A Framework for Quadratic Form Maximization over Convex Sets through Nonconvex Relaxations

Vijay Bhattiprolu
vspvijay@math.ias.edu
Institute for Advanced Study
Princeton University
USA

Euiwoong Lee
euiwoong@umich.edu
University of Michigan
USA

Assaf Naor
naor@math.princeton.edu
Princeton University
USA

ABSTRACT

We investigate the approximability of the following optimization problem. The input is an $n \times n$ matrix $A = (A_{ij})$ with real entries and an origin-symmetric convex body $K \subseteq \mathbb{R}^n$ that is given by a membership oracle. The task is to compute (or approximate) the maximum of the quadratic form $\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j = \langle x, Ax \rangle$ as x ranges over K. This is a rich and expressive family of optimization problems; for different choices of matrices A and convex bodies K it includes a diverse range of optimization problems like max-cut, Grothendieck/non-commutative Grothendieck inequalities, small set expansion and more. While the literature studied these special cases using case-specific reasoning, here we develop a general methodology for treatment of the approximability and inapproximability aspects of these questions.

The underlying geometry of K plays a critical role; we show under commonly used complexity assumptions that polytime constant-approximability necessitates that K has type-2 constant that grows slowly with n. However, we show that even when the type-2 constant is bounded, this problem sometimes exhibits strong hardness of approximation. Thus, even within the realm of type-2 bodies, the approximability landscape is nuanced and subtle.

However, the link that we establish between optimization and geometry of Banach spaces allows us to devise a generic algorithmic approach to the above problem. We associate to each convex body a new (higher dimensional) auxiliary set that is not convex, but is approximately convex when K has a bounded type-2 constant. If our auxiliary set has an approximate separation oracle, then we design an approximation algorithm for the original quadratic optimization problem, using an approximate version of the ellipsoid method. Even though our hardness result implies that such an oracle does not exist in general, this new question can be solved in specific cases of interest by implementing a range of classical tools from functional analysis, most notably the deep factorization theory of linear operators.

Beyond encompassing the scenarios in the literature for which constant-factor approximation algorithms were found, our generic framework implies that that for convex sets with bounded type-2 constant, constant factor approximability is preserved under the following basic operations: (a) Subspaces, (b) Quotients, (c) Minkowski Sums, (d) Complex Interpolation. This yields a rich family of new examples where constant factor approximations are possible, which were beyond the reach of previous methods. We also show (under commonly used complexity assumptions) that for symmetric norms and unitarily invariant matrix norms the type-2 constant nearly characterizes the approximability of quadratic maximization.

STOC '21, June 21–25, 2021, Virtual, Italy

ACM ISBN 978-1-4503-8053-9/21/06…$15.00
https://doi.org/10.1145/3406325.3451128

CCS CONCEPTS

• **Theory of computation → Rounding techniques**; **Convex optimization**; **Semidefinite programming**; *Random projections and metric embeddings*; • **Mathematics of computing → Continuous optimization**; *Quadratic programming*; **Functional analysis**.

KEYWORDS

Quadratic Maximization, Operator Norms, Grothendieck Inequality, Approximation Algorithms, Inapproximability, Convex Optimization, Continuous Optimization, Functional Analysis, Factorization of Linear Operators

ACM Reference Format:

Vijay Bhattiprolu, Euiwoong Lee, and Assaf Naor. 2021. A Framework for Quadratic Form Maximization over Convex Sets through Nonconvex Relaxations. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3406325.3451128

1 INTRODUCTION

Suppose that $n \in \mathbb{N}$ and that $K \subseteq \mathbb{R}^n$ is a convex body (i.e., K is convex, closed, bounded and has nonempty interior) that is origin-symmetric (i.e., $x \in K$ if and only if $-x \in K$). We will assume throughout that K is given by a membership oracle, so that the efficiency of the ensuing algorithms is measured in terms of the dependence on n and the number of oracle calls.

In this article, we will investigate the approximability of the following optimization problem, special cases of which have been extensively studied in the literature (we will discuss that background after first presenting the problem and our main algorithm). The input is an $n \times n$ matrix with real entries $A = (A_{ij}) \in M_n(\mathbb{R})$, and the task is to evaluate the quantity

$$\mathrm{Q}_K^{\max}(A) \stackrel{\mathrm{def}}{=} \max_{x \in K} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j = \max_{x \in K} \langle x, Ax \rangle \ . \qquad (1)$$

In (1) and throughout this text, $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the standard scalar product on $\mathbb{R}^n$, namely $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$ for every two vectors $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Also, we will adhere throughout to the common convention that even though within any in-line discussion the elements of $\mathbb{R}^n$ are written as row vectors, for the purpose of any linear-algebraic consideration we consider them as column vectors, i.e., members of the $n \times 1$ matrix space $M_{n\times 1}(\mathbb{R})$.

