proved in [VZ23].

For every $\Phi : M_N \to M_N$, we will assign it a function $f_\Phi : \{-1,1\}^{3n} \times \{-1,1\}^{3n} \to \mathbb{C}$ whose Fourier spectrum will be closely related to the one of $\widehat{\Phi}$, as shown in Lemma 6.16, and then we will be able to reduce to Theorem 6.15. The function f_Φ is defined as follows. For $a = (a^1, a^2, a^3),\ b = (b^1, b^2, b^3) \in \{-1,1\}^n \times \{-1,1\}^n \times \{-1,1\}^n$ and $s, t \in \{1,2,3\}^n$, define the following matrices (which are not necessarily states)

$$|a^s\rangle\langle b^t| = \bigotimes_{i\in[n]} |\chi^{s(i)}_{a_i^{s(i)}}\rangle\langle\chi^{t(i)}_{b_i^{t(i)}}|,$$

Here $|\chi^s_a\rangle$ is the eigenvector of σ_s with eigenvalue a. The function $f_\Phi : \{-1,1\}^{3n} \times \{-1,1\}^{3n} \to \mathbb{C}$ is then given by

$$f_\Phi(a,b) = \frac{1}{9^n} \sum_{s,t\in\{1,2,3\}^n} \mathrm{Tr}[\Phi\left(|a^s\rangle\langle b^t|\right)|b^t\rangle\langle a^s|].$$

We recall the reader that any function $\Phi : M_N \to M_N$ can be expressed as

$$\Phi(\rho) = \sum_{x,y\in\{0,1,2,3\}^n} \widehat{\Phi}(x,y)\cdot\sigma_x\rho\sigma_y, \tag{6.11}$$

where $\sigma_x = \otimes_{i\in[n]}\sigma_{x_i}$ and σ_i for $i \in \{0,1,2,3\}$ are the Pauli matrices. We also recall that, if $|x|$ denotes the number of non-zero entries of $x \in \{0,1,2,3\}^n$, the degree of Φ is the minimum integer d such that $\widehat{\Phi}(x,y) = 0$ if $|x| + |y| > d$.

In the following lemma, the key properties of the function f are presented.

Lemma 6.16. *Let $\Phi : M_N \to M_N$ be a function of degree at most d. Then, f_Φ has also degree d. Moreover, $|f_\Phi(a,b)| \le \|\Phi\|_{S_1\to S_\infty}$ for all a, b and $\|\widehat{\Phi}\|_p \le 3^d\|\widehat{f_\Phi}\|_p$.*

Proof. We first show the bound on $|f_\Phi|$. Given that $|a^s\rangle\langle b^t|$ is a rank one operator such that $\||a^s\rangle\|_2 = \||b^t\rangle\|_2 = 1$, we conclude that

$$\||a^s\rangle\langle b^t|\|_{S_1} = 1. \tag{6.12}$$

Thus, we have that:

$$\begin{aligned}
|f_\Phi(a,b)| &\leq \frac{1}{9^n} \sum_{s,t\in\{1,2,3\}^n} |\operatorname{Tr}[\Phi\left(|a^s\rangle\langle b^t|\right)|b^t\rangle\langle a^s|]| \\
&\leq \frac{1}{9^n} \sum_{s,t\in\{1,2,3\}^n} \|\Phi\left(|a^s\rangle\langle b^t|\right)\|_{S_\infty} \||b^t\rangle\langle a^s|\|_{S_1} \\
&\leq \frac{1}{9^n} \sum_{s,t\in\{1,2,3\}^n} \|\Phi\|_{S_1\to S_\infty} \||a^s\rangle\langle b^t|\|_{S_1} \||b^t\rangle\langle a^s|\|_{S_1} \\
&\leq \frac{1}{9^n} \sum_{s,t\in\{1,2,3\}^n} \|\Phi\|_{S_1\to S_\infty} = \|\Phi\|_{S_1\to S_\infty},
\end{aligned}$$

where in the first inequality we have used the triangle inequality, in the second inequality the duality between S_1 and S_∞, in the third the definition of $S_1 \to S_\infty$ norm and in the fourth inequality we have used Eq. (6.12).

We now prove the estimate $\|\widehat{\Phi}\|_p \leq 3^{-d}\|\widehat{f_\Phi}\|_p$ and also that the degree of f_Φ is d. To this end, it suffices to show that

$$f_\Phi(a,b) = \sum_{x,y\in\{0,1,2,3\}^n} \frac{\widehat{\Phi}(x,y)}{3^{|x|+|y|}} \prod_{i\in\operatorname{supp}(x)} \prod_{j\in\operatorname{supp}(y)} a_i^{x(i)} b_j^{y(j)}, \tag{6.13}$$

where $\operatorname{supp}(x) = \{i \in [n] : x_i \neq 0\}$ and $|x|$ is the size of $\operatorname{supp}(x)$. Indeed, this follows from the fact that $\prod_{i\in\operatorname{supp}(x)} \prod_{j\in\operatorname{supp}(y)} a_i^{x(i)} b_j^{y(j)}$ can be read as $\chi_{S_{x,y}}(a,b)$ for a certain $S_{x,y} \in \{-1,1\}^{6n}$ satisfying that $S_{x,y} \neq S_{x',y'}$ whenever $(x,y) \neq (x',y')$, for for every $x, y \in \{0,1,2,3\}^n$.

To prove Eq. (6.13) the key is observing that for every $s,t \in \{1,2,3\}$, $x,y \in \{0,1,2,3\}$ and $a,b \in \{-1,1\}$, we have that

$$\operatorname{Tr}[\sigma_x|\chi_a^s\rangle\langle\chi_b^t|\sigma_y|\chi_b^t\rangle\langle\chi_a^s|] = \begin{cases} 0 & \text{if } (s\neq x \text{ and } x\neq 0) \text{ or } (t\neq y \text{ and } y\neq 0), \\ 1 & \text{if } x=0 \text{ and } y=0, \\ a & \text{if } s=x \text{ and } y=0, \\ b & \text{if } x=0 \text{ and } t=y, \\ ab & \text{if } s=x \text{ and } y=t. \end{cases}$$

Hence, taking tensor products we have that for every $s,t \in \{1,2,3\}^n$, $x,y \in \{0,1,2,3\}^n$ and $a = (a^1,a^2,a^3)$, $b = (b^1,b^2,b^3) \in \{-1,1\}^n \times \{-1,1\}^n \times \{-1,1\}^n$, it holds that

$$\operatorname{Tr}[\sigma_x|a^s\rangle\langle b^t|\sigma_y|b^t\rangle\langle a^s|] = \langle\chi_a^s|\sigma_x|\chi_a^s\rangle\langle\chi_b^t|\sigma_y|\chi_b^t\rangle = \prod_{i\in\operatorname{supp} x} \prod_{j\in\operatorname{supp} y} a_i^{x(i)} b_j^{y(j)} \delta_{x(i),s(i)} \delta_{y(j),t(j)}.$$

In particular, it follows that

$$f_{\Phi_{x,y}}(a,b) \equiv \frac{1}{9^n} \sum_{s,t\in\{1,2,3\}^n} \mathrm{Tr}[\sigma_x |a^s\rangle\langle b^t| \sigma_y |b^t\rangle\langle a^s|] = \frac{1}{9^n} \prod_{i\in \mathrm{supp}\, x} \prod_{j\in \mathrm{supp}\, y} a_i^{x(i)} b_j^{y(j)} \sum_{s\in\mathcal{X}, t\in\mathcal{Y}} 1,$$

where $\mathcal{X} = \{s \in \{1,2,3\}^n : s(i) = x(i)\ \forall\ i \in \mathrm{supp}(x)\}$. Since $|\mathcal{X}| = 3^{n-|x|}$, Eq. (6.13) follows for $\Phi_{x,y}$. Finally, Eq. (6.13) follows in general because

$$f_\Phi(a,b) = \sum_{x,y\in\{0,1,2,3\}^n} \widehat{\Phi}(x,y) f_{\Phi_{x,y}}(a,b).$$

□

Proof of Theorem 6.3. Let $\Re f_\Phi : \{-1,1\}^{6n} \to \mathbb{R}$ be defined as $(\Re f_\phi)(x) = \Re(f_\Phi(x))$ and $\Im f_\Phi : \{-1,1\}^{6n} \to \mathbb{R}$ as $(\Im f_\phi)(x) = \Im(f_\Phi(x))$. Note that we have that $\widehat{f_\phi} = \widehat{\Re f}_\Phi + i\widehat{\Im f}_\Phi$. By Lemma 6.16,

$$|(\Re f_\phi)(a,b)|, |(\Im f_\phi)(a,b)| \leq |f_\Phi(x)| \leq \|\Phi\|_{S_1\to S_\infty},$$

and that the degree of both the real and imaginary part is at most d. Hence, by the triangle inequality and Theorem 6.15 we have

$$\|\widehat{f_\Phi}\|_{\frac{2d}{d+1}} \leq \|\widehat{\Re f}_\Phi\|_{\frac{2d}{d+1}} + \|\widehat{\Im f}_\Phi\|_{\frac{2d}{d+1}} \leq C^{\sqrt{d\log d}} \|\Phi\|_{S_1\to S_\infty}.$$

Thus, as $\|\widehat{\Phi}\|_{2d/(d+1)} \leq 3^d \|\widehat{f_\Phi}\|_{2d/(d+1)}$, we have that $\|\widehat{\Phi}\|_{2d/(d+1)} \leq C^d \|\Phi\|_{S_1\to S_\infty}$. □

Corollary 6.17. *Let $\Phi : M_N \to M_N$ be an n-qubit quantum channel of degree at most d. Then*

$$\|\widehat{\Phi}\|_{2d/(d+1)} \leq C^d,$$

Proof. We just have to show that if Φ is a quantum channel, then $\|\Phi\|_{S_1\to S_\infty} \leq 1$. This is true since $\|\Phi\|_{S_1\to S_\infty} \leq \|\Phi\|_{S_1\to S_1}$ and $\Phi^\dagger$ is a completely positive and unital map between C*-algebras, so we have $\|\Phi\|_{S_1\to S_1} = \|\Phi^\dagger\|_{S_\infty\to S_\infty} = 1$ [Pau03, Proposition 3.2]. □

Remark 6.18. Theorem 6.3 generalizes the non-commutative BH inequality proved by Volberg and Zhang in [VZ23]. Indeed, given $M = \sum_x \widehat{M}(x)\sigma_x \in M_N$ the main result

of [VZ23] is recovered when one applies Theorem 6.3 to $\Phi_M(\cdot) = (\cdot)M$, which satisfies $\widehat{\Phi}_M(x,y) = \delta_{x,0^n}\widehat{M}(y)$ and $\|\Phi_M\|_{S_1\to S_\infty} = \|M\|$.

6.4 Learning low-degree quantum objects

This section is devoted to explaining the applications of the results developed in the previous section to learning theory.

Why are BH inequalities useful for learning?

We start by recalling a classical problem in learning theory which includes some of the results we present next and serves as motivation for other problems that are explained further below. Consider a function $f : \{-1,1\}^n \to \mathbb{R}$ to which we only have access through random samples. Here, a random sample means that we have access to $(x, f(x))$ for an element x chosen uniformly at random from $\{-1,1\}^n$. Assume that we fix $\varepsilon > 0$ and $\delta > 0$. Then, we want to devise an algorithm such that, by having access to $T(n,\varepsilon,\delta)$ random samples, produces another function $f' : \{-1,1\}^n \to \mathbb{R}$ satisfying, with probability at least $1-\delta$, that $\|f - f'\|_2 < \varepsilon$. In this case, we say that f can be learned within L_2-error ε by using $T(n,\varepsilon,\delta)$ samples.[1] The goal is to minimize the number of samples needed to learn the function.

A relevant instance of the problem we just have introduced is learning a bounded function $f : \{-1,1\}^n \to [-1,1]$ of degree at most d. The seminal low-degree algorithm by Linial, Mansour and Nisan solves it with $O_{d,\varepsilon}(n^d)$ samples [LMN93].[2] Their algorithm is based on learning the relevant part of the Fourier spectrum of the function f which, thanks to Parseval's identity, allows us to learn the function. More precisely, if f' has also degree at most d, we then have that

$$\|f - f'\|_2^2 = \sum_{s\in\{0,1\}^n, |s|\leq d} |\widehat{f}(s) - \widehat{f'}(s)|^2.$$

Hence, in order to learn f up to error ε, it suffices to learn each of its Fourier coefficients $\widehat{f}(s)$ with $|s| \leq d$ up to error $\varepsilon/\sqrt{n^d}$. Indeed, since there are at most $O(n^d)$ of these coefficients, this immediately implies that $\|f - f'\|_2^2 < \epsilon^2$.

Now, we explain how to learn the Fourier coefficients $\widehat{f}(s)$ for $|s| \leq d$ with probability $\geq 1-\delta$ and by just using $T = O(n^d \log(n^d/\delta)/\varepsilon^2)$ random samples $(x_i, f(x_i))_{i\in[T]}$.

[1] Despite we don't mention δ explicitly, this parameter is implicit in the problem. Sometimes, one fixes $\delta = 2/3$.

[2] Here and below, we use $O_{d,\varepsilon}$ to hide factors that depend on d and $1/\varepsilon$ and are independent of n.

To this end, let us consider the *empirical Fourier coefficients*, defined as

$$\widehat{f}'(s) = \frac{1}{T}\sum_{i\in[T]} f(x_i)\chi_s(x_i).$$

Note that, for a fixed s, $\widehat{f}'(s)$ can be seen as the average of T independent random variables distributed identically to the random variable $h_s : \{-1,1\}^n \to [-1,1]$ given by $h_s(\cdot) = f(\cdot)\chi_s(\cdot)$. Fixing s, since $\mathbb{E}\, h_s = \widehat{f}(s)$, we can apply the Hoeffding bound to state that

$$\Pr\Big(|\widehat{f}'(s)-\widehat{f}(s)| > \frac{\varepsilon}{\sqrt{n^d}}\Big) = \Pr\Big(\frac{1}{T}\Big|\sum_{i\in[T]}\big(f(x_i)\chi_s(x_i)-\widehat{f}(s)\big)\Big| > \frac{\varepsilon}{\sqrt{n^d}}\Big) \le \exp\Big(-\frac{T\epsilon^2}{2n^d}\Big).$$

A union bound can then be applied to upper bound the probability that $|\widehat{f}'(s)-\widehat{f}(s)| \le \frac{\varepsilon}{\sqrt{n^d}}$ for every $|s| \le d$ by

$$1 - \exp\Big(-\frac{T\epsilon^2}{2n^d} + d\log n\Big).$$

Hence, by choosing $T = 2n^d \log(n^d/\delta)/\varepsilon^2$, we make this upper bound equal to $1-\delta$ as we wanted.

The algorithm by Linial et al. was the state of the art until recently, when Eskenazis and Ivanisvili showed that a function of degree d can actually be learnt by using only $O_{d,\varepsilon}(\log n)$ random samples [EI22]. Their key insight was to use a Bohnenblust and Hille inequality for functions defined on the hypercube $f : \{-1,1\}^n \to \mathbb{R}$, proved in [DMP19], which can be used to upper bound the contribution of the *small Fourier coefficients*. To illustrate this, we consider the sum of the squares of the Fourier coefficients which are smaller than a certain parameter ε', which will be fixed later; namely

$$\sum_{s\in\{0,1\}^n,\ |\widehat{f}(s)|\le\varepsilon'} |\widehat{f}(s)|^2.$$

To upper bound this quantity, one can use that $2 = 2/(d+1) + 2d/(d+1)$, so

$$\sum_{s\in\{0,1\}^n,\ |\widehat{f}(s)|\le\varepsilon'} |\widehat{f}(s)|^2 \le \varepsilon'^{\frac{2}{d+1}} \sum_{s\in\{0,1\}^n,\ |\widehat{f}(s)|\le\varepsilon'} |\widehat{f}(s)|^{\frac{2d}{d+1}}.$$

Now one can use the aforementioned BH inequality, which states that $\|\widehat{f}\|_{2d/(d+1)} \le$

$C^{\sqrt{d \log d}} \|f\|_\infty$, to obtain

$$\sum_{s \in \{0,1\}^n,\ |\widehat{f}(s)| \leq \varepsilon'} |\widehat{f}(s)|^{\frac{2d}{d+1}} \leq \varepsilon'^{\frac{2}{d+1}} C^{\sqrt{d \log d}}.$$

Therefore, by setting $\varepsilon' = \varepsilon^{d+1} C^{-(d+1)\sqrt{d \log d}/2}$, it follows that

$$\sum_{s \in \{0,1\}^n,\ |\widehat{f}(s)| \leq \varepsilon'} |\widehat{f}(s)|^2 \leq \varepsilon^2. \tag{6.14}$$

From Eq. (6.14), Eskenazis and Ivanisvili essentially followed the ideas of Linial et al., but now they just needed to learn every low-degree Fourier coefficient up to error $\varepsilon' = \varepsilon^{d+1} C^{-(d+1)\sqrt{d \log d}/2}$, which is *much bigger* than $\varepsilon/\sqrt{n^d}$ and, in particular, independent of n. Using this approach, they proved that these functions can be learned with L_2-error ε and confidence $1 - \delta$ by using

$$O\left(\varepsilon^{-2(d+1)} \|\widehat{f}\|_{\frac{2d}{d+1}}^{2d} d^2 \log\left(\frac{n}{\delta}\right)\right) \tag{6.15}$$

random samples.

Learning quantum query algorithms

In particular, the result of Eskenazis and Ivanisvili applies to the amplitudes of quantum query algorithms as in Eq. (6.2) which, since the early days of quantum query complexity, are known to be bounded d-linear forms $T : \{-1,1\}^n \times \{-1,1\}^n \to [-1,1]$ [EI22, BBC$^+$01]. In addition, for d-linear forms it is known that the BH inequality holds with a polynomial constant, $\|T\|_{2d/(d+1)} \leq \mathrm{poly}(d)\|T\|_\infty$ [BPSS14]. Hence, it follows from Eq. (6.15) that the amplitudes of quantum query algorithms can be learned from

$$O(\varepsilon^{-2(d+1)} \mathrm{poly}(d)^d \log(n/\delta)) \tag{6.16}$$

samples.

A key observation here, proved in [ABP19], is that those d-linear forms arising from quantum algorithms actually satisfy that $\|T\|_{\mathrm{cb}} \leq 1$ [BBC$^+$01]. Hence, Theorem 6.1 implies the following improvement with respect to Eq. (6.16).

Corollary 6.2. *Consider a quantum algorithm that makes d queries as explained above. Then, its amplitudes can be learned with L_2^2-accuracy ε and confidence $1 - \delta$ from $O(\varepsilon^{-2(d+1)} d^2 \log(n/\delta))$ uniform random samples.*

Note that this result requires a number of samples that is polynomial in n when ε and δ are constants and $d = \log(n)$, while using (6.16) one would get $O(n^{\log\log n})$ samples as an upper bound.

