Quantum computing, norms and polynomials

Escudero Gutiérrez, F.

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

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

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

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

□

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 $s \in [d]$,

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

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 $\{e_0, \ldots, e_d\}$ 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

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

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

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

where the sup is taken over all (real/complex) contractions $Z_1, \ldots, Z_{n+1} \in M_m$, 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

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

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

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

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

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

We see that we can iterate this process to obtain

$$\Big\| \sum_{i_1\in[N]} R^\dagger_{i_1} R_{i_1} \Big\| \le N^{d-1} \Big\| \sum_{i_d\in[N]} (U^d_{i_d})^\dagger U^d_{i_d} \Big\| = N^d.$$

Therefore, we conclude that $\Big\| \sum_{\mathbf{i}\in[N]^d} a_{(i_1,i_2)} \dots a_{(i_{d-1},i_d)} U^1_{i_1} \dots U^d_{i_d} \Big\| \le N^{\frac{d+1}{2}}$.

□

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 $f_1, \dots, f_d : \{0,1\}^n \to \{-1,1\}$, its d-forrelation is defined as

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

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

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

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

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

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

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

□

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

$$f(x) = \sum_{a\in\{-1,1\}^{d-1}} \underbrace{\frac{x_1(1) - a_1 x_1(2)}{2} \cdots \frac{x_{d-1}(1) - a_{d-1}x_{d-1}(2)}{2}}_{g_a(x_1,\dots,x_{d-1})} x_d(a), \qquad (6.10)$$

where we identify $\{-1,1\}^{d-1}$ with $[2^{d-1}]$ in the canonical way. The address function is Boolean because for every $(x_1,\dots,x_{d-1}) \in (\{-1,1\}^n)^{d-1}$ 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 ± 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

$$\|f - f'\|_2^2 = \sum_{s\in\{0,1\}^n, |s|\le d} |\widehat{f}(s) - \widehat{f'}(s)|^2.$$

Hence, in order to learn f up to error ε, it suffices to learn each of its Fourier coefficients $\widehat{f}(s)$ with $|s| \le d$ 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 $|s| \le d$ with probability $\ge 1-\delta$ 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

$$\sum_{s\in\{0,1\}^n,\ |\widehat{f}(s)|\le\varepsilon'} |\widehat{f}(s)|^{\frac{2d}{d+1}} \le \varepsilon'^{\frac{2}{d+1}} C^{\sqrt{d\log d}}.$$

Therefore, by setting $\varepsilon' = \varepsilon^{d+1}C^{-(d+1)\sqrt{d\log d}/2}$, it follows that

$$\sum_{s\in\{0,1\}^n,\ |\widehat{f}(s)|\le\varepsilon'} |\widehat{f}(s)|^2 \le \varepsilon^2. \tag{6.14}$$

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, $\|T\|_{2d/(d+1)} \le \mathrm{poly}(d)\|T\|_\infty$ [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 $\|T\|_{\mathrm{cb}} \le 1$ [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 $O\left(4^d d \log(1/\delta)\right)$ uniform quantum samples. Also, there is a classical algorithm that uses $O\left(4^d d \log(n/\delta)\right)$ 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 $s_1, \dots, s_N$ are the N samples, then the probability that we have $s_i \neq s$ for every $i = 1, \dots, N$ 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}. $\square$

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

$$O\big(C^{4d(d+1)} (1/\varepsilon)^{8d+10} \log(C^{2d(d+1)} (1/\varepsilon)^{4d+4} (1/\delta))\big).$$

□