The literature also considers a bilinear variant of (1) in which one is given $m, n \in \mathbb{N}$, two convex origin-symmetric bodies $K \subseteq \mathbb{R}^n$ and $L \subseteq \mathbb{R}^m$, and an $n \times m$ matrix $B = (B_{ij}) \in M_{n\times m}(\mathbb{R})$, and the task is to evaluate (or estimate) the quantity

$$\begin{aligned}\mathrm{Op}^{\max}_{K,L}(B) &\stackrel{\mathrm{def}}{=} \max_{\substack{x\in K\\ y\in L}} \sum_{i=1}^{n}\sum_{j=1}^{m} B_{ij}x_i y_j = \max_{\substack{x\in K\\ y\in L}} \langle x, By\rangle \\ &= \frac{1}{2}\max_{z\in K\times L}\left\langle z, \begin{pmatrix} 0 & B\\ B^* & 0\end{pmatrix} z\right\rangle, \end{aligned} \tag{2}$$

where $B^* = (B_{ji}) \in M_{m\times n}(\mathbb{R})$ is the transpose of B. The final equality in (2) shows that (2) is a special case of (1), which is why we will mostly focus on (1). But, it is beneficial to consider the bilinear variant separately because sometimes it exhibits better approximation properties than what is possible in the quadratic setting (a notable example is Grothendieck's inequality; see below).

Another important special case of (1) which the literature sometimes treats separately is when the input matrix A is symmetric and positive semidefinite (PSD). In that case

$$\begin{aligned}\mathrm{Q}^{\max}_K(A) &= \max_{x\in K} \|A^{\frac12}x\|^2_{\ell_2^n} = \max_{\substack{x\in K\\ y\in \mathrm{Ball}(\ell_2^n)}} \left\langle A^{\frac12}x, y\right\rangle^2 \\ &= \left(\mathrm{Op}^{\max}_{K,\mathrm{Ball}(\ell_2^n)}\left(A^{\frac12}\right)\right)^2,\end{aligned}$$

where $\|\cdot\|_{\ell_2^n}$ is the standard Euclidean norm on $\mathbb{R}^n$ and $\mathrm{Ball}(\ell_2^n) = \{x \in \mathbb{R}^n : x_1^2 + \ldots + x_n^2 \leqslant 1\}$ is the corresponding Euclidean ball of radius 1. Thus, the PSD case of (1) is a special case of the aforementioned bilinear variant of (1), which explains why it has better properties (another reason is that in this case L is a Euclidean ball rather than a more general convex body).

The above framework is a rich and expressive family of optimization problems which contains many discrete and continuous optimization problems as special cases (corresponding to choices of matrices and convex bodies) that occur in several areas, including combinatorial optimization, computational complexity, graph theory, quantum information theory, statistical physics, machine learning, game theory and functional analysis. In fact, we suspect that many readers have already spotted familiar questions as such special cases, but in order to first discuss the contribution of the present work, we will defer specifying a variety of such examples to Section 4.5.

While the literature contains investigations of such special cases using case-specific reasoning, here we develop a general methodology for treatment of the approximability and inapproximability aspects of these questions. We devise an overarching method for obtaining constant factor approximation algorithms that includes the prior cases in the literature for which this was achieved, as well as many more new cases.

The precursor (and inspiration) of the present article is the manuscript [51] that has not yet been published but was circulated widely over the years and will be published soon (it is available on request). The goal of [51] was to broach the same issue of finding an algorithmic approach to the optimization problem (1) which treats a class of convex bodies K that is more general than the special cases that have been previously studied, as an extension of the study of the ball of ℓ_p^n that was conducted in [40] (see [32] for the corresponding hardness result under a weaker hypothesis than that of [40]). The success of [51] was partial, as it pertains only to a certain subclass of convex bodies K that satisfies the following symmetry condition.

$$\forall (x_1, \ldots, x_n) \in \mathbb{R}^n,\ (x_1, \ldots, x_n) \in K \iff (|x_1|, \ldots, |x_n|) \in K. \tag{3}$$

When (3) holds, there is an obvious vector relaxation of (1) that is given by the maximization

$$\max_{\substack{x_1,\ldots,x_n\in\mathbb{R}^n\\ (\|x_1\|_{\ell_2^n},\ldots,\|x_n\|_{\ell_2^n})\in K}} \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left\langle x_i, x_j\right\rangle. \tag{4}$$

The utility of such a relaxation was investigated in [40, 51, 52], where further geometric assumptions on K were isolated that guarantee that (4) is a convex program that has bounded integrality gap (see below). Note that (3) probes only the intersection $K \cap [0,\infty)^n$ of K with the positive orthant, which is why it is natural to study it only when (4) holds; otherwise K need not be determined by the region of space to which the relaxation (4) is sensitive.