Learning low-degree Boolean functions

In this section we propose almost optimal classical and quantum algorithms to learn low-degree Boolean functions. While we have already explained the classical access model (via random samples), we will also need to know what we mean by a quantum access model. The quantum counterpart of these samples are the quantum uniform samples, defined via the $(n+1)$-qubit states

$$|f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{-1,1\}^n} |x\rangle \otimes |f(x)\rangle \in (\mathbb{C}^2)^n \otimes \mathbb{C}^2 = (\mathbb{C}^2)^{n+1},$$

where $\{|-1\rangle, |1\rangle\}$ is the canonical (or computational) basis of $\mathbb{C}^2$ and we have denoted $|x\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle \in (\mathbb{C}^2)^n$ for every $x \in \{-1,1\}^n$. Quantum uniform samples are at least as powerful as classical samples. Indeed, if one measures the first n qubits of $|f\rangle$ in the basis $\{|x\rangle\}_x$, then the last qubit collapses to $|f(x)\rangle$ for a uniformly random x. However, they are actually strictly more powerful, as they allow one to sample from the Fourier distribution $(|\widehat{f}(s)|^2)_s$. For a proof of this well-known result, see for instance [ACL+21, Lemma 4].

Lemma 6.19 (Fourier sampling)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. There is an algorithm that inputs $|f\rangle$, succeeds with probability 1/2 and, in this case, samples a string $s \in \{0,1\}^n$ according to the probability distribution $(|\widehat{f}(s)|^2)_s$.*

We now state the main result of this section on Boolean functions.

Proposition 6.20. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a degree-d function. There is a quantum algorithm that learns f exactly with probability $1-\delta$ using $O\left(4^d d \log\left(1/\delta\right)\right)$ uniform quantum samples. Also, there is a classical algorithm that uses $O\left(4^d d \log\left(n/\delta\right)\right)$ uniform samples for this task.*

Despite the simplicity of the proof of Proposition 6.20, we include it for completeness and because it seems not to be well-known. See for instance [NPVY23, Corollary 34], which proposes a quantum algorithm for the same problem that requires $O(n^d)$ samples, or [EIS22, Corollary 4] that proposes a classical algorithm that requires $O(2^{d^2}\log n)$ samples. Proposition 6.20 highly improves those estimates. We

also remark that a classical lower bound of $\Omega(2^d \log n)$ samples was recently proved, making our classical result nearly optimal [EIS22]. Regarding the tightness of our quantum result, since learning functions $f : \{-1,1\}^d \to \{-1,1\}$ of degree d requires $\Omega(2^d)$ uniform quantum samples (which is folklore and follows from example from [AdW18]), our quantum estimate is almost optimal too.

Proof. For the classical upper bound we propose the following algorithm. We take $T = 2 \cdot 4^d \log(n^d/\delta)$ uniform samples $(x_i, f(x_i))$ and use them to define the empirical Fourier coefficients as

$$\widehat{f}'(s) = \frac{1}{T} \sum_{i \in [T]} f(x_i)\chi_s(x_i),$$

for every $s \subseteq [n]$. Define now the event $\mathcal{E} = \{|\widehat{f}(s) - \widehat{f}'(s)| < 2^{-d} \; \forall \, |s| \leq d\}$. Then, one can argue exactly in the same way as in Section 6.4 to conclude that $\Pr[\mathcal{E}] \geq 1 - \delta$.

Once we have computed the coefficients $\widehat{f}'(s)$, we round every of them to the closest number $\widehat{f}''(s) \in 2^{d-1}\mathbb{Z}$. If $\mathcal{E}$ occurs, by granularity we have that $\widehat{f}''(s) = \widehat{f}(s)$ for every $|s| \leq d$, so $f = \sum_s \widehat{f}''(s)\chi_s$, as desired.

For the quantum upper bound we begin by sampling $N = 4^d \log(4^d/\delta)$ times from $(\widehat{f}(s)^2)_{s \in \{0,1\}^n}$. This can be done, with probability $\geq 1 - \delta$, by using $T_1 = O(4^d \log(4^d/\delta))$ quantum uniform samples, thanks to Lemma 6.19 and a Hoeffding bound. Now, given s such that $\widehat{f}(s) \neq 0$, the probability that a sample s' according to the distribution $(\widehat{f}(s)^2)_{s \in \{0,1\}^n}$ satisfies $s' \neq s$ is given by $1 - \widehat{f}(s)^2 \leq 1 - 4^{1-d}$, where we have used that $\widehat{f}(s)^2 \geq 4^{1-d}$ by Lemma 6.13. Hence, if $s_1, \ldots, s_N$ are the N samples, then the probability that we have $s_i \neq s$ for every $i = 1, \ldots, N$ is upper bounded by

$$(1 - 4^{1-d})^N \leq \frac{\delta}{4^d}.$$

Thus, taking a union bound over the at most 4^{d-1} non-zero Fourier coefficients (due to Lemma 6.13 and $\sum_s |\widehat{f}(s)|^2 = 1$), it follows that, with probability $1 - \delta$, we will have sampled every non-zero Fourier coefficient.

In the second part of the algorithm we use $T_2 = O(4^d \log(4^d/\delta))$ quantum uniform samples and measure them in the computational basis, which generates classical uniform samples. From here, we can argue as in the classical upper bound and learn f exactly. The quantum advantage comes from Fourier sampling, that allows us to detect the non-zero Fourier coefficients, and apply the union bound only over those, that are at most 4^{d-1}. □

Learning low-degree quantum channels

First of all, we define the access model we use. Given a channel Φ, a learning algorithm is allowed to make queries to Φ as follows: it can choose a state ρ, feed ρ to the channel to obtain $\Phi(\rho)$ and measure $\Phi(\rho)$ in any basis.

The goal here, as in the previous sections, is to produce a classical description of a map $\widetilde{\Phi}$ that is close to Φ in the ℓ_2-distance defined by the usual inner product for maps from M_N to M_N, i.e., $\langle \Phi, \widetilde{\Phi} \rangle = \mathrm{Tr}[J(\Phi)J(\widetilde{\Phi})]/4^n$, where $J(\Phi)$ is the Choi-Jamiolkowski (CJ) representation of Φ.

For a reader not familiar with quantum computing, we remark that the proof of the main result of this section does not require prior knowledge of quantum computing, if one uses Lemmas 6.21 and 6.23 in a black-box manner. The reader can find in [NC10] an excellent reference to learn about quantum computing.

An important fact for our learning algorithm is that $\widehat{\Phi} = (\widehat{\Phi}(x,y))_{x,y}$ is a state that can be prepared with 1 query to Φ (see [BY23, Lemma 8]). This is the content of the following statement.

Lemma 6.21. *If Φ is a quantum channel, then $\widehat{\Phi}$ is a state unitarily equivalent to $v(\Phi)$. In particular, one query to Φ suffices to sample once from $(\widehat{\Phi}(x,x))_x$, which is a probability distribution.*

We will also make use of the following lemma, proved in [KMY03, Proposition 7].

Lemma 6.22. *Let ρ, ρ' be two states. Then, one can estimate $\mathrm{Tr}[\rho\rho']$ up to error ε with probability $1-\delta$, by using $O((1/\varepsilon)^2 \log(1/\delta))$ copies of ρ and ρ'.*

Before proving the main theorem of the section, we show that for a given $x, y \in \{0,1,2,3\}^n$, the corresponding Pauli coefficient $\widehat{\Phi}(x,y)$ can be efficiently learned.

Lemma 6.23 (Pauli coefficient estimation for channels)**.** *Let $\Phi : M_N \to M_N$ be a quantum channel and let $x, y \in \{0,1,2,3\}^n$. Then, $\widehat{\Phi}(x,y)$ can be estimated with error ε and probability $1-\delta$ using $O((1/\varepsilon)^2 \log(1/\delta))$ queries to Φ.*

Proof. If $x = y$, we just have to prepare $\widehat{\Phi}$ and apply Lemma 6.22 to $\widehat{\Phi}$ and the state $\rho = |x\rangle\langle x|$. If $x \neq y$, one first learns $\widehat{\Phi}(x,x)$ and $\widehat{\Phi}(y,y)$ with error ε as before. On the one hand, one can learn $\widehat{\Phi}(x,x) + \widehat{\Phi}(y,y) + 2\Re\widehat{\Phi}(x,y)$, with error ε by applying Lemma 6.22 to $\widehat{\Phi}$ and $|\xi\rangle\langle\xi|$, where $|\xi\rangle = 1/\sqrt{2}(|x\rangle + |y\rangle)$. Hence, one learns $\Re\widehat{\Phi}(x,y)$ with error $3\varepsilon/2$. On the other hand, one can learn $\widehat{\Phi}(x,x) + \widehat{\Phi}(y,y) + 2\Im\widehat{\Phi}(x,y)$, with error ε by applying Lemma 6.22 to $\widehat{\Phi}$ and $|\eta\rangle\langle\eta|$, where $|\eta\rangle = 1/\sqrt{2}(|x\rangle + i|y\rangle)$, and one can then learn $\Im\widehat{\Phi}(x,y)$ with error $3\varepsilon/2$. □

Now, we are ready to prove Theorem 6.4, which we restate for the convenience of the reader.

Theorem 6.4. *Let Φ be a n-qubit degree-d quantum channel. Then it can be learned in L_2-accuracy ε and confidence $\geq 1-\delta$ by making $\exp\left(\widetilde{O}(d^2 + d\log(1/\varepsilon))\right) \cdot \log(1/\delta)$ queries to Φ. Here, we use the notation $\widetilde{O}$ to hide logarithmic factors in d, $1/\varepsilon$, and $1/\delta$.*

Proof. The algorithm consists of 2 steps. In the first one we detect the relevant Pauli coefficients, while in the second step we learn the few Pauli coefficients detected as relevant.

Step 1. Detect the big Pauli coefficients. Let $c > 0$ be a parameter to be determined later. We invoke Lemma 6.21 to sample T_1 times from $(\widehat{\Phi}(x,x))_x$ by making T_1 queries to Φ. Let $(\widehat{\Phi}'(x,x))_x$ be the empirical distribution obtained from these samples. We store the big Pauli coefficients in the set $\mathcal{X}_c = \{x : \widehat{\Phi}'(x,x) \geq c\}$. Note that, since $\sum_{x\in\mathcal{X}_c} \widehat{\Phi}'(x,x) \leq 1$, we know that

$$|\mathcal{X}_c| \leq \frac{1}{c}. \tag{6.17}$$

Step 2. Learn the big Pauli coefficients. We invoke Lemma 6.23 to state that, by querying Φ just

$$T_2 = O((1/c)^4(1/\varepsilon)^2 \log((1/c)^2(1/\delta)))$$

times, we can find approximations $\widehat{\Phi}''(x,y)$ of $\Phi(x,y)$ for the at most $(1/c)^2$ pairs $(x,y) \in \mathcal{X}_c$, such that

$$\sup_{(x,y)\in\mathcal{X}_c\times\mathcal{X}_c} |\widehat{\Phi}(x,y) - \widehat{\Phi}''(x,y)| \leq c\varepsilon. \tag{6.18}$$

happens with probability $\geq 1-\delta$.

Output. We output $\Phi''(\cdot) = \sum_{x,y\in\mathcal{X}_c} \widehat{\Phi}''(x,y)\sigma_x(\cdot)\sigma_y$ as our approximation for Φ.

Correctness analysis. We consider the event $\mathcal{E} = \{|\widehat{\Phi}(x,x) - \widehat{\Phi}'(x,x)| \leq c \ \forall x \in \{0,1,2,3\}^n\}$. By Lemma 2.6, taking $T_1 = O((1/c)^2\log(1/\delta))$ ensures that

$$\Pr[\mathcal{E}] \geq 1-\delta.$$

Assuming the event $\mathcal{E}$ holds, we have that

$$x \notin \mathcal{X}_c \implies |\widehat{\Phi}(x,x)| \leq |\widehat{\Phi}'(x,x)| + ||\widehat{\Phi}(x,x)| - |\widehat{\Phi}'(x,x)|| \leq 2c. \tag{6.19}$$

In particular, it follows that

$$x \notin \mathcal{X}_c \implies |\widehat{\Phi}(x,y)| \leq \sqrt{|\widehat{\Phi}(x,x)||\widehat{\Phi}(y,y)|} \leq \sqrt{2c} \quad \forall\, y \in \{0,1,2,3\}^n, \tag{6.20}$$

where in the first inequality we have used that $\widehat{\Phi}$ is positive semidefinite and in the second inequality we have used Eq. (6.19) and that $\widehat{\Phi}(y,y) \leq 1$.

Assuming that both parts of the algorithm succeed, we have that Φ'' is close to Φ. Indeed,

$$\begin{aligned}
\|\Phi - \Phi''\|_2^2 &= \sum_{x,y \in \mathcal{X}_c} |\widehat{\Phi}(x,y) - \widehat{\Phi}''(x,y)|^2 + \sum_{x \vee y \notin \mathcal{X}_c} |\widehat{\Phi}(x,y)|^2 \\
&\leq \varepsilon^2 + \sum_{x \vee y \notin \mathcal{X}_c} |\widehat{\Phi}(x,y)|^{\frac{2}{d+1}} |\widehat{\Phi}(x,y)|^{\frac{2d}{d+1}} \\
&\leq \varepsilon^2 + (2c)^{\frac{1}{d+1}} \|\widehat{\Phi}\|_{\frac{2d}{d+1}}^{\frac{2d}{d+1}} \\
&\leq \varepsilon^2 + c^{\frac{1}{d+1}} C^d.
\end{aligned}$$

Here, in the equality we have used Parseval's identity, in the first inequality we used Eq. (6.17), Eq. (6.18) and that $2 = 1/(d+1/2) + 2d/(d+1/2)$; in the second inequality we have used Eq. (6.20) and in the third inequality we used the Bohnenblust-Hille inequality for channels (Corollary 6.17). Hence, by choosing

$$c = \varepsilon^{2d+2} C^{-d(d+1)}$$

we obtain the desired result.

Complexity analysis. Note that $T_2 > T_1$, so the complexity T_2 dominates the complexity of the first part of the algorithm. Hence, the total number of queries made is

$$O\big(C^{4d(d+1)}(1/\varepsilon)^{8d+10} \log(C^{2d(d+1)}(1/\varepsilon)^{4d+4}(1/\delta))\big).$$

□

Chapter 7

Testing and learning quantum Hamiltonians

7.1 Introduction

A fundamental and important challenge with building quantum devices is being able to characterize and calibrate its behavior. One approach to do so is *Hamiltonian learning* which seeks to learn the Hamiltonian governing the dynamics of a quantum system given finite classical and quantum resources. Beyond system characterization, it is also carried out during validation of physical systems and designing control strategies for implementing quantum gates [IBF+20]. However, learning an n-qubit Hamiltonian is known to be difficult, requiring complexity that scales exponential in the number of qubits unless a coarse metric is used [Car23].

In practice, however, prior knowledge on the structure of Hamiltonians is available e.g., those of engineered quantum devices [SMCG16] where the underlying Hamiltonians primarily involve local interactions with few non-local interactions, and even naturally occurring physical quantum systems such as those with translationally invariant Hamiltonians. To highlight these structural properties, consider an n-qubit Hamiltonian H (which is a self-adjoint operator acting on $(\mathbb{C}^2)^{\otimes n}$) expanded in terms of the n-qubit Pauli operators:

$$H = \sum_{x \in \{0,1,2,3\}^n} \lambda_x \sigma_x,$$

We call the set of Paulis with non-zero coefficients λ_x as the Pauli spectrum of the Hamiltonian denoted by $\mathcal{S} = \{x \in \{0,1,2,3\}^n : \lambda_x \neq 0\}$. Of particular relevance are *k-local* Hamiltonians which involve Pauli operators that act non-trivially on all but at most k qubits and *s-sparse* Hamiltonians whose Pauli expansion contains at most s non-zero Pauli operators i.e., $|\mathcal{S}| \leq s$.

There has thus been a growing suite of Hamiltonian learning results that have shown that when the underlying n-qubit Hamiltonian H satisfies these structural properties, learning can be performed with only $\text{poly}(n)$ query complexity, either by making "queries" to the unitary evolution operator $U(t) = \exp(-iHt)$ [dSLCP11, HBCP15, ZYLB21, HKT22, YSHY23, DOS23, HTFS23, LTN$^+$23, SFMD$^+$24, GCC24, Zha24, HMG$^+$25], or by assuming one has access to Gibbs state [AAKS21, HKT22, RSF23, ORSFW23, BLMT23, GCC24]. Notably, [BLMT24] considered the problem of learning Hamiltonians that are both local and sparse, without prior knowledge of the support. Several of the learning algorithms mentioned above however require assumptions on the support of the Hamiltonian beyond locality or sparsity, such as [HTFS23] which considers *geometrically-local* Hamiltonians (a subset of local Hamiltonians) and [YSHY23] which requires assumptions on the support.

Moreover, before learning, it might be desirable to uncover *what is the structure* of an unknown Hamiltonian in order to choose specialized learning algorithms. Even deciding if a Hamiltonian has a particular structure is a fundamental challenge and constitutes the problem of *testing* if an unknown Hamiltonian satisfies a certain structural property. This line of investigation is nascent with only a few works on Hamiltonian *property* testing [SY23, ACQ22, LW22] with Blum et al. [BCO24b] having considered the problem of testing local Hamiltonians and the problem of testing sparse Hamiltonians yet to be tackled. This leads us to the motivating question of this chapter:

What is the query complexity of learning and testing structured Hamiltonians?

Problem statement

Before we state our results answering the question above, we clearly mention our learning and testing problems first. If H is the Hamiltonian describing the dynamics of a certain physical system, then the state of that system evolves according to the *time evolution operator* $U(t) = e^{-iHt}$. This means that if $\rho(0)$ is the state at time 0, at time t the state would have evolved to $\rho(t) = U(t)\rho(0)U^\dagger(t)$. Hence, to test and

learn a Hamiltonian one can do the following: prepare a desired state, apply $U(t)$ or tensor products of $U(t)$ with identity to the state, and finally measure in a chosen basis. From here onwards, this is what we mean by *querying* the unitary $U(t)$. It is usual to impose the normalization condition $\|H\|_{\text{op}} \leq 1$ (i.e., that the eigenvalues of H are bounded in absolute value by 1). We will assume this normalization unless otherwise stated, but we will also work out the dependence on $\|H\|_{\text{op}}$ for our learning algorithms. Throughout this paper, we will consider the normalized Frobenius norm as the distance between Hamiltonians, unless otherwise stated. This distance is

$$d(H, H') = \|H - H'\|_2 = \sqrt{\frac{\text{Tr}[(H - H')^2]}{2^n}},$$

and it equals the ℓ_2-norm of the Pauli spectrum, $d(H, H') = \sqrt{\sum |\lambda_x - \lambda'_x|^2}$. A *property* of a Hamiltonian, denoted $\mathcal{H}$ is a class of Hamiltonians that satisfy the property (here we will be interested in sparse and local properties). We say that H is ε-far from having a property $\mathcal{H}$ if $d(H, H') > \varepsilon$ for every $H' \in \mathcal{H}$, and otherwise is ε-close. Now, we are ready to state the testing and learning problems.

