Efficient rounding for the noncommutative Grothendieck inequality

Assaf Naor *　　Oded Regev †　　Thomas Vidick ‡

Abstract

The classical Grothendieck inequality has applications to the design of approximation algorithms for NP-hard optimization problems. We show that an algorithmic interpretation may also be given for a *noncommutative* generalization of the Grothendieck inequality due to Pisier and Haagerup. Our main result, an efficient rounding procedure for this inequality, leads to a constant-factor polynomial time approximation algorithm for an optimization problem which generalizes the Cut Norm problem of Frieze and Kannan, and is shown here to have additional applications to robust principle component analysis and the orthogonal Procrustes problem.

Contents

*Courant Institute, New York University, 251 Mercer Street, New York NY 10012, USA. Supported by NSF grant CCF-0832795, BSF grant 2010021, the Packard Foundation and the Simons Foundation. Part of this work was completed while A. N. was visiting Université de Paris Est Marne-la-Vallée.

†Courant Institute, New York University. Supported by a European Research Council (ERC) Starting Grant. Part of the work done while the author was with the CNRS, DI, ENS, Paris.

‡Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Supported by the National Science Foundation under Grant No. 0844626.

1 Introduction

In what follows, the standard scalar product on $\mathbb{C}^n$ is denoted $\langle\cdot,\cdot\rangle$, i.e, $\langle x, y\rangle = \sum_{i=1}^n x_i\overline{y_i}$ for all $x, y \in \mathbb{C}^n$. We always think of $\mathbb{R}^n$ as canonically embedded in $\mathbb{C}^n$; in particular the restriction of $\langle\cdot,\cdot\rangle$ to $\mathbb{R}^n$ is the standard scalar product on $\mathbb{R}^n$. Given a set S, the space $M_n(S)$ stands for all the matrices $M = (M_{ij})_{i,j=1}^n$ with $M_{ij} \in S$ for all $i, j \in \{1, \ldots, n\}$. Thus, $M_n(M_n(\mathbb{R}))$ is naturally identified with the n^4-dimensional space of all 4-tensors $M = (M_{ijkl})_{i,j,k,l=1}^n$ with $M_{ijkl} \in \mathbb{R}$ for all $i, j, k, l \in \{1, \ldots, n\}$. The set of all $n \times n$ orthogonal matrices is denoted $\mathcal{O}_n \subseteq M_n(\mathbb{R})$, and the set of all $n \times n$ unitary matrices is denoted $\mathcal{U}_n \subseteq M_n(\mathbb{C})$.

Given $M = (M_{ijkl}) \in M_n(M_n(\mathbb{R}))$ denote

$$\mathrm{Opt}_{\mathbb{R}}(M) \stackrel{\mathrm{def}}{=} \sup_{U,V\in\mathcal{O}_n} \sum_{i,j,k,l=1}^n M_{ijkl}U_{ij}V_{kl}.$$

and similarly, for $M = (M_{ijkl}) \in M_n(M_n(\mathbb{C}))$ denote

$$\mathrm{Opt}_{\mathbb{C}}(M) \stackrel{\mathrm{def}}{=} \sup_{U,V\in\mathcal{U}_n} \Big| \sum_{i,j,k,l=1}^n M_{ijkl}U_{ij}\overline{V_{kl}}\Big|.$$

Theorem 1. *There exists a polynomial time algorithm that takes as input $M \in M_n(M_n(\mathbb{R}))$ and outputs $U, V \in \mathcal{O}_n$ such that*

$$\mathrm{Opt}_{\mathbb{R}}(M) \leqslant O(1) \sum_{i,j,k,l=1}^n M_{ijkl}U_{ij}V_{kl}.$$

Respectively, there exists a polynomial time algorithm that takes as input $M \in M_n(M_n(\mathbb{C}))$ and outputs $U, V \in \mathcal{U}_n$ such that

$$\mathrm{Opt}_{\mathbb{C}}(M) \leqslant O(1)\Big| \sum_{i,j,k,l=1}^n M_{ijkl}U_{ij}\overline{V_{kl}}\Big|.$$

We will explain the ideas that go into the proof of Theorem 1 later, and it suffices to say at this juncture that our algorithm is based on a rounding procedure for semidefinite programs that is markedly different from rounding algorithms that have been previously used in the optimization literature, and as such it indicates the availability of techniques that have thus far remained untapped for the purpose of algorithm design. Prior to explaining the proof of Theorem 1 we list below some of its applications as an indication of its usefulness.

Remark 2. The implied constants in the $O(1)$ terms of Theorem 1 can be taken to be any number greater than $2\sqrt{2}$ in the real case, and any number greater than 2 in the complex case. There is no reason to believe that the factor $2\sqrt{2}$ in the real case is optimal, but the factor 2 in the complex case is sharp in a certain natural sense that will become clear later. The main content of Theorem 1 is the availability of a constant factor algorithm rather than the value of the constant itself. In particular, the novelty of the applications to combinatorial optimization that are described below is the mere existence of a constant-factor approximation algorithm.

1.1 Applications of Theorem 1

We now describe some examples demonstrating the usefulness of Theorem 1. The first example does not lead to a new result, and is meant to put Theorem 1 in context. All the other examples lead to new algorithmic results. Many of the applications below follow from a more versatile reformulation of Theorem 1 that is presented in Section 5 (see Proposition 20).

1.1.1 The Grothendieck problem

The Grothendieck optimization problem takes as input a matrix $A \in M_n(\mathbb{R})$ and aims to efficiently compute (or estimate) the quantity

$$\max_{\varepsilon,\delta\in\{-1,1\}^n} \sum_{i,j=1}^{n} A_{ij}\varepsilon_i\delta_j. \tag{1}$$

This problem falls into the framework of Theorem 1 by considering the 4-tensor $M \in M_n(M_n(\mathbb{R}))$ given by $M_{iijj} = A_{ij}$ and $M_{ijkl} = 0$ if either $i \neq j$ or $k \neq l$. Indeed,

$$\mathrm{Opt}_{\mathbb{R}}(M) = \max_{U,V\in\mathcal{O}_n} \sum_{i,j=1}^{n} A_{ij}U_{ii}V_{jj} = \max_{x,y\in[-1,1]^n} \sum_{i,j=1}^{n} A_{ij}x_iy_j = \max_{\varepsilon,\delta\in\{-1,1\}^n} \sum_{i,j=1}^{n} A_{ij}\varepsilon_i\delta_j.$$

A constant-factor polynomial time approximation algorithm for the Grothendieck problem was designed in [AN04], where it was also shown that it is NP-hard to approximate this problem within a factor less that $1+\varepsilon_0$ for some $\varepsilon_0 \in (0,1)$. A simple transformation [AN04] relates the Grothendieck problem to the Frieze-Kannan Cut Norm problem [FK99] (this transformation can be made to have no loss in the approximation guarantee [KN12, Sec. 2.1]), and as such the constant-factor approximation algorithm for the Grothendieck problem has found a variety of applications in combinatorial optimization; see the survey [KN12] for much more on this topic. In another direction, based on important work of Tsirelson [Tsi87], the Grothendieck problem has found applications to quantum information theory [CHTW04]. Since the problem of computing $\mathrm{Opt}_{\mathbb{R}}(\cdot)$ contains the Grothendieck problem as a special case, Theorem 1 encompasses all of these applications, albeit with the approximation factor being a larger constant.

1.1.2 Robust PCA

The input to the classical principal component analysis (PCA) problem is $K, n \in \mathbb{N}$ a set of points $a_1, \ldots, a_N \in \mathbb{R}^n$. The goal is to find a K-dimensional subspace maximizing the sum of the squared ℓ_2 norms of the projections of the a_i on the subspace. Equivalently, the problem is to find the maximizing vectors in

$$\max_{\substack{y_1,\ldots,y_K \in \mathbb{R}^n \\ \langle y_i, y_j \rangle = \delta_{ij}}} \sum_{i=1}^{N} \sum_{j=1}^{K} |\langle a_i, y_j \rangle|^2, \tag{2}$$

where here, and in what follows, δ_{ij} is the Kronecker delta. This question has a closed-form solution in terms of the singular values of the $N \times n$ matrix whose i-th row contains the coefficients of the point a_i.

The fact that the quantity appearing in (2) is the maximum of the sum of the *squared* norms of the projected points makes it somewhat non-robust to outliers, in the sense that a single long vector can have a large effect on the maximum. Several more robust versions of PCA were suggested in the literature. One variant, known as "R1-PCA," is due to Ding, Zhou, He, and Zha [DZHZ06], and aims to maximize the sum of the Euclidean norms of the projected points, namely,

$$\max_{\substack{y_1,\ldots,y_K \in \mathbb{R}^n \\ \langle y_i, y_j \rangle = \delta_{ij}}} \sum_{i=1}^{N} \Big(\sum_{j=1}^{K} |\langle a_i, y_j \rangle|^2 \Big)^{1/2}. \tag{3}$$

We are not aware of any prior efficient algorithm for this problem that achieves a guaranteed approximation factor. Another robust variant of PCA, known as "L1-PCA", was suggested by Kwak [Kwa08], and further studied by McCoy and Tropp [MT12] (see Section 2.7 in [MT12] in particular). Here the goal is to maximize the sum of the ℓ_1 norms of the projected points, namely,

$$\max_{\substack{y_1,\ldots,y_K \in \mathbb{R}^n \\ \langle y_i, y_j \rangle = \delta_{ij}}} \sum_{i=1}^{N} \sum_{j=1}^{K} |\langle a_i, y_j \rangle|. \tag{4}$$

In [MT12] a constant factor approximation algorithm for the above problem is obtained for $K = 1$ based on [AN04], and for general K an approximation algorithm with an approximation guarantee of $O(\log n)$ is obtained based on prior work by So [So11].

In Section 5.1 we show that both of the above robust versions of PCA can be cast as special cases of Theorem 1, thus yielding constant-factor approximation algorithms for both problems and all $K \in \{1, \ldots, n\}$.

1.1.3 The orthogonal Procrustes problem

Let $n, d \geqslant 1$ and $K \geqslant 2$ be integers. Suppose that $S_1, \ldots, S_K \subseteq \mathbb{R}^d$ are n-point subsets of $\mathbb{R}^d$. The goal of the *generalized orthogonal Procrustes problem* is to rotate each of the S_k separately so as to best align them. Formally, write $S_k = \{x_1^k, x_2^k, \ldots, x_n^k\}$. The goal is to find K orthogonal matrices $U_1, \ldots, U_K \in \mathcal{O}_d$ that maximize the quantity

$$\sum_{i=1}^{n} \Big\| \sum_{k=1}^{K} U_k x_i^k \Big\|_2^2. \tag{5}$$

If one focuses on a single summand appearing in (5), say $\sum_{k=1}^{K} U_k x_1^k$, then it is clear that in order to maximize its length one would want to rotate each of the x_1^k so that they would all point in the same direction, i.e., they would all be positive multiples of the same vector. The above problem aims to achieve the best possible such alignment (in aggregate) for multiple summands of this type. We note that by expanding the squares one sees that $U_1, \ldots, U_K \in \mathcal{O}_d$ maximize the quantity appearing in (5) if and only if they minimize the quantity $\sum_{i=1}^{n} \sum_{k,l=1}^{K} \|U_k x_i^k - U_l x_i^l\|_2^2$.

The term "generalized" was used above because the *orthogonal Procrustes problem* refers to the case $K = 2$, which has a closed-form solution. The name "Procustes" is a (macabre) reference to Greek mythology (see `http://en.wikipedia.org/wiki/Procrustes`).

The generalized orthogonal Procrustes problem has been extensively studied since the 1970s, initially in the psychometric literature (see, e.g., [CC71, Gow75, TB77]), and more recent applications of it are to areas such as image and shape analysis, market research and biometric identification; see the books [GD04, DM98], the lecture notes [SG02], and [MP07] for much more information on this topic.

The generalized orthogonal Procrustes problem is known to be intractable, and it has been investigated algorithmically in, e.g., [TB77, Bou82, SB88]. A rigorous analysis of a polynomial-time approximation algorithm for this problem appears in the work of Nemirovski [Nem07], where the generalized orthogonal Procrustes problem is treated as an important special case of a more general family of problems called "quadratic optimization under orthogonality constraints", for which he obtains a $O(\sqrt[3]{n+d} + \log K)$ approximation algorithm. This was subsequently improved by So [So11] to $O(\log(n + d + K))$. In Section 5.2 we use Theorem 1 to improve the approximation guarantee for the generalized orthogonal Procrustes problem as defined above to a constant approximation factor. See also Section 5.2 for a more complete discussion of variants of this problem considered in [Nem07, So11] and how they compare to our work.

1.1.4 A Frieze-Kannan decomposition for 4-tensors

In [FK99] Frieze and Kannan designed an algorithm which decomposes every (appropriately defined) "dense" matrix into a sum of a few "cut matrices" plus an error matrix that has small cut-norm. We refer to [FK99] and also Section 2.1.2 in the survey [KN12] for a precise formulation of this statement, as well as its extension, due to [AN04], to an algorithm that allows sub-constant errors. In Section 5.3 we apply Theorem 1 to prove the following result, which can be viewed as a noncommutative variant of the Frieze-Kannan decomposition. For the purpose of the statement below it is convenient to identify the space $M_n(M_n(\mathbb{C}))$ of all 4-tensors with $M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$. Also, for $M \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ we denote from now on its Frobenius (Hilbert-Schmidt) norm by

$$\|M\|_2 \stackrel{\mathrm{def}}{=} \sqrt{\sum_{i,j,k,l=1}^{n} |M_{ijkl}|^2}.$$

Theorem 3. *There exists a universal constant $c \in (0, \infty)$ with the following property. Suppose that $M \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ and $0 < \varepsilon \leqslant 1/2$, and let*

$$T \stackrel{\mathrm{def}}{=} \left\lceil \frac{cn^2 \|M\|_2^2}{\varepsilon^2 \mathrm{Opt}_{\mathbb{C}}(M)^2} \right\rceil . \tag{6}$$

One can compute in time $\mathrm{poly}(n, 1/\varepsilon)$ *a decomposition*

$$M = \sum_{t=1}^{T} \alpha_t (A_t \otimes B_t) + E, \tag{7}$$

such that $A_t, B_t \in \mathcal{U}_n$*, the coefficients* $\alpha_t \in \mathbb{C}$ *satisfy* $|\alpha_t| = O(\|M\|_2/n)$*, and* $\mathrm{Opt}_{\mathbb{C}}(E) \leqslant \varepsilon \mathrm{Opt}_{\mathbb{C}}(M)$*. Moreover, if* $M \in M_n(\mathbb{R}) \otimes M_n(\mathbb{R})$ *then one can replace* $\mathrm{Opt}_{\mathbb{C}}(M)$ *in* (6) *by* $\mathrm{Opt}_{\mathbb{R}}(M)$*, take the coefficients* α_t *to be real,* $A_t, B_t \in \mathcal{O}_n$ *and* E *such that* $\mathrm{Opt}_{\mathbb{R}}(E) \leqslant \varepsilon \mathrm{Opt}_{\mathbb{R}}(M)$*.*

Theorem 3 contains as a special case its commutative counterpart, as studied in [FK99, AN04]. Here we are given $A \in M_n(\mathbb{R})$ with $|a_{ij}| \leqslant 1$ for all $i, j \in \{1, \ldots, n\}$, and we aim for an error εn^2. Define $M_{iijj} = a_{ij}$ and $M_{ijkl} = 0$ if $i \neq j$ or $k \neq l$. Then $\|M\|_2 \leqslant n$. An application of Theorem 3 (in the real case) with ε replaced by $\varepsilon n^2/\mathrm{Opt}_{\mathbb{R}}(M)$ yields a decomposition $A = \sum_{t=1}^{T} \alpha_t (a_t b_t^*) + E$ with $a_t, b_t \in [-1, 1]^n$ and $E \in M_n(\mathbb{R})$ satisfying $\sup_{\varepsilon,\delta \in \{-1,1\}} \sum_{ij=1}^{n} E_{ij} \varepsilon_i \delta_j \leqslant \varepsilon n^2$. Moreover, the number of terms is $T = O(1/\varepsilon^2)$.

Theorem 3 is proved in Section 5.3 via an iterative application of Theorem 1, following the "energy decrement" strategy as formulated by Lovász and Szegedy [LS07] in the context of general weak regularity lemmas. Other than being a structural statement of interest in its own right, we show in Section 5.3 that Theorem 3 can be used to enhance the constant factor approximation of Theorem 1 to a PTAS for computing $\mathrm{Opt}_{\mathbb{C}}(M)$ when $\mathrm{Opt}_{\mathbb{C}}(M) = \Omega(n\|M\|_2)$. Specifically, if $\mathrm{Opt}_{\mathbb{C}}(M) \geqslant \kappa n \|M\|_2$ then one can compute a $(1+\varepsilon)$-factor approximation to $\mathrm{Opt}_{\mathbb{C}}(M)$ in time $2^{\mathrm{poly}(1/(\kappa\varepsilon))}\mathrm{poly}(n)$. This is reminiscent of the Frieze-Kannan algorithmic framework [FK99] for dense graph and matrix problems.