This was the starting point of our work. Namely, for convex bodies that do not satisfy the symmetry assumption (3), there is no longer an obvious vector relaxation. Note that (3) is a stringent assumption that fails for many norms of interest; e.g. for unit balls of matrix norms such as the Schatten–von Neumann trace classes (see below) where the norm of the entry-wise absolute value $(|A_{ij}|)$ of a given matrix $A = (A_{ij}) \in M_n(\mathbb{R})$ can be drastically different from the norm of A. To overcome this conceptual obstacle, we devise an entirely different algorithmic methodology. Before proceeding, we set some notation and record some basic definitions.

2 NOTATION AND PRELIMINARIES

2.1 Normed Spaces

It is most natural to present our approach in the (equivalent) setting of normed spaces rather than origin-symmetric convex bodies. Specifically, let $\|\cdot\|_X : \mathbb{R}^n \to [0,\infty)$ be a norm on $\mathbb{R}^n$ and denote the corresponding normed space $(\mathbb{R}^n, \|\cdot\|_X)$ by X. The (closed) unit ball of X will be denoted throughout what follows by

$$\mathrm{Ball}(X) \stackrel{\mathrm{def}}{=} \{x \in \mathbb{R}^n : \|x\|_X \leqslant 1\}.$$

The standard correspondence is that $\mathrm{Ball}(X)$ is an origin-symmetric convex body, and conversely any $K \subseteq \mathbb{R}^n$ as above is equal to $\mathrm{Ball}(X)$ for some $X = (\mathbb{R}^n, \|\cdot\|_X)$, where the norm $\|x\|_X$ of each $x \in \mathbb{R}^n \setminus \{0\}$ is the unique scaling factor $s > 0$ for which $\frac{1}{s}x$ belongs to the boundary of K.

In accordance with the above convention for convex bodies, we will tacitly assume throughout the ensuing discussion that all normed spaces $X = (\mathbb{R}^n, \|\cdot\|_X)$ are given by a membership oracle for $\mathrm{Ball}(X)$. By binary search for the smallest $r \geqslant 0$ such that $x \in \mathbb{R}^n$ belongs to $r \cdot \mathrm{Ball}(X)$, such an oracle directly yields also a norm-evaluation oracle.

So, given a normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$ and a matrix $A \in M_n(\mathbb{R})$, denote

$$\mathrm{Q}_X^{\max}(A) \stackrel{\mathrm{def}}{=} \mathrm{Q}_{\mathrm{Ball}(X)}^{\max}(A).$$

Observe in passing that the bilinear variant (2) when $K = \mathrm{Ball}(X)$ and $L = \mathrm{Ball}(Y)$ for normed spaces $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $Y = (\mathbb{R}^m, \|\cdot\|_Y)$, respectively, is nothing more that the operator norm of the matrix $B \in M_{n\times m}(\mathbb{R})$ when it is viewed as an operator from Y to the dual X^* of X. Namely,

$$\mathrm{Op}_{K,L}^{\max}(B) = \|B\|_{Y\to X^*} = \|B^*\|_{X\to Y^*}, \tag{5}$$

where the first equality in (5) can be taken to be the definition of the corresponding operator norm and it is equal to the more common definition $\|B\|_{Y\to X^*} = \max_{y\in\mathrm{Ball}(Y)} \|By\|_{X^*}$ by duality (Hahn–Banach). The second equality in (5) is the fact that the norm of an operator between Banach spaces is equal to the norm of its adjoint. See e.g. the textbook [61] for this standard material.