Let $\mathcal{H}$ be a property and let H be an unknown Hamiltonian with $\|H\|_{\text{op}} \leq 1$ and $\text{Tr}[H] = 0$.

Problem 7.1 (Tolerant testing)**.** Promised H is either ε_1-close or ε_2-far from satisfying property $\mathcal{H}$, decide which is the case by making queries to $U(t)$.

Problem 7.2 (Hamiltonian learning)**.** Promised $H \in \mathcal{H}$, output a classical description of $\widetilde{H} \in \mathcal{H}$ such that $\|H - \widetilde{H}\|_2 \leq \varepsilon$ by making queries to $U(t)$.

Summary of results

The main results of this chapter are query-efficient algorithms for testing and learning Hamiltonians that are local and/or sparse. We summarize our results in Table 7.1 (for simplicity we state our results for constant accuracy). Before we discuss our results in more detail, we make a few remarks about our main results.

	Testing	**Learning**
s-sparse	poly(s)	poly(s)
k-local	$O(1)$	exp(k^2)

Table 7.1: Query complexity for learning and testing n-qubit structured Hamiltonians.

(*i*) As far as we know, by the time of the publication, the results of this chapter are the first: (*a*) with complexities that are *independent* of n [1], and (*b*) that does not assume knowledge of the support.[2]

(*iv*) We give the first learning algorithm for Hamiltonians that are only promised to be sparse, and not necessarily local. Similarly, our local Hamiltonian learning problem doesn't assume geometric locality which was assumed in several prior works.

(*iii*) Our testing algorithms are tolerant, i.e., they can handle the setting where $\varepsilon_1 \neq 0$. As far as we know, there are only a handful of polynomial-time tolerant testers for quantum objects.

(*iv*) Our learning algorithms are based on a subroutine that learns arbitrary n-qubit Hamilotmians with $O(1/\varepsilon^4)$ queries, albeit in the coarser metric of the ℓ_∞-norm of the Pauli coefficients. As far as we know, this is the only best result for unstructured Hamiltonians. Notably, it is also the first time-efficient proposal for this problem.

We remark that most previous works on Hamiltonian learning (that we highlighted earlier) are done under the distance induced by the supremum norm of the Pauli spectrum and with extra constraints apart from locality [dSLCP11, HBCP15, ZYLB21, HKT22, WKR$^+$22, YSHY23, Car23, DOS23, HTFS23, LTN$^+$23, SFMD$^+$24, GCC24]. When transformed into learning algorithms under the finer distance induced by the ℓ_2-norm of the Pauli spectrum, these proposals yield complexities that depend polynomially on n^k and only work for a restricted family of k-local Hamiltonians. The works that explicitly consider the problem of learning under the ℓ_2-norm have complexities depending on n and assume a stronger access model [CW23, BLMT24].

Results

Testing. Recently, Bluhm, Caro and Oufkir proposed a non-tolerant testing algorithm, meaning that it only works for the case $\varepsilon_1 = 0$, whose query complexity is $O(n^{2k+2}/(\varepsilon_2 - \varepsilon_1)^4)$ and with total evolution time $O(n^{k+1}/(\varepsilon_2 - \varepsilon_1)^3)$. They posed as

[1] There are a few works that achieve n-independent complexities for learning local Hamiltonians in the ∞-norm of the Pauli coefficients, but when transformed into 2-norm learners they yield complexities depending on n^k.

[2] Soon after [Esc24b], Bakshi et al. [BLMT24] presented a learning algorithm that does not require prior knowledge of the support, achieving Heisenberg scaling using heavy machinery.

open questions whether the dependence on n could be removed and whether an efficient tolerant-tester was possible [BCO24a, Section 1.5]. Our first result gives positive answer to both questions.

Result 7.3. *There is an algorithm that solves Problem 7.1 for k-local Hamiltonians by making* $\mathrm{poly}(1/(\varepsilon_2-\varepsilon_1))$ *queries to the evolution operator and with* $\mathrm{poly}(1/(\varepsilon_2-\varepsilon_1))$ *total evolution time.*

See Theorem 7.12 for a formal statement of this result. Our algorithm to test for locality is simple. It consists of repeating the following process $1/(\varepsilon_2-\varepsilon_1)^4$ times: prepare n EPR pairs, apply $U(\varepsilon_2-\varepsilon_1)\otimes \mathrm{Id}_{2^n}$ to them and measure in the Bell basis. Each time that we repeat this process, we sample from the Pauli sprectrum of $U(\varepsilon_2-\varepsilon_1)$.[3] As $\varepsilon_2-\varepsilon_1$ is small, Taylor expansion ensures that $U(\varepsilon_2-\varepsilon_1)\approx \mathrm{Id}_{2^n}-i(\varepsilon_2-\varepsilon_1)H$, so sampling from the Pauli spectrum of $U(\varepsilon_2-\varepsilon_1)$ allows us to estimate the weight of the non-local terms of H. If that weight is big, we output that H is far from k-local, and otherwise we conclude that H is close to k-local.

Despite the numerous papers in the classical literature studying the problems of testing and learning sparse Boolean functions [GOS+11, NS12, YZ20, EIS22], there are not many results on learning Hamiltonians that are sparse (and not necessarily local) and the only testing result that we are aware of requires $O(sn)$ queries [BCO24b, Remark B.2]. Here, we present the first sparsity testing algorithm whose complexity does not depend on n and the first learning algorithm for sparse Hamiltonians which does not make any assumptions regarding the support of the Hamiltonian beyond sparsity.

Result 7.4. *There is an algorithm that solves Problem 7.1 for s-sparse Hamiltonians by making* $\mathrm{poly}(s/(\varepsilon_2-\varepsilon_1))$ *queries to the evolution operator and with* $\mathrm{poly}(s/(\varepsilon_2-\varepsilon_1))$ *total evolution time.*

See Theorem 7.15 for a formal statement. This testing algorithm consists on performing Pauli sampling of $U(\sqrt{(\varepsilon_2^2-\varepsilon_1^2)/s})$ a total of $O(s^4/(\varepsilon_2^2-\varepsilon_1^2)^4)$ times. From these samples one can estimate the sum of the squares of the top s Pauli coefficients of U. If this quantity is big enough, we output that the Hamiltonian is close to s-sparse, and otherwise that is far. Although from this high-level description the algorithm seems similar to the locality testing one, the analysis is more involved and requires taking the second order Taylor expansion, which is the reason why the dependence on $(\varepsilon_2-\varepsilon_1)$ is worse in this case.

[3] The Pauli spectrum of a unitary $U=\sum_x \widehat{U}_x\sigma_x$ determines a probability distribution because $\sum_x|\widehat{U}_x|^2=1$.

Additionally, we provide a sparsity tester (Theorem 7.16) that only makes $O(s^2/\varepsilon_2^4)$ queries with $O(s^{1.5}/\varepsilon_2^3)$ total evolution time, but only works in the regime $\varepsilon_1 = O(\varepsilon_2/\sqrt{s})$.

Learning. We first propose a protocol to learn unstructured Hamiltonians efficiently in the coarser ℓ_∞ norm of the Pauli coefficients. Then, we turn it into a learner in the ℓ_2 norm for local and sparse Hamiltonians.

Result 7.5. *There is an algorithm that outputs estimates $\widetilde{\lambda}_x$ such that $|\lambda_x - \widetilde{\lambda}_x| \leq \varepsilon$ for every $x \in \{0,1,2,3\}^n$ by making $O(1/\varepsilon^4)$ queries to the evolution operator with $O(1/\varepsilon^3)$ total evolution time.*

See Theorem 7.18 for a formal result. The learning algorithm has two stages. In the first stage one samples from the Pauli distribution of $U(\varepsilon)$, as in the testing algorithm, and from that one can detect which are the big Pauli coefficients of H. In the second stage we learn the big Pauli coefficients via a novel subroutine based on Clifford Shadows (see Lemma 7.17) and set the small to 0.

For Hamiltonians that are k-local, we have the following learning result in the ℓ_2-norm.

Result 7.6. *There is an algorithm that solves Problem 7.2 for k-local Hamiltonians by making $\exp(k^2 + k\log(1/\varepsilon))$ queries to the evolution operator with $\exp(k^2 + k\log(1/\varepsilon))$ total evolution time.*

See Theorem 7.19 for a formal statement of this result. In the case that the Hamiltonian is k-local, one can ensure that the coefficients not detected as big in the first stage of the algorithm of Result 7.5 have a neglectable contribution to the ℓ_2-norm, from which Result 7.6 follows. To argue this formally, we use the non-commutative Bohnenblust-Hille inequality, which has been used recently for various quantum learning algorithms [HCP23b, VZ23].

For Hamiltonians that are s-sparse, we have the following learning result in the ℓ_2-norm.

Result 7.7. *There is an algorithm that solves Problem 7.2 for s-sparse Hamiltonians by making $\mathrm{poly}(s/\varepsilon)$ queries to the evolution operator with $\mathrm{poly}(s/\varepsilon)$ total evolution time.*

See Theorem 7.21 for a formal statement. Result 7.7 follows by adding a rounding step to the algorithm of Result 7.5 that ensures that all zero coefficients of the Hamiltonians are also zero for the approximating Hamiltonian.

Direct comparison to previous work. Comparing the plethora of Hamiltonian learning algorithms can be challenging due to the different assumptions on the structure of the Hamiltonians (local, sparse, geometrical structures, etc.), the different distances to measure the error (ℓ_∞ norm of the coefficients, ℓ_2 norm, etc.), the different complexity measures (queries, total evolution time, number of experiments, etc.), the different access models (coherent/non-coherent queries, with/without memory, etc.) and the different goals of the algorithm (minimizing the dependence on the dimension parameters like n, s, k, achieving the Heisenberg scaling $1/\varepsilon$, etc.). Thus, we only include a direct comparison in Table 7.2 with the works that explicitly consider the same structure and the same error metric as us. As a summary, one can say that for constant ε our results achieve better dependence on the parameters n, s, k than previous work, while also using the weaker model of incoherent queries, where one can perform only one query before measuring, as opposed to the coherent query model. We also want to remark that our result for learning unstructured Hamiltonian is time efficient, while the, to the best of our knowledge, only previous one is not [Car23].

Hamiltonians	Reference	t_{total}	Queries	Access model
Unstructured, ℓ_∞ error	[Car23]	n/ϵ^4	n/ϵ^4	Coherent queries
	Theorem 7.18	$1/\varepsilon^3$	$1/\epsilon^4$	Incoherent queries
s-sparse, ℓ_∞ error	[Zha24]*	$1/\epsilon^4$	$1/\epsilon^8$	Coherent queries
	[HMG+25]†	s^2/ϵ	s^2/ϵ	Coherent queries
	Theorem 7.21	$1/\varepsilon^3$	$1/\epsilon^4$	Incoherent queries
k-local, ℓ_2 error	[CW23]	n^k/ϵ^2	n^k/ϵ^2	Controlled and inverse queries
	[MFPT24]°	$(9n)^k/\epsilon$	$(27n^3)^k/\epsilon^2$	Coherent queries
	Theorem 7.19	$\exp(k^2)/\varepsilon^k$	$\exp(k^2)/\varepsilon^k$	Incoherent queries

Table 7.2: Comparison of algorithms for learning Hamiltonians with $\|H\|_{\text{op}} \leq 1$.
* It can be improved to $O(1/\varepsilon^{2+o(1)})$ total evolution time and $O(1/\varepsilon^{6+o(1)})$ queries by paying huge constant factors.
† This algorithm works for Hamiltonians with $\sup_x |\lambda_x| \leq 1$, a weaker constraint than $\|H\|_{\text{op}} \leq 1$.
° This algorithm is the only one in the table that uses no quantum memory. We provide an algorithm with no quantum memory for k-local learning that performs as the one in the last row, but with an extra factor $\log n$.

Note added. After sharing Theorem 7.12 with Bluhm et al., they independently improved the analysis of their testing algorithm and showed that it only requires $O(1/(\varepsilon_2 - \varepsilon_1)^3 \varepsilon_2)$ queries and $O(1/(\varepsilon_2 - \varepsilon_1)^{2.5} \varepsilon_2^{0.5})$ total evolution time, which is very similar to our Theorem 7.12 [BCO24b]. In addition, for a wide range of $k = O(n)$, their algorithm does not require the use of auxiliary qubits.

7.2 Preliminaries

Notation

Every n-qubit operator H can be written down in its Pauli decomposition as

$$H = \sum_{x \in \{0,1,2,3\}^n} \lambda_x \sigma_x,$$

where the real-valued coefficients λ_x are given by $\lambda_x = \frac{1}{2^n} \operatorname{Tr}(H\sigma_x)$. Parseval's identity states that the normalized Frobenius norm of H (denoted as $\|H\|_2$) equals the ℓ_2-norm of its Pauli spectrum, i.e.,

$$\|H\|_2 = \sqrt{\frac{\operatorname{Tr}[H^\dagger H]}{2^n}} = \sqrt{\sum_{x \in \{0,1,2,3\}^n} |\lambda_x|^2}.$$

We will repeatedly use that $\|H\|_2 \leq \|H\|_{\text{op}}$, which holds because $\|H\|_2^2$ is the average of the squares of the eigenvalues of H. We will also consider the ℓ_∞ norm of the Pauli coefficients of an operator, which is given by

$$\|H\|_{\ell_\infty} = \sup_x |h_x|.$$

Additionally, we will use $\|H\| := \max\{\|H\|_{\text{op}}, 1\}$.

Given $x \in \{0,1,2,3\}^n$, define $|x|$ as the number of indices $i \in [n]$ where $x_i \neq 0$, define

$$H_{>k} = \sum_{|x|>k} \lambda_x \sigma_x$$

and $H_{\leq k}$ as $\sum_{|x| \leq k} \lambda_x \sigma_x$. From the formulation of the 2-norm in terms of the Pauli coefficients it follows that $\|H_{>k}\|_2 \leq \|H\|_2$. We note that the distance of a Hamiltonian H from the space of k-local Hamiltonians is given by $\|H_{>k}\|_2$, as $H_{\leq k}$ is the k-local Hamiltonian closest to H. The ℓ_2-distance of H to being s-sparse also has a nice expression. Assign labels from $[4^n]$ to $x \in \{0,1,2,3\}^n$ in a way that and $|\lambda_{x_1}| \geq |\lambda_{x_2}| \cdots \geq |\lambda_{x_{4^n}}|$. Then, $\sum_{i \in [s]} \lambda_{x_i} \sigma_{x_i}$ is the closest s-sparse Hamiltonian to H, so the ℓ_2-distance of H to the space of s-sparse Hamiltonians is $\sqrt{\sum_{i>s} |\lambda_{x_i}|^2}$.

Necessary subroutines

Suppose U is a unitary and we write out its Pauli decomposition as $U = \sum_x \widehat{U}_x \sigma_x$, then by Parseval's identity $\sum |\widehat{U}_x|^2 = \operatorname{Tr}[U^\dagger U]/2^n = 1$, i.e., $\{|\widehat{U}_x|^2\}_x$ is a *probability*

distribution. We will be using the fact below extensively.

Fact 7.8. Given access to a unitary U, one can sample from the distribution $\{|\widehat{U}_x|^2\}_x$.

Proof. The proof simply follows by applying $U \otimes \mathrm{Id}_{2^n}$ to n EPR pairs (i.e., preparing the Choi-Jamiolkowski state of U) and measuring in the Bell basis, because

$$U \otimes \mathrm{Id}_{2^n} |\mathrm{EPR}_n\rangle = \sum_{x \in \{0,1,2,3\}^n} \widehat{U}_x \bigotimes_{i \in [n]} (\sigma_{x_i} \otimes \mathrm{Id}_2 |\mathrm{EPR}\rangle),$$

and the Bell states can be written as $\sigma_x \otimes \mathrm{Id}_2 |\mathrm{EPR}\rangle$ for $x \in \{0,1,2,3\}$. □

We will also use that given a Hamiltonian H, the Taylor expansion of the exponential allows us to approximate the time evolution operator as

$$U(t) = e^{-itH} = \mathrm{Id}_{2^n} - itH + ct^2 R_1(t) \|H\|_{\mathrm{op}}^2 \tag{7.1}$$

for $t \leq 1/2$, where the first order remainder $R_1(t)$ is bounded $\|R_1(t)\|_{\mathrm{op}} \leq 1$ and $c > 1$ is a universal constant.

We will also use the celebrated Classical Shadows by Huang, Chen and Preskill.

Theorem 7.9 (Clifford shadows [HKP20])**.** *Let ρ be an n-qubit state and let $\{O_i\}_{i \in [M]}$ be n-qubit traceless observables. Assume that $\sup_i \mathrm{Tr}[O_i^2] = O(1)$. Then, Algorithm 1 obtains estimates $\widetilde{O}_{i,\rho}$ such that, with probability $1 - \delta$, satisfy*

$$|\mathrm{Tr}[O_i \rho] - \widetilde{O}_{i,\rho}]| \leq \varepsilon$$

for every $i \in [M]$. The algorithm uses $O\left(\frac{\log(M/\delta)}{\varepsilon^2}\right)$ copies of ρ.

7.3 Technical results

In this section, we will first prove our main structural theorems for Hamiltonians and provide subroutines which will be used later for testing and learning these structured Hamiltonians.

Structural lemma for local Hamiltonians

First, we prove a lemma regarding the discrepancy on the weights of non-local terms of the short-time evolution operator for close-to-local and far-from-local Hamiltonians.

