On the complexity of approximating the diamond norm

Avraham Ben-Aroya[*]   Amnon Ta-Shma[†]

Abstract

The *diamond norm* is a norm defined over the space of quantum transformations. This norm has a natural operational interpretation: it measures how well one can distinguish between two transformations by applying them to a state of arbitrarily large dimension. This interpretation makes this norm useful in the study of quantum interactive proof systems.

In this note we exhibit an efficient algorithm for computing this norm using convex programming. Independently of us, Watrous [Wat09] recently showed a different algorithm to compute this norm. An immediate corollary of this algorithm is a slight simplification of the argument of Kitaev and Watrous [KW00] that QIP ⊆ EXP.

1   Introduction

How well can one distinguish two quantum transformations? Imagine we have access to some unknown admissible super-operator T and we want to distinguish the case it is T_1 from the case it is T_2 (T_1 and T_2 are known). Suppose that T_1 and T_2 take as input a state from a Hilbert space $\mathcal{V}$. One possible test to distinguish T_1 from T_2 is preparing an input state $\rho \in D(\mathcal{V})$ (where $D(\mathcal{V})$ denotes the set of density matrices over $\mathcal{V}$), applying T on ρ and measuring the result. This corresponds to:

$$\sup \left\{ \|T_1\rho - T_2\rho\|_{\mathrm{tr}} \ : \ \rho \in D(\mathcal{V}) \right\}.$$

However, somewhat surprisingly, it turns out that often one can distinguish T_1 and T_2 better, by taking an auxiliary Hilbert space $\mathcal{A}$, preparing an *entangled* input state $\rho \in D(\mathcal{V} \otimes \mathcal{A})$, applying T on the $\mathcal{V}$ register of ρ and then measuring the global result. Therefore, we define:

$$\mathrm{dist}(\rho_1, \rho_2) \ = \ \sup \left\{ \left\|(T_1 \otimes I_{L(\mathcal{A})})\rho - (T_2 \otimes I_{L(\mathcal{A})})\rho\right\|_{\mathrm{tr}} \ : \ \dim(\mathcal{A}) < \infty, \ \rho \in D(\mathcal{V} \otimes \mathcal{A}) \right\}.$$

Kitaev [Kit97] proved that this phenomena is restricted by dimension and the maximum is attained already with an auxiliary Hilbert space $\mathcal{A}$ of dimension $\dim(\mathcal{A}) \leq \dim(\mathcal{V})$. Define the following functions on general (not necessarily admissible) super-operators $T : L(\mathcal{V}) \to L(\mathcal{W})$:

$$\|T\|_{\mathrm{tr}} \ = \ \sup \left\{ \|T(X)\|_{\mathrm{tr}} : X \in L(\mathcal{V}), \|X\|_{\mathrm{tr}} = 1 \right\}, \text{ and,}$$
$$\|T\|_{\diamond} \ = \ \left\|T \otimes I_{L(\mathcal{V})}\right\|_{\mathrm{tr}}.$$

Kitaev showed that both $\|\cdot\|_{\mathrm{tr}}$ and $\|\cdot\|_{\diamond}$ are norms. Furthermore, Rosgen and Watrous [RW05, Lemma 2.4] showed $\mathrm{dist}(T_1, T_2) = \|T_1 - T_2\|_{\diamond}$ for T_1 and T_2 that are completely positive.

The diamond norm naturally appears when studying the class QIP of languages having a single-prover, multi-round interactive proof protocol between an all-powerful prover and an efficient quantum

---

verifier. Kitaev and Watrous [KW00] showed that, without loss of generality, perfect completeness can be achieved and three rounds suffice (starting with the verifier). They also showed that the value of a three round quantum interactive protocol can be expressed as $\|T\|_\diamond$, for some super-operator T that is naturally defined given the protocol of the verifier. They used this characterization, and the fact that $\|T_1 \otimes T_2\|_\diamond = \|T_1\|_\diamond \cdot \|T_2\|_\diamond$ to show perfect parallel amplification for QIP protocols. Finally, they showed that QIP $\subseteq$ EXP by reducing the problem to an exponential size semi-definite programming problem. Thus QIP is somewhere between PSPACE and EXP (the containment PSPACE = IP $\subseteq$ QIP is immediate). Very recently, Jain et. al. [JJUW09] showed that QIP = PSPACE, by showing a space efficient solution to a semi-definite program that captures the complexity of the class QIP.

Another connection between QIP and the diamond norm was given by Rosgen and Watrous [RW05]. They defined the promise problem $\mathrm{QCD}_{a,b}$ (quantum circuit distinguishability) whose input is two admissible super-operators T_1 and T_2, the "yes" instances are pairs (T_1, T_2) for which $\|T_1 - T_2\|_\diamond \geq a$ and the "no" instances are the pairs for which $\|T_1 - T_2\|_\diamond \leq b$. Rosgen and Watrous [RW05] proved that for every $a < b$ the problem $\mathrm{QCD}_{a,b}$ is QIP-complete (see also [Ros08]).