2.2 Type and Cotype

It is beneficial to introduce the following convention regarding random variables that will be used extensively in what follows. We will work throughout with the families of random variables $\{\varepsilon_i : i \in \mathbb{N}\}$, $\{\mathrm{g}_i : i \in \mathbb{N}\}$ and $\{\mathrm{g}_{ij} : i, j \in \mathbb{N}\}$, where it will always be tacitly understood that they are independent, $\{\varepsilon_i : i \in \mathbb{N}\}$ are ±1 Bernoulli random variables, i.e., distributed uniformly over $\{-1, 1\}$, and $\{\mathrm{g}_i : i \in \mathbb{N}\}$ and $\{\mathrm{g}_{ij} : i, j \in \mathbb{N}\}$ are standard Gaussian random variables. All the expectations that appear below are with respect to the joint distribution of these random variables. We will always denote the standard Gaussian random vector in $\mathbb{R}^n$ by $\mathrm{g} = (\mathrm{g}_1, \ldots, \mathrm{g}_n)$.

The (Rademacher) type 2 constant [26] of a normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$, denoted $T_2(X)$, is the smallest $T > 0$ such that for every $m \in \mathbb{N}$, every $x_1, \ldots, x_m \in \mathbb{R}^n$ satisfy

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^m \varepsilon_i x_i\Big\|_X^2\Big] \leqslant T^2 \cdot \sum_{i=1}^m \|x_i\|_X^2, \tag{6}$$

Correspondingly, the (Rademacher) cotype 2 constant of X, denoted $C_2(X)$, is the smallest $C > 0$ such that for every $m \in \mathbb{N}$, every choice of vectors $x_1, \ldots, x_m \in \mathbb{R}^n$ satisfies

$$\sum_{i=1}^m \|x_i\|_X^2 \leqslant C^2 \cdot \mathbb{E}\Big[\Big\|\sum_{i=1}^m \varepsilon_i x_i\Big\|_X^2\Big]. \tag{7}$$

These invariants of normed spaces are of immense importance to various areas; see the survey [46] for an indication of (part of) this body of work, as well as its history. Here we show that they are closely related to the computational complexity of the quadratic optimization problem (1), and, in fact, under common complexity assumptions, they govern it in a sense that will be made precise later. For concreteness, we record the following asymptotic evaluations[1] of these constants when $X = \ell_p^n$ for some integer $n \geqslant 2$ and $p \in [1, \infty]$, all of which can be found in [48].

$$T_2(\ell_p^n) \asymp \begin{cases} n^{\frac{1}{p}-\frac{1}{2}} & \text{if } 1 \leqslant p \leqslant 2, \\ \sqrt{\min\{p, \log n\}} & \text{if } 2 \leqslant p \leqslant \infty, \end{cases} \quad \text{and}$$

$$C_2(\ell_p^n) \asymp \begin{cases} 1 & \text{if } 1 \leqslant p \leqslant 2, \\ n^{\frac{1}{2}-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty. \end{cases} \tag{8}$$

We also record the following duality relations that hold for any normed space X.

$$C_2(X^*) \leqslant T_2(X) \lesssim C_2(X^*) \log(\dim(X) + 1). \tag{9}$$

The first inequality in (9) is straightforward [47] and the second inequality in (9) is from [54].

3 A GENERIC FRAMEWORK

We are now ready to describe our algorithmic approach, starting with a simpler "warm-up" algorithm which covers many new instances of (1). Fix an integer $n \geqslant 2$. The set of symmetric positive definite matrices with real entries will be denoted by $\mathbb{PSD}^n \subseteq M_n(\mathbb{R})$. For $\mathbb{W} \in M_n(\mathbb{R})$ we will use the notation $\mathbb{W} \succcurlyeq 0$ to indicate that $\mathbb{W} \in \mathbb{PSD}^n$. We associate to a normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$ the following subset $\mathcal{U}(X)$ of $\mathbb{PSD}^n$ that we call the *upper covariance body of* X.

$$\mathcal{U}(X) \stackrel{\mathrm{def}}{=} \bigcup_{m=1}^{\infty} \Big\{ \sum_{i=1}^m w_i w_i^* : \ w_1, \ldots, w_m \in \mathbb{R}^n \ \text{ and } \ \mathbb{E}\Big[\Big\|\sum_{i=1}^m \mathrm{g}_i w_i\Big\|_X^2\Big] \leqslant 1\Big\}. \tag{10}$$

For every $w_1, \ldots, w_m \in \mathbb{R}^n$, the random vectors $\sum_{i=1}^m \mathrm{g}_i w_i$ and $\mathbb{W}^{\frac{1}{2}}\mathrm{g}$, where $\mathbb{W} = \sum_{i=1}^m w_i w_i^* \succcurlyeq 0$, are equi-distributed, since they are both Gaussian vectors whose covariance matrix is $\mathbb{W}$. Thus,

$$\mathcal{U}(X) = \Big\{\mathbb{W} \in \mathbb{PSD}^n : \ \mathbb{E}\Big[\Big\|\mathbb{W}^{\frac{1}{2}}\mathrm{g}\Big\|_X^2\Big] \leqslant 1\Big\}. \tag{11}$$