1.1.5 Quantum XOR games

As we already noted, the Grothendieck problem (recall Section 1.1.1) also has consequences in quantum information theory [CHTW04], and more specifically to bounding the power of entanglement in so-called "XOR games", which are two-player one-round games in which the players each answer with a bit and the referee bases her decision on the XOR of the two bits. As will be explained in detail in Section 1.2 below, the literature on the Grothendieck problem relies on a classical inequality of Grothendieck [Gro53], while our work relies on a more recent yet by now classical noncommutative Grothendieck inequality of Pisier [Pis78] (and its sharp form due to Haagerup [Haa85]). Even more recently, the Grothendieck inequality has been generalized to another setting, that of *completely bounded* linear maps defined on operator spaces [PS02, HM08]. While we do not discuss the operator space Grothendieck inequality here, we remark that in [RV12a] the operator space Grothendieck inequality is proved by reducing it to the Pisier-Haagerup noncommutative Grothendieck inequality. Without going into details, we note that this reduction is also algorithmic. Combined with our results, it leads to an algorithmic proof of the operator space Grothendieck inequality, together with an accompanying rounding procedure.

In the preprint [RV12b] written by the last two named authors, the noncommutative and operator space Grothendieck inequalities are shown to have consequences in a setting that generalizes that of classical XOR games, called "quantum XOR games": in such games, the questions to the players may be quantum states (and the answers are still a single classical bit). The results in [RV12b] derive an efficient factor-2 approximation algorithm for the maximum success probability of players in such a game, in three settings: players sharing an arbitrary quantum state, players sharing a maximally entangled state, and players not sharing any entanglement. Theorem 1 implies that in all three cases a good strategy for the players, achieving a success that is a factor 2 from optimal, may be found in polynomial time. These matters are taken up in [RV12b] and will not be discussed further here.

1.2 The noncommutative Grothendieck inequality

The natural semidefinite relaxation of (1) is

$$\sup_{d\in\mathbb{N}}\ \sup_{x,y\in(S^{d-1})^n}\sum_{i,j=1}^n A_{ij}\langle x_i,y_j\rangle, \tag{8}$$

where S^{d-1} is the unit sphere of $\mathbb{R}^d$. Since, being a semidefinite program (SDP), the quantity appearing in (8) can be computed in polynomial time with arbitrarily good precision (see [GLS93]), the fact that the Grothendieck optimization problem admits a constant-factor polynomial time approximation algorithm follows from the following inequality, which is a classical inequality of Grothendieck of major importance to several mathematical disciplines (see Pisier's survey [Pis12] and the references therein for much more on this topic; the formulation of the inequality as below is due to Lindenstrauss and Pełczyński [LP68]).

$$\sup_{d\in\mathbb{N}}\ \sup_{x,y\in(S^{d-1})^n}\sum_{i,j=1}^n A_{ij}\langle x_i,y_j\rangle \leqslant K_G \sup_{\varepsilon,\delta\in\{-1,1\}^n}\sum_{i,j=1}^n A_{ij}\varepsilon_i\delta_j. \tag{9}$$

Here $K_G\in(0,\infty)$, which is understood to be the infimum over those constants for which (9) holds true for all $n\in\mathbb{N}$ and all $A\in M_n(\mathbb{R})$, is a universal constant known as the (real) Grothendieck constant. Its exact value remains unknown, the best available bounds [Ree91, BMMN11] being $1.676<K_G<1.783$. In order to actually find an assignment ε,δ to (1) that is within a constant factor of the optimum one needs to argue that a proof of (9) can be turned into an efficient rounding algorithm; this is done in [AN04].

If one wishes to mimic the above algorithmic success of the Grothendieck inequality in the context of efficient computation of $\mathrm{Opt}_{\mathbb{R}}(\cdot)$, the following natural strategy presents itself: one should replace real entries of matrices by vectors in ℓ_2, i.e., consider elements of $M_n(\ell_2)$, and replace the orthogonality constraints underlying the inclusion $U\in\mathcal{O}_n$, namely,

$$\forall\, i,j\in\{1,\ldots,n\},\quad \sum_{k=1}^n U_{ik}U_{jk}=\sum_{k=1}^n U_{ki}U_{kj}=\delta_{ij},$$

by the corresponding constraints using scalar product. Specifically, given an $n\times n$ vector-valued matrix $X\in M_n(\ell_2)$ define two real matrices $XX^*,X^*X\in M_n(\mathbb{R})$ by

$$\forall\, i,j\in\{1,\ldots,n\},\quad (XX^*)_{ij}\stackrel{\mathrm{def}}{=}\sum_{k=1}^n\langle X_{ik},X_{jk}\rangle,\quad\text{and}\quad (X^*X)_{ij}\stackrel{\mathrm{def}}{=}\sum_{k=1}^n\langle X_{ki},X_{kj}\rangle, \tag{10}$$

and let the set of d-dimensional vector-valued orthogonal matrices be given by

$$\mathcal{O}_n(\mathbb{R}^d)\stackrel{\mathrm{def}}{=}\left\{X\in M_n(\mathbb{R}^d):\ XX^*=X^*X=I\right\}. \tag{11}$$

One then considers the following quantity associated to every $M \in M_n(M_n(\mathbb{R}))$,

$$\mathrm{SDP}_{\mathbb{R}}(M) \stackrel{\mathrm{def}}{=} \sup_{d\in\mathbb{N}} \sup_{X,Y\in\mathcal{O}_n(\mathbb{R}^d)} \sum_{i,j,k,l=1}^{n} M_{ijkl}\left\langle X_{ij}, Y_{kl}\right\rangle. \tag{12}$$

Since the constraints that underlie the inclusion $X, Y \in \mathcal{O}_n(\mathbb{R}^d)$ are linear equalities in the pairwise scalar products of the entries of X and Y, the quantity $\mathrm{SDP}_{\mathbb{R}}(M)$ is a semidefinite program and can therefore be computed in polynomial time with arbitrarily good precision. One would therefore aim to prove the following noncommutative variant of the Grothendieck inequality (9),

$$\forall\, n \in \mathbb{N},\ \forall\, M \in M_n(M_n(\mathbb{R})), \quad \mathrm{SDP}_{\mathbb{R}}(M) \leqslant O(1)\cdot \mathrm{Opt}_{\mathbb{R}}(M). \tag{13}$$

The term "noncommutative" refers here to the fact that $\mathrm{Opt}_{\mathbb{R}}(M)$ is an optimization problem over the noncommutative group $\mathcal{O}_n$, while the classical Grothendieck inequality addresses an optimization problem over the commutative group $\{-1,1\}^n$. In the same vein, noncommutativity is manifested by the fact that the classical Grothendieck inequality corresponds to the special case of "diagonal" 4-tensors $M \in M_n(M_n(\mathbb{R}))$, i.e., those that satisfy $M_{ijkl} = 0$ whenever $i \neq j$ or $k \neq l$.

Grothendieck conjectured [Gro53] the validity of (13) in 1953, a conjecture that remained open until its 1978 affirmative solution by Pisier [Pis78]. A simpler, yet still highly nontrivial proof of the noncommutative Grothendieck inequality (13) was obtained by Kaijser [Kai83]. In Section 4 we design a rounding algorithm corresponding to (13) based on Kaijser's approach. This settles the case of real 4-tensors of Theorem 1, albeit with worse approximation guarantee than the one claimed in Remark 2. The algorithm modeled on Kaijser's proof is interesting in its own right, and seems to be versatile and applicable to other problems, such as possible non-bipartite extensions of the noncommutative Grothendieck inequality in the spirit of [AMMN06]; we shall not pursue this direction here.

A better approximation guarantee, and arguably an even more striking rounding algorithm, arises from the work of Haagerup [Haa85] on the complex version of (13). In Section 3 we show how the real case of Theorem 1 follows formally from our results on its complex counterpart, so from now on we focus our attention on the complex case.

1.2.1 The complex case

In what follows we let $S_{\mathbb{C}}^{d-1}$ denote the unit sphere of $\mathbb{C}^d$ (thus $S_{\mathbb{C}}^{0}$ can be identified with the unit circle $S^1 \subseteq \mathbb{R}^2$). The classical complex Grothendieck inequality [Gro53, LP68] asserts that there exists $K \in (0,\infty)$ such that

$$\forall\, n \in \mathbb{N},\ \forall\, A \in M_n(\mathbb{C}), \qquad \sup_{x,y\in (S_{\mathbb{C}}^{2n-1})^n} \Big| \sum_{i,j=1}^{n} A_{ij}\langle x_i, y_j\rangle \Big| \leqslant O(1) \sup_{\alpha,\beta\in(S_{\mathbb{C}}^{0})^n} \Big| \sum_{i,j=1}^{n} A_{ij}\alpha_i\overline{\beta_j}\Big|. \tag{14}$$

Let $K_G^{\mathbb{C}}$ denote the infimum over those $K \in (0,\infty)$ for which (14) holds true. The exact value of $K_G^{\mathbb{C}}$ remains unknown, the best available bounds being $1.338 < K_G^{\mathbb{C}} < 1.4049$ (the left inequality is due to unpublished work of Davie, and the right one is due to Haagerup [Haa87]).

For $M \in M_n(M_n(\mathbb{C}))$ we define

$$\mathrm{SDP}_{\mathbb{C}}(M) \stackrel{\mathrm{def}}{=} \sup_{d\in\mathbb{N}} \sup_{X,Y\in\mathcal{U}_n(\mathbb{C}^d)} \Big| \sum_{i,j,k,l=1}^{n} M_{ijkl}\left\langle X_{ij}, Y_{kl}\right\rangle \Big|, \tag{15}$$

where analogously to (11) we set

$$\mathcal{U}_n(\mathbb{C}^d) \stackrel{\mathrm{def}}{=} \left\{X \in M_n(\mathbb{C}^d): \ XX^* = X^*X = I\right\}.$$

Here for $X \in M_n(\mathbb{C}^d)$ the complex matrices $XX^*, X^*X \in M_n(\mathbb{C})$ are defined exactly as in (10), with the scalar product being the complex scalar product. Haagerup proved [Haa85] that

$$\forall\, n \in \mathbb{N},\ \forall\, M \in M_n(M_n(\mathbb{C})), \quad \mathrm{SDP}_{\mathbb{C}}(M) \leqslant 2 \cdot \mathrm{Opt}_{\mathbb{C}}(M). \tag{16}$$

Our main algorithm is an efficient rounding scheme corresponding to inequality (16). The constant 2 in (16) is sharp, as shown in [HI95] (see also [Pis12, Sec. 12]).

We note that the noncommutative Grothendieck inequality, as it usually appears in the literature, involves a slightly more relaxed semidefinite program. In order to describe it, we first remark that instead of maximizing over $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$ in (15) we could equivalently maximize over $X, Y \in M_n(\mathbb{C}^d)$ satisfying $XX^*, X^*X, YY^*, Y^*Y \leqslant I$, which is the same as the requirement $\|XX^*\|, \|X^*X\|, \|YY^*\|, \|Y^*Y\| \leqslant 1$, where here and in what follows $\|\cdot\|$ denotes the operator norm of matrices. This fact is made formal in Lemma 6 below. By relaxing the constraints to $\|XX^*\| + \|X^*X\| \leqslant 2$ and $\|YY^*\| + \|Y^*Y\| \leqslant 2$, we obtain the following quantity, which can be shown to still be a semidefinite program.

$$\|M\|_{nc} \stackrel{\mathrm{def}}{=} \sup_{d\in\mathbb{N}} \sup_{\substack{X,Y \in M_n(\mathbb{C}^d) \\ \|XX^*\|+\|X^*X\|\leqslant 2 \\ \|YY^*\|+\|Y^*Y\|\leqslant 2}} \Big| \sum_{i,j,k,l=1}^{n} M_{ijkl} \left\langle X_{ij}, Y_{kl} \right\rangle \Big|. \tag{17}$$

Clearly $\|M\|_{nc} \geqslant \mathrm{SDP}_{\mathbb{C}}(M)$ for all $M \in M_n(M_n(\mathbb{C}))$. Haagerup proved [Haa85] that the following stronger inequality holds true for all $n \in \mathbb{N}$ and $M \in M_n(M_n(\mathbb{C}))$.

$$\|M\|_{nc} \leqslant 2 \cdot \mathrm{Opt}_{\mathbb{C}}(M). \tag{18}$$

As our main focus is algorithmic, in the following discussion we will establish a rounding algorithm for the tightest relaxation (16). In Section 2.3 we show that the same rounding procedure can be used to obtain an algorithmic analogue of (18) as well.

1.2.2 The rounding algorithm

Our main algorithm is an efficient rounding scheme corresponding to (16). In order to describe it, we first introduce the following notation. Let $\varphi : \mathbb{R} \to \mathbb{R}^+$ be given by

$$\varphi(t) \stackrel{\mathrm{def}}{=} \frac{1}{2}\mathrm{sech}\Big(\frac{\pi}{2}t\Big) = \frac{1}{e^{\pi t/2} + e^{-\pi t/2}}. \tag{19}$$

One computes that $\int_{\mathbb{R}} \varphi(t)\mathrm{d}t = 1$, so φ is a density of a probability measure μ on $\mathbb{R}$, known as the *hyperbolic secant distribution*. By [JKB95, Sec. 23.11] we have

$$\forall\, a \in (0, \infty), \quad \int_{\mathbb{R}} a^{it} \varphi(t)\mathrm{d}t = \frac{2a}{1+a^2}. \tag{20}$$

It is possible to efficiently sample from μ using standard techniques; see, e.g., [Dev86, Ch. IX.7].

In what follows, given $X \in M_n(\mathbb{C}^d)$ and $z \in \mathbb{C}^d$ we denote by $\langle X, z\rangle \in M_n(\mathbb{C})$ the matrix whose entries are $\langle X, z\rangle_{jk} = \langle X_{jk}, z\rangle$.

Rounding procedure

1. Let $X, Y \in M_n(\mathbb{C}^d)$ be given as input. Choose $z \in \{1, -1, i, -i\}^d$ uniformly at random, and sample $t \in \mathbb{R}$ according to the hyperbolic secant distribution μ.

2. Set $X_z \stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{2}}\langle X, z\rangle \in M_n(\mathbb{C})$ and $Y_z \stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{2}}\langle Y, z\rangle \in M_n(\mathbb{C})$.

3. Output the pair of matrices $(A, B) = (A(z,t), B(z,t)) \stackrel{\mathrm{def}}{=} (U_z|X_z|^{it}, V_z|Y_z|^{-it}) \in \mathcal{U}_n \times \mathcal{U}_n$ where $X_z = U_z|X_z|$ and $Y_z = V_z|Y_z|$ are the polar decompositions of X_z and Y_z, respectively.

Figure 1: The rounding algorithm takes as input a pair of vector-valued matrices $X, Y \in M_n(\mathbb{C}^d)$. It outputs two matrices $A, B \in \mathcal{U}_n(\mathbb{C})$.

Theorem 4. *Fix $n, d \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Suppose that $M \in M_n(M_n(\mathbb{C}))$ and that $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$ are such that*

$$\Big| \sum_{i,j,k,l=1}^{n} M_{ijkl}\langle X_{ij}, Y_{kl}\rangle \Big| \geqslant (1-\varepsilon)\mathrm{SDP}_{\mathbb{C}}(M), \tag{21}$$

where $\mathrm{SDP}_{\mathbb{C}}(M)$ *is given in* (15). *Then the rounding procedure described in Figure 1 outputs a pair of matrices $A, B \in \mathcal{U}_n$ such that*

$$\mathbb{E}\Big[\Big| \sum_{i,j,k,l=1}^{n} M_{ijkl}\, A_{ij}\overline{B_{kl}} \Big| \Big] \geqslant \Big(\frac{1}{2} - \varepsilon\Big)\mathrm{SDP}_{\mathbb{C}}(M). \tag{22}$$

Moreover, rounding can be performed in time polynomial in n and $\log(1/\varepsilon)$, *and can be derandomized in time* $\mathrm{poly}(n, 1/\varepsilon)$.

While the rounding procedure of Figure 1 and the proof of Theorem 4 (contained in Section 2 below) appear to be different from Haagerup's original proof of (18) in [Haa85], we derived them using Haagerup's ideas. One source of difference arises from changes that we introduced in order to work with the quantity $\mathrm{SDP}_{\mathbb{C}}(M)$, while Haagerup's argument treats the quantity $\|M\|_{nc}$. A second source of difference is that Haagerup's proof of (18) is rather indirect and nonconstructive, while it is crucial to the algorithmic applications that were already mentioned in Section 1.1 for us to formulate a polynomial-time rounding procedure. Specifically, Haagerup establishes the *dual* formulation of (18), through a repeated use of duality, and he uses a bootstrapping argument that relies on nonconstructive tools from complex analysis. The third step in Figure 1 originates from Haagerup's complex-analytic considerations. Readers who are accustomed to semidefinite rounding techniques will immediately notice that this step is unusual; we give intuition for it in Section 1.2.3 below, focusing for simplicity on applying the rounding procedure to vectors rather than matrices (i.e., the more familiar setting of the classical Grothendieck inequality).