Algorithm 1 Clifford shadows

Input: Copies of a quantum state ρ, target set of observables $\{O_i\}_{i\in[M]}$, error parameter $\varepsilon \in (0,1)$, and failure parameter $\delta \in (0,1)$

Set $T = O(\log(M/\delta)/\varepsilon^2)$ and $J = O(\log(M/\delta))$
for $j \in [J]$ **do**
  for $k \in [T/J]$ **do**
    Apply a uniformly random Clifford gate C to a copy of ρ
    Measure in the computational basis. Let $|b_{j,k}\rangle$ be the outcome
    for $i \in [M]$ **do**
      Let $\widetilde{O}_{i,j,k} = (2^n+1)\langle b_{j,k}|C^{-1}O_iC|b_{j,k}\rangle$
    end for
  end for
  for $i \in [M]$ **do**
    Let $\widetilde{O}_{i,j} = \mathrm{Mean}((\widetilde{O}_{i,j,k})_k)$
  end for
end for
for $i \in [M]$ **do**
  Set $\widetilde{O}_i := \mathrm{Median}((O_{i,j})_j)$
end for

Output: $(\widetilde{O}_i)_{i\in[M]}$

Lemma 7.10. *Let $0 \le \varepsilon_1 < \varepsilon_2$. Let $\alpha = (\varepsilon_2-\varepsilon_1)/(3c)$ and H be an n-qubit Hamiltonian with $\|H\|_{\mathrm{op}} \le 1$, where c is the constant appearing in Taylor expansion (see Eq. (7.1)). If H is ε_1-close k-local, then*

$$\|U(\alpha)_{>k}\|_2 \le (\varepsilon_2-\varepsilon_1)\frac{2\varepsilon_1+\varepsilon_2}{9c},$$

and if H is ε_2-far from being k-local, then

$$\|U(\alpha)_{>k}\|_2 \ge (\varepsilon_2-\varepsilon_1)\frac{\varepsilon_1+2\varepsilon_2}{9c}.$$

Proof. Recall that $U(\alpha) = \mathrm{Id}_{2^n} - i\alpha H + c\alpha^2 R(\alpha)$ by Eq (7.1) where $\|R\|_{\mathrm{op}} \le 1$. For simplicity, we set $U = U(\alpha)$ and $R = R_1(\alpha)$. First, assume that H is ε_1-close k-local, then by definition we have that $\|H_{>k}\|_2 \le \varepsilon_1$. Then

$$\|U_{>k}\|_2 \le \alpha\|H_{>k}\|_2 + c\alpha^2\|R_{>k}\|_2 \le \frac{\varepsilon_2-\varepsilon_1}{3c}\varepsilon_1 + c\left(\frac{\varepsilon_2-\varepsilon_1}{3c}\right)^2 = (\varepsilon_2-\varepsilon_1)\frac{2\varepsilon_1+\varepsilon_2}{9c},$$

where in the first inequality we have used the triangle inequality, and in the second that H is ε_1-close to k-local and that $\|R_{>k}\|_2 \le \|R\|_2 \le \|R\|_{\mathrm{op}} \le 1$. Now, assume

that H is ε_2-far from being k-local (i.e., $\|H_{>k}\|_2 \geq \varepsilon_2$). Then

$$\|U_{>k}\|_2 \geq \alpha\|H_{>k}\|_2 - c\alpha^2\|R_{>k}\|_2 \geq \frac{\varepsilon_2-\varepsilon_1}{3c}\varepsilon_2 - c\left(\frac{\varepsilon_2-\varepsilon_1}{3c}\right)^2 \geq (\varepsilon_2-\varepsilon_1)\frac{\varepsilon_1+2\varepsilon_2}{9c},$$

where in first inequality we have used triangle inequality on $i\alpha H = ct^2R(\alpha) - U(\alpha)$ to conclude $\alpha\|H_{>k}\|_2 \leq \|U_{>k}\|_2 + c\alpha^2\|R_{>k}\|_2$, and in the second the fact that H is ε_2-far from k-local. □

Structural lemma for sparse Hamiltonians

Similar to local Hamiltonians, we show a discrepancy in the sum of the top Pauli coefficients of the short-time evolution operator for close-to-sparse and far-from-sparse Hamiltonians. To formally state this result we need to introduce the concept of *top energy*. Let $U(t)$ the time evolution operator at time t and let $\{\widehat{U}(t)\}_x$ be its Pauli coefficients. We assign labels from $\{x_0, \ldots, x_{4^n-1}\}$ to $x \in \{0,1,2,3\}^n$ in a way that $\widehat{U}_{x_0} = \widehat{U}_{0^n}$ and $|\widehat{U}_{x_1}| \geq |\widehat{U}_{x_2}| \geq \cdots \geq |\widehat{U}_{x_{4^n-1}}|$. Now, we define the top energy at time t as

$$\mathsf{TopEnergy}(t;s) := |\widehat{U}_{x_0}(t)|^2 + \sum_{i\in[s]} |\widehat{U}_{x_i}(t)|^2,$$

Lemma 7.11. *Let H be a n-qubit Hamiltonian with $\|H\|_{\mathrm{op}} \leq 1$ and $\mathrm{Tr}[H] = 0$. Let $t \in (0,1)$. On the one hand, if H is ε_1-close to s-sparse, then*

$$\mathsf{TopEnergy}(t;s) \geq 1 - \varepsilon_1^2 t^2 - O(t^3 s).$$

On the other hand, if H is ε_2-far from s-sparse, then

$$\mathsf{TopEnergy}(t;s) \leq 1 - \varepsilon_2^2 t^2 + O(t^3 s).$$

Proof. For this proof we need to consider the 2nd order Taylor expansion of $U(t)$,

$$U(t) = \mathrm{Id} - itH - t^2H^2/2 + O(t^3)R_2,$$

where R_2 is the remainder of the series of order 2 that satisfies $\|R_2\|_{\mathrm{op}} \leq 1$, because $\|H\|_{\mathrm{op}} \leq 1$. Since $\mathrm{Tr}[H] = 0$ (so $\lambda_{0^n} = 0$), we have

$$\widehat{U}_0(t) = 1 - \frac{t^2}{2} \cdot \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2 + O(t^3),$$

so, using that $|a^2 - b^2| = |a - b||a + b|$, we have that

$$\Big||\widehat{U}_0(t)|^2 - \big(1 - t^2 \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2\big)\Big| = O(t^3). \tag{7.2}$$

To control $|U_x(t)|$ for $x \neq 0^{2n}$, we use the first order Taylor expansion of $U(t) = \mathrm{Id}_{2^n} - itH + ct^2 R_1(t)$ and get

$$\big||\widehat{U}_x(t)| - |t\lambda_x|\big| \leq |\widehat{U}_x(t) - (-it\lambda_x)| \leq \|U(t) - (-itH)\|_2 \leq O(t^2)\|R_1\|_2 \leq O(t^2), \tag{7.3}$$

where we again used that $\|R_1\|_2 \leq 1$. From this it follows that

$$\begin{aligned}\big||\widehat{U}_x(t)|^2 - t^2\lambda_x^2\big| = \Big|\big(|\widehat{U}_x(t)| - |t\lambda_x|\big) \cdot \big(|\widehat{U}_x(t)| + |t\lambda_x|\big)\Big| &= O(t^2)(|U_x| + |t\lambda_x|) \\ &= O(t^2)(2|t\lambda_x| + O(t^2)) = O(t^3),\end{aligned} \tag{7.4}$$

where the second and third equality both used Eq. (7.3); and in the last line used $|\lambda_x| \leq \|H\|_{\mathrm{op}} \leq 1$. In particular, the above implies that

$$|\widehat{U}_x(t)|^2 \geq t^2|\lambda_x|^2 - O(t^3) \tag{7.5}$$

Now we will define a quantity similar to the top energy, but now we will define the top coefficients as the top coefficients of H. To be precise, we assign labels to $\{y_0, \ldots, y_{4^n-1}\}$ to the elements of $\{0,1,2,3\}^n$ in a way such that $y_0 = 0^{2n}$ and $|\lambda_{y_1}| \geq \cdots \geq |\lambda_{y_{4^n-1}}|$. We now define

$$\mathsf{TopEnergy}_H(t;s) := \Big(1 - t^2 \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2\Big) + \sum_{i\in[s]} (t\lambda_{y_i})^2.$$

If the top s Pauli coefficients of H coincided with the ones of $U(t)$ and there was no error in the Taylor expansion, then $\mathsf{TopEnergy}_H(t;s)(t) = \mathsf{TopEnergy}(t;s)$. However, this may not be true in general. Nevertheless, we show that both quantities are close

to each other. To this end,

$$\begin{aligned}\mathsf{TopEnergy}(t;s) &= |\widehat{U}_{x_0}(t)|^2 + \sum_{i\in[s]} |\widehat{U}_{x_i}(t)|^2 \\ &\geq |\widehat{U}_{y_0}(t)|^2 + \sum_{i\in[s]} |\widehat{U}_{y_i}(t)|^2 \\ &\geq \big(1 - t^2 \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2\big) + \sum_{i\in[s]} (t\lambda_{y_i})^2 - (s+1)O(t^3) \\ &= \mathsf{TopEnergy}_H(t;s) - (s+1)O(t^3),\end{aligned}$$

where in the first inequality we used that $x_1, \ldots, x_s$ correspond to the s largest coefficients of $U(t)$, so $\sum_{i\in[s]} |\widehat{U}_{x_i}(t)|^2$ is larger than the sum of the squares of any other s coefficients of U; in the second inequality we used Eqs. (7.2) and (7.5). Similarly, one can check that $\mathsf{TopEnergy}_H(t;s) \geq \mathsf{TopEnergy}(t;s) - (s+1)O(t^3)$, so

$$|\mathsf{TopEnergy}_H(t;s) - \mathsf{TopEnergy}(t;s)| \leq O(st^3).$$

Now, the claimed result follows by noticing that

$$\mathsf{TopEnergy}_H(t;s) = 1 - t^2 \sum_{i>s} |\lambda_{y_i}|^2,$$

and that $\sum_{i>s} |\lambda_{y_i}|^2$ is the square of the ℓ_2-distance of H to the space of s-sparse Hamiltonians, because $\sum_{i\in[s]} \lambda_{y_i}\sigma_{y_i}$ is the s-sparse Hamiltonian closest to H. □

7.4 Testing Hamiltonians

In this section, we give our testing algorithms for local Hamiltonians.

7.4.1 Testing local Hamiltonians

We now state our locality testing algorithm and prove its guarantees.

Theorem 7.12. *Algorithm 2 solves the locality testing problem (Problem 7.1 with the property of being k-local) with probability $\geq 1-\delta$, by making $O(1/(\varepsilon_2-\varepsilon_1)^4 \cdot \log(1/\delta))$ queries to the evolution operator and with $O(1/(\varepsilon_2-\varepsilon_1)^3 \cdot \log(1/\delta))$ total evolution time.*

Proof. Let $t = (\varepsilon_2 - \varepsilon_1)/(3c)$ and let $U = U(t)$. For notational simplicity, let $\alpha_k :=$

Algorithm 2 Locality tester

Input: Query access to the time evolution of $U(t) = e^{-itH}$, closeness and farness parameters $\varepsilon_1,\ \varepsilon_2 \in (0,1)$, locality parameter $k \in \mathbb{N}$ and failure parameter $\delta \in (0,1)$

Set $T = O(\log(1/\delta)/(\varepsilon_2 - \varepsilon_1)^4)$
Let $t = (\varepsilon_2 - \varepsilon_1)/(3c)$ and $U = U(t)$
Initialize $\alpha'_k = 0$
for $i = 1, ..., T$ **do**
    Perform Pauli sampling from U. Let $x \in \{0,1,2,3\}^n$ be the outcome.
    if $|x| > k$ **then**
        $\alpha'_k \leftarrow \alpha'_k + 1/T$
    end if
end for Set $\alpha''_k = 0$
for $i = 1, ..., T$ **do**
    Perform Pauli sampling from U. Let $x \in \{0,1,2,3\}^n$ be the outcome.
    If $|x| > k$, $\alpha''_k \leftarrow \alpha''_k + 1/T$
end for

Output: If $\alpha'_k \geq (3/4)(\varepsilon_2 - \varepsilon_1)^2$ or $\alpha''_k \geq (\varepsilon_2 - \varepsilon_1)(\varepsilon_1 + 2\varepsilon_2)/(9c) - (\varepsilon_2 - \varepsilon_1)^2/(18c)$ output that H is far from local, and close to local otherwise

$\|U_{>k}\|_2^2$. We will first estimate α_k upto error $(\varepsilon_2 - \varepsilon_1)^2/4$. To do that we sample from $\{|\widehat{U}_x|^2\}_x$ using Fact 7.8 a total of $T = O(1/(\varepsilon_2 - \varepsilon_1)^4 \log(1/\delta))$ times, which can be done with T queries. If $x_1, \ldots, x_T$ are the outcomes of those samples, we define our estimate as

$$\alpha'_k := \frac{1}{T} \sum_{i \in [T]} [|x_i| > k].$$

By the Hoeffding bound, we have that indeed $|\alpha'_k - \alpha_k| \leq (\varepsilon_2 - \varepsilon_1)^2/4$ with probability $\geq 1 - \delta/2$.

If $\alpha'_k \geq (3/4)(\varepsilon_2 - \varepsilon_1)^2$, then $\alpha_k \geq (\varepsilon_2 - \varepsilon_1)^2/2$, so by Lemma 7.10 we conclude that H is far from k-local. Otherwise, if $\alpha'_k \leq (3/4)(\varepsilon_2 - \varepsilon_1)^2$, then $\alpha_k \leq (\varepsilon_2 - \varepsilon_1)^2$. Now we take again T samples from $y_1, \ldots, y_T$ from $\{|\widehat{U}_x|^2\}_x$ and define a new estimate

$$\alpha''_k = \frac{1}{T} \sum_{i \in [T]} [|y_i| > k].$$

By definition α''_k equals α_k in expectation. Furthermore, α_k is the empirical average of random variables whose variance is considerably small, because

$$\mathbb{E}[[|y| > k]^2] = \mathbb{E}[[|y| > k]] = \|U_{>k}\|_2^2 \leq (\varepsilon_2 - \varepsilon_1)^2.$$

Then, an application of Bernstein's inequality (Lemma 2.22) shows that α''_k approximates $\|U_{>k}\|_2^2$ up to error $((\varepsilon_2-\varepsilon_1)^2/(18c))^2$ with success probability $1-\delta/2$. At this point, using our structure Lemma 7.10, this is sufficient for testing k-locality. □

Remark 7.13. We remark that the algorithm for testing locality can be used in more generality for testing if the support of the Hamiltonians is a given $\mathcal{S} \subseteq \{0,1,2,3\}^n$. Also, by a union bound one can test for M supports $\mathcal{S}_1, \ldots, \mathcal{S}_M$ by paying a factor $\log(M)$.

Theorem 7.14. *Let H be a n-qubit Hamiltonian and let $\mathcal{S}_1, \ldots, \mathcal{S}_M \subseteq \{0,1,2,3\}^n$. Then, with $O(1/(\varepsilon_2-\varepsilon_1)^4 \log(M/\delta))$ queries and $O(1/(\varepsilon_2-\varepsilon_1)^3 \log(M/\delta))$ total evolution time one can simultaneously, for every $i \in [M]$, test if H is ε_1-close or or ε_2-far from being supported on $\mathcal{S}_i$.*

Theorem 7.12 is one case of Theorem 7.14 where $M = 1$ and $\mathcal{S}_1 = \{x \in \{0,1,2,3\}^n : |x| \leq k\}$.

7.4.2 Testing sparse Hamiltonians

Now we state our sparsity testing algorithm and prove its guarantees.

Algorithm 3 Fully tolerant sparsity tester

Input: Query access to the time evolution of $U(t) = e^{-itH}$, closeness and farness parameters $\varepsilon_1,\ \varepsilon_2 \in (0,1)$, sparsity parameter $s \in \mathbb{N}$ and failure parameter $\delta \in (0,1)$

1: Set $T = O(s^6/(\varepsilon_2^2-\varepsilon_1^2)^6 \cdot \log(1/\delta))$
2: Let $t = O((\varepsilon_2^2-\varepsilon_1^2)/s)$ and $U = U(t)$
3: Perform Pauli sampling from U a total of T times. Let $(|\alpha_x|^2)_{x\in\{0,1,2,3\}^n}$ the empirical estimate of $(|\widehat{U}_x|^2)_x$ obtained this way.
4: Let $|\alpha_{x_1}|^2, \ldots, |\alpha_{x_s}|^2$ the s-biggest elements of $(|\alpha_x|^2)_{x\in\{0,1,2,3\}^n-\{0^n\}}$
5: Set $\Gamma = |\alpha_{0^n}|^2 + \sum_{i\in[s]} |\alpha_{x_i}|^2$.

Output: If $\Gamma \geq 1 - \varepsilon_1^2 \frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2} - \frac{1}{2}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2}$ output that H is close to sparse, and far from sparse otherwise

Theorem 7.15. *Algorithm 3 solves the s-sparsity testing problem with probability $\geq 1-\delta$, by making $O(s^6/(\varepsilon_2^2-\varepsilon_1^2)^6 \cdot \log(1/\delta))$ queries to the evolution operator and with $O(s^5/(\varepsilon_2^2-\varepsilon_1^2)^5 \cdot \log(1/\delta))$ total evolution time.*

Proof. Let $t = O((\varepsilon_2^2-\varepsilon_1^2)/s)$. By Lemma 7.11 we have that if H is ε_1-close to being sparse, then

$$\mathsf{TopEnergy}(t;s) \geq 1 - \varepsilon_1^2 \frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2} - \frac{1}{3}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2},$$

while if H is ε_2-far from s-sparse, then

$$\mathsf{TopEnergy}(t;s) \leq 1 - \varepsilon_2^2 \frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2} + \frac{1}{3}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2}.$$

From here, it follows that to test it suffices to estimate $\mathsf{TopEnergy}(t;s)$ up to error

$$\begin{aligned}\varepsilon &= \frac{1}{2}\left(1 - \varepsilon_1^2\frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2} - \frac{1}{3}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2} - \left\{1 - \varepsilon_2^2\frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2} + \frac{1}{3}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2}\right\}\right)\\ &= \frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{6s^2}.\end{aligned}$$

To do that we will obtain an estimate $(\{|\alpha_x|^2\}_x$ of $\{|\widehat{U}_x|^2\}_x$ and use it to approximate $\mathsf{TopEnergy}(t;s)$. Using Fact 2.6, we obtain an empirical distribution $\{|\alpha_x|^2\}_x$ that is obtained after $T = O(s^2\log(1/\delta)/\varepsilon^2)$ samples from $\{|\widehat{U}_x|^2\}_x$ (which can be performed with T queries to $U(t)$ thanks to Fact 7.8) satisfies that

$$\left||\alpha_x|^2 - |\widehat{U}_x|^2\right| \leq \frac{\varepsilon}{2s+1} \tag{7.6}$$

for all $x \in \{0,1,2,3\}^n$ with probability $\geq 1-\delta$. We assign new labels $y_0, y_1, \ldots, y_{4^n-1}$ to $\{0,1,2,3\}^n$ in a way such that $|\alpha_{y_0}|^2 = |\alpha_{0^n}|^2$ and $|\alpha_{y_1}|^2 \geq \cdots \geq |\alpha_{y_{4^n-1}}|^2$. Now, we define our estimate for $\mathsf{TopEnergy}(t;s)$ as

$$\mathsf{TopEnergy}'(t;s) = |\alpha_{y_0}(t)|^2 + 2\sum_{i\in[s]}|\alpha_{y_i}(t)|^2.$$

It only remains to show that $\mathsf{TopEnergy}'(t;s)$ ε-approximates $\mathsf{TopEnergy}(t;s)$. We will see that in two steps. First,

$$\begin{aligned}\mathsf{TopEnergy}'(t;s) &= |\alpha_{y_0}(t)|^2 + 2\sum_{i\in[s]}|\alpha_{y_i}(t)|^2\\ &\geq |\alpha_{x_0}(t)|^2 + 2\sum_{i\in[s]}|\alpha_{x_i}(t)|^2\\ &\geq |u_{x_0}(t)|^2 + 2\sum_{i\in[s]}|u_{x_i}(t)|^2 - \varepsilon\\ &= \mathsf{TopEnergy}(t;s) - \varepsilon,\end{aligned}$$

where the second line is true by definition of $y_0, \ldots, y_{4^n-1}$ and the third line is true because Eq. (7.6). Switching the roles of $\mathsf{TopEnergy}'(t;s)$ and $\mathsf{TopEnergy}(t;s)$, one can

prove that $\mathsf{TopEnergy}(t;s) \geq \mathsf{TopEnergy}'(t;s) - \varepsilon$.