The work of Kitaev and Watrous, as well as the work of Rosgen and Watrous do not imply that approximating the diamond norm itself can be done in P. In this note we prove that the diamond norm can be computed by solving a convex optimization problem, and therefore it is in P. More precisely, if we are given as input a description of $T : L(\mathcal{V}) \to L(\mathcal{V})$, e.g., written as a matrix of dimensions $N^2 \times N^2$ (where $N = \dim(\mathcal{V})$), and we are given $\epsilon > 0$, then we can approximate $\|T\|_\diamond$ to within ϵ additive accuracy in time $\mathrm{poly}(N, \log \epsilon^{-1})$. Independently of us, Watrous [Wat09] recently showed a similar result using a semi-definite program.

This claim can also be used to simplify the (somewhat more complicated) proof given in [KW00] that QIP $\subseteq$ EXP. To see this, notice that Kitaev and Watrous already proved that the value of a three round quantum interactive proof system can be captured as the diamond norm of a natural super-operator T. Thus, given such a proof system, all we need to do is to explicitly write down the description of T (which can be done in PSPACE and therefore in time exponential in $\mathrm{poly}(n)$, where n is the input length of the QIP protocol) and then approximate its diamond norm, in time polynomial in $\exp(\mathrm{poly}(n))$.

Our proof is surprisingly simple. We use an equivalent formulation of the diamond norm, proved by Kitaev, and we notice that it gives a convex program using the joint concavity of the fidelity function. We use a representation for density matrices suggested by Liu [Liu06] in a different context for a similar purpose.

2   Preliminaries

Let $\mathcal{V}, \mathcal{W}$ be two Hilbert spaces. $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ denotes the set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$ and is a vector space of dimension $\dim(\mathcal{V}) \cdot \dim(\mathcal{W})$ equipped with the Hilbert-Schmidt inner product $\langle T_1, T_2 \rangle = \mathrm{Tr}(T_1^\dagger T_2)$. $L(\mathcal{V})$ denotes $\mathrm{Hom}(\mathcal{V}, \mathcal{V})$. Let $\{|i\rangle\}$ denote the standard basis for $\mathcal{V}$. The set

$$\{|i\rangle \langle j| \ : \ 1 \leq i, j \leq \dim(\mathcal{V})\}$$

is an orthonormal basis of $L(\mathcal{V})$. When $\dim(\mathcal{V}) = 2^n$, tensor products of Pauli operators form another natural basis for $L(\mathcal{V})$. The Pauli operators are

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

The set $\{\sigma_{i_1} \otimes \ldots \otimes \sigma_{i_n} \ : \ 0 \leq i_1, \ldots, i_n \leq 3\}$ is an orthogonal basis of $L(\mathcal{V})$, and all basis elements have eigenvalues ± 1.

For a linear operator $A \in \mathrm{Hom}(\mathcal{V}, \mathcal{W})$, the spectral norm of A is

$$\|A\| \stackrel{\mathrm{def}}{=} \sup_{x : \|x\| = 1} x^\dagger A^\dagger A x$$

and is equal to the largest singular value of A. For any Pauli operator P, $\|P\| = 1$. The ℓ_2 norm of A is $\|A\|_2 = \mathrm{Tr}(A^\dagger A)$ and is equal to the ℓ_2 norm of the singular values of A.

A pure state is a unit vector in some Hilbert space. A general quantum system is in a *mixed state*–a probability distribution over pure states. Let $\{p_i, |\phi_i\rangle\}$ denote the mixed state in which the pure state $|\phi_i\rangle$ occurs with probability p_i. The behavior of the mixed-state $\{p_i, |\phi_i\rangle\}$ is completely characterized by its *density matrix* $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$, in the sense that two mixed states with the same density matrix behave the same under any physical operation. Notice that a density matrix over a Hilbert space $\mathcal{V}$ belongs to $L(\mathcal{V})$. Density matrices are positive semi-definite operators and have trace 1. We denote the set of density matrices over $\mathcal{V}$ by $D(\mathcal{V})$.

Trace norm and fidelity. The *trace norm* of a matrix A is defined by

$$\|A\|_{\mathrm{tr}} = \mathrm{Tr}(|A|) = \mathrm{Tr}\left(\sqrt{A^\dagger A}\right),$$

which is the sum of the magnitudes of the singular values of A. One way to measure the distance between two density matrices ρ_1 and ρ_2 is by their trace distance $\|\rho_1 - \rho_2\|_{\mathrm{tr}}$. Another useful alternative to the trace metric as a measure of closeness of density matrices is the *fidelity*. For two positive semi-definite operators ρ_1, ρ_2 on the same finite dimensional space $\mathcal{V}$ (not necessarily having trace 1) we define

$$F(\rho_1, \rho_2) = \left[\mathrm{Tr}\left(\sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}}\right)\right]^2 = \|\sqrt{\rho_1}\sqrt{\rho_2}\|_{\mathrm{tr}}^2.$$