1.2.3 An intuitive description of the rounding procedure in the commutative case

Consider the effect of the rounding procedure in the commutative case, i.e., when $X, Y \in M_n(\mathbb{C}^d)$ are diagonal matrices. Let the diagonals of X, Y be $x, y \in (\mathbb{C}^d)^n$, respectively. The first step consists in performing a random projection: for $j \in \{1, \ldots, n\}$ let $\alpha_j = \langle x_j, z \rangle / \sqrt{2} \in \mathbb{C}$ and $\beta_j = \langle y_j, z \rangle / \sqrt{2} \in \mathbb{C}$, where z is chosen uniformly at random from $\{1, -1, i, -i\}^n$ (alternatively, with minor modifications to the proof one may choose i.i.d. z_j uniformly from the unit circle, as was done by Haagerup [Haa85], or use standard complex Gaussians). This step results in sequences of complex numbers whose pairwise products $\alpha_k \overline{\beta_j}$, in expectation, exactly reproduce the pairwise scalar products $\langle x_k, y_j \rangle$. However, in general the resulting complex numbers α_k and β_j may have modulus larger than 1. Extending the "sign" rounding performed in, say, the Goemans-Williamson algorithm for MAXCUT [GW95] to the complex domain, one could then round each α_k and β_j independently by simply replacing them by their respective complex phase.

The procedure that we consider differs from this standard practice by taking into account potential information contained in the modulus of the random complex numbers α_k, β_j. Writing in polar coordinates $\alpha_k = r_k e^{i\theta_k}$ and $\beta_j = s_j e^{i\phi_j}$ we sample a real t according to a specific distribution (the hyperbolic secant distribution μ), and round each α_k and each β_j to

$$a_k \stackrel{\text{def}}{=} e^{i(\theta_k + t\log r_k)} \in S^0_{\mathbb{C}}, \quad \text{and} \quad b_j \stackrel{\text{def}}{=} e^{i(\phi_j - t\log s_j)} \in S^0_{\mathbb{C}},$$

respectively. Observe that this step performs a *correlated* rounding: the parameter t is the same for all $j, k \in \{1, \ldots, n\}$.

The proof presented in [Haa85] uses the maximum modulus principle to show the *existence* of a real t for which a_k, b_j as defined above provide a good assignment. Intuition for the existence of such a good t can be given as follows. Varying t along the real line corresponds to rotating the phases of the complex numbers α_j, β_k at a speed proportional to the logarithm of their modulus: elements with very small modulus vary very fast, those with modulus 1 are left unchanged, and elements with relatively large modulus are again varied at (logarithmically) increasing speeds. This means that the rounding procedure takes into account the fact that an element with modulus away from 1 is a "miss": that particular element's phase is probably irrelevant, and should be changed. However, elements with modulus close to 1 are "good": their phase can be kept essentially unchanged.

We identify a specific distribution μ such that a random t distributed according to μ is good, in expectation. This results in a variation on the usual "sign" rounding technique: instead of directly keeping the phases obtained in the initial step of random projection, they are synchronously rotated for a random time t, at speeds depending on the associated moduli, resulting in a provably good pair of sequences a_k, b_j of complex numbers with modulus 1.

Roadmap: In Section 2 we prove Theorem 4 both as stated in Section 1.2.2 and in a form based on (18). The real case as well as a closely related Hermitian case are treated next, first in Section 3 as a corollary of Theorem 4, and then using an alternative direct rounding procedure in Section 4. Section 5 presents the applications that were outlined in Section 1.1.

2 Proof of Theorem 4

In this section we prove Theorem 4. The rounding procedure described in Figure 1 is analyzed in Section 2.1, while the derandomized version is presented in Section 2.2. The efficiency of the procedure is clear; we refer to Section 2.2 for a discussion on how to discretize the choice of t. In Section 2.3 we show how the analysis can be modified to the case of $\|M\|_{nc}$ and (18).

In what follows, it will be convenient to use the following notation. Given $M \in M_n(M_n(\mathbb{C}))$ and $X, Y \in M_n(\mathbb{C}^d)$, define

$$M(X,Y) \stackrel{\mathrm{def}}{=} \sum_{i,j,k,l=1}^{n} M_{ijkl} \left\langle X_{ij}, Y_{kl} \right\rangle \in \mathbb{C}. \tag{23}$$

Thus $M(\cdot,\cdot)$ is a sesquilinear form on $M_n(\mathbb{C}^d) \times M_n(\mathbb{C}^d)$, i.e., $M(\alpha X, \beta Y) = \alpha\overline{\beta} M(X,Y)$ for all $X, Y \in M_n(\mathbb{C}^d)$ and $\alpha, \beta \in \mathbb{C}$. Observe that if $A, B \in M_n(\mathbb{C})$ then

$$M(A,B) = \sum_{i,j,k,l=1}^{n} M_{ijkl} A_{ij} \overline{B_{kl}} = \sum_{i,j,k,l=1}^{n} M_{ijkl} \left(A \otimes \overline{B} \right)_{(ij),(kl)}. \tag{24}$$

2.1 Analysis of the rounding procedure

Proof of (22). Let $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$ be vector-valued matrices satisfying (21). Let $z \in \{1, -1, i, -i\}^d$ be chosen uniformly at random, and

$$X_z \stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{2}} \langle X, z \rangle \qquad \text{and} \qquad Y_z \stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{2}} \langle Y, z \rangle$$

be random variables taking values in $M_n(\mathbb{C})$ defined as in the second step of the rounding procedure (see Figure 1). Then,

$$\mathbb{E}_z \big[M(X_z, Y_z) \big] = \frac{1}{2} \mathbb{E}_z \Big[\sum_{r,s=1}^{d} \overline{z_r} z_s \sum_{i,j,k,l=1}^{n} M_{ijkl} (X_{ij})_r \overline{(Y_{kl})_s} \Big] = \frac{1}{2} M(X,Y), \tag{25}$$

where we used the fact that $\mathbb{E}[\overline{z_r} z_s] = \delta_{rs}$ for every $r, s \in \{1, \ldots, d\}$.

Observe that (20) implies that

$$\forall\, a \in (0,\infty), \quad \mathbb{E}_t[a^{it}] = 2a - \mathbb{E}_t[a^{2+it}].$$

Applying this identity to the nonzero singular values of $X_z \otimes \overline{Y_z}$, we deduce the matrix equality

$$\begin{aligned} \mathbb{E}_t \left(A \otimes \overline{B} \right) &= \mathbb{E}_t \left[\left(U_z |X_z|^{it} \right) \otimes \left(\overline{V_z} |Y_z|^{it} \right) \right] \\ &= 2 X_z \otimes \overline{Y_z} - \mathbb{E}_t \left[\left(U_z |X_z|^{2+it} \right) \otimes \left(\overline{V_z} |Y_z|^{2+it} \right) \right] \\ &= 2 X_z \otimes \overline{Y_z} - \mathbb{E}_t \left[\left(U_z |X_z|^{2+it} \right) \otimes \left(\overline{V_z |Y_z|^{2-it}} \right) \right], \end{aligned} \tag{26}$$

where $U_z, V_z \in \mathcal{U}_n$ are such that $X_z = U_z|X_z|$ and $Y_z = V_z|Y_z|$ are the polar decompositions of X_z and Y_z, respectively (and therefore the polar decomposition of $\overline{Y_z}$ is $\overline{Y_z} = \overline{V_z}|Y_z|$), and we recall that the output of our rounding scheme as described in Figure 1 is $A = U_z|X_z|^{it}$ and $B = V_z|Y_z|^{-it}$.

It follows from (23), (24), (25) and (26) that

$$\mathbb{E}_{z,t} \big[M(A,B) \big] = M(X,Y) - \mathbb{E}_{z,t} \big[M\big(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it} \big) \big]. \tag{27}$$

Our goal from now on is to bound the second, "error" term on the right-hand side of (27). Specifically, the rest of the proof is devoted to showing that for any fixed $t \in \mathbb{R}$ we have

$$\Big| \mathbb{E}_z \big[M\big(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it} \big) \big] \Big| \leqslant \frac{1}{2} \mathrm{SDP}_{\mathbb{C}}(M). \tag{28}$$

Once established, the estimate (28) completes the proof of the desired expectation bound (22) since

$$\mathbb{E}_{z,t}\big[\,|M(A,B)|\,\big] \overset{(27)\wedge(28)}{\geqslant} M(X,Y) - \frac{1}{2}\mathrm{SDP}_{\mathbb{C}}(M) \overset{(21)}{\geqslant} \left(\frac{1}{2}-\varepsilon\right)\mathrm{SDP}_{\mathbb{C}}(M).$$

So, for the rest of the proof, fix some $t \in \mathbb{R}$. As a first step towards (28) we state the following claim.

Claim 5. *Let $W \in M_n(\mathbb{C}^d)$ be a vector-valued matrix, and for every $r \in \{1,\ldots,d\}$ define $W_r \in M_n(\mathbb{C})$ by $(W_r)_{ij} = (W_{ij})_r$. Let $z \in \{1,-1,i,-i\}^d$ be chosen uniformly at random. Writing $W_z = \langle W, z\rangle \in M_n(\mathbb{C})$, we have*

$$\mathbb{E}_z\left[(W_zW_z^*)^2\right] = (WW^*)^2 + \sum_{r=1}^d W_r(W^*W - W_r^*W_r)W_r^*, \tag{29}$$

$$\mathbb{E}_z\left[(W_z^*W_z)^2\right] = (W^*W)^2 + \sum_{r=1}^d W_r^*(WW^* - W_rW_r^*)W_r. \tag{30}$$

Proof. By definition $W_z = \sum_{r=1}^d \overline{z_r}W_r$, and recalling (10) we have $WW^* = \sum_{r=1}^d W_rW_r^*$ and $W^*W = \sum_{r=1}^d W_r^*W_r$. Consequently,

$$\begin{aligned}
\mathbb{E}_z\left[(W_zW_z^*)^2\right] &= \mathbb{E}_z\Big[\sum_{p,q,r,s=1}^d \overline{z_p}z_q\overline{z_r}z_s W_pW_q^*W_rW_s^*\Big]\\
&= \sum_{p=1}^d W_pW_p^*W_pW_p^* + \sum_{\substack{p,q\in\{1,\ldots,d\}\\ p\neq q}}\Big(W_pW_p^*W_qW_q^* + W_pW_q^*W_qW_p^*\Big)\\
&= \sum_{p,q=1}^d W_pW_p^*W_qW_q^* + \sum_{p,q=1}^d W_pW_q^*W_qW_p^* - \sum_{p=1}^d W_pW_p^*W_pW_p^*\\
&= \Big(\sum_{p=1}^d W_pW_p^*\Big)^2 + \sum_{p=1}^d W_p\Big(\sum_{q=1}^d W_q^*W_q\Big)W_p^* - \sum_{p=1}^d W_pW_p^*W_pW_p^*,
\end{aligned}$$

proving (29). A similar calculation yields (30). $\square$

Now, for every $t \in \mathbb{R}$ define two vector-valued matrices

$$F(t), G(t) \in M_n\left(\mathbb{C}^{\{1,-1,i,-i\}^d}\right)$$

by setting for every $j,k \in \{1,\ldots,n\}$ and $z \in \{1,-1,i,-i\}^d$,

$$(F(t)_{jk})_z \overset{\mathrm{def}}{=} \frac{1}{2^d}\left(U_z|X_z|^{2+it}\right)_{jk} \qquad \text{and} \qquad (G(t)_{jk})_z \overset{\mathrm{def}}{=} \frac{1}{2^d}\left(V_z|Y_z|^{2-it}\right)_{jk}. \tag{31}$$

Thus,

$$M(F(t),G(t)) = \frac{1}{4^d}\sum_{z\in\{1,-1,i,-i\}^d} M\big(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it}\big) = \mathbb{E}_z\big[\,M\big(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it}\big)\,\big]. \tag{32}$$

Moreover, recalling that $X_z = U_z|X_z|$ is the polar decomposition of X_z, we have

$$F(t)F(t)^* = \frac{1}{4^d} \sum_{z\in\{1,-1,i,-i\}^d} U_z|X_z|^4 U_z^* = \mathbb{E}_z\left[U_z|X_z|^4U_z^*\right] = \mathbb{E}_z\left[(X_zX_z^*)^2\right]. \tag{33}$$

Similarly $F(t)^*F(t) = \mathbb{E}_z\left[(X_z^*X_z)^2\right]$, so that an application of Claim 5 with $W = \frac{1}{\sqrt{2}}X$ yields, using $XX^* = X^*X = I$ since $X \in \mathcal{U}_n(\mathbb{C}^d)$,

$$F(t)F(t)^* + \frac{1}{4}\sum_{r=1}^d X_rX_r^*X_rX_r^* = F(t)^*F(t) + \frac{1}{4}\sum_{r=1}^d X_r^*X_rX_r^*X_r = \frac{1}{2}I. \tag{34}$$

Analogously,

$$G(t)G(t)^* + \frac{1}{4}\sum_{r=1}^d Y_rY_r^*Y_rY_r^* = G(t)^*G(t) + \frac{1}{4}\sum_{r=1}^d Y_r^*Y_rY_r^*Y_r = \frac{1}{2}I. \tag{35}$$

The two equations above imply that $F(t), G(t)$ satisfy the norm bounds

$$\max\left\{\|F(t)F(t)^*\|,\ \|F(t)^*F(t)\|,\ \|G(t)G(t)^*\|,\ \|G(t)^*G(t)\|\right\} \leqslant \frac{1}{2}. \tag{36}$$

As shown in Lemma 6 below, (36) implies that there exists a pair of vector-valued matrices

$$R(t), S(t) \in \mathcal{U}_n(\mathbb{C}^{d+2n^2})$$

such that

$$M(R(t), S(t)) = M(\sqrt{2}F(t), \sqrt{2}G(t)). \tag{37}$$

(This fact can also be derived directly using (34) and (35).) Recalling the definition of $\mathrm{SDP}_{\mathbb{C}}(M)$ in (15), it follows that for every $t \in \mathbb{R}$,

$$\left|\mathbb{E}_z\left[\, M\big(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it}\big)\,\right]\right| \overset{(32)}{=} |M(F(t),G(t))| \overset{(37)}{=} \frac{1}{2}|M(R(t),S(t))| \leqslant \frac{1}{2}\mathrm{SDP}_{\mathbb{C}}(M), \tag{38}$$

completing the proof of (28). □

Lemma 6. *Let $X, Y \in M_n(\mathbb{C}^d)$ be such that* $\max(\|X^*X\|, \|XX^*\|, \|Y^*Y\|, \|YY^*\|) \leqslant 1$. *Then there exist $R, S \in \mathcal{U}_n(\mathbb{C}^{d+2n^2})$ such that for every $M \in M_n(M_n(\mathbb{C}))$ we have $M(R,S) = M(X,Y)$. Moreover, R and S can be computed from X and Y in time* $\mathrm{poly}(n,d)$.

Proof. Let $A = I - XX^*$ and $B = I - X^*X$, and note that $A, B \geqslant 0$ and $\mathrm{Tr}(A) = \mathrm{Tr}(B)$. Write the spectral decompositions of A and B as $A = \sum_{i=1}^n \lambda_i(u_iu_i^*)$ and $B = \sum_{j=1}^n \mu_j(v_jv_j^*)$ respectively. Set $\sigma = \sum_{i=1}^n \lambda_i = \sum_{j=1}^n \mu_j$, and define

$$R \overset{\mathrm{def}}{=} X \oplus \Big(\bigoplus_{i,j=1}^n \sqrt{\frac{\lambda_i\mu_j}{\sigma}}\,(u_iv_j^*)\Big) \oplus \Big(0_{M_n(\mathbb{C}^{n^2})}\Big) \in M_n(\mathbb{C}^d \oplus \mathbb{C}^{n^2} \oplus \mathbb{C}^{n^2}).$$

With this definition we have $RR^* = XX^* + A = I$ and $R^*R = X^*X + B = I$, so $R \in \mathcal{U}_n(\mathbb{C}^{d+2n^2})$. Let $S \in \mathcal{U}_n(\mathbb{C}^{d+2n^2})$ be defined analogously from Y, with the last two blocks of n^2 coordinates permuted. One checks that $M(R,S) = M(X,Y)$, as required.

Finally, A, B, their spectral decomposition, and the resulting R, S can all be computed in time $\mathrm{poly}(n,d)$ from X, Y. □

2.2 Derandomized rounding

Note that we can always assume that $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$, where $d \leqslant 2n^2$. We start by slightly changing the projection step. Define X'_z to be the projection $X_z = \frac{1}{\sqrt{2}}\langle X, z\rangle$, after we replace all singular values of X_z that are smaller than ε with ε. Then, writing

$$2X'_z \otimes \overline{Y_z}' = 2X_z \otimes \overline{Y_z} + 2(X'_z - X_z) \otimes \overline{Y_z} + 2X'_z \otimes (\overline{Y'_z} - \overline{Y_z}),$$

we see that in the analogue of (27) the first term is at least $M(X, Y) - 4\varepsilon d \mathrm{SDP}_{\mathbb{C}}(M)$. Here we use that $\|X_z\|, \|Y_z\| \leqslant d$ which follows by the triangle inequality and $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$. For the second term in (27), the previous analysis remains unchanged, provided we prove an analogue of (36). Using (33) and the analogous equations for $F(t)^*F(t), G(t)G(t)^*$ and $G(t)^*G(t)$, it will suffice to bound four expressions such as $\big\|\mathbb{E}_z\big[\,(X'_z(X'_z)^*)^2\,\big]\big\|$. One checks that the modification to the rounding we did can only increase this by ε^4 (even for each z), hence following the previous analysis we get the bound in (38) with $\mathrm{SDP}_{\mathbb{C}}(M)/2$ replaced by $(1+\varepsilon^4)\mathrm{SDP}_{\mathbb{C}}(M)/2$.