This observation explains our choice of nomenclature, namely $\mathcal{U}(X)$ consists of those covariance matrices of Gaussian vectors in $\mathbb{R}^n$ whose expected squared X-norm is bounded from above by 1. An important property of $\mathcal{U}(X)$ is that one can relate quadratic optimization over $\mathrm{Ball}(X)$ to linear optimization over $\mathcal{U}(X)$:

3.1 From Quadratic Optimization to Linear Optimization

Observe that for any $A = (A_{ij}) \in M_n(\mathbb{R})$, any normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$ satisfies

$$\mathrm{Q}_X^{\max}(A) = \max_{\mathbb{W}=(W_{ij})\in\mathcal{U}(X)} \sum_{i=1}^n \sum_{j=1}^n A_{ij} W_{ij} = \max_{\mathbb{W}\in\mathcal{U}(X)} \langle A, \mathbb{W}\rangle\,. \tag{12}$$

Indeed, if $\mathbb{W} \in \mathbb{PSD}^n$ satisfies $\mathbb{E}\big[\big\|\mathbb{W}^{\frac{1}{2}}\mathrm{g}\big\|_X^2\big] \leqslant 1$, then

$$\begin{aligned}\langle A, \mathbb{W}\rangle = \mathbf{Tr}(A\mathbb{W}) = \mathbf{Tr}\big(\mathbb{W}^{\frac{1}{2}} A \mathbb{W}^{\frac{1}{2}}\big) = \mathbb{E}\big[\big\langle \mathbb{W}^{\frac{1}{2}}\mathrm{g}, A\mathbb{W}^{\frac{1}{2}}\mathrm{g}\big\rangle\big] \\ \leqslant \mathbb{E}\Big[\mathrm{Q}_X^{\max}(A) \cdot \big\|\mathbb{W}^{\frac{1}{2}}\mathrm{g}\big\|_X^2\Big] \leqslant \mathrm{Q}_X^{\max}(A).\end{aligned}$$

This shows that right hand side of (12) is at most the left hand side of (12). The reverse inequality follows by noting that if $w \in \mathrm{Ball}(X)$, then $ww^* \in \mathcal{U}(X)$ and $\langle A, ww^*\rangle = \langle w, Aw\rangle$.

[1] In addition to the usual $o(\cdot), O(\cdot), \Omega(\cdot), \Theta(\cdot)$ notation for asymptotic relations, we will also use throughout the following (standard) asymptotic notation. For $P, Q > 0$, the notations $P \lesssim Q$ and $Q \gtrsim P$ mean that $P \leqslant KQ$ for a universal constant $K > 0$. The notation $P \asymp Q$ stands for $(P \lesssim Q) \wedge (Q \lesssim P)$.

3.2 Approximate Convexity of $\mathcal{U}(X)$

The body $\mathcal{U}(X)$ need not be convex, but it is $T_2(X)^2$-*approximately convex* in the sense that

$$\mathcal{U}(X) \subseteq \mathbf{conv}\big(\mathcal{U}(X)\big) \subseteq T_2(X)^2 \cdot \mathcal{U}(X), \tag{13}$$

where, given a subset S of some $\mathbb{R}^d$, we denote the convex hull of S by $\mathbf{conv}(S)$. To justify (13), fix $k \in \mathbb{N}$ and suppose that $\mathbb{W}_1, \dots \mathbb{W}_k \in \mathcal{U}(X)$ and $s_1, \dots, s_k \in [0,1]$ satisfy $\sum_{j=1}^k s_j = 1$. The goal is to demonstrate that $T_2(X)^{-2} \sum_{j=1}^k s_j \mathbb{W}_j \in \mathcal{U}(X)$. For each $j \in \{1, \dots, k\}$, the assumption $\mathbb{W}_j \in \mathcal{U}(X)$ means that for some $m(j) \in \mathbb{N}$ there are vectors $w_{1,j}, \dots, w_{m(j),j} \in \mathbb{R}^n$ such that

$$\mathbb{W}_j = \sum_{i=1}^{m(j)} w_{ij} w_{ij}^* \qquad \text{and} \qquad \mathbb{E}\Big[\Big\| \sum_{i=1}^{m(j)} \mathsf{g}_{ij} w_{ij} \Big\|_X^2\Big] \leqslant 1.$$