We remark that some authors define $\sqrt{F} = \|\sqrt{\rho_1}\sqrt{\rho_2}\|_{\mathrm{tr}}$ as the fidelity. Our definition is consistent with [KSV02]. $\sqrt{F}$ is jointly concave, i.e., for every set $\{(\rho_i, \xi_i)\}_{i=1}^k$ of pairs of density matrices and every $0 \le \lambda_1, \ldots, \lambda_k \le 1$ such that $\sum_{i=1}^k \lambda_i = 1$,

$$\sqrt{F}(\sum_{i=1}^k \lambda_i \rho_i, \sum_{i=1}^k \lambda_i \xi_i) \ge \sum_{i=1}^k \lambda_i \sqrt{F}(\rho_i, \xi_i).$$

A proof of this fact appears, e.g., in [NC00, Exercise 9.19].[1] We remark that F is not jointly concave (see [MPH+08, Section 2] for a short survey on what is known about the fidelity function).

The diamond norm. Kitaev gave a different equivalent characterization of the diamond norm as follows. Any $T : L(\mathcal{V}) \to L(\mathcal{V})$ can be written in a *Stinespring representation*, i.e., as

$$T(X) = \mathrm{Tr}_{\mathcal{A}}(BXC^\dagger),$$

where $B, C \in \mathrm{Hom}(\mathcal{V}, \mathcal{V} \otimes \mathcal{A})$ and $\dim(\mathcal{A}) \le (\dim(\mathcal{V}))^2$ (see, e.g., [KSV02, page 110] or [Wat04, Lecture 4]). Define two completely positive super-operators $T_1, T_2 : L(\mathcal{V}) \to L(\mathcal{A})$:

$$T_1(X) = \mathrm{Tr}_{\mathcal{V}}(BXB^\dagger), \tag{1}$$
$$T_2(X) = \mathrm{Tr}_{\mathcal{V}}(CXC^\dagger). \tag{2}$$

Then, the diamond norm of T can be written as

$$\|T\|_{\diamond} = \max\left\{\sqrt{F}(T_1(\rho), T_2(\xi)) \ : \ \rho, \xi \in D(\mathcal{V})\right\}.$$

The proof of this characterization can be found in [KSV02, Problem 11.10] or in Watrous' lecture notes [Wat04, Lecture 22, Theorem 22.2] (and notice that Watrous defines the fidelity function to be

---

[1] Note that in [NC00] the fidelity function is defined to be $\sqrt{F}$. In particular, the joint concavity of the fidelity function proved in [NC00, Exercise 9.19] proves joint concavity of $\sqrt{F}$ according to our notation.

$\sqrt{F}$). Further information on the trace norm and the diamond norm of super-operators can be found in [KSV02].

Convex programming. Maximizing a convex function over a convex domain is, in general, NP-hard (see [FV95] for a survey). In sharp contrast to this, *convex programming*, which is the problem of *minimizing* a convex function over a convex domain, is in P. One of the reasons that convex programming is easier to solve is due to the fact that in a convex program any *local* optimum equals the *global* optimum. Special cases of convex programming are semi-definite programming and linear programming. Convex programming can be solved in polynomial time using the ellipsoid algorithm [Kha79] or interior-point methods. Often, these algorithms assume a *separation oracle*, i.e., an efficient procedure that given a point tells whether it belongs to the convex set, and if not, gives a half-space that separates the point from the convex set. However, the problem can also be solved using a *membership oracle* [YN76, GLS88] (a randomized algorithm is given in [BV04]).

For $a \in \mathbb{R}^n$ and $R > 0$ we define $B_n(a, R) = \{x \in \mathbb{R}^n : \|x - a\|_2 \leq R\}$. For a set $K \subseteq \mathbb{R}^n$ we define

$$
\begin{aligned}
K_{-\epsilon} &= \{x \in \mathbb{R}^n : B_n(x, \epsilon) \subseteq K\} \\
K_{+\epsilon} &= \{x \in \mathbb{R}^n : \exists y \in K \text{ such that } x \in B_n(y, \epsilon)\}
\end{aligned}
$$

That is, $K_{-\epsilon}$ is the set of points ϵ-deep in K and $\mathbb{R}^n \setminus K_{+\epsilon}$ is the set of points ϵ-deep in the complement of K.