Next, we observe that the coordinates of z need not be independent, and it suffices if they are chosen from a four-wise independent distribution. As a result, there are only $\mathrm{poly}(n)$ possible values of z and they can be enumerated efficiently. Therefore, we can assume that we have a value $z \in \{1, -1, i, -i\}^d$ for which

$$\big|\mathbb{E}_t\big[\,M\big(U_z|X'_z|^{it}, V_z|Y'_z|^{-it}\big)\,\big]\big| \;\geqslant\; \Big(\frac{1}{2} - \varepsilon\Big)\mathrm{SDP}_{\mathbb{C}}(M). \tag{39}$$

Notice that for some universal constant $c > 0$, with probability at least $1 - \varepsilon$, a sample from the hyperbolic secant distribution is at most $c\log(1/\varepsilon)$ in absolute value. Therefore, denoting the restriction of the hyperbolic secant distribution to the interval $[-c\log(1/\varepsilon), c\log(1/\varepsilon)]$ by μ', and using the fact that the expression inside the expectation in (39) is never larger than $\mathrm{SDP}_{\mathbb{C}}(M)$, for t' distributed according to μ' we have

$$\big|\mathbb{E}_{t'}\big[\,M\big(U_z|X'_z|^{it'}, V_z|Y'_z|^{-it'}\big)\,\big]\big| \;\geqslant\; \Big(\frac{1}{2} - 2\varepsilon\Big)\mathrm{SDP}_{\mathbb{C}}(M).$$

Moreover, for any $t, t' \in \mathbb{R}$,

$$\begin{aligned}
&\big|M\big(U_z|X'_z|^{it}, V_z|Y'_z|^{-it}\big) - M\big(U_z|X'_z|^{it'}, V_z|Y'_z|^{-it'}\big)\big| \\
&\quad = \big|M\big(U_z(|X'_z|^{it} - |X'_z|^{it'}), V_z|Y'_z|^{-it}\big) + M\big(U_z|X'_z|^{it'}, V_z(|Y'_z|^{-it} - |Y'_z|^{-it'})\big)\big| \\
&\quad \leqslant \big|M\big(U_z(|X'_z|^{it} - |X'_z|^{it'}), V_z|Y'_z|^{-it}\big)\big| + \big|M\big(U_z|X'_z|^{it'}, V_z(|Y'_z|^{-it} - |Y'_z|^{-it'})\big)\big|.
\end{aligned} \tag{40}$$

The first absolute value in (40) is at most

$$\begin{aligned}
\mathrm{SDP}_{\mathbb{C}}(M) \cdot \|U_z(|X'_z|^{it} - |X'_z|^{it'})\| \cdot \|V_z|Y'_z|^{-it}\| &= \mathrm{SDP}_{\mathbb{C}}(M) \cdot \||X'_z|^{it} - |X'_z|^{it'}\| \\
&\leqslant \mathrm{SDP}_{\mathbb{C}}(M) \cdot \log(\max\{\|X'_z\|, \|(X'_z)^{-1}\|\}) \cdot |t - t'|.
\end{aligned}$$

We have $\|X'_z\| \leqslant d$ as explained above, and $\|(X'_z)^{-1}\| \leqslant 1/\varepsilon$ by the way our modified rounding procedure was defined. Similar bounds hold for Y'_z. It therefore suffices to pick t from a grid of size $O(\log(1/\varepsilon)\max(1/\varepsilon^2, d/\varepsilon))$.

2.3 The rounding procedure in the case of (18)

Theorem 4 addressed the performance of the rounding procedure described in Figure 1 with respect to inequality (16). Here we prove that this rounding procedure has the same performance with respect to the noncommutative Grothendieck inequality (18) as well. This is the content of the following theorem.

Theorem 7. *Fix $n, d \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Suppose that $M \in M_n(M_n(\mathbb{C}))$ and that $X, Y \in M_n(\mathbb{C}^d)$ satisfy*

$$\max\left\{\|XX^*\| + \|X^*X\|, \|YY^*\| + \|Y^*Y\|\right\} \leqslant 2, \tag{41}$$

and

$$\Big|\sum_{i,j,k,l=1}^{n} M_{ijkl}\langle X_{ij}, Y_{kl}\rangle\Big| \geqslant (1-\varepsilon)\|M\|_{nc}, \tag{42}$$

where $\|M\|_{nc}$ was defined in (18). *Then the rounding procedure described in Figure 1 outputs a pair of matrices $A, B \in \mathcal{U}_n$ such that*

$$\mathbb{E}\Big[\Big|\sum_{i,j,k,l=1}^{n} M_{ijkl}\, A_{ij}\overline{B_{kl}}\Big|\Big] \geqslant \Big(\frac{1}{2} - \varepsilon\Big)\|M\|_{nc}. \tag{43}$$

Proof. We shall explain how to modify the argument presented in Section 2.1, relying on the notation that was introduced there. All we need to do is to replace (38) by the assertion

$$\forall\, t \in \mathbb{R}, \qquad |M(F(t), G(t))| \leqslant \frac{1}{2}\|M\|_{nc}. \tag{44}$$

To this end we use the following corollary of Claim 5, which is a slight variant of [Haa85, Lem. 4.1].

Corollary 8. *Let $W \in M_n(\mathbb{C}^d)$ be a vector-valued matrix and $z \in \{1, -1, i, -i\}^d$ chosen uniformly at random. Writing $W_z = \langle W, z\rangle \in M_n(\mathbb{C})$, we have*

$$\big\|\mathbb{E}_z\left[(W_zW_z^*)^2\right]\big\| + \big\|\mathbb{E}_z\left[(W_z^*W_z)^2\right]\big\| \leqslant \big(\|WW^*\| + \|W^*W\|\big)^2. \tag{45}$$

Proof. Starting from (29) and noting that

$$\Big\|\sum_{r=1}^{d} W_r(W^*W - W_r^*W_r)W_r^*\Big\| \leqslant \|W^*W\| \cdot \Big\|\sum_{r=1}^{d} W_rW_r^*\Big\| = \|W^*W\| \cdot \|WW^*\|,$$

we obtain the inequality

$$\big\|\mathbb{E}_z\left[(W_zW_z^*)^2\right]\big\| \leqslant \|WW^*\|^2 + \|W^*W\| \cdot \|WW^*\|. \tag{46}$$

Similarly, from (30) we get

$$\big\|\mathbb{E}_z\left[(W_z^*W_z)^2\right]\big\| \leqslant \|W^*W\|^2 + \|WW^*\| \cdot \|W^*W\|. \tag{47}$$

By summing (46) and (47) one deduces (45). □

Combining (33) with (45) for $W = X/\sqrt{2}$, we have

$$\|F(t)F(t)^*\| + \|F(t)^*F(t)\| \leqslant \frac{1}{4}(\|XX^*\| + \|X^*X\|)^2 \overset{(41)}{\leqslant} 1. \tag{48}$$

Analogously,

$$\|G(t)G(t)^*\| + \|G(t)^*G(t)\| \leqslant \frac{1}{4}(\|YY^*\| + \|Y^*Y\|)^2 \leqslant 1. \tag{49}$$

Recalling the definition of $\|M\|_{nc}$, it follows from (48) and (49) that for every $t \in \mathbb{R}$,

$$\left|\mathbb{E}_z\left[\, M\big(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it}\big)\,\right]\right| \overset{(32)}{=} |M(F(t), G(t))| \leqslant \frac{1}{2}\|M\|_{nc}. \tag{50}$$

Hence,

$$\mathbb{E}_{z,t}\left[\, |M(A,B)|\,\right] \overset{(27)\wedge(50)}{\geqslant} M(X,Y) - \frac{1}{2}\|M\|_{nc} \overset{(42)}{\geqslant} \left(\frac{1}{2} - \varepsilon\right)\|M\|_{nc},$$

completing the proof of the desired expectation bound (43). □

Remark 9. The following example, due to Haagerup [Haa85], shows that the factor 2 approximation guarantee obtained in Theorem 7 is optimal: the best constant in (18) equals 2. Let $M \in M_n(M_n(\mathbb{C}))$ be given by $M_{1jk1} = \delta_{jk}$, and $M_{ijkl} = 0$ if $(i,l) \neq (1,1)$. A direct computation shows that $\mathrm{Opt}_{\mathbb{C}}(M) = 1$. Define $X, Y \in M_n(\mathbb{C}^n)$ by $X_{1j} = Y_{j1} = \sqrt{2/(n+1)}\, e_j \in \mathbb{C}^n$ for $j \in \{1,\ldots,n\}$ and all other entries of X and Y vanish (here $e_1,\ldots,e_n$ is the standard basis of $\mathbb{C}^n$). Using these two vector-valued matrices one shows that $\|M\|_{nc} \geqslant 2n/(n+1)$. Recall that in the Introduction we mentioned that it was shown in [HI95] that the best constant in the weaker inequality (16) is also 2, but the example exhibiting this stronger fact is more involved.

3 The real and Hermitian cases

The $n \times n$ Hermitian matrices are denoted H_n. A 4-tensor $M \in M_n(M_n(\mathbb{C})) \cong M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ is said to be Hermitian if $M_{ijkl} = \overline{M_{jilk}}$ for all $i,j,k,l \in \{1,\ldots,n\}$. Investigating the noncommutative Grothendieck inequality in the setting of Hermitian M is most natural in applications to quantum information, while problems in real optimization as described in the Introduction lead to real $M \in M_n(M_n(\mathbb{R}))$. These special cases are treated in this section.

Consider the following Hermitian analogue of the quantity $\mathrm{Opt}_{\mathbb{C}}(M)$.

$$\mathrm{Opt}^*_{\mathbb{C}}(M) \overset{\mathrm{def}}{=} \sup_{\substack{A,B \in H_n \\ \|A\|,\|B\| \leqslant 1}} \Big|\sum_{i,j,k,l=1}^n M_{ijkl} A_{ij} \overline{B_{kl}}\Big|.$$

Note that the convex hull of $\mathcal{U}_n$ consists of all the matrices $A \in M_n(\mathbb{C})$ with $\|A\| \leqslant 1$, so by convexity for every $M \in M_n(M_n(\mathbb{C}))$ we have

$$\mathrm{Opt}_{\mathbb{C}}(M) = \sup_{\substack{A,B \in M_n(\mathbb{C}) \\ \|A\|,\|B\| \leqslant 1}} \Big|\sum_{i,j,k,l=1}^n M_{ijkl} A_{ij} \overline{B_{kl}}\Big|. \tag{51}$$

This explains why $\mathrm{Opt}^*_{\mathbb{C}}(M)$ should indeed be viewed as a Hermitian analogue of $\mathrm{Opt}_{\mathbb{C}}(M)$. The real analogue of (51) is that, due to the fact that the convex hull of $\mathcal{O}_n$ consists of all the matrices $A \in M_n(\mathbb{R})$ with $\|A\| \leqslant 1$, for every $M \in M_n(M_n(\mathbb{R}))$ we have

$$\mathrm{Opt}_{\mathbb{R}}(M) = \sup_{\substack{A,B\in M_n(\mathbb{R})\\ \|A\|,\|B\|\leqslant 1}} \Big| \sum_{i,j,k,l=1}^{n} M_{ijkl} A_{ij} B_{kl} \Big|. \tag{52}$$

The following theorem establishes an algorithmic equivalence between the problems of approximating either of these two quantities.

Theorem 10. *For every $K \in [1,\infty)$ the following two assertions are equivalent.*

1. *There exists a polynomial time algorithm* Alg^* *that takes as input a Hermitian $M \in M_n(M_n(\mathbb{C}))$ and outputs $A, B \in H_n$ with* $\max\{\|A\|,\|B\|\} \leqslant 1$ *and* $\mathrm{Opt}^*_{\mathbb{C}}(M) \leqslant K|M(A,B)|$.
2. *There exists a polynomial time algorithm* Alg *that takes as input $M \in M_n(M_n(\mathbb{R}))$ and outputs $U, V \in \mathcal{O}_n$ such that* $\mathrm{Opt}_{\mathbb{R}}(M) \leqslant KM(U,V)$.

In Section 3.2 we show that for every $K > 2\sqrt{2}$ there exists an algorithm Alg^* as in part 1) of Theorem 10. Consequently, we obtain the algorithm for computing $\mathrm{Opt}_{\mathbb{R}}(M)$ whose existence was claimed in Theorem 1. The implication 1) $\Longrightarrow$ 2) of Theorem 10 is the only part of Theorem 10 that will be used in this article; the reverse direction 2) $\Longrightarrow$ 1) is included here for completeness. Both directions of Theorem 10 are proved in Section 3.3.

3.1 Two-dimensional rounding

In this section we give an algorithmic version of Krivine's proof [Kri79] that the 2-dimensional real Grothendieck constant satisfies $K_G(2) \leqslant \sqrt{2}$. The following theorem is implicit in the proof of [Kri79, Thm. 1].

Theorem 11 (Krivine). *Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = \mathrm{sign}(\cos(x))$, and let $f : [0,\pi/2) \to \mathbb{R}$ be given by*

$$f(t) \stackrel{\mathrm{def}}{=} \begin{cases} 1 & \text{if } 0 \leqslant t \leqslant \frac{\pi}{4}, \\ \frac{6}{\pi}\left(\frac{\pi}{2} - t\right) - \frac{1}{2}\left(\frac{4}{\pi}\right)^3\left(\frac{\pi}{2} - t\right)^3 & \text{if } \frac{\pi}{4} \leqslant t < \frac{\pi}{2}. \end{cases}$$

Extend f to a function defined on all of $\mathbb{R}$ by requiring that it is even and $f(x+\pi) = -f(x)$ for all $x \in \mathbb{R}$. There exists a sequence $\{b_{2\ell+1}\}_{\ell=0}^{\infty} \in \mathbb{R}^{\mathbb{N}}$ such that for every $L \in \mathbb{N}$ the numbers $\{b_0, \ldots, b_{2L+1}\}$ can be computed in $\mathrm{poly}(L)$ *time, $\sum_{\ell=L+1}^{\infty} |b_{2\ell+1}| \leqslant C/L$ for some universal constant C, $\sum_{\ell=0}^{\infty} |b_{2\ell+1}| = 1$, and*

$$\forall x, y \in \mathbb{R}, \qquad \cos(x-y) \;=\; \sqrt{2} \sum_{\ell=0}^{\infty} b_{2\ell+1} \frac{1}{2\pi} \int_{-\pi}^{\pi} f\big((2\ell+1)x - t\big) g\big(t - (2\ell+1)y\big)\mathrm{d}t.$$

An explicit formula for the sequence $\{b_{2\ell+1}\}_{\ell=0}^{\infty}$ can be extracted as follows from the proof of [Kri79, Thm. 1]. For any $\ell \geqslant 0$, define $a_{2\ell} = 0$,

$$a_{2\ell+1} \;=\; (-1)^{\ell} \cos\Big(\frac{(2\ell+1)\pi}{4}\Big) \frac{16}{\pi^2(2\ell+1)^4} \Big(\frac{1}{2\ell+1} - (-1)^{\ell}\frac{\pi}{4}\Big),$$

$b_1 = \sqrt{2}(\pi/4)^3/(3a_1)$, and for $\ell > 0$,

$$b_{2\ell+1} \;=\; -\frac{1}{a_1} \sum_{\substack{d|(2\ell+1)\\ d\neq 1}} a_d b_{\frac{2\ell+1}{d}}.$$

Then $|a_{2\ell+1}| = O(1/\ell^4)$, from which one deduces the crude bound $|b_{2\ell+1}| = O(1/\ell^2)$.

Two-dimensional rounding procedure

1. Let $\varepsilon > 0$ and, for $j,k \in \{1,\ldots,n\}$ let $x_j, y_k \in \mathbb{C}$ with $|x_j| = |y_k| = 1$, be given as input. Let f, g, C and $\{b_{2\ell+1}\}_{\ell=0}^{\infty}$ be as in Theorem 11.
2. For every j,k let $\theta_j \in [0,2\pi)$ (resp. $\phi_k \in [0,2\pi)$) be the angle that x_j (resp. y_k) makes with the x-axis.
3. Select $t \in [-\pi,\pi]$ uniformly at random. Let $L = \lceil C/\varepsilon \rceil$ and $p = 1 - \sum_{\ell=L+1}^{\infty} |b_{2\ell+1}|$. Select $\ell \in \{-1,0,\ldots,L\}$ with probability $\Pr(-1) = 1-p$ and $\Pr(\ell) = |b_{2\ell+1}|$ for $\ell \in \{0,\ldots,L\}$.
4. For every j,k, if $\ell \geqslant 0$ then set $\lambda_j \stackrel{\text{def}}{=} \operatorname{sign}(b_{2\ell+1}) f((2\ell+1)\theta_j - t)$ and $\mu_k \stackrel{\text{def}}{=} g(t-(2\ell+1)\phi_k)$. Otherwise, set $\lambda_j = 0$, $\mu_k = 0$.
5. Return $(\lambda_j)_{j\in\{1,\ldots,n\}}$ and $(\mu_k)_{k\in\{1,\ldots,n\}}$.