Complexity analysis. We have queried $U(t)$ a total of $T = O(s^2 \log(1/\delta)/\varepsilon^2)$ times with $\varepsilon = (\varepsilon_2^2 - \varepsilon_1^2)^3/6s^2$ and $t = O((\varepsilon_2^2 - \varepsilon_1^2)/s)$, so the number of queries is

$$O\left(\frac{s^6}{(\varepsilon_2^2 - \varepsilon_1^2)^6} \log(1/\delta)\right)$$

and the total evolution time

$$O\left(\frac{s^5}{(\varepsilon_2^2 - \varepsilon_1^2)^5} \log(1/\delta)\right).$$

□

Furthermore, for the regime where $\varepsilon_1 = O(\varepsilon_2/s^{0.5})$ we propose a more efficient testing algorithm.

Algorithm 4 Not that tolerant sparsity tester

Input: Query access to the time evolution of $U(t) = e^{-itH}$, sparsity parameter $s \in \mathbb{N}$, closeness and farness parameters $\varepsilon_1,\ \varepsilon_2 \in (0,1)$ satisfying $\varepsilon_1 = O(\varepsilon_2/\sqrt{s})$ and failure parameter $\delta \in (0,1)$

1: Set $T = O(s^2/\varepsilon_2^4 \cdot \log(1/\delta))$
2: Let $t = \Omega(\varepsilon_2/\sqrt{s})$ and $U = U(t)$
3: Perform Pauli sampling from U a total of T times. Let $\mathcal{X}$ the set of sampled Paulis.

Output: If $|\mathcal{X} - \{0^{2n}\}| \leq s$ output that H is close to sparse, and far from sparse otherwise

Theorem 7.16. *Let H be a traceless Hamiltonian with $\|H\|_{\mathrm{op}} \leq 1$. Provided that $\varepsilon_1 = O(\varepsilon_2/s^{0.5})$, Algorithm 4 solves the s-sparsity testing problem with probability $\geq 1 - \delta$. The algorithm makes $O(s^2/\varepsilon_2^4 \cdot \log(1/\delta))$ queries to the evolution operator and uses $O(s^{1.5}/\varepsilon_2^3 \cdot \log(1/\delta))$ total evolution time.*

Proof. Let $C > 1$ be a constant that appears in the first-order Taylor expansion,

$$U(t) = \mathrm{Id} - itH + Ct^2 R_1(t)$$

with $\|R_1\|_{\mathrm{op}} \leq 1$ for $t \in (0,1)$. We will assume that $\delta = 1/3$, as the case $\delta \in (0, 1/3)$ follows by a standard majority voting argument. Algorithm 4 is simple. One just performs Pauli sampling of $U = U(t)$ a number of T times, for some t and T to be determined later. Let $\mathcal{X}$ be the labels of the Pauli strings sampled in this process. If

$|\mathcal{X} - \{0^{2n}\}| \leq s$ we output that H is sparse, and otherwise we output that is far from sparse. It remains to analyze the correctness.

Correctness. In the case that H is ε_1-close s-sparse, there exists $\mathcal{S} \subset \{0,1,2,3\}^n$ of size s where H is ε_1-concentrated. Then, by Taylor expansion,

$$\sqrt{\sum_{x \notin (\mathcal{S} \cup \{0^{2n}\})} |\widehat{U}_x|^2} \leq t \sqrt{\sum_{x \notin (\mathcal{S} \cup \{0^{2n}\})} |\lambda_x|^2} + Ct^2 \leq t\varepsilon_1 + Ct^2 \leq 2Ct^2,$$

where in the last inequality we have assumed that

$$\varepsilon_1 \leq Ct. \tag{7.7}$$

Hence, the probability of sampling an element outside $\mathcal{S} \cup \{0^{2n}\}$ in one sample is at most $4C^2t^4$. Thus, the probability of not sampling an element outside $\mathcal{S} \cup \{0^{2n}\}$ in T samples is at least

$$(1 - 4C^2t^2)^T \geq 1 - 4C^2t^4T.$$

In particular, if

$$T \leq \frac{1}{3}\frac{1}{4C^2t^4} \tag{7.8}$$

it will be satisfied that $|\mathcal{X} - \{0^{2^n}| \leq s$ with probability $\geq 2/3$, as desired.

In the case that H is ε_2-far from s-sparse, we will perform an analysis similar to the coupon collector problem. By Taylor expansion we have that for every set $\mathcal{S}$ of size s,

$$\sqrt{\sum_{x \notin (\mathcal{S} - \{0^{2n}\})} |\widehat{U}_x|^2} \geq \varepsilon_2 t - Ct^2 \geq \frac{\varepsilon_2 t}{2}, \tag{7.9}$$

where we have assumed that

$$Ct \leq \varepsilon_2/2. \tag{7.10}$$

Let X_i the random variable that accounts for the number of samples between the $(i-1)$-th sampled non-0^{2n}-Pauli and the i-th sampled non-0^{2n}-Pauli. Applying Eq. (7.9) to every $\mathcal{X}_i$, it follows that $\mathbb{E}[X_i] \leq 4/\varepsilon_2^2 t^2$ for every $i \in [s+1]$, so

$$\mathbb{E}[X_1 + \cdots + X_{s+1}] \leq \frac{4(s+1)}{\varepsilon_2^2 t^2}.$$

Hence, by Markov's inequality, if

$$T \geq \frac{\sqrt{34}(s+1)}{\varepsilon_2^2 t^2} \tag{7.11}$$

it will be satisfied that $|\mathcal{X} - \{0^{2n}\}| \geq s+1$ with probability $\geq 2/3$, as desired.

Finally, we note that we have assumed conditions Eqs. (7.7), (7.8), (7.10) and (7.11) to ensure the correctness of the algorithm. All these equations are satisfied provided that

$$\begin{aligned} t &= \frac{\varepsilon_2}{\sqrt{50C^2(s+1)}} = \Omega\left(\frac{\varepsilon_2}{\sqrt{s}}\right), \\ T &= \frac{1}{12C^2t^4} = O\left(\frac{s^2}{\varepsilon_2^4}\right), \\ \varepsilon_1 &\leq \frac{\varepsilon_2}{\sqrt{50(s+1)}} = O\left(\frac{\varepsilon_2}{\sqrt{s}}\right). \end{aligned}$$

□

7.5 Learning Hamiltonians

7.5.1 Learning unstructured Hamiltonians

We start by showing how to efficiently learn an arbitrary n-qubit Hamiltonian in ℓ_∞ error. To do that, we propose a protocol to estimate a given set of Pauli coefficients $\mathcal{X}$ of a Hamiltonian via Shadow tomography. To describe the protocol, we introduce the following $2n$-qubit observables. Given $x \in \{0,1,2,3\}^n$, we define

$$\begin{aligned} \mathcal{R}_x &:= \frac{1}{2}(|\mathrm{Bell}_{0^{2n}}\rangle\langle\mathrm{Bell}_x| + |\mathrm{Bell}_x\rangle\langle\mathrm{Bell}_{0^{2n}}|), \\ \mathcal{I}_x &:= \frac{1}{2}(-i|\mathrm{Bell}_{0^{2n}}\rangle\langle\mathrm{Bell}_x| + i|\mathrm{Bell}_x\rangle\langle\mathrm{Bell}_{0^{2n}}|). \end{aligned}$$

Lemma 7.17. *Let H be an n-qubit traceless Hamiltonian and $\mathcal{X} \subseteq \{0,1,2,3\}^n$. Then, Algorithm 5 allows one to estimate the Pauli coefficients corresponding to $\mathcal{X}$ with success probability $\geq 1-\delta$. It uses $O((\log|\mathcal{X}|/\delta)\|H\|^4/\varepsilon^4)$ queries and $O(\log(|\mathcal{X}|/\delta)\|H\|^2/\varepsilon^3)$ total evolution time. The minimum evolution time is $\varepsilon/\|H\|^2$, the number of ancillas is n, and the time complexity is $O(\mathrm{poly}(n)|\mathcal{X}|\|H\|^4/\varepsilon^4 \cdot \log(|\mathcal{X}|/\delta))$.*

Proof. **Correctness of the algorithm:** Let $t_0 = \Theta(\varepsilon/\|H\|^2)$ and $U = U(t_0)$. As

Algorithm 5 Estimating a given set of Pauli coefficients of a Hamiltonian

Input: Query access to the time evolution of $U(t) = e^{-itH}$, target set of Pauli coefficients $\mathcal{X} \subseteq \{0,1,2,3\}^n - \{0^n\}$, error parameter $\varepsilon \in (0,1)$, and failure parameter $\delta \in (0,1)$

Set $T = O(\|H\|^4/\varepsilon^4 \cdot \log(|\mathcal{X}|/\delta))$ and $t_0 = \Theta(\varepsilon/\|H\|^2)$
Set $U = U(t_0)$
for $j \in [T]$ **do**
    Prepare $|J(U)\rangle = (U \otimes \mathrm{Id}_{2^n})|\mathrm{Bell}_n\rangle$
    Apply a uniformly random Clifford gate C
    Measure in the computational basis. Let $|b_j\rangle$ be the outcome
    for $x \in \mathcal{X}$ **do**
        Let $\mathcal{R}_{x,j} = (2^n+1)\langle b_j|C^{-1}\mathcal{R}_x C|b_j\rangle$ and $\mathcal{I}_{x,j} = (2^n+1)\langle b_j|C^{-1}\mathcal{I}_x C|b_j\rangle$
    end for
end for
for $x \in \mathcal{X}$ **do**
    Set $\widetilde{R}_x := \mathrm{MedianOfMeans}(\mathcal{R}_{x,j})_j$ and $\widetilde{I}_x := \mathrm{MedianOfMeans}(\mathcal{I}_{x,j})_j$
end for

Output: $((\widetilde{R}_x + i\widetilde{I}_x)/(-it))_{x\in\mathcal{X}}$

$\mathrm{Tr}[\mathcal{R}_x^2] = \mathrm{Tr}[\mathcal{I}_x^2] = 2$, by Theorem 7.9, the numbers $\widetilde{R}_x$ and $\widetilde{I}_x$ that Algorithm 5 outputs satisfy

$$|\,\mathrm{Tr}[\mathcal{R}_x|J(U)\rangle\langle J(U)|] - \widetilde{R}_x| \le \frac{\varepsilon^2}{\|H\|^2}, \quad |\,\mathrm{Tr}[\mathcal{I}_x|J(U)\rangle\langle J(U)|] - \widetilde{I}_x| \le \frac{\varepsilon^2}{\|H\|^2}, \tag{7.12}$$

for every $x \in \mathcal{X}$ with probability $\ge 1-\delta$. By Taylor expansion, as $\lambda_{0^{2n}} = 0$, we have that $|\widehat{U}_{0^{2n}} - 1| \le O(t_0^2\|H\|^2)$. Thus,

$$\mathrm{Tr}[\mathcal{R}_x|J(U)\rangle\langle J(U)|] = \frac{1}{2}(\widehat{U}_x\widehat{U}_{0^{2n}}^* + \widehat{U}_{0^{2n}}\widehat{U}_x^*) = \mathrm{Re}(\widehat{U}_x\widehat{U}_0^*) = \mathrm{Re}(\widehat{U}_x) \pm O(t_0^2\|H\|^2), \tag{7.13}$$

and similarly $\mathrm{Tr}[\mathcal{I}_x|J(U)\rangle\langle J(U)|] = \mathrm{Im}(\widehat{U}_x) \pm O(t_0^2\|H\|^2)$. Hence, combining Eqs. (7.12) and (7.13) we have that

$$|\widehat{U}_x - (\widetilde{R}_x + i\widetilde{I}_x)| \le \frac{\varepsilon^2}{\|H\|^2} + O(t_0^2\|H\|^2) \le O\left(\frac{\varepsilon^2}{\|H\|^2}\right),$$

for every $x \in \mathcal{X}$. Finally, by Taylor expansion we have that $|\widehat{U}_x/(-it_0) - \lambda_x| \le O(t_0\|H\|^2)$, so

$$\left|\lambda_x - \frac{\widetilde{R}_x + i\widetilde{I}_x}{-it_0}\right| \le O\left(\frac{\varepsilon^2}{t_0\|H\|^2}\right) + O(t_0\|H\|^2) = O(\varepsilon),$$

for every $x \in \mathcal{X}$, as claimed.

Time complexity: The time complexity is dominated by the first loop in Algorithm 5, whose time complexity is $O(|\mathcal{X}| \cdot T \cdot (t_{est} + \text{poly}(n))$, where the poly$(n)$ comes from applying a random Clifford gate and t_{est} is the time taken to compute $\langle b|C^{-1}\mathcal{R}_x C|b\rangle$ for an n-qubit Clifford gate C and a computational basis state $|b\rangle$. Now, expanding R_x one can write $\langle b|C^{-1}\mathcal{R}_x C|b\rangle$ as an algebraic expression of a finite number of terms of the kind $\langle y|D|z\rangle$, where $|y\rangle$ and $|z\rangle$ are computational basis states and D a Clifford gate. Hence, via Gottesman-Knill theorem [Got98, AG04] follows that $t_{est} = O(n^2)$, so the total time complexity is $O(\text{poly}(n)|\mathcal{X}|\|H\|^4/\varepsilon^4 \cdot \log(|\mathcal{X}|/\delta))$. □

Now, we are ready to present our learning algorithm for arbitrary Hamiltonians with no promise about its structure.

Algorithm 6 Learning unstructured Hamiltonians

Input: Query access to the time evolution of $U(t) = e^{-itH}$, error parameter $\varepsilon \in (0,1)$, and failure parameter $\delta \in (0,1)$

1: Set $T = O(\|H\|^4/\varepsilon^4 \cdot \log(\|H\|^2/\varepsilon^2\delta))$ and $t_0 = \Theta(\varepsilon/\|H\|^2)$
2: Set $U = U(t_0)$
3: Set $\mathcal{X} = \emptyset$
4: **for** $j \in [T]$ **do**
5: Prepare $|J(U)\rangle = (U \otimes \text{Id}_{2^n})|\text{Bell}_n\rangle$
6: Measure in the Bell basis and add the outcome $x \in \{0,1,2,3\}^n$ to $\mathcal{X}$ if $x \neq 0^{2^n}$
7: **end for**
8: Run Algorithm 5 run with $U(t)$, $\mathcal{X}$, ε and δ as inputs. Let $(\widetilde{\lambda}_x)_{x\in\mathcal{X}}$ the output.

Output: $\widetilde{H} = \sum_{x\in\mathcal{X}} \widetilde{\lambda}_x \sigma_x$

Theorem 7.18 (Learning unstructured Hamiltonians)**.** *Let H be an n-qubit and traceless Hamiltonian. Then, Algorithm 6 ε-learns all Pauli coefficients of H with success probability $\geq 1-\delta$. It uses $\widetilde{O}((\|H\|/\varepsilon)^4)$ queries to the evolution operator and $\widetilde{O}(\|H\|^2/\varepsilon^3)$ total evolution time. The minimum evolution time is $\Theta(\varepsilon/\|H\|^2)$, the algorithm uses n ancilla qubits and only one round of adaptivity, and the time complexity is* poly$(n, 1/\varepsilon, \|H\|)$.

Proof. Let $t_0 = \Theta(\varepsilon/\|H\|^2)$ and $U = U(t_0)$ and let $T = O(\|H\|^4/\varepsilon^4 \cdot \log(\|H\|^2/\varepsilon^2\delta))$, as in Algorithm 6.

Correctness: We claim that with probability $\geq 1-\delta$ the set $\mathcal{X}$ generated in Algorithm 6 contains all x such that

$$|\lambda_x| \geq \varepsilon, \tag{7.14}$$

and that

$$|\mathcal{X}| \leq \widetilde{O}\left(\frac{\|H\|^4}{\varepsilon^4}\right). \tag{7.15}$$

To show Eq. (7.14) we note that by Taylor expansion, if $|\lambda_x| \geq \varepsilon$, then $|\widehat{U}_x| = \Omega((\varepsilon^2/\|H\|^2))$, so $|\widehat{U}_x|^2 = \Omega((\varepsilon^4/\|H\|^4))$. Hence, the probability that such an x does not belong to $\mathcal{X}$, which stores the non-0^{2n} outcomes of sampling from $(|\widehat{U}_x|^2)_x$, is at most

$$\left(1 - |\widehat{U}_x|^2\right)^T \leq e^{-T|\widehat{U}_x|^2} \leq \frac{\varepsilon^2\delta}{\|H\|^2}.$$

Hence, as there is at most $\|H\|^2/\varepsilon^2$ coefficients with $|\lambda_x| \geq \varepsilon$, because $\sum_x |\lambda_x|^2 \leq \|H\|^2$, Eq. (7.14) follows from a union bound. Eq. (7.15) holds because $|\mathcal{X}| \leq T$.

Now, if Eqs. (7.14) and (7.15) are satisfied, Algorithm 5 provides estimates of the coefficients of $\mathcal{X}$, which contains all labels x of coefficients $|\lambda_x| \geq \varepsilon$.

Complexities: The query complexity is $2T = \tilde{O}(\|H\|^4/\varepsilon^4)$, the minimum evolution time $t_0 = \Theta(\varepsilon/\|H\|^2)$ and the total time evolution $2Tt_0 = \widetilde{O}(\|H\|^2/\varepsilon^3)$. Additionally, the time complexity of Algorithm 6 is dominated by the call to Algorithm 5, which runs in time $O(\text{poly}(n)|\mathcal{X}|\|H\|^2/\varepsilon^2)$, which thanks to Eq. (7.15) is $\text{poly}(n, 1/\varepsilon, \|H\|)$. □

7.5.2 Learning local Hamiltonians

We now introduce our local Hamiltonian learner and prove its guarantees.

Algorithm 7 Local Hamiltonian learner

Input: Query access to the time evolution of $U(t) = e^{-itH}$, error parameter $\varepsilon \in (0, 1)$, locality parameter $k \in \mathbb{N}$ and failure parameter $\delta \in (0, 1)$

Set $T = \exp(O(k^2 + k\log(1/\varepsilon))\log(1/\delta)$
Let $t = \varepsilon^{k+1}\exp(-k(k+1)/2)$ and $U = U(t)$
Set $\gamma = (\varepsilon/\|H\|^2)^{k+1}\exp(-k(k+1)/2)$ and $\beta = \gamma\varepsilon/\|H\|$
Learn β-estimates λ'_x of λ_x via Algorithm 6
for $|x| \leq k$ **do**
    if $|\lambda'_x| \leq \gamma$ **then**
        $\widetilde{\lambda}_x = 0$
    else
        $\widetilde{\lambda}_x = \lambda'_x$
    end if
end for

Output: $\sum_{x \leq k} \widetilde{\lambda}_x \sigma_x$

Theorem 7.19. *Given a n-qubit k-local Hamiltonian H, Algorithm 7 outputs $\widetilde{H}$ such that with probability $\geq 1-\delta$ satisfies $\|H-\widetilde{H}\|_{\ell_2} \leq \varepsilon$. The algorithm makes $\exp(O(k^2 + k\log(\|H\|^2/\varepsilon))\log(1/\delta)$ queries to the evolution operator with $\exp(O(k^2 + k\log(\|H\|^2/\varepsilon))\log(1/\delta)$ total evolution time.*

To prove this theorem, we use the non-commutative Bohnenblust-Hille inequality by Volberg and Zhang [VZ23].