Definition 2.1. *A function $O_K : \mathbb{R}^n \times \mathbb{R}^+ \to \{0, 1\}$ is a* membership oracle *for $K \subseteq \mathbb{R}^n$ if for every $\epsilon > 0$, $O_K(x, \epsilon) = 1$ for any $x \in K_{-\epsilon}$ and $O_K(x, \epsilon) = 0$ for any $x \notin K_{+\epsilon}$. O_K is* efficient, *if it runs in time polynomial in its input length.*

Definition 2.2. *A function $O_f : K \times \mathbb{R}^+ \to \mathbb{R}$ is an* evaluation oracle *computing f over K, if for every $x \in K$ and every $\epsilon > 0$, $|f(x) - O_f(x, \epsilon)| \leq \epsilon$. O_f is* efficient, *if it runs in time polynomial in its input length.*

Theorem 2.1 ([YN76],[GLS88, Theorem 4.3.13])**.** *There exists an algorithm that solves the following problem:*

Input *: 1. A convex body K given by an efficient membership oracle.*

> *2. An integer n, rational numbers $R, r > 0$ and a vector $a_0 \in \mathbb{R}^n$ such that*
>
> $$B_n(a_0, r) \subseteq K \subseteq B_n(\overline{0}, R) \subseteq \mathbb{R}^n.$$

> *3. A rational number $\epsilon > 0$.*

> *4. A convex function $g : K_{+\epsilon} \to \mathbb{R}$ given by an efficient evaluation oracle.*

Output *: A value $x \in K_{+\epsilon}$ such that $|g(x) - \widetilde{\mathrm{opt}}| \leq \epsilon$, where $\widetilde{\mathrm{opt}} = \min_{x \in K_{-\epsilon}} g(x)$.*

The algorithm runs in time $\mathrm{poly}(n, \log \epsilon^{-1}, \log(R/r))$.

Remark 2.1. *The theorem is a slight variant of the one appearing in [GLS88]. There g is required to be defined and convex over the whole of $\mathbb{R}^n$, whereas we only require that it is defined over $K_{+\epsilon}$.*

To see why our variant is correct, notice that the proof given in [GLS88] works by a Turing reduction to the weak nonemptiness problem *for sets of the form $K(t) = K \bigcap \{x \in \mathbb{R}^n : g(x) \leq t\}$ for some values $t \in \mathbb{R}$ that are chosen with a binary search strategy (see [GLS88, page 106]). All we need from g is that for any $t \in \mathbb{R}$ the set $K(t)$ is convex and has an efficient membership oracle.*

Now, clearly, for $K(t)$ to be convex we only need g to be convex over K. Furthermore, we can build a membership oracle for $K(t)$ by querying the membership oracles of K and $\{x : g(x) \leq t\}$ (with certain accuracy parameters) on the same point, see [GLS88, page 129]. The membership oracle answers yes iff both answers are yes. If x is ϵ-far from K it is also ϵ-far from $K(t)$ and we can safely reject. Hence, we only need to query g on inputs that are in $K_{+\epsilon}$.

3 Approximating the diamond norm in P

3.1 Representing density matrices

We follow [Liu06] in the way we represent density matrices as vectors. This is due to that fact that we need the set of vectors representing the density matrices to contain and to be contained in balls of appropriate radii around the origin.

We represent $\rho \in D(\mathcal{V})$ by its Pauli-basis coefficients, but excluding the identity coefficient which is always 1. Thus, we represent $\rho \in D(\mathcal{V})$ as a vector $v(\rho) \in \mathbb{R}^{N^2-1}$, where the ith coordinate of this vector is given by $v_i(\rho) = \mathrm{Tr}(P_{i+1}\rho)$, where P_i is the ith Pauli operator and $P_1 = I$. (Notice that $\mathrm{Tr}(P\rho) \in \mathbb{R}$ for Hermitian P and positive semi-definite ρ.) We let

$$K^{(1)} = \{v(\rho) \ : \ \rho \in D(\mathcal{V})\}.$$

The converse transformation $\Phi : K^{(1)} \to D(\mathcal{V})$ is defined by

$$\Phi(x) = \frac{1}{N}\left(I + \sum_{i=1}^{N^2-1} x_i P_{i+1}\right) \in D(\mathcal{V}).$$

Notice that for any $\rho \in D(\mathcal{V})$, $\Phi(v(\rho)) = \rho$ and similarly, for any $x \in K^{(1)}$, $v(\Phi(x)) = x$. Also for every $x \in \mathbb{R}^{N^2-1}$ (not necessarily in $K^{(1)}$) we have that $\mathrm{Tr}(\Phi(x)) = 1$, and for every $x, y \in \mathbb{R}^{N^2-1}$, $\|\Phi(x) - \Phi(y)\|_2 = \frac{1}{\sqrt{N}}\|x - y\|_2$ where the first norm is over $L(\mathcal{V})$ and the second over $\mathbb{R}^{N^2-1}$.