Figure 2: The two-dimensional rounding algorithm takes as input real 2-dimensional unit vectors. It returns real numbers of absolute value at most 1.

Figure 2 describes a two-dimensional rounding scheme derived from Theorem 11. The following claim states its correctness in a way that will be useful for us later.

Claim 12. *Let $\varepsilon > 0$ and for every $j,k \in \{1,\ldots,n\}$ let $x_j, y_k \in \mathbb{C}$ satisfy $|x_j| = |y_k| = 1$. Then the rounding procedure described in Figure 2 runs in time $\mathrm{poly}(n, 1/\varepsilon)$ and returns $\lambda_j, \mu_k \in \mathbb{R}$ with $|\lambda_j|, |\mu_k| \leqslant 1$ for every $j,k \in \{1,\ldots,n\}$, and*

$$\mathbb{E}\big[\,\lambda_j\,\mu_k\,\big] \;=\; \frac{1}{\sqrt{2}}\,\Re\left(x_j\overline{y_k}\right) + \varepsilon\langle x'_j, y'_k\rangle, \tag{53}$$

where $x'_j, y'_k \in L_2(\mathbb{R})$ are such that $\|x'_j\|_2, \|y'_k\|_2 \leqslant 1$.

Proof. Fix $j,k \in \{1,\ldots,n\}$ and let θ_j, ϕ_k and λ_j, μ_k be as defined in Steps 2 and 4 of the rounding procedure, respectively. Applying Theorem 11,

$$\mathbb{E}\big[\,\lambda_j\,\mu_k\,\big] = \sum_{\ell=0}^{L} b_{2\ell+1}\frac{1}{2\pi}\int_{-\pi}^{\pi} f\big((2\ell+1)\theta_j - t\big)g\big(t-(2\ell+1)\phi_k\big)\mathrm{d}t = \frac{1}{\sqrt{2}}\cos(\theta_j-\phi_k) + \eta_{jk},$$

where

$$\eta_{jk} \stackrel{\text{def}}{=} -\sum_{\ell=L+1}^{\infty} b_{2\ell+1}\frac{1}{2\pi}\int_{-\pi}^{\pi} f\big((2\ell+1)\theta_j - t\big)g\big(t-(2\ell+1)\phi_k\big)\mathrm{d}t.$$

By definition, $\cos(\theta_j - \phi_k) = \Re\left(x_j\overline{y_k}\right)$. Using $1-\varepsilon \leqslant p \leqslant 1$, which follows from the bound stated in Theorem 11, η_{jk} equals $1-p \leqslant \varepsilon$ times a weighted average of the product of certain values taken by f and g, the former only depending on θ_j and the latter on ϕ_k. Equivalently, this weighted average can be written as the inner product of two vectors x'_j and y'_k of norm at most 1. Finally, all steps of the rounding procedure can be performed in time polynomial in n and $1/\varepsilon$. □

3.2 Rounding in the Hermitian case

Let $M \in M_n(M_n(\mathbb{C}))$ be Hermitian, and $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$. For every $r \in \{1,\ldots,d\}$ define as usual $X_r, Y_r \in M_n(\mathbb{C})$ by $(X_r)_{jk} = (X_{jk})_r$ and $(Y_r)_{jk} = (Y_{jk})_r$. Define $X', Y' \in M_n(\mathbb{C}^{2d})$ by

$$X'_{jk} \stackrel{\text{def}}{=} \sum_{p=1}^{d} \left(\left(\frac{X_p + X_p^*}{2} \right)_{jk} e_{2p-1} + i \left(\frac{X_p - X_p^*}{2} \right)_{jk} e_{2p} \right) \in \mathbb{C}^{2d},$$

and

$$Y'_{jk} \stackrel{\text{def}}{=} \sum_{p=1}^{d} \left(\left(\frac{Y_p + Y_p^*}{2} \right)_{jk} e_{2p-1} + i \left(\frac{Y_p - Y_p^*}{2} \right)_{jk} e_{2p} \right) \in \mathbb{C}^{2d}.$$

Then $(X')(X')^* = (X')^*(X') = (XX^* + X^*X)/2 = I$, so $X' \in \mathcal{U}_n(\mathbb{C}^{2d})$ and similarly $Y' \in \mathcal{U}_n(\mathbb{C}^{2d})$. Moreover, since M is Hermitian, $|M(X,Y)| = |M(X',Y')|$. This shows that for the purpose of proving the noncommutative Grothendieck inequality for Hermitian M we may assume without loss of generality that the "component matrices" of X, Y are Hermitian themselves. Nevertheless, even in this case the rounding algorithm described in Figure 1 returns unitary matrices A, B that are not necessarily Hermitian. The following simple argument shows how Krivine's two-dimensional rounding scheme can be applied on the eigenvalues of A, B to obtain Hermitian matrices of norm 1, at the loss of a factor $\sqrt{2}$ in the approximation. A similar argument, albeit not explicitly algorithmic, also appears in [RV12b, Claim 4.7].

Theorem 13. *Let n be an integer, $M \in M_n(M_n(\mathbb{C}))$ Hermitian, $\varepsilon \in (0,1)$ and $X, Y \in \mathcal{U}_n(\mathbb{C}^d)$ such that*

$$\big|\, M\big(X,\, Y\big) \,\big| \geqslant (1-\varepsilon)\mathrm{SDP}_{\mathbb{C}}(M).$$

Then the rounding procedure described in Figure 3 runs in time polynomial in n and $1/\varepsilon$ and outputs a pair of Hermitian matrices $A', B' \in M_n(\mathbb{C})$ with norm at most 1 such that

$$\mathbb{E}\Big[\, \big|\, M\big(A',\, B'\big) \,\big| \,\Big] \geqslant \Big(\frac{1}{2\sqrt{2}} - \Big(1 + \frac{1}{\sqrt{2}}\Big)\varepsilon\Big)\mathrm{SDP}_{\mathbb{C}}(M).$$

Proof. Let $A, B \in M_n(\mathbb{C})$ be as defined as in Step 2 of Figure 3, and assume as in Figure 3 that $M(A,B)$ is real. By Theorem 4 we have $\mathbb{E}[|M(A,B)|] \geqslant \left(\frac{1}{2} - \varepsilon\right)\mathrm{SDP}_{\mathbb{C}}(M)$. Hence to conclude it will suffice to show that for any fixed pair of matrices $A, B \in M_n(\mathbb{C})$,

$$\mathbb{E}\Big[\, \big|M(A',B')\big| \,\Big] \geqslant \frac{1}{\sqrt{2}}\big|M(A,B)\big| - \varepsilon\,\mathrm{SDP}_{\mathbb{C}}(M), \tag{54}$$

Hermitian rounding procedure

1. Let $X, Y \in M_n(\mathbb{C}^d)$ and $\varepsilon > 0$ be given as input.
2. Let $A, B \in M_n(\mathbb{C})$ be the unitary matrices returned by the complex rounding procedure described in Figure 1. If necessary, multiply A by a complex phase to ensure that $M(A,B)$ is real. Write the spectral decompositions of A, B as
$$A = \sum_{j=1}^n e^{i\theta_j} u_j u_j^* \qquad \text{and} \qquad B = \sum_{k=1}^n e^{i\phi_k} v_k v_k^*,$$
where $\theta_j, \phi_k \in \mathbb{R}$ and $u_j, v_k \in \mathbb{C}^n$.
3. Apply the two-dimensional rounding algorithm from Figure 2 to the vectors $x_j \stackrel{\mathrm{def}}{=} e^{i\theta_j}$ and $y_k \stackrel{\mathrm{def}}{=} e^{i\phi_k}$. Let λ_j, μ_k be the results.
4. Output
$$A' \stackrel{\mathrm{def}}{=} \sum_{j=1}^n \lambda_j u_j u_j^* \qquad \text{and} \qquad B' \stackrel{\mathrm{def}}{=} \sum_{k=1}^n \mu_k v_k v_k^*.$$

Figure 3: The Hermitian rounding algorithm takes as input a pair of vector-valued matrices $X, Y \in M_n(\mathbb{C}^d)$. It outputs two Hermitian matrices $A', B' \in M_n(\mathbb{C})$ of norm at most 1.

where A', B' are as returned by the rounding procedure and the expectation is over the random choices made in the two-dimensional rounding step, i.e., Step 3 of Figure 3. Applying Claim 12,

$$\begin{aligned}
\mathbb{E}\big[\,\big|M(A',B')\big|\,\big] &\geqslant \Big|\sum_{j,k=1}^n M(u_j u_j^*, v_k v_k^*)\,\mathbb{E}\big[\lambda_j \mu_k\big]\Big| \\
&\stackrel{(53)}{\geqslant} \Big|\frac{1}{\sqrt{2}}\sum_{j,k=1}^n M(u_j u_j^*, v_k v_k^*)\Re\big(e^{i(\theta_j-\phi_k)}\big)\,\Big| - \varepsilon\Big|\sum_{j,k=1}^n \langle x_j', y_k'\rangle\, M(u_j u_j^*, v_k v_k^*)\Big| \\
&= \frac{1}{\sqrt{2}}\big|M(A,B)\big| - \varepsilon\big|M(W,Z)\big|, \qquad (55)
\end{aligned}$$

where in the last inequality, for the first term we used that $M(u_j u_j^*, v_k v_k^*)$ is real since M is Hermitian, and for the second term we defined the vector-valued matrices

$$W \stackrel{\mathrm{def}}{=} \sum_{j=1}^n x_j'\, u_j u_j^* \in M_n(L_2(\mathbb{C})) \qquad \text{and} \qquad Z \stackrel{\mathrm{def}}{=} \sum_{k=1}^n y_k'\, v_k v_k^* \in M_n(L_2(\mathbb{C})).$$

One checks that $WW^* = \sum_{j=1}^n \|x_j'\|_2^2 u_j u_j^*$. Since $\|x_j'\|_2 \leqslant 1$ for all $j \in \{1,\ldots,n\}$, it follows that $\|WW^*\| \leqslant 1$. Similarly $\max\{\|W^*W\|, \|ZZ^*\|, \|Z^*Z\|\} \leqslant 1$. Applying Lemma 6 we obtain $R, S \in \mathcal{U}_n(\ell_2(\mathbb{C}))$ such that $M(W,Z) = M(R,S)$, hence $|M(W,Z)| \leqslant \mathrm{SDP}_{\mathbb{C}}(M)$. Eq. (55) then implies the desired estimate (54). $\square$

3.3 Proof of Theorem 10

In this section we prove Theorem 10. We first record for future use the following simple lemma, which is an algorithmic version of (52).

Lemma 14. *There exists a polynomial time algorithm that takes as input a* 4*-tensor* $M \in M_n(M_n(\mathbb{R}))$ *and two matrices* $A, B \in M_n(\mathbb{R})$ *with* $\max\{\|A\|, \|B\|\} \leqslant 1$ *and outputs two orthogonal matrices* $U, V \in \mathcal{O}_n$ *such that*

$$\sum_{i,j,k,l=1}^{n} M_{ijkl} A_{ij} B_{kl} \leqslant \sum_{i,j,k,l=1}^{n} M_{ijkl} U_{ij} V_{kl}.$$

Proof. Write the singular value decompositions of A, B as $A = \sum_{i=1}^{n} \sigma_i e_i f_i^*$ and $B = \sum_{i=1}^{n} \tau_i g_i h_i^*$, where each of the sequences $(e_i)_{i=1}^n, (f_i)_{i=1}^n, (g_i)_{i=1}^n, (h_i)_{i=1}^n \subseteq \mathbb{R}^n$ is orthonormal, and $\sigma, \tau \in [0,1]^n$ (since $\max\{\|A\|, \|B\|\} \leqslant 1$). Now, $M(A,B)$ is given by

$$\sum_{i,j=1}^{n} \sigma_i \tau_j M(e_i f_i^*, g_j h_j^*). \tag{56}$$

Fixing all σ_i, τ_j but, say, σ_1, (56) is a linear function of σ_1, and thus we can shift σ_1 to either -1 or 1, without decreasing (56). Proceeding in this way with the other variables $\sigma_2, \ldots, \sigma_n, \tau_1, \ldots, \tau_n$, each one in its turn, we obtain $\varepsilon, \delta \in \{-1,1\}^n$ such that if we define $U, V \in \mathcal{O}_n$ by $U = \sum_{i=1}^{n} \varepsilon_i (e_i f_i^*)$ and $V = \sum_{i=1}^{n} \delta_i (g_i h_i^*)$ then $M(A,B) \leqslant M(U,V)$, as required. $\square$

Proof of Theorem 10. We first prove the implication 1) $\implies$ 2). For any $A \in M_{2n}(\mathbb{C})$ define $A_1, A_2, A_3, A_4 \in M_n(\mathbb{C})$ through the block decomposition $A = \left(\begin{smallmatrix} A_1 & A_2 \\ A_3 & A_4 \end{smallmatrix}\right)$. Given $M \in M_n(M_n(\mathbb{R}))$ let $M' \in M_{2n}(M_{2n}(\mathbb{R}))$ be such that $M'(A,B) = M(\Re A_2, \Re B_2)$ for every $A, B \in H_{2n}$. Formally, for every $i,j,k,l \in \{1, \ldots, 2n\}$ we have

$$M'_{ijkl} = \frac{1}{4} \begin{cases} M_{i,j-n,k,l-n} & \text{if } (i,k) \in \{1,\ldots,n\}^2 \text{ and } (j,l) \in \{n+1,\ldots,2n\}^2, \\ M_{j,i-n,l,k-n} & \text{if } (j,l) \in \{1,\ldots,n\}^2 \text{ and } (i,k) \in \{n+1,\ldots,2n\}^2, \\ M_{i,j-n,l,k-n} & \text{if } (i,l) \in \{1,\ldots,n\}^2 \text{ and } (j,k) \in \{n+1,\ldots,2n\}^2, \\ M_{j,i-n,k,l-n} & \text{if } (j,k) \in \{1,\ldots,n\}^2 \text{ and } (i,l) \in \{n+1,\ldots,2n\}^2, \\ 0 & \text{otherwise.} \end{cases}$$

Then M' is Hermitian. Apply the algorithm Alg^* (whose existence is the premise of part 1) of Theorem 10) to M' to get $A, B \in H_{2n}$ such that $\mathrm{Opt}^*_{\mathbb{C}}(M) \leqslant K|M(A,B)|$. Since $A, B \in H_{2n}$ we have $A_3 = A_2^*$ and $B_3 = B_2^*$. Because $\max\{\|A\|, \|B\|\} \leqslant 1$ also $\max\{\|\Re A\|, \|\Re B\|\} \leqslant 1$. By Lemma 14 we can therefore efficiently find $U, V \in \mathcal{O}_n$ such that

$$\begin{aligned} KM(U,V) &\geqslant K|M(\Re A, \Re B)| = K|M'(A,B)| \geqslant \mathrm{Opt}^*_{\mathbb{C}}(M') \\ &= \sup_{\substack{C,D \in H_{2n} \\ \|C\|,\|D\| \leqslant 1}} |M'(C,D)| \geqslant \sup_{S,T \in \mathcal{O}_n} \left| M'\left(\left(\begin{smallmatrix} 0 & S \\ S^* & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & T \\ T^* & 0 \end{smallmatrix}\right) \right) \right| = \sup_{S,T \in O_n} M(S,T) = \mathrm{Opt}_{\mathbb{R}}(M). \end{aligned}$$

To prove the reverse implication 2) $\implies$ 1), define $\psi : M_{2n}(\mathbb{R}) \to H_n$ by

$$\psi\left(\left(\begin{smallmatrix} A_1 & A_2 \\ A_3 & A_4 \end{smallmatrix}\right) \right) \stackrel{\mathrm{def}}{=} \frac{1}{4}(A_1 + A_1^* + A_4 + A_4^*) + \frac{i}{4}(A_2 - A_2^* + A_3^* - A_3) \in H_n. \tag{57}$$

Suppose that $M \in M_n(M_n(\mathbb{C}))$ is Hermitian and define $M'' \in M_{2n}(M_{2n}(\mathbb{R}))$ by requiring that $M''(A,B) = M(\psi(A),\psi(B))$ for every $A,B \in M_{2n}(\mathbb{R})$. Note that $M(\psi(A),\psi(B)) \in \mathbb{R}$ since $M,\psi(A),\psi(B)$ are all Hermitian. Apply the algorithm Alg of part 2) to M'', obtaining two orthogonal matrices $U,V \in \mathcal{O}_{2n}$ satisfying $\mathrm{Opt}_{\mathbb{R}}(M'') \leqslant KM''(U,V) = KM(\psi(U),\psi(V))$. By Lemma 15 below we have $\max\{\|\psi(U)\|,\|\psi(V)\|\} \leqslant 1$. Moreover, using Lemma 16 below we have

$$\mathrm{Opt}_{\mathbb{R}}(M'') = \sup_{\substack{A,B\in M_{2n}(\mathbb{R})\\ \|A\|,\|B\|\leqslant 1}} \big|M(\psi(A),\psi(B))\big| \geqslant \sup_{\substack{X,Y\in H_n\\ \|X\|,\|Y\|\leqslant 1}} \left|M\left(\psi\left(\left(\begin{smallmatrix} \Re X & \Im X \\ -\Im X & \Re X\end{smallmatrix}\right)\right),\psi\left(\left(\begin{smallmatrix} \Re Y & \Im Y \\ -\Im Y & \Im Y\end{smallmatrix}\right)\right)\right)\right|. \quad (58)$$