Hence,

$$\mathbb{E}\Big[\Big\| \sum_{j=1}^{k} \sum_{i=1}^{m(j)} \mathsf{g}_{ij} \sqrt{s_j} w_{ij} \Big\|_X^2\Big] = \mathbb{E}\Big[\Big\| \sum_{j=1}^{k} \varepsilon_j \sqrt{s_j} \sum_{i=1}^{m(j)} \mathsf{g}_{ij} w_{ij} \Big\|_X^2\Big]$$

$$\leqslant T_2(X)^2 \cdot \sum_{j=1}^{k} s_j \mathbb{E}\Big[\Big\| \sum_{i=1}^{m(j)} \mathsf{g}_{ij} w_{ij} \Big\|_X^2\Big] \leqslant T_2(X)^2.$$

Therefore

$$T_2(X)^{-2} \cdot \sum_{j=1}^{k} \sum_{i=1}^{m(j)} \big(\sqrt{s_j} w_{ij}\big)\big(\sqrt{s_j} w_{ij}\big)^* = T_2(X)^{-2} \cdot \sum_{j=1}^{k} s_j \mathbb{W}_j$$

indeed belongs to $\mathcal{U}(X)$.

Motivated by (13), we set the following terminology.

Definition 3.1. Suppose that $S \subseteq \mathbb{R}^n$ is star-shaped with respect to the origin, i.e., $tx \in S$ for every $x \in S$ and $t \in [0,1]$. Given $\alpha \in [1, \infty)$, we say that S is α-approximately convex if $\mathbf{conv}(S) \subseteq \alpha S$.

The two observations (12) and (13) highlight the following important facts. Firstly, the relaxation of $\mathrm{Ball}(X) \subseteq \mathbb{R}^n$ to the upper covariance body $\mathcal{U}(X) \subseteq M_n(\mathbb{R})$ is lossless, i.e., it reduces the maximization over $\mathrm{Ball}(X)$ of a quadratic form to a maximization over $\mathcal{U}(X)$ of a linear function. Secondly, the geometry of X, through the extent to which it has type 2, plays a role by ensuring that the potentially complicated set $\mathcal{U}(X)$ is at the very least approximately convex. It is thus natural to investigate the efficient optimization of linear functions over approximately convex sets. However, the following theorem (see Section 9 in [13]) shows that this is a subtle matter, because even when the type-2 constant of X is small, the computational complexity of approximating $\mathrm{Q}_X^{\max}(A)$ could be poor.

3.3 Impossibility Results

Theorem 3.2 (Impossibility of quadratic maximization assuming only bounded type-2).
For every $n \in \mathbb{N}$ and $0 < \varepsilon < 1$ there exists a distribution $\mathbb{P} = \mathbb{P}_{n,\varepsilon}$ over random normed spaces $\mathsf{X} = (\mathbb{R}^n, \|\cdot\|_{\mathsf{X}})$ and $p_n \in (0,1)$ with $\lim_{n\to\infty} p_n = 1$, such that the following properties hold.

(1) $\mathbb{P}_{n,\varepsilon}[T_2(\mathsf{X}) \lesssim 1] = 1$.
(2) $\mathbb{P}_{n,\varepsilon}\big[S \cap \mathrm{Ball}(\mathsf{X}) = S \cap \mathrm{Ball}(\ell_2^n)\big] \geqslant p_n$ *for every* $S \subseteq \mathbb{R}^n$ *with* $|S| \leqslant \exp(n^\varepsilon)$.
(3) $\mathbb{P}_{n,\varepsilon}\big[\mathrm{Q}_{\mathsf{X}}^{\max}(\mathrm{I}_n) \gtrsim n^{1-\varepsilon}\big] \geqslant p_n$, *where* $\mathrm{I}_n \in M_n(\mathbb{R})$ *is the identity matrix.*

Theorem 3.2 demonstrates that if there were an algorithm that takes as input a normed space X whose type-2 constant is $O(1)$ and outputs a number that is guaranteed to be within a factor that is $o(n^{1-\varepsilon})$ of $\mathrm{Q}_X^{\max}(\mathrm{I}_n)$, then that algorithm must necessarily make more than $\exp(n^\varepsilon)$ membership queries to $\mathrm{Ball}(X)$. Indeed, $\mathrm{Q}_X^{\max}(\mathrm{I}_n) = 1$ when $X = \ell_2^n$, while if X is the random normed space of Theorem 3.2, then $T_2(\mathsf{X}) \lesssim 1$ and with high probability $\mathrm{Q}_{\mathsf{X}}^{\max}(\mathrm{I}_n) \gtrsim n^{1-\varepsilon}$. However, if S is the set of points that the algorithm queried, then with high probability the algorithm did not obtain any information that distinguishes X from ℓ_2^n.