Theorem 7.20 (Non-Commutative Bohnenblust-Hille inequality)**.** *Let $H = \sum_x \lambda_x \sigma_x$ be a k-local Hamiltonian. Then, there is a universal constant C such that*

$$\widetilde{H} = \sum_{x\in\{0,1,2,3\}^n} |\lambda_x|^{\frac{2k}{k+1}} \leq C^k \|H\|.$$

Proof of Theorem 7.19. We only analyze the correctness of Algorithm 7, as the complexity quickly follows from Theorem 7.18. In this proof we also use the notation of Algorithm 7. The ℓ_2-error of approximating H with $\widetilde{H}$ is

$$\|\widetilde{H}-H\|_{\ell_2}^2 = \sum_{|\lambda'_x|\leq\gamma} |\lambda_x|^2 + \sum_{|\lambda'_x|\geq\gamma, |x|\leq k} |\lambda_x - \lambda'_x|^2. \tag{7.16}$$

We show separately that the two terms are at most $O(\varepsilon^2)$. To bound the contribution of the small Pauli coefficients, we first note that by Theorem 7.18 we have that

$$|\lambda'_x| \leq \gamma \quad \Longrightarrow \quad |\lambda_x| \leq \gamma + \beta = O(\gamma). \tag{7.17}$$

Hence,

$$\sum_{|\lambda'_x|\leq\gamma} |\lambda_x|^2 \leq \sum_{|\lambda_x|\leq O(\gamma)} |\lambda_x|^2 \leq O(\gamma^{\frac{2}{k+1}}) \sum_{x\in\{0,1,2,3\}^n} |\lambda_x|^{\frac{2k}{k+1}} \leq \gamma^{\frac{2}{k+1}} (C^k\|H\|^2)^{\frac{2k}{k+1}} = O(\varepsilon), \tag{7.18}$$

where in the first inequality we have used Eq. (7.17), in the third inequality we have used Theorem 7.20 and in the last inequality that $\gamma = (\varepsilon/\|H\|^2)^{k+1}\exp(-k(k+1)/2)$. To bound the contribution of the coefficients $|\lambda_x| \geq \gamma$ we notice that there is at most $\|H\|^2/\gamma^2$ of them, because $\sum_x |\lambda_x|^2 \leq \|H\|^2$. Thus,

$$\sum_{|\lambda'_x|\geq\gamma, |x|\leq k} |\lambda_x - \lambda'_x|^2 \leq \frac{\|H\|^2}{\gamma^2} \sup_x |\lambda_x - \lambda'_x|^2 \leq \frac{\|H\|^2\beta^2}{\gamma^2} = \varepsilon^2,$$

where in the second inequality we use the λ'_x are β-estimates of λ_x and in the last equality we use that $\beta = \gamma\varepsilon/\|H\|$. □

7.5.3 Learning sparse Hamiltonians

In this section we introduce our sparse Hamiltonian learner and prove its guarantees.

Algorithm 8 Sparse Hamiltonian learner

Input: Query access to the time evolution of $U(t) = e^{-itH}$, error parameter $\varepsilon \in (0,1)$, sparsity parameter $s \in \mathbb{N}$ and failure parameter $\delta \in (0,1)$

1: Learn $(\varepsilon/2)$-estimates λ'_x of λ_x via Algorithm 6
2: **for** $x \in \{x : \lambda_x \neq 0\}$ **do**
3: **if** $\lambda'_x \leq \varepsilon/2$ **then**
4: $\widetilde{\lambda}_x = 0$
5: **else**$\lambda_x > \varepsilon/2$
6: $\widetilde{\lambda}_x = \lambda'_x$
7: **end if**
8: **end for**

Output: $\widetilde{H} = \sum_x \widetilde{\lambda}_x \sigma_x$

Theorem 7.21 (Sparse Hamiltonian learning)**.** *Given an n-qubit, s-sparse Hamiltonian H, Algorithm 8 outputs another Hamiltonian $\widetilde{H} = \sum \widetilde{\lambda}_x \sigma_x$ such that with probability $\geq 1-\delta$ satisfies $\|H - \widetilde{H}\|_{\ell_\infty} \leq \varepsilon$, The algorithms uses $\widetilde{O}(\|H\|^4/\varepsilon^4)$ queries and $\widetilde{O}(\|H\|^2/\varepsilon^3)$ total evolution time.*

Furthermore, if $\lambda_x = 0$, then $\widetilde{\lambda}_x = 0$. This implies that running Algorithm 8 with $\varepsilon = \varepsilon'/\sqrt{s}$ outputs $\widetilde{H}$ such that $\|H - \widetilde{H}\|_{\ell_2} \leq \varepsilon'$. In this case, the algorithm uses $\widetilde{O}(\|H\|^4 s^2/\varepsilon'^4)$ queries and $\widetilde{O}(\|H\|^2 s^{1.5}/\varepsilon'^3)$ total evolution time.

Proof. The first part, concerning learning in the ℓ_∞ error follows from Theorem 7.18. The fact that $\lambda_x = 0$, then $\widetilde{\lambda}_x = 0$ follows from Line 3 of Algorithm 8. Finally, we note that having $\lambda_x = 0 \implies \widetilde{\lambda}_x = 0$ and $|\lambda_x - \lambda_x| \leq \varepsilon'/\sqrt{s}$, implies $\|H - \widetilde{H}\|_{\ell_2} \leq \varepsilon'$. Indeed,

$$\|H - \widetilde{H}\|_{\ell_2} = \sum_{\lambda_x \neq 0} |\lambda_x - \widetilde{\lambda}_x|^2 \leq s \sup_x |\lambda_x - \widetilde{\lambda}_x|^2 = \varepsilon'^2,$$

where in the first step we have used that $\lambda_x = 0 \implies \widetilde{\lambda}_x = 0$, in the second that $|\lambda_x - \lambda_x| \leq \varepsilon'/\sqrt{s}$ and in the third that H is s-sparse. □

Part III

Bonus

Chapter 8

Cute remarks

In this chapter, we gather new proofs of known results that we find elegant and concise. All of them relate to functions defined on the Boolean cube.[1]

8.1 Generalizing a work of Kalai and Schulman

Kalai and Schulman studied the influences of multilinear polynomials with $\{-1, 0, 1\}$-valued coefficients [KS19] (we refer to their work for motivation). They showed an upper bound of $\sum_{i\in[n]} \sqrt{\mathrm{Inf}_i[p]}$ in terms of $\|p\|_\infty$. They proved that

$$\sum_{i\in[n]} \sqrt{\mathrm{Inf}_i[p]} \leq 3^d d^{5/2} \|p\|_\infty$$

for unimodular polynomials. We can improve this bound, generalize it to arbitrary polynomials (not necessarily unimodular), and show that the exponential dependence on d is necessary. Our proof is simple, short and based on hypercontractivity [Bon70] and a bound on the sum of L_1 influences [BB14, FHKL16].

Before diving into the proof of the main result of this section, Proposition 8.3, we need to the define the L_q influences and state two results that we use as lemmas. The L_q influence is defined by

$$\mathrm{Inf}_i^q[p] = \mathbb{E}_x \left[\left| \frac{p(x^{i\to 1}) - p(x^{i\to -1})}{2} \right|^q \right],$$

[1] The results of Section 8.1 were derived in a conversation with Miquel Saucedo during a research stay in Hausdorff Institute for Mathematics, in Bonn, Germany.

where the expectation is taken with respect to the uniform distribution on $\{-1,1\}^n$. Note that $\mathrm{Inf}^2[p]$ equals the $\mathrm{Inf}[p]$ defined in Section 2.5. In the Boolean case, $\mathrm{Inf}_i^2[p] = \mathrm{Inf}_i^q[p]$ for every $q \in [1,\infty)$.

Theorem 8.1 (Hypercontractivity)**.** *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most d. Then,*

$$\sqrt{\mathbb{E}_x\,|p(x)|^2} \le e^d \mathbb{E}_x |p(x)|.$$

Theorem 8.2 (Bound on sum of L_1 influences)**.** *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most d. Then,*

$$\sum_{i\in[n]} \mathrm{Inf}_i^1[p] \le d^2 \|p\|_\infty.$$

Proposition 8.3. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree d. Then,*

$$\sum_{i\in[n]} \sqrt{\mathrm{Inf}_i^2[p]} \le e^d d^2 \|p\|_\infty.$$

In addition, there is a unimodular degree d polynomial p such that

$$\sum_{i\in[n]} \sqrt{\mathrm{Inf}_i^2[p]} \ge \sqrt{2}^{d-2} \|p\|_\infty.$$

Proof. By Theorem 8.1 it follows that for every $i \in [n]$

$$\sqrt{\mathrm{Inf}_i^2[p]} \le e^d \mathrm{Inf}_i^1[p].$$

Now, taking the sum over $i \in [n]$ and applying Theorem 8.2 we arrive at the claimed result.

Let $n = 2^{d-1}$. The (unnormalized) address function of $p : (\{-1,1\})^n)^d \to \mathbb{R}$ of degree d is defined as

$$p(x) = \sum_{a\in\{-1,1\}^{d-1}} \underbrace{(x_1(1) - a_1 x_1(2)) \dots (x_{d-1}(1) - a_{d-1} x_{d-1}(2))}_{g_a(x_1,\dots,x_{d-1})} x_d(a), \tag{8.1}$$

where we identify $\{-1,1\}^{d-1}$ with $[2^{d-1}]$. It is satisfied that $\|p\|_\infty = 2^{d-1}$, because given $(x_1,\dots,x_{d-1}) \in (\{-1,1\}^n)^{d-1}$ there is only one $a \in \{-1,1\}^{d-1}$ such that $g_a(x_1,\dots,x_{d-1})$ is not 0, in which case it takes the value $\pm 2^{d-1}$. For every of the 2^{d-1} variables $x_d(a)$ we have that $\mathrm{Inf}_{d,a}^2[p] = 2^{d-2}$, so $\sum \sqrt{\mathrm{Inf}_i^2[p]} \ge (2^{d-2})^{3/2}$. □

8.2 The adversary method via Grothendieck's inequality

In the first part of this thesis, we have focused on the polynomial method. Here, we will revisit the other main method to prove quantum query lower bounds: the adversary method [Amb00, HLv07] (see [LS21] for a survey). To define the adversary bound we must introduce some notation. Let $f : \{-1,1\}^n \to \{-1,1\}$. A matrix $\Gamma \in M_{2^n}(\mathbb{C})$ is an *adversary matrix for* f if it is Hermitian and for every $x, y \in \{-1,1\}^n$ such that $f(x) = f(y)$ we have that

$$\langle x|\Gamma|y\rangle = 0,$$

where $\{|x\rangle\}$ is an orthonormal basis of $\mathbb{C}^{2^n}$. Given $i \in [n]$, $D_i \in M_{2^n}$ is the matrix defined by

$$\langle x|D_i|y\rangle = \begin{cases} 0 & \text{if } x_i = y_i, \\ 1 & \text{if } x_i \neq y_i. \end{cases}$$

The *adversary bound of* f is defined by

$$\mathrm{Adv}(f) := \sup_{\Gamma} \frac{\|\Gamma\|_{\mathrm{op}}}{\max_{i\in[n]} \|\Gamma \circ D_i\|_{\mathrm{op}}}, \tag{8.2}$$

where the supremum runs over all adversary matrices Γ and $\circ$ denotes the entry-wise matrix product, namely $(A \circ B)_{ij} = A_{ij}B_{ij}$. In this section, we will give a, to the best of our knowledge, novel proof of the following result via Grothendieck's inequality (see Section 2.7.1).

Proposition 8.4. *Let* $f : \{-1,1\}^n \to \{-1,1\}$. *Then,* $Q(f) = \Omega(\mathrm{Adv}(f))$.

Before diving into the proof, we give an intuition of why such a result holds. Consider an algorithm whose bias approximates f with high probability. Before making any query, the algorithm prepares a state that does not depend on the input x. In terms of adversary matrices Γ, this will mean that some closeness measure, to be defined below, will have value $\|\Gamma\|_{\mathrm{op}}$ before making any query. Also, at the end of the algorithm, the state prepared on a pair of inputs x and y such that $f(x) \neq f(y)$ must be *far away*, so the algorithm can distinguish them with a measurement. In terms of Γ, this will be formalized via Grothendieck's inequality and will mean that the closeness measure will have value $\leq K_G/5 \cdot \|\Gamma\|_{\mathrm{op}}$ at the end of the algorithm. Finally, it will follow from a simple argument that the algorithm can only separate the states prepared when querying x and y a bounded amount per query. In terms of

Γ, this will mean that the closeness measure can only decrease $2\max_i \|\Gamma \circ D_i\|_{\mathrm{op}}$ per query. Putting everything together we have that the algorithm must make at least

$$\frac{\|\Gamma\|_{\mathrm{op}} - \frac{K_G}{5}\|\Gamma\|_{\mathrm{op}}}{2\max_{i\in[n]} \|\Gamma \circ D_i\|_{\mathrm{op}}}$$

queries.

Proof. We will show that $Q_{2/100}(f) = \Omega(\mathrm{Adv}(f))$. Let $\mathcal{A}$ be an algorithm that makes t queries and whose bias $2/100$ approximates $f(x)$. This means, that $\mathcal{A}$ fails with probability $\|\Pi_{-f(x)}|\psi_x^t\rangle\|_2^2 \leq 1/100$. Let Γ be an adversary matrix for f. Let $|\delta\rangle$ be such that $|\langle\delta|\Gamma|\delta\rangle| = \|\Gamma\|_{\mathrm{op}}$. We define the closeness measure at step $s \in \{0,\ldots,t\}$ as

$$\mathcal{C}^s := |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^s|\psi_y^s\rangle|,$$

where $|\psi_x^s\rangle$ is the state prepared by the algorithm on input x just after the sth query. We divide the rest of the proof in three steps. First, we note that

$$\mathcal{C}^0 = |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^0|\psi_y^0\rangle| = |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y| = |\langle\delta|\Gamma|\delta\rangle| = \|\Gamma\|_{\mathrm{op}},$$

where we have used that $|\psi_x^0\rangle$ does not depend on x because no queries have been made.

Second, we claim that

$$\mathcal{C}^t \leq \frac{K_G}{5}\|\Gamma\|_{\mathrm{op}}$$

where K_G is the (complex) Grothendieck's constant, which is strictly smaller than 5. Indeed, let Π_{-1}, Π_1 be the measurement performed by the algorithm, then

$$\begin{aligned}
\mathcal{C}^t &= |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^t|\psi_y^t\rangle| = |\sum_{x,y:f(x)\neq f(y)} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^t|(\Pi_{-1}+\Pi_1)|\psi_y^t\rangle| \\
&\leq |\sum_{x,y:f(x)\neq f(y)} \Gamma_{xy}\delta_x^*\delta_y(\langle\psi_x^t|\Pi_{-f(x)})|\psi_y^t\rangle| + |\sum_{x,y:f(x)\neq f(y)} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^t|(\Pi_{-f(y)}|\psi_y^t\rangle)| \\
&\leq \frac{2}{10}\sup_{\|u_x\|_2,\|v_y\|_2\leq 1} |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\langle u_x, v_y\rangle| \\
&\leq \frac{2K_G}{10}\sup_{\alpha_x,\beta_y\in\{-1,1\}} |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\alpha_x\beta_y| \\
&= \frac{K_G}{5}\|\Gamma\|_{\mathrm{op}},
\end{aligned}$$

where in the first line we have used that Γ is an adversary matrix; in the third line that $\|\Pi_{-f(x)}|\psi_x^t\rangle\|_2^2$ is the failure probability, so it is smaller than $1/100$; and in the fourth line we have used Grothendieck's inequality, Theorem 2.19.