The convex set that we optimize over is $K = K^{(1)} \times K^{(1)}$. We claim:

Claim 3.1. *K is convex and $B_{2N^2-2}(\overline{0}, \frac{1}{2\sqrt{N}}) \subseteq K \subseteq B_{2N^2-2}(\overline{0}, 2N)$.*

Proof. $K^{(1)}$ is convex since the set of density matrices is convex. Hence K is also convex. Next we show $B_{N^2-1}(\overline{0}, \frac{1}{2\sqrt{N}}) \subseteq K^{(1)}$ which implies $B_{2N^2-2}(\overline{0}, \frac{1}{2\sqrt{N}}) \subseteq K$. Indeed, let $x \in \mathbb{R}^{N^2-1}$ be such that $\|x\|_2 \le \frac{1}{2\sqrt{N}}$ and let

$$\rho = \Phi(x) = \frac{1}{N}\left(I + \sum_{i=2}^{N^2} x_i P_i\right).$$

Clearly ρ is Hermitian and has trace 1. We are left to verify that ρ is positive semi-definite. Fix a unit vector $u \in \mathbb{R}^N$. Then,

$$
\begin{aligned}
u^\dagger \rho u &= \frac{1}{N}\left(u^\dagger I u + \sum_{i=1}^{N^2-1} x_i u^\dagger P_{i+1} u\right) \ge \frac{1}{N}\left(1 - \Big|\sum_{i=1}^{N^2-1} x_i u^\dagger P_{i+1} u\Big|\right) \\
&\ge \frac{1}{N}\left(1 - \sum_{i=1}^{N^2-1} |x_i| \cdot \|P_{i+1}\|\right) \ge \frac{1}{N}\left(1 - \sqrt{N}\|x\|_2\right) > 0.
\end{aligned}
$$

In order to show $K \subseteq B_{2N^2-2}(\overline{0}, 2N)$ it is enough to show $K^{(1)} \subseteq B_{N^2-1}(\overline{0}, N)$. Let $x \in K^{(1)}$. Then $\rho = \Phi(x) \in D(\mathcal{V})$ and for any $1 \le i \le N^2 - 1$,

$$v_i(\rho) = |\mathrm{Tr}(\rho P_{i+1})| \le \mathrm{Tr}(|\rho P_{i+1}|) \le \|P_{i+1}\|\,\mathrm{Tr}(\rho) \le 1,$$

and so $\|x\|_2 = \|v(\rho)\|_2 \le N$. $\qquad\square$

Claim 3.2. *There exists an efficient membership oracle for K.*

Proof. Clearly it is enough to give an efficient membership oracle for $K^{(1)}$. Given an input $x \in \mathbb{R}^{N^2-1}$ and an $\epsilon > 0$ we construct the Hermitian matrix $\rho = \Phi(x)$ and approximate its eigenvalues with accuracy $\zeta = \frac{\epsilon}{10N^{3/2}}$ in the ℓ_∞ norm. We then look at its smallest eigenvalue and we return 1 if it is positive and 0 otherwise.

Given x, let $\sum_i \lambda_i |v_i\rangle\langle v_i|$ with $\lambda_1 \geq \ldots \geq \lambda_N$ be the spectral decomposition of $\rho = \Phi(x)$. The correctness of the membership oracle follows from the following two claims:

- If $x \in K_{-\epsilon}^{(1)}$ then $\lambda_N \geq \frac{\epsilon}{10\sqrt{N}} > \zeta$.

- If $x \notin K_{+\epsilon}^{(1)}$ then $\lambda_N \leq -\frac{\epsilon}{2N^{3/2}} < -\zeta$.

For the first item, assume $x \in K_{-\epsilon}^{(1)}$ but $\lambda_N \leq \frac{\epsilon}{10\sqrt{N}}$. Define $\sigma = (1+\alpha)\rho - \alpha|v_N\rangle\langle v_N|$ for $\alpha = \frac{2\lambda_N}{1-\lambda_N}$. Then $v(\sigma) \notin K^{(1)}$ because $|v_N\rangle$ is an eigenvector of σ with negative eigenvalue, but

$$\|x - v(\sigma)\|_2 = \sqrt{N}\|\rho - \sigma\|_2 \leq \sqrt{N}\|\rho - \sigma\|_{\mathrm{tr}} \leq 2\sqrt{N}\alpha \leq 10\sqrt{N}\lambda_N \leq \epsilon,$$

and so $x \notin K_{-\epsilon}^{(1)}$. A contradiction.

For the second item, assume $x \notin K_{+\epsilon}^{(1)}$ and $0 > \lambda_N \geq -\frac{\epsilon}{2N^{3/2}}$. Define $\sigma = \frac{1}{1+\Delta}\sum_{i:\lambda_i>0}\lambda_i|v_i\rangle\langle v_i|$ for $\Delta = -\sum_{i:\lambda_i<0}\lambda_i$. Clearly, $v(\sigma) \in K^{(1)}$. Also,

$$\|x - v(\sigma)\|_2 = \sqrt{N}\|\rho - \sigma\|_2 \leq \sqrt{N}\|\rho - \sigma\|_{\mathrm{tr}} = 2\sqrt{N}\Delta \leq 2\sqrt{N}N|\lambda_N| \leq \epsilon.$$