Observe that for every $X \in H_n$ we have $\psi\left(\left(\begin{smallmatrix} \Re X & \Im X \\ -\Im X & \Re X\end{smallmatrix}\right)\right) = X$. Consequently the rightmost term in (58) equals $\mathrm{Opt}^*_{\mathbb{C}}(M)$. Therefore $\mathrm{Opt}^*_{\mathbb{C}}(M) \leqslant K|M(\psi(U),\psi(V))|$, so that the algorithm that outputs $\psi(U),\psi(V)$ has the desired approximation factor. $\square$

Lemma 15. *Let $\psi : M_{2n}(\mathbb{R}) \to H_n$ be given as in* (57). *Then $\|\psi(Y)\| \leqslant \|Y\|$ for all $Y \in M_{2n}(\mathbb{R})$.*

Proof. For $Y \in M_{2n}(\mathbb{R})$ write $Z = (Y+Y^*)/2$. Setting $Z = \left(\begin{smallmatrix} Z_1 & Z_2 \\ Z_3 & Z_4\end{smallmatrix}\right)$, where $Z_1,Z_2,Z_3,Z_4 \in M_n(\mathbb{R})$, we then have $\psi(Y) = (Z_1+Z_4)/2 + i(Z_2-Z_3)/2$. Take $z \in \mathbb{C}^n$ and write $z = x+iy$, where $x,y \in \mathbb{R}^n$. Then

$$\begin{aligned}
&\Re\big(\langle \psi(Y)z,z\rangle\big)\\
&= \frac{\langle Z_1x,x\rangle + \langle Z_4y,y\rangle - \langle Z_2y,x\rangle - \langle Z_3x,y\rangle}{2} + \frac{\langle Z_1y,y\rangle + \langle Z_4x,x\rangle + \langle Z_2x,y\rangle + \langle Z_3y,x\rangle}{2}\\
&= \frac{1}{2}\left\langle Z\begin{pmatrix} x\\ -y\end{pmatrix}, \begin{pmatrix} x\\ -y\end{pmatrix}\right\rangle + \frac{1}{2}\left\langle Z\begin{pmatrix} y\\ x\end{pmatrix}, \begin{pmatrix} y\\ x\end{pmatrix}\right\rangle.
\end{aligned}$$

Since $\psi(Y)$ is Hermitian, it follows that $\|\psi(Y)\| \leqslant \|Z\| \leqslant \|Y\|$. $\square$

Lemma 16. *For every $X \in H_n$ we have $\left\|\left(\begin{smallmatrix} \Re X & \Im X \\ -\Im X & \Re X\end{smallmatrix}\right)\right\| = \|X\|$.*

Proof. Write $Z = \left(\begin{smallmatrix} \Re X & \Im X \\ -\Im X & \Re X\end{smallmatrix}\right) \in M_{2n}(\mathbb{R})$. Since X is Hermitian, Z is symmetric. For every $a,b \in \mathbb{R}^n$,

$$\begin{aligned}
\left\langle Z\begin{pmatrix} a\\ b\end{pmatrix}, \begin{pmatrix} a\\ b\end{pmatrix}\right\rangle &= \langle(\Re X)a,a\rangle + \langle(\Im X)b,a\rangle - \langle(\Im X)a,b\rangle + \langle(\Re X)b,b\rangle\\
&= \Re\left(\langle X(a-ib),a-ib\rangle\right).
\end{aligned}$$

Since Z is symmetric and X is Hermitian, it follows that $\|Z\| = \|X\|$. $\square$

4 Direct rounding in the real case

We now describe and analyze a different rounding procedure for the real case of the noncommutative Grothendieck inequality. The argument is based on the proof of the noncommutative Grothendieck inequality due to Kaijser [Kai83], which itself has a structure similar to the proof of the classical Grothendieck inequality due to Krivine [Kri74] (see also [Jam85], the "Notes and Remarks" section of Chapter 1 of [DJT95], and the survey article [JL01]), and uses ideas from [Pis78] to extend that proof to the non-commutative setting.

Real rounding procedure

1. Let $X, Y \in M_n(\mathbb{R}^d)$ be given as input, and let $\varepsilon \in \{-1,1\}^d$ be chosen uniformly at random.
2. Set $Y_\varepsilon \stackrel{\mathrm{def}}{=} \langle \varepsilon, Y\rangle$. Write the singular value decomposition of Y_ε as $Y_\varepsilon = \sum_{i=1}^n t_i(\varepsilon)\, u_i(\varepsilon) v_i(\varepsilon)^*$, where $t_i(\varepsilon) \in [0,\infty)$ and $u_i(\varepsilon), v_i(\varepsilon) \in \mathbb{R}^n$. Define
$$(Y_\varepsilon)_\tau \stackrel{\mathrm{def}}{=} \sum_{i=1}^n \min\{t_i(\varepsilon), \tau\} u_i(\varepsilon) v_i(\varepsilon)^*,$$
where $\tau \stackrel{\mathrm{def}}{=} \sqrt{3}/2$.
3. Let $X(\varepsilon) \in M_n(\mathbb{R})$ of norm at most 1 be such that
$$M(X(\varepsilon), (Y_\varepsilon)_\tau) = \max_{\substack{X \in M_n(\mathbb{R})\\ \|X\| \leqslant 1}} |M(X, (Y_\varepsilon)_\tau)|.$$
4. Output the pair $A = X(\varepsilon)$ and $B = \frac{1}{\tau}(Y_\varepsilon)_\tau$.

Figure 4: The real rounding algorithm takes as input $X, Y \in M_n(\mathbb{R}^d)$. It outputs two real matrices $A, B \in M_n(\mathbb{R})$ of norm at most 1.

Theorem 17. *Given $n \in \mathbb{N}$, $M \in M_n(M_n(\mathbb{R}))$ and $\eta \in (0, 1/2)$, suppose that $X, Y \in \mathcal{O}_n(\mathbb{R}^d)$ are such that*
$$\big|\, M\big(X,\, Y\big) \,\big| \;\geqslant\; (1-\eta)\mathrm{SDP}_{\mathbb{R}}(M), \tag{59}$$
where $\mathrm{SDP}_{\mathbb{R}}(M)$ is defined in (12)*. Then the rounding procedure described in Figure 4 runs in time polynomial in n and outputs a pair of real matrices $A, B \in M_n(\mathbb{R})$ with norm at most* 1 *such that*
$$\mathbb{E}\Big[\,\big|\, M\big(A,\, B\big) \,\big|\,\Big] \;\geqslant\; \frac{(1-2\eta)^2}{3\sqrt{3}}\mathrm{SDP}_{\mathbb{R}}(M). \tag{60}$$

Note that by Lemma 14 we can also efficiently convert the matrices A, B to orthogonal matrices $U, V \in \mathcal{O}_n$ without changing the approximation guarantee.

The proof of Theorem 17 relies on two claims that are used to bound the error that results from the truncation step (step 2) in the rounding procedure in Figure 4. The first claim plays the same role as Claim 5 did in the complex case.

Claim 18. *Fix $X \in M_n(\mathbb{R}^d)$ and let $\varepsilon \in \{-1,1\}^d$ be chosen uniformly at random. Set $X_\varepsilon = \langle \varepsilon, X\rangle$. Then*
$$\mathbb{E}_\varepsilon\big[(X_\varepsilon X_\varepsilon^*)^2\big] \leqslant (XX^*)^2 + 2\|X^*X\|XX^*,$$
$$\mathbb{E}_\varepsilon\big[(X_\varepsilon^* X_\varepsilon)^2\big] \leqslant (X^*X)^2 + 2\|XX^*\|X^*X.$$

Proof. Define the symmetric real vector-valued matrix Z by
$$Z \stackrel{\mathrm{def}}{=} \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}.$$

Following the proof of Lemma 1.1 in [Pis78] (which establishes a bound analogous to the one proved in Claim 8 for the case of i.i.d. $\{\pm 1\}$ random variables) and defining as usual $Z_r \in M_n(\mathbb{R})$ by $(Z_r)_{ij} = (Z_{ij})_r$ for every $r \in \{1, \ldots, d\}$, we have

$$\begin{aligned}\mathbb{E}_\varepsilon\big[\langle \varepsilon, Z\rangle^4\big] &= \sum_{r=1}^{d} Z_r^4 + \sum_{\substack{r,s\in\{1,\ldots,d\}\\ r\neq s}} \big(Z_r^2 Z_s^2 + Z_r Z_s Z_r Z_s + Z_r Z_s^2 Z_r\big)\\ &= \Big(\sum_{r=1}^{d} Z_r^2\Big)^2 + \sum_{\substack{r,s\in\{1,\ldots,d\}\\ r< s}} \big(Z_r Z_s + Z_s Z_r\big)^2. \end{aligned} \tag{61}$$

Using the inequality $(A+B)(A+B)^* \leqslant 2(AA^* + BB^*)$, which holds for all $A, B \in M_n(\mathbb{R})$, we can bound the second sum in (61) as follows.

$$\begin{aligned}\sum_{\substack{r,s\in\{1,\ldots,d\}\\ r< s}} \big(Z_r Z_s + Z_s Z_r\big)^2 &\leqslant 2\Big(\sum_{r=1}^{d} Z_r\Big(\sum_{\substack{s\in\{1,\ldots,d\}\\ s>r}} Z_s^2\Big) Z_r + \sum_{s=1}^{d} Z_s\Big(\sum_{\substack{r\in\{1,\ldots,d\}\\ r<s}} Z_r^2\Big) Z_s\Big)\\ &= 2\sum_{r=1}^{d} Z_r\Big(\sum_{\substack{s\in\{1,\ldots,d\}\\ s\neq r}} Z_s^2\Big) Z_r\\ &\leqslant 2\sum_{r=1}^{d} Z_r\Big(\sum_{s=1}^{d} Z_s^2\Big) Z_r.\end{aligned}$$

Replacing Z_r by its definition and using $ABA^* \leqslant \|B\| AA^*$, which holds true for every positive semidefinite $B \in M_n(\mathbb{R})$, we arrive at the following matrix inequality:

$$\begin{aligned}\mathbb{E}_\varepsilon\big[\langle \varepsilon, Z\rangle^4\big] &= \begin{pmatrix} \mathbb{E}_\varepsilon\big[(X_\varepsilon X_\varepsilon^*)^2\big] & 0\\ 0 & \mathbb{E}_\varepsilon\big[(X_\varepsilon^* X_\varepsilon)^2\big]\end{pmatrix}\\ &\leqslant \begin{pmatrix} (XX^*)^2 & 0\\ 0 & (X^*X)^2\end{pmatrix} + 2\begin{pmatrix} \|X^*X\|\, XX^* & 0\\ 0 & \|XX^*\|\, X^*X\end{pmatrix}.\end{aligned}$$

The inequality above implies a separate matrix inequality for both diagonal blocks, proving the claim. □

The second claim appears as Lemma 2.3 in [Kai83] (in the complex case). We include a short proof for the sake of completeness.

Claim 19. *Let $Y \in M_n(\mathbb{R})$. For any $\tau > 0$, there exists a decomposition $Y = Y_\tau + Y^\tau$ such that $\|Y_\tau\|_\infty \leqslant \tau$ and*

$$Y^\tau (Y^\tau)^* \leqslant \frac{1}{(4\tau)^2}(YY^*)^2 \qquad \text{and} \qquad (Y^\tau)^* Y^\tau \leqslant \frac{1}{(4\tau)^2}(Y^*Y)^2.$$

Proof. Define Y_τ by "truncating" the singular values of Y at τ, as is done in step 2 of the rounding procedure described in Figure 4, so that $\|Y_\tau\| \leqslant \tau$. Define $Y^\tau = Y - Y_\tau$. By definition, Y, Y_τ and Y^τ all have the same singular vectors, and the non-zero singular values of Y^τ are of the form $\mu - \tau$, where $\mu \geqslant \tau$ is a singular value of Y. Using the bound $|\mu - \tau| \leqslant \mu^2/(4\tau)$ valid for any $\mu \geqslant \tau$, any nonzero eigenvalue $\lambda = (\mu - \tau)^2$ of $Y^\tau (Y^\tau)^*$ (resp. $(Y^\tau)^* Y^\tau$) satisfies $\lambda \leqslant \mu^4/(4\tau)^2$, which proves the claim. □

Proof of Theorem 17. We shall continue using the notation introduced in Figure 4, Claim 18 and Claim 19. Let $X, Y \in \mathcal{O}_n(\mathbb{R}^d)$ satisfying (59). For every $\varepsilon \in \{-1,1\}^d$ and any $\tau > 0$ let

$$Y_\varepsilon = \langle \varepsilon, Y\rangle = (Y_\varepsilon)_\tau + Y_\varepsilon^\tau$$

be the decomposition promised by Claim 19. Combining the bound from Claim 18 with the one from Claim 19, we see that

$$\big\|\mathbb{E}_\varepsilon\big[Y_\varepsilon^\tau (Y_\varepsilon^\tau)^*\big]\big\| \leqslant \frac{1}{(4\tau)^2}\big(\|YY^*\|^2 + 2\|Y^*Y\|\|YY^*\|\big) = \frac{3}{(4\tau)^2}, \tag{62}$$

where the final step of (62) follows from $Y \in \mathcal{O}_n(\mathbb{R}^d)$, and the same bound holds on $\|\mathbb{E}_\varepsilon\big[(Y_\varepsilon^\tau)^* Y_\varepsilon^\tau\big]\|$. We have

$$\begin{aligned} \big|M(X,\, Y)\big| = \big|\mathbb{E}_\varepsilon\big[M(X_\varepsilon,\, Y_\varepsilon)\big]\big| &\leqslant \big|\mathbb{E}_\varepsilon\big[M(X_\varepsilon,\, (Y_\varepsilon)_\tau)\big]\big| + \big|\mathbb{E}_\varepsilon\big[M(X_\varepsilon,\, Y_\varepsilon^\tau)\big]\big| \\ &\leqslant \big|\mathbb{E}_\varepsilon\big[M(X_\varepsilon,\, (Y_\varepsilon)_\tau)\big]\big| + \frac{\sqrt{3}}{4\tau}\mathrm{SDP}_{\mathbb{R}}(M), \end{aligned} \tag{63}$$

where the last inequality in (63) follows from the definition of $\mathrm{SDP}_{\mathbb{R}}(M)$ and (62).