Finally, we claim that

$$\mathcal{C}^s - \mathcal{C}^{s-1} \leq 2 \max \|\Gamma \circ D_i\|_{\text{op}}$$

for $s \in [t]$. Let U^s be the unitary in between the $(s-1)$th and the sth queries. Recall that O_x acts as the controlled version of $|i\rangle \to x_i|i\rangle$, so $O_x|0\rangle|i\rangle = |0\rangle|i\rangle$ and $O_x|1\rangle|i\rangle = x_i|1\rangle|i\rangle$. Let d be the extra dimensions of the algorithm as in Eq. (2.3). Then,

$$\begin{aligned}
\mathcal{C}^s - \mathcal{C}^{s-1} &\leq |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y(\langle\psi_x^s|\psi_y^s\rangle - \langle\psi_x^{s-1}|\psi_y^{s-1}\rangle)| \\
&= |\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^{s-1}|(O_x \otimes \mathrm{Id}_d)\underbrace{U_s^\dagger U_s}_{\mathrm{Id}_{2nd}}(O_y \otimes \mathrm{Id}_d) - (\mathrm{Id}_{2nd})|\psi_y^{s-1}\rangle| \\
&= |\sum_{ij}\sum_{x,y} \Gamma_{xy}\delta_x^*\delta_y\langle\psi_x^{s-1}|(|1i\rangle\langle 1i| \otimes \mathrm{Id}_d)\underbrace{(x_i y_j - 1)}_{-2(D_i)_{xy}}(|1j\rangle\langle 1j| \otimes \mathrm{Id}_d)|\psi_y^{s-1}\rangle| \\
&= 2|\sum_i \sum_{x,y} (\Gamma \circ D_i)_{xy}\delta_x^*\langle\psi_x^{s-1}|(|1i\rangle \otimes \mathrm{Id}_d)(\langle 1i| \otimes \mathrm{Id}_d)|\psi_y^{s-1}\rangle\delta_y|.
\end{aligned}$$

Now, if we define $\widetilde{\Gamma}$ as the block diagonal matrix with $\Gamma \circ D_i$ as diagonal blocks for $i \in [n]$, and G as the block diagonal matrix whose blocks are the Gram matrices of $\{\delta_x(\langle 1i| \otimes \mathrm{Id}_d)|\psi_x^{s-1}\rangle\}_x$, we have that

$$\begin{aligned}
\mathcal{C}^s - \mathcal{C}^{s-1} &= 2|\langle\widetilde{\Gamma}, G\rangle| \leq 2\|\widetilde{\Gamma}\|_{\text{op}}\|G\|_{\text{tr}} \\
&= 2\max_{i\in[n]}\|\Gamma \circ D_i\|_{\text{op}} \sum_i \mathrm{tr}[\mathrm{Gram}(\{\delta_x(\langle 1i| \otimes \mathrm{Id}_d)|\psi_x^{s-1}\rangle\}_x)] \\
&= 2\max_{i\in[n]}\|\Gamma \circ D_i\|_{\text{op}} \underbrace{\sum_x |\delta_x|^2}_{=\langle\delta,\delta\rangle=1} \underbrace{\sum_i \langle\psi_x^{s-1}|(|1i\rangle\langle 1i| \otimes \mathrm{Id}_d)|\psi_x^{s-1}\rangle}_{\leq\langle\psi_x^{s-1}|\psi_x^{s-1}\rangle=1} \\
&= 2\max_{i\in[n]}\|\Gamma \circ D_i\|_{\text{op}}.
\end{aligned}$$

Putting everything together, it follows that

$$t \geq \frac{\mathcal{C}^t - \mathcal{C}^0}{2\max_{i\in[n]}\|\Gamma \circ D_i\|_{\text{op}}} = \Omega\left(\frac{\|\Gamma\|_{\text{op}}}{\max_{i\in[n]}\|\Gamma \circ D_i\|_{\text{op}}}\right). \qquad \square$$

8.3 Average sensitivity lower bounds all reasonable complexity measures

We will show that the average sensitivity $\overline{s}(f)$ of a Boolean function lower bounds all the *reasonable* complexity measures of a Boolean function, which is the list of well-studied complexity measures considered in [ABDK+21]. For total Boolean functions, all of these measures are polynomially related to classical and quantum query complexity. In particular, we will show that the average sensitivity lower bounds the spectral sensitivity of a Boolean function $\lambda(f)$. This is enough, as $\lambda(f)$ lower bounds, up to constant factors, all the reasonable complexity measures. From there, we can easily show that all reasonable complexity measures are $\Omega(n)$ for almost all Boolean functions, concisely reproving previous results such as $Q(f) = \Omega(n)$ for almost all total Boolean functions [Amb99, ABSdW13]. More formally, given $f : \{-1,1\}^n \to \{-1,1\}$ its average sensitivity is defined by

$$\overline{s}(f) := \mathbb{E}_x \sum_{i\in[n]} \left[\left(\frac{f(x) - f(x^{\oplus i})}{2} \right)^2 \right],$$

which also equals the sum of the influences, $\sum_{i\in[n]} \mathrm{Inf}_i^2[f]$. Its spectral sensitivity is given by

$$\lambda(f) := \sup_{\Gamma} \frac{\|\Gamma\|}{\max_{i\in[n]} \|\Gamma \circ D_i\|},$$

where the supremum runs over all adversary matrices that satisfy $\Gamma[x,y] = 0$ if the Hamming distance between x and y is not 1 (see Section 8.2 for the definitions of adversary matrix and D_i).

Proposition 8.5. *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then, $\overline{s}(f) \leq \lambda(f)$. Furthermore, the inequality is tight for $f = \chi_{[n]}$.*

Proof. Let Γ be the adversary matrix such that $\Gamma_{x,y} = 1$ if the Hamming distance between x and y is exactly one and $f(x) \neq f(y)$ and 0 in the other case. Note that Γ can be written as

$$\Gamma_{x,y} = \delta_{x^{\oplus i},y}\delta_{f(x),f(x^{\oplus i})} = \delta_{x^{\oplus i},y} \left(\frac{f(x) - f(x^{\oplus i})}{2} \right)^2.$$

For this matrix, we can see that $\|\Gamma\|_{\mathrm{op}} \geq \overline{s}(f)$ and $\|\Gamma \circ D_i\|_{\mathrm{op}} = 1$ for all $i \in [n]$.

Indeed,

$$\begin{aligned}\|\Gamma\|_{\mathrm{op}} &\geq \sum_{x,y\in\{-1,1\}^n} \frac{1}{2^n}\Gamma_{x,y} = \sum_{x,y\in\{-1,1\}^n} \frac{1}{2^n}\delta_{x^{\oplus i},y}\left(\frac{f(x)-f(x^{\oplus i})}{2}\right)^2 \\ &= \sum_{i\in[n]} \mathbb{E}_x\left(\frac{f(x)-f(x^{\oplus i})}{2}\right)^2 = \overline{s}(f).\end{aligned}$$

On the other hand,

$$\begin{aligned}\|\Gamma\circ D_i\|_{\mathrm{op}} &= \sup_{\|u\|_2=1} \sum_{x\in\{-1,1\}^n} u_x u_{x^{\oplus i}}\delta_{f(x),f(x^{\oplus i})} \leq \sup_{\|u\|_2=1} \sum_{x\in\{-1,1\}^n} |u_x u_{x^{\oplus i}}| \\ &\leq \sup_{\|u\|_2=1} \|u\|_2^2 = 1.\end{aligned}$$

Finally, for $f=\chi_{[n]}$ we have that $\lambda(f)=\overline{s}(f)=n$. □

Corollary 8.6. *Let CM be any reasonable complexity measure. For a $1-\exp(-\exp(n))$ fraction of all Boolean functions $f:\{-1,1\}^n\to\{-1,1\}$ we have that $CM(f)=\Omega(n)$.*

Proof. If we pick a uniformly random Boolean function $f:\{-1,1\}^n\to\{-1,1\}$, then

$$\mathbb{E}_f\overline{s}(f) = \mathbb{E}_x\sum_i \mathbb{E}_f\frac{1-f(x)f(x^{\oplus i})}{2} = \mathbb{E}_x\sum_i\frac{1}{2} = \frac{n}{2}.$$

Now, note that changing the value of f on one input makes $\overline{s}(f)$ change at most $2n/2^n$. Then, by McDiarmid's inequality, Lemma 2.23, we have that

$$\Pr\left[\overline{s}(f)\leq\frac{n}{3}\right]\leq\exp(-\exp(n)).$$

Now, the statement follows from Proposition 8.5 and the fact that $\lambda(f)$ lower bounds, up to constant factors, all reasonable complexity measures [ABDK+21]. □

Bibliography

[AA09] Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *arXiv preprint arXiv:0911.0996*, 2009.

[AA15] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 307–316, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2746539.2746547.

[AAI+16] Scott Aaronson, Andris Ambainis, Jānis Iraids, Martins Kokainis, and Juris Smotrovs. Polynomials, quantum query complexity, and Grothendieck's inequality. In *31st Conference on Computational Complexity, CCC 2016*, pages 25:1–25:19, 2016. arXiv:1511.08682. URL: https://doi.org/10.4230/LIPIcs.CCC.2016.25.

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, 2021. doi:10.1038/s41567-021-01232-0.

[Aar21] Scott Aaronson. Open problems related to quantum query complexity. *ACM Transactions on Quantum Computing*, 2(4), 2021. URL: https://doi.org/10.1145/3488559.

[AB23] Andris Ambainis and Aleksandrs Belovs. An exponential separation between quantum query complexity and the polynomial degree. In *Proceedings of the Conference on Proceedings of the 38th Computational Complexity Conference*, CCC '23, Dagstuhl, DEU, 2023. Schloss

Dagstuhl–Leibniz-Zentrum fuer Informatik. URL: https://doi.org/10.4230/LIPIcs.CCC.2023.24.

[ABDK16] Scott Aaronson, Shalev Ben-David, and Robin Kothari. Separations in query complexity using cheat sheets. In *Proceedings of the forty-eighth annual ACM symposium on Theory of computing*, STOC '16, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2897518.2897644.

[ABDK+21] Scott Aaronson, Shalev Ben-David, Robin Kothari, Shravas Rao, and Avishay Tal. Degree vs. approximate degree and quantum implications of huang's sensitivity theorem. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing*, STOC 2021, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451047.

[ABP19] Srinivasan Arunachalam, Jop Briët, and Carlos Palazuelos. Quantum query algorithms are completely bounded forms. *SIAM J. Comput*, 48(3):903–925, 2019. Preliminary version in ITCS'18. URL: https://doi.org/10.1137/18M117563X.

[ABSdW13] Andris Ambainis, Arturs Backurs, Juris Smotrovs, and Ronald de Wolf. Optimal quantum query bounds for almost all Boolean functions. In Natacha Portier and Thomas Wilke, editors, *30th International Symposium on Theoretical Aspects of Computer Science (STACS 2013)*, volume 20 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 446–453, Dagstuhl, Germany, 2013. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.STACS.2013.446, doi:10.4230/LIPIcs.STACS.2013.446.

[ACC+22] Per Austrin, Hao Chung, Kai-Min Chung, Shiuan Fu, Yao-Ting Lin, and Mohammad Mahmoody. On the impossibility of key agreements from quantum random oracles. In *Advances in Cryptology–CRYPTO 2022: 42nd Annual International Cryptology Conference, CRYPTO 2022, Santa Barbara, CA, USA, August 15–18, 2022, Proceedings, Part II*, pages 165–194. Springer, 2022.

[ACE+25] Amira Abbas, Nunzia Cerrato, Francisco Escudero Gutiérrez, Dmitry Grinko, Francesco Anna Mele, and Pulkit Sinha. Nearly optimal algorithms to learn sparse quantum hamiltonians in physically motivated distances. *arXiv preprint arXiv:2509.09813*, 2025.

[ACL+21] Srinivasan Arunachalam, Sourav Chakraborty, Troy Lee, Manaswi Paraashar, and Ronald de Wolf. Two new results about quantum exact learning. *Quantum*, 5:587, 2021.

[ACQ22] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(1):1–9, 2022. doi:10.1038/s41467-021-27922-0.

[ADE25] Srinivasan Arunachalam, Arkopal Dutt, and Francisco Escudero Gutiérrez. Testing and learning structured quantum hamiltonians. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1263–1270, New York, NY, USA, 2025. Association for Computing Machinery. doi:10.1145/3717823.3718289.

[ADEP24] Srinivasan Arunachalam, Arkopal Dutt, Francisco Escudero Gutiérrez, and Carlos Palazuelos. Learning low-degree quantum objects. In *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*, pages 13–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.

[ADEP25] Srinivasan Arunachalam, Arkopal Dutt, Francisco Escudero Gutiérrez, and Carlos Palazuelos. A cb-bohnenblust–hille inequality with constant one and its applications in learning theory. *Mathematische Annalen*, pages 1–30, 2025.

[AdW18] Srinivasan Arunachalam and Ronald de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19(71):1–36, 2018.

[AG04] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004. doi:10.1103/PhysRevA.70.052328.

[Amb99] Andris Ambainis. A note on quantum black-box complexity of almost all boolean functions. *Information Processing Letters*, 71(1):5–7, 1999.

[Amb00] Andris Ambainis. Quantum lower bounds by quantum arguments. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 636–643, 2000.

[Amb03] A. Ambainis. Polynomial degree vs. quantum query complexity. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 230–239, 2003. doi:10.1109/SFCS.2003.1238197.

[Amb06] A. Ambainis. Polynomial degree vs. quantum query complexity. *J. Comput. System Sci.*, 72(2):220–238, 2006. Earlier version in FOCS'03. quant-ph/0305028.

[Amb07] A. Ambainis. Quantum walk algorithm for element distinctness. *SIAM Journal on Computing*, 37(1):210–239, 2007. Earlier version in FOCS'04. arXiv:quant-ph/0311001.

[Amb18] Andris Ambainis. Understanding quantum algorithms via query complexity. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 3265–3285, 2018. URL: https://www.worldscientific.com/doi/abs/10.1142/9789813272880_0181, arXiv:https://www.worldscientific.com/doi/pdf/10.1142/9789813272880_0181, doi:10.1142/9789813272880_0181.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM (JACM)*, 51(4):595–605, 2004.

[BB14] Arturs Backurs and Mohammad Bavarian. On the sum of l1 influences. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 132–143. IEEE, 2014.

[BBB+19] Tom Bannink, Jop Briët, Harry Buhrman, Farrokh Labib, and Troy Lee. Bounding Quantum-Classical Separations for Classes of Nonlocal Games. In *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*, volume 126 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:11, 2019.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[BCE+25] Andreas Bluhm, Matthias C Caro, Francisco Escudero Gutiérrez, Aadil Oufkir, and Cambyse Rouzé. Certifying and learning quantum ising hamiltonians. *arXiv preprint arXiv:2509.10239*, 2025.

[BCO24a] Andreas Bluhm, Matthias C Caro, and Aadil Oufkir. Hamiltonian property testing (version 1). *arXiv preprint 2403.02968v1*, 2024. `arXiv:2403.02968v1`.

[BCO24b] Andreas Bluhm, Matthias C Caro, and Aadil Oufkir. Hamiltonian property testing (version 2). *arXiv preprint 2403.02968v2*, 2024. `arXiv:2403.02968v2`.

[BE22] Jop Briët and Francisco Escudero Gutiérrez. On Converses to the Polynomial Method. In *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, volume 232 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:10. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. URL: `http://doi.org/10.4230/LIPIcs.TQC.2022.6`.

[BE24] Jinge Bao and Francisco Escudero Gutiérrez. Learning junta distributions, quantum junta states, and QAC^0 circuits. *arXiv preprint arXiv:2410.15822*, 2024.

[BEG24] Jop Briët, Francisco Escudero Gutiérrez, and Sander Gribling. Grothendieck inequalities characterize converses to the polynomial method. *Quantum*, 8:1526, 2024.

[Ben80] Paul Benioff. The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines. *Journal of statistical physics*, 22:563–591, 1980.

[BH31] Henri Frédéric Bohnenblust and Einar Hille. On the absolute convergence of dirichlet series. *Annals of Mathematics*, pages 600–622, 1931. URL: `https://doi.org/10.2307/1968255`.

[Bha25] Sreejata Kishor Bhattacharya. Random Restrictions of Bounded Low Degree Polynomials Are Juntas. In Raghu Meka, editor, *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, volume 325 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:21, Dagstuhl, Germany, 2025. Schloss Dagstuhl –

Leibniz-Zentrum für Informatik. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2025.17, doi:10.4230/LIPIcs.ITCS.2025.17.

[BKT20] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *Theory of Computing*, 16(10):1–71, 2020. URL: https://doi.org/10.4086/toc.2020.v016a010.

[Ble79] Ron C Blei. Fractional cartesian products of sets. In *Annales de l'institut Fourier*, volume 29,2, pages 79–105, 1979.

[Ble01] Ron Blei. *Analysis in integer and fractional dimensions*, volume 71. Cambridge University Press, 2001.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013. URL: https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

[BLMT23] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Learning quantum hamiltonians at any temperature in polynomial time, 2023. arXiv:2310.02243.

[BLMT24] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Structure learning of hamiltonians from real-time evolution, 2024. In FOCS'24.

[BMMN13] M. Braverman, K. Makarychev, Y. Makarychev, and A. Naor. The Grothendieck constant is strictly smaller than Krivine's bound. *Forum Math. Pi*, 1:453–462, 2013. Preliminary version in FOCS'11. arXiv:1103.6161.

[Bon70] Aline Bonami. Étude des coefficients de fourier des fonctions de $l^p(g)$. In *Annales de l'institut Fourier*, volume 20,2, pages 335–402, 1970.

[BP19] Jop Briët and Carlos Palazuelos. Failure of the trilinear operator space Grothendieck inequality. *Discrete Analysis*, 2019. Paper No. 8.

[BPSS14] Frédéric Bayart, Daniel Pellegrino, and Juan B Seoane-Sepúlveda. The Bohr radius of the n-dimensional polydisk is equivalent to $(\log n)/n$. *Advances in Mathematics*, 264:726–746, 2014.

[Bri19] Jop. Briët. Note on chebychev polynomials. *unpublished*, 2019. URL: https://zwebmail.cwi.nl/service/home/~/?auth=co&loc=en&id=462&part=2.

[BS21] Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 1303–1316, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451040.

[BSdW22] Nikhil Bansal, Makrand Sinha, and Ronald de Wolf. Influence in Completely Bounded Block-Multilinear Forms and Classical Simulation of Quantum Algorithms. In *37th Computational Complexity Conference (CCC 2022)*, volume 234, pages 28:1–28:21, 2022. doi:10.4230/LIPIcs.CCC.2022.28.

[BSS03] H. Barnum, M. Saks, and M. Szegedy. Quantum query complexity and semi-definite programming. In *18th IEEE Annual Conference on Computational Complexity, 2003. Proceedings.*, pages 179–193, 2003. doi:10.1109/CCC.2003.1214419.

[BV93] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 11–20, 1993.

[BY23] Zongbo Bao and Penghui Yao. On Testing and Learning Quantum Junta Channels. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1064–1094. PMLR, 12–15 Jul 2023. URL: https://proceedings.mlr.press/v195/bao23b.html.

[Can20] Clément L Canonne. A short note on learning discrete distributions. *arXiv preprint arXiv:2002.11457*, 2020.

[Car23] Matthias C. Caro. Learning quantum processes and hamiltonians via the pauli transfer matrix, 2023. arXiv:2212.04471.

[CPGSV21] J. Ignacio Cirac, David Pérez-García, Norbert Schuch, and Frank Verstraete. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Rev. Mod. Phys.*, 93:045003, Dec 2021.

URL: https://link.aps.org/doi/10.1103/RevModPhys.93.045003, doi:10.1103/RevModPhys.93.045003.

[CS87] Erik Christensen and Allan M Sinclair. Representations of completely bounded multilinear operators. *Journal of Functional analysis*, 72(1):151–181, 1987.

[CW23] Juan Castaneda and Nathan Wiebe. Hamiltonian learning via shadow tomography of pseudo-choi states, 2023. arXiv:2308.13020.

[Dav84] A. Davie. Lower bound for K_G. Unpublished, 1984.

[Dav06] AM Davie. Matrix norms related to Grothendieck's inequality. In *Banach Spaces: Proceedings of the Missouri Conference held in Columbia, USA, June 24–29, 1984*, pages 22–26. Springer, 2006.

[DFOC+11] Andreas Defant, Leonhard Frerick, Joaquim Ortega-Cerdà, Myriam Ounaïes, and Kristian Seip. The Bohnenblust–Hille inequality for homogeneous polynomials is hypercontractive. *Annals of mathematics*, pages 485–497, 2011.

[DGMP19] Andreas Defant, Doming García, Manuel Maestre, and Pablo Peris. *Dirichlet series and holomorphic functions in high dimensions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 07 2019.

[DJ92] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992. doi:10.1098/rspa.1992.0167.

[DMFPSS14] Diogo Diniz, Gustavo Muñoz-Fernández, Daniel Pellegrino, and J Seoane-Sepúlveda. Lower bounds for the constants in the Bohnenblust–Hille inequality: The case of real scalars. *Proceedings of the American Mathematical Society*, 142(2):575–580, 2014.

[DMP19] Andreas Defant, Mieczysław Mastyło, and Antonio Pérez. On the fourier spectrum of functions on boolean cubes. *Mathematische Annalen*, 374(1):653–680, 2019.