Thus, $x \in K_{+\epsilon}^{(1)}$. A contradiction. $\square$

3.2 The target function

Let $\mathcal{V}$ be a Hilbert space of dimension N. Let $T : L(\mathcal{V}) \to L(\mathcal{V})$ be a linear operator given in a Stinespring representation, i.e., as a pair of operators (B, C) such that

$$T(X) = \mathrm{Tr}_{\mathcal{A}}(BXC^\dagger),$$

and let $\epsilon > 0$. We assume that N is a power of 2. From B and C we can compute T_1 and T_2 as in Equations (1) and (2). We define a target function $g : K \to [-1, 0]$ by

$$g(x, y) = -\sqrt{F}(T_1(\Phi(x)), T_2(\Phi(y))),$$

Claim 3.3. *g is convex over K.*

Proof. For every $0 \leq \lambda_1, \ldots, \lambda_k \leq 1$ such that $\sum_{j=1}^k \lambda_j = 1$,

$$g(\sum_{j=1}^k \lambda_j(x_j, y_j)) = g(\sum_{j=1}^k \lambda_j x_j, \sum_{j=1}^k \lambda_j y_j) = -\sqrt{F}(T_1(\Phi(\sum_{j=1}^k \lambda_j x_j)), T_2(\Phi(\sum_{j=1}^k \lambda_j y_j)))$$

$$= -\sqrt{F}(T_1(\sum_{j=1}^k \lambda_j \rho_j), T_2(\sum_{j=1}^k \lambda_j \xi_j)),$$

where $\rho_j = \Phi(x_j) \in D(\mathcal{V})$, $\xi_j = \Phi(y_j) \in D(\mathcal{V})$, and we used the fact that Φ is linear for convex sums, i.e., $\Phi(\sum \lambda_j v_j) = \sum \lambda_j \Phi(v_j)$. Now, by the joint concavity of $\sqrt{F}$,

$$g(\sum_{j=1}^k \lambda_j(x_j, y_j)) = -\sqrt{F}(\sum_{j=1}^k \lambda_j T_1(\rho_j), \sum_{j=1}^k \lambda_j T_2(\xi_j))$$

$$\leq -\sum_{j=1}^k \lambda_j \sqrt{F}(T_1(\rho_j), T_2(\xi_j)) = \sum_{j=1}^k \lambda_j g(x_j, y_j).$$

$\square$

Claim 3.4. *There exists an efficient evaluation oracle for g over K.*

Proof. We are given as input $(x_1, x_2) \in K$ and $\epsilon > 0$. We compute $M_1 = T_1(\Phi(x_1))$ and $M_2 = T_2(\Phi(x_2))$ and this is done with no error. We would like to compute $g(x_1, x_2) = \left\| \sqrt{M_1}\sqrt{M_2} \right\|_{\mathrm{tr}}$. We approximate $\sqrt{M_i}$ with $\zeta/2$ accuracy in the operator norm (it will turn out that $\zeta = \frac{\epsilon}{2N \cdot (\|B\| + \|C\| + 1)}$ suffices), and then we change each negative eigenvalue (if there are any) to zero. We get positive semi-definite S_i such that $\left\| S_i - \sqrt{M_i} \right\| \leq \zeta$. We output an approximation of $\|S_1 S_2\|_{\mathrm{tr}}$ with $\epsilon/2$ accuracy.

By Claims 3.5 and 3.6 below:

$$\left| \|S_1 S_2\|_{\mathrm{tr}} - \left\| \sqrt{M_1}\sqrt{M_2} \right\|_{\mathrm{tr}} \right| \leq N\zeta \left(\|S_1\| + \left\| \sqrt{M_2} \right\| \right) \leq N\zeta(\|B\| + \|C\| + \zeta) \leq \epsilon/2$$

Thus, our output is ϵ-close to $g(x_1, x_2)$ as required. Also, observe that $\log(\zeta^{-1})$ is polynomial in the input length, since $\log(\|B\|)$ and $\log(\|C\|)$ are polynomial in the input length. Therefore, the evaluation oracle is efficient. $\square$

Claim 3.5. *If $\rho_1, \rho_2, \sigma_1, \sigma_2 \in L(\mathcal{V})$ are positive semi-definite and $\|\rho_i - \sigma_i\| \leq \zeta$ for $i \in \{1, 2\}$ then*

$$\left| \|\rho_1 \rho_2\|_{\mathrm{tr}} - \|\sigma_1 \sigma_2\|_{\mathrm{tr}} \right| \leq N\zeta(\|\rho_1\| + \|\sigma_2\|).$$

Proof.

$$\begin{aligned}
\left| \|\rho_1 \rho_2\|_{\mathrm{tr}} - \|\sigma_1 \sigma_2\|_{\mathrm{tr}} \right| &\leq \left| \|\rho_1 \rho_2\|_{\mathrm{tr}} - \|\rho_1 \sigma_2\|_{\mathrm{tr}} \right| + \left| \|\rho_1 \sigma_2\|_{\mathrm{tr}} - \|\sigma_1 \sigma_2\|_{\mathrm{tr}} \right| \\
&\leq \|\rho_1(\rho_2 - \sigma_2)\|_{\mathrm{tr}} + \|(\rho_1 - \sigma_1)\sigma_2\|_{\mathrm{tr}} \\
&\leq \|\rho_1\| \|\rho_2 - \sigma_2\|_{\mathrm{tr}} + \|\sigma_2\| \|\rho_1 - \sigma_1\|_{\mathrm{tr}} \\
&\leq N\zeta(\|\rho_1\| + \|\sigma_2\|).
\end{aligned}$$

$\square$

Claim 3.6. *For any $\rho \in D(\mathcal{V})$:* $\left\| \sqrt{T_1(\rho)} \right\| \leq \|B\|$, $\left\| \sqrt{T_2(\rho)} \right\| \leq \|C\|$.