Thus, even if X has a small type-2 constant, this does not suffice for the existence of an efficient algorithm for approximating $\mathrm{Q}_X^{\max}(\cdot)$, but, as we have seen, requiring this property is a good place to start because it ensures that the upper covariance body is approximately convex. The following theorem establishes a further connection between type 2 and the computational complexity of approximating $\mathrm{Q}_X^{\max}(\cdot)$ by providing evidence (under a commonly used complexity assumption, namely the Small Set Expansion Hypothesis) that if the type 2 constant of X is very large, then there is no polynomial time algorithm that obtains a $O(1)$-approximation to $\mathrm{Q}_X^{\max}(\cdot)$. Further hardness results (with and without non-uniform complexity assumptions and with weaker assumptions on the growth of the type-2 constant assuming (necessarily) the Exponential Time Hypothesis), are derived in the full version [13].

Theorem 3.3 (Impossibility of quadratic maximization whenever type-2 is growing polynomially).
Fix a sequence of normed spaces $\{X^n = (\mathbb{R}^n, \|\cdot\|_{X^n})\}_{n=1}^\infty$ satisfying $T_2(X^n) = n^{\Omega(1)}$. We assume that they are given to us algorithmically in the sense that there is a polynomial time algorithm that takes as input $x \in \mathbb{R}^n$ and determines whether or not $x \in \mathrm{Ball}(X^n)$. Then, assuming the Small Set Expansion Hypothesis and that $\mathrm{NP} \nsubseteq \mathrm{P}_{/\mathrm{poly}}$, *there is no polynomial time algorithm that takes as input a matrix $A \in M_n(\mathbb{R})$ and approximates $\mathrm{Q}_{X^n}^{\max}(A)$ up to a universal constant factor.*

Remark 1. *The Small Set Expansion Hypothesis (SSEH) is a commonly used hardness assumption that was formulated in [59] and is recalled in the full version [13]. Of course, the SSEH is less standard than, say,* $\mathrm{NP} \nsubseteq \mathrm{P}_{/\mathrm{poly}}$, *so one should take Theorem 3.3 as evidence that if the underlying norm has large type-2 constant, then it is unlikely that there is an efficient constant-factor algorithm for* (1), *namely by designing such an algorithm one would refute the SSEH, thus making a major breakthrough in complexity theory.*

Remark 2. *Recalling* (8), *Theorem 3.3 applies in particular to $X^n = \ell_p^n$ when $1 \leqslant p < 2$, thus demonstrating the computational difficulty of the ℓ_p Grothendieck problem, which was left open in [40], where it was shown that this problem does have a $O_p(1)$ approximation algorithm when $2 \leqslant p < \infty$. In the unpublished manuscript [1] it was proved that a $O(1)$ approximation algorithm exists when $p = 1$ provided that all of the diagonal entries of the input matrix A vanish; see the exposition in [38]. In the full version [13] we show that if $X^n = \ell_p^n$ and $1 < p < 2$, then the hardness statement of Theorem 3.3 holds even when the diagonal of A vanishes, so in this setting we obtain* *rigorous evidence for an interesting complexity theoretic terrain: The ℓ_p Grothendieck problem is approximable when $p = 1$ or $2 \leqslant p < \infty$, but likely hard to approximate when $1 < p < 2$ or $p = \infty$ (see [7] for hardness when $p = \infty$).*

3.4 Approximation Algorithms from Upper Covariance Separation Oracle

Recall that Theorem 3.2 implies that even though the (random) upper covariance body $\mathcal{U}(\mathsf{X})$ is $O(1)$-approximately convex (as X has bounded type 2 constant), with high probability one cannot optimize linear functionals over $\mathcal{U}(\mathsf{X})$ efficiently. It turns out that the issue at hand is that even if one permits the algorithm to make oracle norm-evaluation queries for X, the auxiliary body $\mathcal{U}(\mathsf{X})$ need not even have an efficient "approximate separation oracle," which we define as follows.