To bound the first term in (63), let $X_\varepsilon = \sum_{i=1}^n s_i(\varepsilon) w_i(\varepsilon) z_i(\varepsilon)^*$, $(Y_\varepsilon)_\tau = \sum_{i=1}^n t_i'(\varepsilon) u_i(\varepsilon) v_i(\varepsilon)^*$, where $s_i(\varepsilon), t_i'(\varepsilon) \geqslant 0$ and $u_i(\varepsilon), v_i(\varepsilon), w_i(\varepsilon), z_i(\varepsilon) \in \mathbb{R}^n$, be the singular value decompositions of $X_\varepsilon, (Y_\varepsilon)_\tau$, respectively. Set $M_{i,j}(\varepsilon) \stackrel{\mathrm{def}}{=} M(w_i(\varepsilon) z_i(\varepsilon)^*, u_j(\varepsilon) v_j(\varepsilon)^*)$. With these definitions,

$$\begin{aligned} \big|\mathbb{E}_\varepsilon\left[M(X_\varepsilon,\, (Y_\varepsilon)_\tau)\right]\big| = \Big|\mathbb{E}_\varepsilon\Big[\sum_{i,j=1}^n M_{i,j}(\varepsilon) s_i(\varepsilon) t_j'(\varepsilon)\Big]\Big| &\leqslant \mathbb{E}_\varepsilon\Big[\sum_{i=1}^n s_i(\varepsilon)\Big|\sum_{j=1}^n M_{i,j}(\varepsilon) t_j'(\varepsilon)\Big|\Big] \\ &\leqslant \Big(\mathbb{E}_\varepsilon\Big[\sum_{i=1}^n s_i(\varepsilon)^2\Big|\sum_{j=1}^n M_{i,j}(\varepsilon) t_j'(\varepsilon)\Big|\Big]\Big)^{1/2}\Big(\mathbb{E}_\varepsilon\Big[\sum_{i=1}^n\Big|\sum_{j=1}^n M_{i,j}(\varepsilon) t_j'(\varepsilon)\Big|\Big]\Big)^{1/2}. \end{aligned} \tag{64}$$

Note that the rightmost term in (64) is at most $\Big(\tau\mathbb{E}_\varepsilon\Big[M(X(\varepsilon), B(\varepsilon))\Big]\Big)^{1/2}$, where $B(\varepsilon) = \frac{1}{\tau}(Y_\varepsilon)_\tau$ and $X(\varepsilon)$ is as defined in the rounding procedure in Figure 4. To bound the leftmost term in (64), define $W_\varepsilon \stackrel{\mathrm{def}}{=} \sum_{i=1}^n (r_i(\varepsilon) s_i(\varepsilon)^2) w_i(\varepsilon) z_i(\varepsilon)^*$, where $r_i(\varepsilon)$ is the sign of $\sum_{j=1}^n M_{i,j}(\varepsilon) t_j'(\varepsilon)$, so that

$$\sum_{i=1}^n s_i(\varepsilon)^2 \Big|\sum_{j=1}^n M_{i,j}(\varepsilon) t_j'(\varepsilon)\Big| = M(W_\varepsilon, (Y_\varepsilon)_\tau).$$

Moreover, by definition $W_\varepsilon W_\varepsilon^* = (X_\varepsilon X_\varepsilon^*)^2$, so using Claim 18 we have

$$\big\|\mathbb{E}_\varepsilon\big[W_\varepsilon^* W_\varepsilon\big]\big\| = \big\|\mathbb{E}_\varepsilon\left[(X_\varepsilon^* X_\varepsilon)^2\right]\big\| \leqslant \|X^*X\|^2 + 2\|XX^*\|\|X^*X\| = 3,$$

and the same bound holds for $\|\mathbb{E}_\varepsilon\big[W_\varepsilon W_\varepsilon^*\big]\|$. By the definition of $(Y_\varepsilon)_\tau$ we also have

$$\max\Big\{\big\|\mathbb{E}_\varepsilon\left[(Y_\varepsilon)_\tau^* (Y_\varepsilon)_\tau\right]\big\|,\, \big\|\mathbb{E}_\varepsilon\left[(Y_\varepsilon)_\tau (Y_\varepsilon)_\tau^*\right]\big\|\Big\} \leqslant \tau^2.$$

Hence

$$\mathbb{E}_\varepsilon\Big[\sum_{i=1}^n s_i(\varepsilon)^2\Big|\sum_{j=1}^n M_{i,j}(\varepsilon) t_j'(\varepsilon)\Big|\Big] \;=\; \sqrt{3}\tau\,\mathbb{E}_\varepsilon\Big[M\Big(\frac{W_\varepsilon}{\sqrt{3}}, \frac{(Y_\varepsilon)_\tau}{\tau}\Big)\Big] \;\leqslant\; \sqrt{3}\tau\,\mathrm{SDP}_{\mathbb{R}}(M),$$

where we used the definition of $\mathrm{SDP}_{\mathbb{R}}(M)$. Finally, combining (63) and (64) with the bounds shown above we obtain

$$(1-\eta)\mathrm{SDP}_{\mathbb{R}}(M) \overset{(59)}{\leqslant} \big|M\big(X,Y\big)\big| \leqslant \sqrt[4]{3}\tau\sqrt{\mathrm{SDP}_{\mathbb{R}}(M)\cdot \mathbb{E}_\varepsilon\left[M(X(\varepsilon),B(\varepsilon))\right]}+\frac{\sqrt{3}}{4\tau}\mathrm{SDP}_{\mathbb{R}}(M). \quad (65)$$

Setting $\tau=\sqrt{3}/2$ in (65) and simplifying leads to the desired bound (60).

All steps in the rounding procedure can be performed efficiently. The calculation of $X(\varepsilon)$ in the third step of Figure 4 can be expressed as a semidefinite program and solved in time polynomial in n. Alternatively, one may directly compute $X(\varepsilon)$ by writing the polar decomposition

$$\mathrm{Tr}_2(M^*(I\otimes B)) = QP \in M_n(\mathbb{R}),$$

where the partial trace is taken with respect to the second tensor, Q is an orthogonal matrix and P is positive semidefinite, and taking $X(\varepsilon)=Q^*$. □

5 Some applications

Before presenting the details of our applications of Theorem 1, we observe that the problem of computing $\mathrm{Opt}_{\mathbb{R}}(M)$ is a rather versatile optimization problem, perhaps more so than what one might initially guess from its definition. The main observation is that by considering matrices M which only act non-trivially on certain diagonal blocks of the two variables U, V that appear in the definition of $\mathrm{Opt}_{\mathbb{R}}(M)$, these variables can each be thought of as a sequence of multiple matrix variables, possibly of different shapes but all with operator norm at most 1. This allows for some flexibility in adapting the noncommutative Grothendieck optimization problem to concrete settings, and we explain the transformation in detail next.

For every $n,m\geqslant 1$, let $M_{m,n}(\mathbb{R})$ be the vector space of real $m\times n$ matrices. Given integers $k,\ell\geqslant 1$ and sequences of integers $(m_i),(n_i)\in\mathbb{N}^k$, $(p_j),(q_j)\in\mathbb{N}^\ell$, we define $\mathrm{Bil}_{\mathbb{R}}(k,\ell;(m_i),(n_i),(p_j),(q_j))$, or simply $\mathrm{Bil}_{\mathbb{R}}(k,\ell)$ when the remaining sequences are clear from context, as the set of all

$$f:\Big(\bigoplus_{i=1}^k M_{m_i,n_i}(\mathbb{R})\Big)\times\Big(\bigoplus_{j=1}^\ell M_{p_j,q_j}(\mathbb{R})\Big)\to\mathbb{R}$$

that are linear in both arguments. Concretely, $f\in\mathrm{Bil}_{\mathbb{R}}(k,\ell)$ if and only if there exists real coefficients $\alpha_{irs,juv}$ such that for every $(A_i)\in\bigoplus_{i=1}^k M_{m_i,n_i}(\mathbb{R})$ and $(B_j)\in\bigoplus_{j=1}^\ell M_{p_j,q_j}(\mathbb{R})$,

$$f\big((A_i)_{i\in\{1,\ldots,k\}},(B_j)_{j\in\{1,\ldots,\ell\}}\big)=\sum_{i=1}^k\sum_{j=1}^\ell\sum_{r=1}^{m_i}\sum_{s=1}^{n_i}\sum_{u=1}^{p_j}\sum_{v=1}^{q_j}\alpha_{irs,juv}\,(A_i)_{rs}(B_j)_{uv}. \quad (66)$$

For integers $m,n\geqslant 1$, let $\mathcal{O}_{m,n}\subset M_{m,n}(\mathbb{R})$ denote the set of all $m\times n$ real matrices U such that $UU^*=I$ if $m\leqslant n$ and $U^*U=I$ if $m\geqslant n$. If $m=n$ then $\mathcal{O}_{n,n}=\mathcal{O}_n$ is the set of orthogonal matrices; $\mathcal{O}_{n,1}$ is the set of all n-dimensional unit vectors; $\mathcal{O}_{1,1}$ is simply the set $\{-1,1\}$. Given $f\in\mathrm{Bil}_{\mathbb{R}}(k,\ell)$, consider the quantity

$$\mathrm{Opt}_{\mathbb{R}}(f)\stackrel{\mathrm{def}}{=}\sup_{\substack{(U_i)\in\bigoplus_{i=1}^k\mathcal{O}_{m_i,n_i}\\(V_j)\in\bigoplus_{j=1}^\ell\mathcal{O}_{p_j,q_j}}} f\big((U_i),(V_j)\big).$$

Note that this definition coincides with the definition of $\mathrm{Opt}_{\mathbb{R}}(f)$ given in the introduction whenever $f \in \mathrm{Bil}_{\mathbb{R}}(1,1;n,n,n,n)$. The proof of the following proposition shows that the new optimization problem still belongs the framework of the noncommutative Grothendieck problem.

Proposition 20. *There exists a polynomial time algorithm that takes as input $k,\ell \in \mathbb{N}$, $(m_i),(n_i) \in \mathbb{N}^k$, $(p_j),(q_j) \in \mathbb{N}^\ell$ and $f \in Bil_{\mathbb{R}}(k,\ell;(m_i),(n_i),(p_j),(q_j))$ and outputs $(U_i) \in \bigoplus_{i=1}^k \mathcal{O}_{m_i,n_i}$ and $(V_j) \in \bigoplus_{j=1}^\ell \mathcal{O}_{p_j,q_j}$ such that*

$$\mathrm{Opt}_{\mathbb{R}}(f) \leqslant O(1)\cdot f\big((U_i),(V_j)\big).$$

Moreover, the implied constant in the $O(1)$ term can be taken to be any number larger than $2\sqrt{2}$.

Proof. Let $k,\ell \in \mathbb{N}$, $(m_i),(n_i) \in \mathbb{N}^k$, $(p_j),(q_j) \in \mathbb{N}^\ell$ be given, and define

$$m \stackrel{\mathrm{def}}{=} \sum_{i=1}^k m_i, \qquad n \stackrel{\mathrm{def}}{=} \sum_{i=1}^k n_i, \qquad p \stackrel{\mathrm{def}}{=} \sum_{j=1}^\ell p_j, \qquad q \stackrel{\mathrm{def}}{=} \sum_{j=1}^\ell q_j,$$

and $t \stackrel{\mathrm{def}}{=} \max\{m,n,p,q\}$. We first describe how $\bigoplus_{i=1}^k M_{m_i,n_i}(\mathbb{R})$ (respectively $\bigoplus_{j=1}^\ell M_{p_j,q_j}(\mathbb{R})$) can be identified with a subset of $M_t(\mathbb{R})$ consisting of block diagonal matrices. For any $i \in \{1,\ldots,k\}$ and $r \in \{1,\ldots,m_i\}$, $s \in \{1,\ldots,n_i\}$, let $F^i_{r,s} \in M_t(\mathbb{R})$ be the matrix that has all entries equal to 0 except the entry in position $(r+\sum_{j<i} m_j, s+\sum_{j<i} n_j)$, which equals 1. Similarly, for any $j \in \{1,\ldots,\ell\}$ and $u \in \{1,\ldots,p_j\}$, $v \in \{1,\ldots,q_j\}$ we let $G^j_{u,v} \in M_t(\mathbb{R})$ be the matrix that has all entries equal to 0 except the entry in position $(u+\sum_{i<j} p_i, v+\sum_{i<j} q_i)$, which equals 1. Define linear maps $\Phi : \bigoplus_{i=1}^k M_{m_i,n_i}(\mathbb{R}) \to M_t(\mathbb{R})$ and $\Psi : \bigoplus_{j=1}^\ell M_{p_j,q_j}(\mathbb{R}) \to M_t(\mathbb{R})$ by

$$\Phi\big((A_i)\big) \stackrel{\mathrm{def}}{=} \sum_{i=1}^k \sum_{r=1}^{m_i} \sum_{s=1}^{n_i} (A_i)_{r,s} F^i_{r,s} \qquad \text{and} \qquad \Psi\big((B_j)\big) \stackrel{\mathrm{def}}{=} \sum_{j=1}^\ell \sum_{u=1}^{p_j} \sum_{v=1}^{q_j} (B_j)_{u,v} G^j_{u,v}.$$

From the definition, one verifies that

$$\|\Phi((A_i))\| = \max_{i\in\{1,\ldots,k\}} \|A_i\| \qquad \text{and} \qquad \|\Psi((B_j))\| = \max_{j\in\{1,\ldots,\ell\}} \|B_j\|. \tag{67}$$

Let $f \in \mathrm{Bil}_{\mathbb{R}}(k,\ell)$ and $(\alpha_{irs,juv})$ real coefficients as in (66). Define $M \in M_t(M_t(\mathbb{R}))$ by

$$M(F^i_{r,s}, G^j_{u,v}) \stackrel{\mathrm{def}}{=} \alpha_{irs,juv},$$

and $M(A,B) \stackrel{\mathrm{def}}{=} 0$ if A or B are not in the linear span of the $F^i_{r,s}$ or $G^j_{u,v}$ respectively. (Recall that the notation $M(A,B)$ was introduced at the beginning of Section 2.) With this definition, for any $(A_i) \in \bigoplus_{i=1}^k M_{m_i,n_i}(\mathbb{R})$ and $(B_j) \in \bigoplus_{j=1}^\ell M_{p_j,q_j}(\mathbb{R})$ we have

$$M\big(\Phi((A_i)),\Psi((B_j))\big) = f\big((A_i),(B_j)\big).$$

We claim that $\mathrm{Opt}_{\mathbb{R}}(f) = \mathrm{Opt}_{\mathbb{R}}(M)$, where $\mathrm{Opt}_{\mathbb{R}}(M)$ is as defined in the Introduction. Indeed, by (67) for any $(U_i) \in \bigoplus_{i=1}^k \mathcal{O}_{m_i,n_i}$ and $(V_j) \in \bigoplus_{j=1}^\ell \mathcal{O}_{p_j,q_j}$ we have $\|\Phi((U_i))\|, \|\Psi((V_j))\| \leqslant 1$, hence $\mathrm{Opt}_{\mathbb{R}}(f) \leqslant \mathrm{Opt}_{\mathbb{R}}(M)$. Conversely, let $U, V \in \mathcal{O}_t$ be arbitrary, and U', V' their orthogonal projections on $\mathrm{Im}(\Phi)$, $\mathrm{Im}(\Psi)$ respectively. Then $M(U', V') = M(U, V)$. Moreover, if

$$(U_i') \in \bigoplus_{i=1}^{k} M_{m_i,n_i}(\mathbb{R}) \qquad \text{and} \qquad (V_j') \in \bigoplus_{j=1}^{\ell} M_{p_j,q_j}(\mathbb{R})$$

are such that $\Phi((U_i')) = U'$ and $\Psi((V_j')) = V'$ then by (67) $\max_{i\in\{1,\ldots,k\}} \|U_i'\| = \|U'\| \leqslant \|U\| = 1$, and similarly $\max_{j\in\{1,\ldots,\ell\}} \|V_j'\| \leqslant 1$. As in the proof of Lemma 14, we may then argue that there exists $U_i \in \mathcal{O}_{m_i,n_i}$, $V_j \in \mathcal{O}_{p_j,q_j}$ such that

$$f((U_i),(V_j)) \geqslant f((U_i'),(V_j')) = M(U', V') = M(U, V),$$

proving the reverse inequality $\mathrm{Opt}_{\mathbb{R}}(f) \geqslant \mathrm{Opt}_{\mathbb{R}}(M)$.

To conclude the proof of Proposition 20 it remains to note that the algorithm of Theorem 1, when applied to M, produces in polynomial time $U, V \in \mathcal{O}_t$ such that $\mathrm{Opt}_{\mathbb{R}}(M) \leqslant O(1) \cdot M(U, V)$. Arguing as in Lemma 14, (U_i) and (V_j) can be computed from U, V in polynomial time and constitute the output of the algorithm. □

5.1 Constant-factor algorithm for robust PCA problems

We start with the R1-PCA problem, as described in (3). Let $a_1, \ldots, a_N \in \mathbb{R}^n$ be given, and define $f \in \mathrm{Bil}_{\mathbb{R}}(1, N; (K), (n), (1, \ldots, 1), (K, \ldots, K))$ by

$$f\big(Y, (Z_1, \ldots, Z_N)\big) \stackrel{\text{def}}{=} \sum_{i=1}^{N} \sum_{k=1}^{K} (Z_i)_{1k} \Big(\sum_{j=1}^{n} Y_{kj}(a_i)_j \Big).$$

The condition $Z_i \in \mathcal{O}_{1,K}$ is equivalent to Z_i being a unit vector, while $Y \in \mathcal{O}_{K,n}$ is equivalent to the K rows of Y being orthonormal. Using that the ℓ_2 norm of a vector $u \in \mathbb{R}^K$ is equal to $\max_{v\in S^{K-1}} \langle u, v\rangle$, the quantity appearing in (3) is equal to

$$\sup_{\substack{Y\in\mathcal{O}_{K,n} \\ Z_1,\ldots,Z_N\in\mathcal{O}_{1,K}}} f\big(Y, (Z_1, \ldots, Z_N)\big),$$

which by definition is equal to $\mathrm{Opt}_{\mathbb{R}}(f)$. An approximation algorithm for the R1-PCA problem then follows immediately from Proposition 20. The algorithm for L1-PCA follows similarly. Letting $g \in \mathrm{Bil}_{\mathbb{R}}(1, NK; (K), (n), (1, \ldots, 1), (1, \ldots, 1))$ be defined as

$$g\big(Y, (Z_{1,1}, \ldots, Z_{N,K})\big) = \sum_{i=1}^{N} \sum_{k=1}^{K} Z_{ik} \Big(\sum_{j=1}^{n} Y_{kj}(a_i)_j \Big),$$

and using that $Z_{ik} \in \mathcal{O}_{1,1}$ is equivalent to $Z_{ik} \in \{-1, 1\}$, the quantity appearing in (4) is equal to

$$\sup_{\substack{Y\in\mathcal{O}_{K,n} \\ Z_{1,1},\ldots,Z_{N,K}\in\mathcal{O}_{1,1}}} g\big(Y, (Z_{1,1}, \ldots, Z_{N,K})\big) = \mathrm{Opt}_{\mathbb{R}}(g),$$

which again fits into the framework of Proposition 20.