Proof. T_1 is completely positive and so $T_1(\rho)$ is positive semi-definite and $\left\| \sqrt{T_1(\rho)} \right\| = \sqrt{\|T_1(\rho)\|}$. Express $\rho = \sum_i \lambda_i |v_i\rangle\langle v_i|$ with $\{|v_i\rangle\}$ being an orthonormal basis, $\lambda_i > 0$ and $\sum_i \lambda_i = 1$. Denote $|w_i\rangle = B|v_i\rangle$. Then,

$$\|T_1(\rho)\| = \left\| \sum_i \lambda_i \mathrm{Tr}_{\mathcal{V}}(B|v_i\rangle\langle v_i| B^\dagger) \right\| \leq \sum_i \lambda_i \|\mathrm{Tr}_{\mathcal{V}}(|w_i\rangle\langle w_i|)\| \leq \sum_i \lambda_i \||w_i\rangle\|_2^2,$$

where we have used $\|\mathrm{Tr}_{\mathcal{V}}(|w\rangle\langle w|)\| \leq \||w_i\rangle\|_2^2$. Thus, $\|T_1(\rho)\| \leq \|B\|^2 \sum_i \lambda_i = \|B\|^2$. A similar argument applies for T_2. $\square$

3.3 The algorithm

To compute the diamond norm of a given super-operator, the algorithm essentially solves the convex program that finds the minimum value of g over the convex set. The last thing that we need is to show that g is indeed defined and can be evaluated over points that are at most ϵ-far from this set. However the set K is not good enough for this purpose since matrices that lie outside this set (but still close to it) have negative eigenvalues and it is not clear how one should define the fidelity for such matrices. To overcome this problem we define a new convex set S that is just a shrinking of K. This ensures that matrices that are ϵ-close to the boundary are still positive.

We set $M = -N\sqrt{\|T_1\| \|T_2\|}$, where $\|T_i\|$ is the spectral norm of T_i when viewed as a linear operator in $\mathrm{Hom}(L(\mathcal{V}), L(\mathcal{A}))$. It can be verified that $\min_{x \in K} g(x) \geq -M$. Given $\epsilon > 0$, we define $\alpha = \frac{\epsilon}{4M}$ and $\epsilon' = \frac{\alpha}{\sqrt{N}}$. We define

$$S^{(1)} = (1 - \alpha)K^{(1)}.$$

Claim 3.7. $S^{(1)} = \left\{ x \in K \ : \ \lambda_N(\Phi(x)) \geq \frac{\alpha}{N} \right\}$. *Furthermore, $S^{(1)}$ is convex, has an efficient membership oracle and $S^{(1)}_{+\epsilon'} \subseteq K^{(1)}$.*

Proof.

$$
\begin{aligned}
z \in S^{(1)} \quad &\Leftrightarrow \quad z = (1 - \alpha)x \text{ for some } x \in K^{(1)} \\
&\Leftrightarrow \quad \Phi(z) = (1 - \alpha)\Phi(x) + \alpha\frac{I}{N} \ \text{ for some } \Phi(x) \in D(\mathcal{V}) \\
&\Leftrightarrow \quad \lambda_N(\Phi(z)) \geq \frac{\alpha}{N}.
\end{aligned}
$$

$S^{(1)}$ is convex and has an efficient membership oracle because $K^{(1)}$ does. Also, $S^{(1)}_{+\epsilon'} \subseteq K^{(1)}$ because if $z \in S^{(1)}$ and $\|x - z\|_2 \leq \epsilon'$ then

$$
\lambda_N(\phi(z)) \geq \lambda_N(\phi(x)) - \|\Phi(x) - \Phi(z)\| = \frac{\alpha}{N} - \frac{\|x - z\|}{\sqrt{N}} \geq \frac{\alpha}{N} - \frac{\epsilon'}{\sqrt{N}} = 0.
$$

$\square$

We are now ready to prove:

Theorem 3.1. *Let $\mathcal{V}$ be a Hilbert space of dimension N. Let $T : L(\mathcal{V}) \to L(\mathcal{V})$ be a linear operator given in a Stinespring representation, i.e., as a pair of operators (B, C) such that*

$$
T(X) = \mathrm{Tr}_\mathcal{A}(BXC^\dagger),
$$

and let $\epsilon > 0$. Then there exists a polynomial time algorithm (in the input length of T and $\log \epsilon^{-1}$) that outputs a value c such that $\mid c - \|T\|_\diamond \mid \leq \epsilon$.