[DOS23] Alicja Dutkiewicz, Thomas E. O'Brien, and Thomas Schuster. The advantage of quantum control in many-body hamiltonian learning, 2023. arXiv:2304.07172.

[dSLCP11] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin. Practical characterization of quantum devices without tomography. *Physical Review Letters*, 107(21):210404, 2011. doi:10.1103/PhysRevLett.107.210404.

[EFFJ+23] Francisco Escudero Gutiérrez, David Fernández-Fernández, Gabriel Jaumà, Guillermo F Peñas, and Luciano Pereira. Hardware-efficient entangled measurements for variational quantum algorithms. *Physical Review Applied*, 20(3):034044, 2023.

[EI22] Alexandros Eskenazis and Paata Ivanisvili. Learning low-degree functions from a logarithmic number of random queries. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 203–207, 2022.

[EIS22] Alexandros Eskenazis, Paata Ivanisvili, and Lauritz Streck. Low-degree learning and the metric entropy of polynomials. *arXiv preprint arXiv:2203.09659*, 2022.

[EM24] Francisco Escudero Gutiérrez and Garazi Muguruza. All S_p notions of quantum expansion are equivalent. *arXiv preprint arXiv:2405.03517*, 2024.

[Esc24a] Francisco Escudero Gutiérrez. Influences of fourier completely bounded polynomials and classical simulation of quantum algorithms. *Chicago Journal of Theoretical Computer Science*, 2024. URL: https://doi.org/10.48550/arXiv.2304.06713.

[Esc24b] Francisco Escudero Gutiérrez. Simple algorithms to test and learn local hamiltonians. *arXiv preprint arXiv:2404.06282*, 2024.

[Esc25] Francisco Escudero Gutiérrez. Christensen-sinclair factorization via semidefinite programming. *Linear Algebra and its Applications*, 714:28–44, 2025.

[FGG07] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum algorithm for the hamiltonian nand tree. *arXiv preprint quant-ph/0702144*, 2007.

[FHKL16] Yuval Filmus, Hamed Hatami, Nathan Keller, and Noam Lifshitz. On the sum of the l 1 influences of bounded functions. *Israel Journal of Mathematics*, 214:167–192, 2016.

[GCC24] Andi Gu, Lukasz Cincio, and Patrick J. Coles. Practical hamiltonian learning with unitary dynamics and gibbs states. *Nature Communications*, 15(1), 2024. doi:10.1038/s41467-023-44008-1.

[GJ14] Gus Gutoski and Nathaniel Johnston. Process tomography for unitary quantum channels. *Journal of Mathematical Physics*, 55(3), 2014.

[GL19] Sander Gribling and Monique Laurent. Semidefinite programming formulations for the completely bounded norm of a tensor. *arxiv,arXiv.1901.04921*, 2019. URL: https://doi.org/10.48550/arXiv.1901.04921.

[GOS+11] Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011. arXiv:https://doi.org/10.1137/100785429.

[Got98] Daniel Gottesman. The heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.

[Gre07] Ben Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007. URL: https://doi.org/10.1090/crmp/043/06.

[Gro53] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1953.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219. ACM, 1996.

[Haa87] Uffe Haagerup. A new upper bound for the complex Grothendieck constant. *Israel Journal of Mathematics*, 60:199–224, 1987.

[Ham25] Yassine Hamoudi. A brief introduction to quantum query complexity. *In preparation*, 2025. URL: `https://yassine-hamoudi.github.io/files/publications/QueryComplexity.pdf`.

[Har72] Lawrence A Harris. Bounds on the derivatives of holomorphic functions of vectors. In *Proc. Colloq. Analysis, Rio de Janeiro*, volume 145, page 163, 1972.

[HBCP15] M Holzäpfel, T Baumgratz, M Cramer, and Martin B Plenio. Scalable reconstruction of unitary processes and hamiltonians. *Physical Review A*, 91(4):042129, 2015.

[HCP23a] Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *PRX Quantum*, 4:040337, Dec 2023. URL: `https://link.aps.org/doi/10.1103/PRXQuantum.4.040337`, `doi:10.1103/PRXQuantum.4.040337`.

[HCP23b] Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *PRX Quantum*, 4(4):040337, 2023. `doi:10.1103/PRXQuantum.4.040337`.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[HKT22] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum hamiltonians from high-temperature gibbs states. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 135–146. IEEE, 2022. `doi:10.1109/FOCS54457.2022.00020`.

[HLv07] Peter Høyer, Troy Lee, and Robert Špalek. Negative weights make adversaries stronger. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 526–535, 2007.

[HMG+25] Hong-Ye Hu, Muzhou Ma, Weiyuan Gong, Qi Ye, Yu Tong, Steven T Flammia, and Susanne F Yelin. Ansatz-free hamiltonian learning with heisenberg-limited scaling. *arXiv preprint arXiv:2502.11900*, 2025.

[HTFS23] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. Learning many-body hamiltonians with heisenberg-limited scaling. *Physical Review Letters*, 130(20):200403, 2023. doi:10.1103/PhysRevLett.130.200403.

[HW+79] Godfrey Harold Hardy, Edward Maitland Wright, et al. *An introduction to the theory of numbers*. Oxford university press, 1979. URL: https://doi.org/10.1126/science.90.2329.158.b.

[IBF+20] Luca Innocenti, Leonardo Banchi, Alessandro Ferraro, Sougato Bose, and Mauro Paternostro. Supervised learning of time-independent Hamiltonians for gate design. *New Journal of Physics*, 22(6):065001, 2020.

[Iva19] Paata Ivanishvili. Aaronson-ambainis conjecture. https://extremal010101.wordpress.com/2019/10/29/aaronson-ambainis-conjecture/, 2019.

[JZ11] Rahul Jain and Shengyu Zhang. The influence lower bound via query elimination. *Electronic Colloquium on Computational Complexity - ECCC*, 7, 02 2011. doi:10.4086/toc.2011.v007a010.

[KM13] Daniel M Kane and Raghu Meka. A PRG for Lipschitz functions of polynomials with applications to sparsest cut. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 1–10, 2013. URL: https://doi.org/10.1145/2488608.2488610.

[KMY03] Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur? In *Algorithms and Computation: 14th International Symposium, ISAAC 2003*, pages 189–198. Springer, 2003.

[KN07] Subhash Khot and Assaf Naor. Linear equations modulo 2 and the l1 diameter of convex bodies. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 318–328. IEEE, 2007. URL: https://doi.org/10.1109/FOCS.2007.20.

[KS19] Gil Kalai and Leonard J Schulman. Quasi-random multilinear polynomials. *Israel Journal of Mathematics*, 230:195–211, 2019.

[Lit30] John E Littlewood. On bounded bilinear forms in an infinite number of variables. *The Quarterly Journal of Mathematics*, pages 164–174, 1930.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

[Lov10] Shachar Lovett. An elementary proof of anti-concentration of polynomials in gaussian variables. In *Electron. Colloquium Comput. Complex.*, volume 17, page 182, 2010.

[LR05] Monique Laurent and Franz Rendl. Semidefinite programming and integer programming. *Handbooks in Operations Research and Management Science*, 12:393–514, 2005.

[LS21] Lily Li and Morgan Shirley. The general adversary bound: A survey. *arXiv preprint arXiv:2104.06380*, 2021.

[LTN+23] Haoya Li, Yu Tong, Hongkang Ni, Tuvia Gefen, and Lexing Ying. Heisenberg-limited hamiltonian learning for interacting bosons, 2023. `arXiv:2307.04690`.

[LW22] Margarite L LaBorde and Mark M Wilde. Quantum algorithms for testing hamiltonian symmetry. *Physical Review Letters*, 129(16):160503, 2022.

[LZ22] Shachar Lovett and Jiapeng Zhang. Fractional certificates for bounded functions. *Electronic Colloquium on Computational Complexity - ECCC*, 107, 2022.

[Man80] Yu I Manin. Computable and uncomputable. *Sov. Radio*, 1980.

[MFPT24] Muzhou Ma, Steven T Flammia, John Preskill, and Yu Tong. Learning k-body hamiltonians via compressed sensing. *arXiv preprint arXiv:2410.18928*, 2024.

[Mid05] Gatis Midrijanis. On randomized and quantum query complexities. *arXiv preprint quant-ph/0501142*, 2005.

[MMFPSS22] Mohammad Sal Moslehian, GA Muñoz-Fernández, AM Peralta, and JB Seoane-Sepúlveda. Similarities and differences between real and complex banach spaces: an overview and recent developments. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 116(2):1–80, 2022. URL: `https://doi.org/10.1007/s13398-022-01222-8`.

[MMG09] Matthew McKague, Michele Mosca, and Nicolas Gisin. Simulating quantum systems using real Hilbert spaces. *Physical review letters*, 102(2):020505, 2009.

[MO08] Ashley Montanaro and Tobias J Osborne. Quantum Boolean functions. *arXiv preprint arXiv:0810.2435*, 2008.

[Mon12] Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12):122206, 2012.

[MS69] Minsky Marvin and A Papert Seymour. Perceptrons. *Cambridge, MA: MIT Press*, 6(318-362):7, 1969.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, UK, 2010. doi:10.1017/CBO9780511976667.

[NGHA15] Miguel Navascués, Yelena Guryanova, Matty J Hoban, and Antonio Acín. Almost quantum correlations. *Nature communications*, 6(1):6288, 2015.

[Nik54] Hukukane Nikaidô. On von Neumann's minimax theorem. *Pacific Journal of Mathematics*, 4:65–72, 1954.

[NPVY23] Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. On the Pauli spectrum of QAC0. *arXiv preprint arXiv:2311.09631*, 2023.

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational complexity*, 4:301–313, 1994.

[NS12] Sahand Negahban and Devavrat Shah. Learning sparse boolean polynomials. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 2032–2036. IEEE, 2012.

[O'D09] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2009. doi:10.1017/cbo9781139814782.

[ORSFW23] Emilio Onorati, Cambyse Rouzé, Daniel Stilck França, and James D. Watson. Efficient learning of ground & thermal states within phases of matter, 2023. arXiv:2301.12946.

[OSSS05] Ryan O'Donnell, Michael Saks, Oded Schramm, and Rocco A Servedio. Every decision tree has an influential variable. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 31–39. IEEE, 2005.

[OZ15] Ryan O'Donnell and Yu Zhao. Polynomial bounds for decoupling, with applications. *arXiv preprint arXiv:1512.01603*, 2015. URL: `https://doi.org/10.4230/LIPIcs.CCC.2016.24`.

[Pau03] Vern Paulsen. *Completely Bounded Maps and Operator Algebras*. arXiv.1901.04921, 02 2003. URL: `https://doi.org/10.1017/CBO9780511546631`.

[PT18] Daniel Pellegrino and Eduardo V Teixeira. Towards sharp Bohnenblust–Hille constants. *Communications in Contemporary Mathematics*, 20(03):1750029, 2018.

[Ree91] J. Reeds. A new lower bound on the real Grothendieck constant. Manuscript (`http://www.dtc.umn.edu/~reedsj/bound2.dvi`), 1991.

[RSF23] Cambyse Rouzé and Daniel Stilck França. Learning quantum many-body systems from a few copies, 2023. `arXiv:2107.03333`.

[RTW+21] Marc-Olivier Renou, David Trillo, Mirjam Weilenmann, Thinh P Le, Armin Tavakoli, Nicolas Gisin, Antonio Acín, and Miguel Navascués. Quantum theory based on real numbers can be experimentally falsified. *Nature*, 600(7890):625–629, 2021.

[RWZ24] Cambyse Rouzé, Melchior Wirth, and Haonan Zhang. Quantum talagrand, kkl and friedguts theorems and the learnability of quantum boolean functions. *Communications in Mathematical Physics*, 405(4):95, 2024.

[SFMD+24] Daniel Stilck França, Liubov A. Markovich, V. V. Dobrovitski, Albert H. Werner, and Johannes Borregaard. Efficient and robust estimation of many-qubit hamiltonians. *Nature Communications*, 15:311, 2024. `doi:10.1038/s41467-023-44012-5`.

[Sho97] Peter Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal of Computing*, 26(5):1484–1509, 1997. Earlier version in FOCS'94.

[Sim97] Daniel R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. arXiv:https://doi.org/10.1137/S0097539796298637, doi:10.1137/S0097539796298637.

[SMCG16] Sarah Sheldon, Easwar Magesan, Jerry M. Chow, and Jay M. Gambetta. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Phys. Rev. A*, 93:060302(R), Jun 2016. URL: https://link.aps.org/doi/10.1103/PhysRevA.93.060302, doi:10.1103/PhysRevA.93.060302.

[SSW21] Alexander A Sherstov, Andrey A Storozhenko, and Pei Wu. An optimal separation of randomized and quantum query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1289–1302, 2021.

[SVZ23a] Joseph Slote, Alexander Volberg, and Haonan Zhang. Bohnenblust–Hille inequality for cyclic groups. *arXiv preprint arXiv:2305.10560*, 2023.

[SVZ23b] Joseph Slote, Alexander Volberg, and Haonan Zhang. Noncommutative Bohnenblust-Hille inequality in the heisenberg-weyl and gell-mann bases with applications to fast learning. *arXiv preprint arXiv:2301.01438*, 2023.

[SY23] Adrian She and Henry Yuen. Unitary Property Testing Lower Bounds by Polynomials. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 96:1–96:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.ITCS.2023.96.

[Tal20] Avishay Tal. Towards optimal separations between quantum and randomized query complexities. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 228–239. IEEE, 2020.

[Tao12] Terence Tao. *Topics in Random Matrix Theory*. Graduate studies in mathematics. American Mathematical Society, 2012.

[Tsi80] B. S. Tsirelson. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 1980. URL: https://doi.org/10.1007/BF00417500.

[TT23] Terence Tao and Joni Teräväinen. Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions. *Journal of the European Mathematical Society*, 2023. doi:https://doi.org/10.4171/jems/1404.

[Val84] Leslie G. Valiant. A Theory of the Learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[Var74] N. Th. Varopoulos. On an inequality of von Neumann and an application of the metric theory of tensor products to operators theory. *J. Functional Analysis*, 16:83–100, 1974. doi:10.1016/0022-1236(74)90071-8.

[VZ23] Alexander Volberg and Haonan Zhang. Noncommutative Bohnenblust–Hille inequalities. *Mathematische Annalen*, pages 1–20, 2023.

[WKR+22] Frederik Wilde, Augustine Kshetrimayum, Ingo Roth, Dominik Hangleiter, Ryan Sweke, and Jens Eisert. Scalably learning quantum many-body hamiltonians from dynamical data, 2022. arXiv:2209.14328.

[YSHY23] Wenjun Yu, Jinzhao Sun, Zeyao Han, and Xiao Yuan. Robust and efficient hamiltonian learning. *Quantum*, 7:1045, 2023. doi:10.22331/q-2023-06-29-1045.

[YZ20] Grigory Yaroslavtsev and Samson Zhou. Fast Fourier Sparsity Testing. In *Symposium on Simplicity in Algorithms*, pages 57–68. SIAM, 2020.

[YZ22] Takashi Yamakawa and Mark Zhandry. Verifiable quantum advantage without structure. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 69–74. IEEE, 2022.

[Zha24] Andrew Zhao. Learning the structure of any hamiltonian from minimal assumptions. *arXiv preprint arXiv:2410.21635*, 2024.

[ZYLB21] Assaf Zubida, Elad Yitzhaki, Netanel H. Lindner, and Eyal Bairey. Optimal short-time measurements for hamiltonian learning, 2021. arXiv:2108.08824.

Abstract

In this thesis, *Quantum computing, norms, and polynomials*, we investigate the interplay between quantum mechanics, complexity theory, and functional analysis, three central areas of physics, computer science, and mathematics, respectively. The unifying theme throughout the thesis is the dynamic exchange between quantum computing and functional analysis: we explore new applications of functional inequalities in quantum computing, and, in the process, establish novel results in functional analysis itself.

In the first part, we study quantum query algorithms through their correspondence with completely bounded polynomials, as established in earlier work. We begin by revisiting this correspondence, extending it, and presenting it in a new form. Building on this foundation, we draw an analogy between quantum query algorithms and the Grothendieck inequality. Finally, we conclude this part by employing completely bounded polynomials to solve a special case of one of the main open problems in quantum query complexity, the Aaronson–Ambainis conjecture.

In the second part, we turn to quantum learning theory, which seeks to determine how much information must be extracted from a quantum system to fully characterize it. We begin by applying existing versions of the Bohnenblust–Hille inequalities and deriving new ones to obtain results in the learning of low-degree quantum objects. We conclude by presenting some of the first results in the emerging area of Hamiltonian testing and learning.

We also include a third part, as a bonus, where we gather three new proofs, that we find elegant and concise, of known results related to the analysis of Boolean functions.

Samenvatting

In dit proefschrift, *Quantum computing, normen en polynomen*, onderzoeken we wisselwerking tussen de kwantummechanica, complexiteitstheorie en functionaalanalyse — drie centrale deelgebieden binnen respectievelijk de natuurkunde, de informatica en de wiskunde. Het overkoepelende thema van dit werk is de dynamische interactie tussen kwantumcomputing en functionaalanalyse: we verkennen nieuwe toepassingen van functionale ongelijkheden binnen de context van kwantumcomputing en bewijzen daarbij tevens nieuwe resultaten in de functionaalanalyse zelf.

In het eerste deel bestuderen we kwantumquery-algoritmen via hun correspondentie met volledig begrensde polynomen, zoals vastgesteld in eerder werk. We beginnen met het herzien, uitbreiden en in een nieuwe vorm presenteren van deze correspondentie. Op basis hiervan leggen we een analogie tussen kwantumquery-algoritmen en de ongelijkheid van Grothendieck. Ten slotte besluiten we dit deel met het gebruik van volledig begrensde polynomen om een speciaal geval te bewijzen van een van de belangrijkste open problemen in de kwantumquery-complexiteit: het Aaronson–Ambainis-vermoeden.

In het tweede deel richten we ons op de kwantumleertheorie, die probeert te bepalen hoeveel informatie uit een kwantumsysteem moet worden gehaald om het volledig te kunnen karakteriseren. We beginnen met het toepassen van bestaande versies van de Bohnenblust–Hille-ongelijkheden en het afleiden van nieuwe versies om resultaten te verkrijgen voor het leren van laaggradige kwantumobjecten. We concluderen dit deel met enkele van de eerste resultaten op het opkomende onderzoeksgebied van Hamiltoniaantesten en -leren.

Tot slot voegen we een derde deel toe, een bonus, waarin we drie nieuwe, naar ons

oordeel elegante en beknopte bewijzen presenteren van bekende resultaten die verband houden met de analyse van Booleaanse functies.

Curriculum Vitae

Francisco was born in 1997 in Madrid. He obtained a double BSc in physics and mathematics from the University Complutense of Madrid in 2020, and an MSc in advanced mathematics from the same institution in 2021. Under the supervision of Jop Briët he completed his PhD at CWI and Leiden University. He will continue his career at INRIA Saclay, where he will be a post-doctoral researcher at Quriosity.