5.2 A constant-factor algorithm for the orthogonal Procrustes problem

The generalized orthogonal Procustes problem was described in Section (1.1.3). Continuing with the notation introduced there, let $A_1, \ldots, A_K$ be $d \times n$ real matrices such that the i-th row of A_k is the vector $x_i^k \in \mathbb{R}^d$. Our goal is then to efficiently approximate

$$P(A_1, \ldots A_K) \stackrel{\mathrm{def}}{=} \max_{U_1, \ldots U_K \in \mathcal{O}_d} \Big\| \sum_{k=1}^{K} U_k A_k \Big\|_2^2 = \max_{U_1, \ldots U_K \in \mathcal{O}_d} \sum_{k,l=1}^{K} \langle U_k A_k, U_l A_l \rangle, \tag{68}$$

where we set $\langle A, B \rangle \stackrel{\mathrm{def}}{=} \sum_{i=1}^{d} \sum_{j=1}^{n} A_{ij} B_{ij}$ for every two $d \times n$ real matrices A, B. We first observe that (68) is equal to

$$\max_{U_1, \ldots U_K \in \mathcal{O}_d} \max_{V_1, \ldots V_K \in \mathcal{O}_d} \Big\langle \sum_{k=1}^{K} U_k A_k, \sum_{l=1}^{K} V_l A_l \Big\rangle. \tag{69}$$

It is clear that (69) is at least as large as (68). For the other direction, using Cauchy-Schwarz,

$$\Big\langle \sum_{k=1}^{K} U_k A_k, \sum_{l=1}^{K} V_l A_l \Big\rangle \leqslant \Big\| \sum_{k=1}^{K} U_k A_k \Big\|_2 \cdot \Big\| \sum_{l=1}^{K} V_l A_l \Big\|_2,$$

so either $U_1, \ldots, U_K$ or $V_1, \ldots, V_K$ achieve a value in (68) that is at least as large as (69). The desired algorithm now follows by noting that (69) falls into the framework of Proposition 20: it suffices to define $f \in \mathrm{Bil}_{\mathbb{R}}(K, K; (d, \ldots, d), (d, \ldots, d), (d, \ldots, d), (d, \ldots, d))$ by

$$f\big((U_1, \ldots, U_K), (V_1, \ldots, V_K)\big) = \Big\langle \sum_{k=1}^{K} U_k A_k, \sum_{l=1}^{K} V_l A_l \Big\rangle.$$

Finally, from an assignment to (69) we can efficiently extract an assignment to (68) achieving at least as high a value by choosing the one among the (U_k) or the (V_l) that leads to a higher value.

Comparison with previous work. To compare the approximation guarantee that we obtained with the literature, we first note that $P(A_1, \ldots, A_K)$ is attained at $U_1, \ldots, U_K \in \mathcal{O}_d$ if and only if $U_1, \ldots, U_K$ are maximizers of the following quantity over all $V_1, \ldots, V_K \in \mathcal{O}_d$.

$$\sum_{\substack{k,l \in \{1, \ldots, K\} \\ k \neq l}} \langle V_k A_k, V_l A_l \rangle \tag{70}$$

Indeed, the diagonal term $\langle V_k A_k, V_k A_k \rangle = \|V_k A_k\|_2^2$ equals $\|A_k\|_2^2$, since V_k is orthogonal. Consequently, the quantity appearing in (70) differs from the quantity defining $P(A_1, \ldots, A_K)$ by an additive term that does not depend on $V_1, \ldots, V_K$. In the same vein, as already mentioned in Section (1.1.3), $P(A_1, \ldots, A_K)$ is attained at $U_1, \ldots, U_K \in \mathcal{O}_d$ if and only if $U_1, \ldots, U_K$ are minimizers of the following quantity over all $V_1, \ldots, V_K \in \mathcal{O}_d$.

$$\sum_{k,l=1}^{K} \|V_k A_k - V_l A_l\|_2^2. \tag{71}$$

While the optimization problems appearing in (68), (70) and (71) have the same exact solutions, this is no longer the case when it comes to approximation algorithms. To the best of our knowledge the polynomial time approximability of the minimization of the quantity appearing in (71) was not investigated in the literature. Nemirovski [Nem07] and So [So11] studied the polynomial time approximability of the maximization of the quantity appearing in (70): the best known algorithm for this problem [So11] has approximation factor $O(\log(n+d+K))$. This immediately translates to the same approximation factor for computing $P(A_1,\ldots,A_K)$, which was the previously best known algorithm for this problem. Our constant-factor approximation algorithm for $P(A_1,\ldots,A_K)$ yields an approximation for the maximization of the quantity appearing in (70) that has a better approximation factor than that of [So11] unless the additive difference $\sum_{k=1}^{K}\|A_k\|_2^2$ is too large. Precisely, our algorithm becomes applicable in this context as long as this term is at most a factor $(1-1/C)$ smaller than $P(A_1,\ldots,A_K)$. This will be the case for typical applications in which one may think of each A_k as obtained from a common A by applying a small perturbation followed by an arbitrary rotation: in that case it is reasonable to expect the optimal solution to satisfy $\langle U_kA_k, U_lA_l\rangle \approx \|A\|_2$ for every $l,k\in\{1,\ldots,K\}$; see, e.g., [Nem07, Sec. 4.3].

5.3 An algorithmic noncommutative dense regularity lemma

Our goal here is to prove Theorem 3, but before doing so we note that it leads to a PTAS for computing $\mathrm{Opt}_{\mathbb{C}}(M)$ whenever $\mathrm{Opt}_{\mathbb{C}}(M)\geqslant \kappa n\|M\|_2$ for some constant $\kappa>0$ (we shall use below the notation introduced in the statement of Theorem 3).

The idea for this is exactly as in Section 3 of [FK99]. The main point is that given such an M and $\varepsilon\in(0,1)$ the decomposition in Theorem 3 only involves $T=O(1/(\kappa\varepsilon)^2)$ terms, which can be computed in polynomial time. Given such a decomposition, we will exhaustively search over a suitably discretized set of values a_t, b_t for $\mathrm{Tr}(A_tX)$ and $\mathrm{Tr}(B_tY)$ respectively. For each such choice of values, verifying whether it is achievable using an X and Y of operator norm at most 1 can be cast as a semidefinite program. The final approximation to $\mathrm{Opt}_{\mathbb{C}}(M)$ is given by the maximum value of $|\sum_{t=1}^{T}\alpha_t a_t b_t|$, among those sequences $(a_t),(b_t)$ that were determined to be feasible.

In slightly more detail, first note that for any X,Y of operator norm at most 1 the values of $\mathrm{Tr}(A_tX)$ and $\mathrm{Tr}(B_tY)$ lie in the complex disc with radius n. Given our assumption on $\mathrm{Opt}_{\mathbb{C}}(M)$, the bound on α_t stated in Theorem 3 implies $|\alpha_t| = O(\mathrm{Opt}_{\mathbb{C}}(M)/(\kappa n^2))$. Hence an approximation of each $\mathrm{Tr}(A_tX)$ and $\mathrm{Tr}(B_tY)$ up to additive error $\varepsilon\kappa n/T$ will translate to an additive approximation error to $M(X,Y)$ of $O(\varepsilon\mathrm{Opt}_{\mathbb{C}}(M))$. As a result, to obtain a multiplicative $(1\pm\varepsilon)$ approximation to $\mathrm{Opt}_{\mathbb{C}}(M)$ it will suffice to exhaustively enumerate among $(O((n\cdot T/(\varepsilon\kappa n))^2))^{2T}$ possible values for the sequences $(a_t),(b_t)$.

Finally, to decide whether a given sequence of values can be achieved, it suffices to decide two independent feasibility problems of the following form: given $n\times n$ complex matrices $(A_t)_{t=1}^T$ of norm at most 1 and $(a_t)_{t=1}^T\in\mathbb{C}^T$, does there exist $X\in M_n(\mathbb{C})$ of norm at most 1 such that

$$\max_{t\in\{1,\ldots,T\}}\max\{|\mathrm{Re}(\mathrm{Tr}(A_tX)-a_t)|, |\mathrm{Im}(\mathrm{Tr}(A_tX)-a_t)|\}\leqslant\frac{\varepsilon\kappa n}{T}\,?$$

This problem can be cast as a semidefinite program, and feasibility can be decided in time that is polynomial in n and T.

Proof of Theorem 3. The argument is iterative. Assume that $\{A_t,B_t\}_{t=1}^{\tau-1}\subseteq\mathcal{U}_n$ have already been constructed. Write $M_1\stackrel{\mathrm{def}}{=}M$ and

$$M_\tau\stackrel{\mathrm{def}}{=}M-\sum_{t=1}^{\tau-1}\alpha_t(A_t\otimes B_t).$$

If $\mathrm{Opt}_{\mathbb{C}}(M_\tau) \leqslant \varepsilon \mathrm{Opt}_{\mathbb{C}}(M)$ then we may stop the construction. Otherwise, by Theorem 1 and multiplying by an appropriate complex phase if necessary, we can find $A_\tau, B_\tau \in \mathcal{U}_n$ such that

$$M_\tau(A_\tau, B_\tau) \geqslant \left(\frac{1}{2} - \frac{\varepsilon}{2}\right)\mathrm{Opt}_{\mathbb{C}}(M_\tau), \tag{72}$$

with the left-hand side of (72) real. Set

$$\alpha_\tau \stackrel{\mathrm{def}}{=} \frac{M_\tau(A_\tau, B_\tau)}{\|A_\tau\|_2^2 \cdot \|B_\tau\|_2^2},$$

and define $M_{\tau+1} \stackrel{\mathrm{def}}{=} M_\tau - \alpha_\tau(A_\tau \otimes B_\tau)$. By expanding the square we have

$$\begin{aligned} \|M_{\tau+1}\|_2^2 &= \|M_\tau\|_2^2 - \frac{M_\tau(A_\tau, B_\tau)^2}{\|A_\tau\|_2^2 \|B_\tau\|_2^2} \\ &\leqslant \|M_\tau\|_2^2 - \frac{(1-\varepsilon)^2}{4n^2}\mathrm{Opt}_{\mathbb{C}}(M_\tau)^2 \\ &\leqslant \|M_\tau\|_2^2 - \frac{\varepsilon^2(1-\varepsilon)^2}{4n^2}\mathrm{Opt}_{\mathbb{C}}(M)^2. \end{aligned} \tag{73}$$

It follows from (73) that as long as this process continues, $\|M_\tau\|_2^2$ decreases by an additive term of at least $\varepsilon^2(1-\varepsilon)^2\mathrm{Opt}_{\mathbb{C}}(M)^2/(4n^2)$ at each step. This process must therefore terminate after at most $T \leqslant 4n^2\|M\|_2^2/(\varepsilon(1-\varepsilon)\mathrm{Opt}_{\mathbb{C}}(M))^2$ steps.

The "moreover" part of Theorem 3 follows immediately by using the part of Theorem 1 pertaining to $\mathrm{Opt}_{\mathbb{R}}(M)$ (note that the specific approximation factor does not matter here). □

Acknowledgements. We thank Daniel Dadush and Raghu Meka for useful discussions.

References

[AMMN06] N. Alon, K. Makarychev, Y. Makarychev, and A. Naor. Quadratic forms on graphs. *Invent. Math.*, 163(3):499–522, 2006.

[AN04] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 72–80 (electronic). ACM, New York, 2004. doi:10.1145/1007352.1007371.

[BMMN11] M. Braverman, K. Makarychev, Y. Makarychev, and A. Naor. The Grothendieck constant is strictly smaller than Krivine's bound. In *Proc. 52nd FOCS*, pages 453–462. 2011.

[Bou82] P. Bourgeois. Extension de la méthode Procuste à la comparaison de nuages de points situés dans des espaces de dimensions différentes. *RAIRO Rech. Opér.*, 16(1):45–63, 1982.

[CC71] E. R. Caianiello and R. M. Capocelli. On form and language: the Procrustes algorithm for feature extraction. *Kybernetik (Berlin)*, 8(6):223–233, 1971.

[CHTW04] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *Proc. 19th IEEE Conf. on Computational Complexity (CCC'04)*, pages 236–249. 2004.

[Dev86] L. Devroye. *Nonuniform random variate generation*. Springer-Verlag, New York, 1986.

[DJT95] J. Diestel, H. Jarchow, and A. Tonge. *Absolutely summing operators*, volume 43 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. ISBN 0-521-43168-9. doi:10.1017/CBO9780511526138.

[DM98] I. L. Dryden and K. V. Mardia. *Statistical shape analysis*. Wiley, 1998.

[DZHZ06] C. Ding, D. Zhou, X. He, and H. Zha. R1-PCA: rotational invariant L1-norm principal component analysis for robust subspace factorization. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 281–288. ACM, New York, USA, 2006.

[FK99] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19:175–220, 1999.

[GD04] J. Gower and G. Dijksterhuis. *Procrustes problems*, volume 30 of *Oxford Statistical Science Series*. Oxford University Press, 2004.

[GLS93] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993.

[Gow75] J. Gower. Generalized Procrustes analysis. *Psychometrika*, 40:33–51, 1975.

[Gro53] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.

[GW95] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995.

[Haa85] U. Haagerup. The Grothendieck inequality for bilinear forms on C^*-algebras. *Adv. in Math.*, 56(2):93 – 116, 1985.

[Haa87] U. Haagerup. A new upper bound for the complex Grothendieck constant. *Israel J. Math.*, 60:199–224, 1987.

[HI95] U. Haagerup and T. Itoh. Grothendieck type norms for bilinear forms on C^*-algebras. *J. Operator Theory*, 34(2):263–283, 1995.

[HM08] U. Haagerup and M. Musat. The Effros-Ruan conjecture for bilinear forms on C^*-algebras. *Invent. Math.*, 174:139–163, 2008.

[Jam85] G. Jameson. The interpolation proof of Grothendieck's inequality. In *Proceedings of the Edinburgh Mathematical Society*, volume 28, pages 217–223. 1985.

[JKB95] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 2. Wiley, 1995.

[JL01] W. B. Johnson and J. Lindenstrauss. Basic concepts in the geometry of Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 1–84. North-Holland, Amsterdam, 2001. doi:10.1016/S1874-5849(01)80003-6.

[Kai83] S. Kaijser. A simple-minded proof of the Pisier-Grothendieck inequality. In R. Blei and S. Sidney, editors, *Banach Spaces, Harmonic Analysis, and Probability Theory*, volume 995 of *Lecture Notes in Mathematics*, pages 33–55. Springer Berlin / Heidelberg, 1983.

[KN12] S. Khot and A. Naor. Grothendieck-type inequalities in combinatorial optimization. *Commun. Pure Appl. Math.*, 65(7):992–1035, 2012.

[Kri74] J. L. Krivine. Théorèmes de factorisation dans les espaces réticulés. In *Séminaire Maurey-Schwartz 1973–1974: Espaces L^p, applications radonifiantes et géométrie des espaces de Banach, Exp. Nos. 22 et 23*, page 22. Centre de Math., École Polytech., Paris, 1974.

[Kri79] J. L. Krivine. Constantes de Grothendieck et fonctions de type positif sur les sphères. *Adv. in Math.*, 31:16–30, 1979.

[Kwa08] N. Kwak. Principal component analysis based on L1-norm maximization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(9):1672–1680, September 2008.

[LP68] J. Lindenstrauss and A. Pełczyński. Absolutely summing operators in L_p-spaces and their applications. *Studia Math.*, 29:275–326, 1968.

[LS07] L. Lovász and B. Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.

[MP07] E. Morand and J. Pagès. Analyse factorielle multiple procustéenne. *J. Soc. Fr. Stat. & Rev. Stat. Appl.*, 148(2):65–97, 2007.

[MT12] M. McCoy and J. A. Tropp. Two proposals for robust PCA using semidefinite programming. *Electron. J. Statist.*, 5:1123–1160, 1012.

[Nem07] A. Nemirovski. Sums of random symmetric matrices and quadratic optimization under orthogonality constraints. *Math. Program.*, 109(2):283–317, January 2007.

[Pis78] G. Pisier. Grothendieck's theorem for noncommutative C^*-algebras, with an appendix on Grothendieck's constants. *J. Funct. Anal.*, 29(3):397–415, 1978.

[Pis12] G. Pisier. Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc.*, 49(2):237–323, 2012. Also available at arXiv:1101.4195.

[PS02] G. Pisier and D. Shlyakhtenko. Grothendieck's theorem for operator spaces. *Invent. Math.*, 150:185–217, 2002.

[Ree91] J. A. Reeds. A new lower bound on the real Grothendieck constant, 1991. Unpublished manuscript, available at `http://www.dtc.umn.edu/reedsj/bound2.dvi`.

[RV12a] O. Regev and T. Vidick. Elementary proofs of Grothendieck theorems for completely bounded norms. *Journal of Operator Theory*, 2012. To appear.

[RV12b] O. Regev and T. Vidick. Quantum XOR games, 2012. Submitted. Available at arXiv:1207.4939.

[SB88] A. Shapiro and J. D. Botha. Dual algorithms for orthogonal Procrustes rotations. *SIAM J. Matrix Anal. Appl.*, 9(3):378–383, 1988.

[SG02] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis, 2002. Available at `http://graphics.stanford.edu/courses/cs164-09-spring/Handouts/paper_shape_spaces_im`

[So11] A. So. Moment inequalities for sums of random matrices and their applications in optimization. *Math. Program.*, 130:125–151, 2011.

[TB77] J. M. F. Ten Berge. Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977.

[Tsi87] B. S. Tsirelson. Quantum analogues of the Bell inequalities. The case of two spatially separated domains. *J. Soviet Math.*, 36:557–570, 1987.