Remark 3.1. *The fact that the input operator T is given in a Stinespring representation is without loss of generality as there exists efficient algorithms to move from such a representation to other standard forms of representing a super-operator (see, e.g., [Wat04, Lecture 5]).*

Proof. We approximate $\|T_i\|$ from above in time polynomial in the representation of T_i, and set M, α, and ϵ' as above. We define $S = S^{(1)} \times S^{(1)}$ and $g : K \to \mathbb{R}$ as above. The target function g has an efficient membership oracle and is convex over K and therefore over $S_{+\epsilon'}$. By Theorem 2.1 we can find a value $\widetilde{opt}$ that approximates $\min_{x \in S_{-\epsilon'}} g(x)$ to within ϵ'.

Now, let $o = (o_1, o_2) \in K$ be a point minimizing g over K, that is, $g(o) = \min_{x \in K} g(x)$. We claim that $o' = (1 - 2\alpha)o$ lies in $S_{-\epsilon'}$. Indeed, fix any $y_i \in B_{N^2 - 1}(o'_i, \epsilon')$. Then,

$$
\lambda_N(\Phi(y_i)) \geq \lambda_N(\Phi(o'_i)) - \frac{\epsilon'}{\sqrt{N}} \geq \frac{2\alpha}{N} - \frac{\epsilon}{\sqrt{N}} = \frac{\alpha}{N},
$$

and therefore $y \in S$. Thus,

$$
g(o) \leq \widetilde{opt} \leq g(o') + \epsilon'.
$$

However,

$$
\begin{aligned}
g(o') \ &= \ g((1 - 2\alpha)o) = -\sqrt{F}\left(T_1\left((1 - 2\alpha)\Phi(o_1) + 2\alpha\frac{I}{N} \right), T_2\left((1 - 2\alpha)\Phi(o_2) + 2\alpha\frac{I}{N} \right) \right) \\
&\leq \ (1 - 2\alpha)\left(-\sqrt{F}\big(T_1(\Phi(o_1)), T_2(\Phi(o_2)) \big) \right) + 2\alpha\left(-\sqrt{F}\left(T_1\left(\frac{I}{N} \right), T_2\left(\frac{I}{N} \right) \right) \right) \\
&\leq \ (1 - 2\alpha)g(o_1, o_2) - 2\frac{\alpha}{N}\sqrt{F}\left(\mathrm{Tr}_\mathcal{V}BB^\dagger, \mathrm{Tr}_\mathcal{V}CC^\dagger \right) \leq (1 - 2\alpha)g(o).
\end{aligned}
$$

Altogether, $|\widetilde{opt} - g(o)| \leq \epsilon' - 2\alpha\, g(o) \leq \epsilon' + 2\alpha M \leq \epsilon$.

$\square$

References

[BV04]   D. Bertsimas and S. Vempala. Solving convex programs by random walks. *Journal of the ACM*, 51(4):540–556, 2004.

[FV95]   C.A. Floudas and V. Visweswaran. Quadratic optimization. *Handbook of Global Optimization*, pages 217–269, 1995.

[GLS88]   M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, New York, 1988.

[JJUW09] R. Jain, Z. Ji, S. Upadhyay, and J. Watrous. QIP = PSPACE. Technical report, quant-ph/0907.4737, 2009.

[Kha79]   L. G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, 20:191–194, 1979.

[Kit97]   A.Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[KSV02]   A.Y. Kitaev, A. Shen, and M.N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, 2002.

[KW00]   A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *ACM Symposium on Theory of Computing (STOC)*, pages 608–617, 2000.

[Liu06]   Y.K. Liu. Consistency of local density matrices is QMA-complete. In *The International Workshop on Randomization and Computation (RANDOM)*, pages 438–449, 2006.

[MPH⁺08] J. A. Miszczak, Z. Puchała, P. Horodecki, A. Uhlmann, and K. Życzkowski. Sub– and super–fidelity as bounds for quantum fidelity. Technical report, quant-ph/0805.2037, 2008.

[NC00]   M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[Ros08]   B. Rosgen. Distinguishing short quantum computations. In *The International Symposium on Theoretical Aspects of Computer Science (ISTCS)*, pages 597–608, 2008.

[RW05]   B. Rosgen and J. Watrous. On the hardness of distinguishing mixed-state quantum computations. In *The Annual IEEE Conference on Computational Complexity (COMP)*, pages 344–354, 2005.

[Wat04]   J. Watrous. Advanced topics in quantum information processing. Lecture notes, 2004. http://www.cs.uwaterloo.ca/~watrous/lecture-notes/701.

[Wat09]   J. Watrous. Semidefinite programs for completely bounded norms. Technical report, quant-ph/0901.4709, 2009.

[YN76]   D. B. Yudin and A. S. Nemirovskii. Informational complexity and efficient methods for the solution of convex extremal problems. *Matekon*, 13(2):3–25, 1976.