Definition 3.4. Fix $\alpha \geqslant 1$. Let $S \subseteq \mathbb{R}^n$ be star shaped with respect to the origin and α-approximately convex. An α-approximate separation oracle for S is a function O defined on $\mathbb{R}^n$ that outputs to each input $x \in \mathbb{R}^n$ either "Inside" or an affine hyperplane of $\mathbb{R}^n$. The requirements for O are as follows.

- If the output $O(x)$ is "Inside," then necessarily $x \in \alpha S$.
- If the output $O(x)$ is a hyperplane $H \subseteq \mathbb{R}^n$, then H must separate x from S, i.e., x and S are contained in different sides of H. Note that this implies in particular that $x \notin \mathbf{conv}(S)$.

Observe that these requirements are not dichotomic, i.e., they are ambiguous when $x \in (\alpha S) \smallsetminus \mathbf{conv}(S)$ (recall that $\mathbf{conv}(S) \subseteq \alpha S$ since S is α-approximately convex). Namely, if $x \in (\alpha S) \smallsetminus \mathbf{conv}(S)$, then the oracle is allowed to either output a hyperplane or output "Inside."

Using a natural approximate version of the ellipsoid method, we prove the following theorem (see the full version [13]).

Theorem 3.5 (Approximate Ellipsoid Method).
Fix $\alpha \geqslant 1$ and $R \geqslant r > 0$. Suppose that $S \subseteq \mathbb{PSD}^n$ is star shaped with respect to the origin, α-approximately convex, and has an α-approximate separation oracle. Suppose also that

$$r \cdot \mathrm{Ball}(\ell_2^{n^2}) \subseteq S \subseteq R \cdot \mathrm{Ball}(\ell_2^{n^2}),$$

where we use the natural identification of $M_n(\mathbb{R})$ with $\mathbb{R}^{n^2}$. Then, there exists an algorithm that takes as input a matrix $A \in M_n(\mathbb{R})$, makes a number of oracle calls that grows polynomially in n, $\log R$, $\log(1/r)$ and the length of the bit description of A, and outputs a matrix $\mathbb{W} \in S$ that satisfies

$$\langle \mathbb{W}, A\rangle \geqslant \frac{1-o(1)}{\alpha} \sup_{\mathbb{V}\in S} \langle \mathbb{V}, A\rangle.$$

For the sake of the discussion within this extended abstract, it will be convenient to always assume tacitly that $X = (\mathbb{R}^n, \|\cdot\|_X)$ is a normed space whose upper covariance body satisfies

$$e^{-n^{O(1)}} \cdot \mathrm{Ball}(\ell_2^{n^2}) \subseteq \mathcal{U}(X) \subseteq e^{n^{O(1)}} \cdot \mathrm{Ball}(\ell_2^{n^2})\,. \tag{14}$$

Such a normalization, which is mechanical to verify in all the cases that we examined, removes the need to state running times in terms of r, R as done in Theorem 3.5. Another simplifying assumption that we will make throughout this extended abstract is that the length of the bit description of all inputs (namely matrices) to algorithms is $n^{O(1)}$.

Using Theorem 3.5 and applying it to $\mathcal{U}(X)$, we readily deduce the following approximation algorithm for quadratic maximization (see the full version [13])

Proposition 3.6 (Quadratic Maximization Given Separation Oracle for Upper Covariance Body).
Given access to an α-approximate separation oracle for $\mathcal{U}(X)$, there is an algorithm that on any input $A \in M_n(\mathbb{R})$ runs in polynomial time and returns a $(1+o(1))\alpha$-approximation to $\mathrm{Q}_X^{\max}(A)$.

The upshot of the above result is that it refocuses our attention to the task of designing an approximate separation oracle for the upper covariance body. Using this approach, we are already able to conclude new results for quadratic maximization by applying tools from classical analysis to design an approximate separation oracle for $\mathcal{U}(X)$. In some cases, however, it is quite difficult to design such an oracle directly for $\mathcal{U}(X)$. Inspired by deep tools from functional analysis, specifically the factorization theory of linear operators (see the monograph [55]), we will prove that under the assumption of having a bounded type-2 constant it suffices to design a separation oracle for the *lower covariance region* of X which we define in (15) below.

To give a couple of examples, it is easy to design a lower covariance separation oracle for the Minkowski sum $\ell_4^n + \ell_5^n$ (see Section 4) or for the quotient ℓ_4^n/ℓ_5^m, while on the other hand it is unclear how to directly describe an upper covariance separation oracle in these cases (see the full version [13] for more details). Another advantage which will become apparent soon is that lower covariance separation oracles allow for provably better approximation factors than the upper covariance separation oracles in the special cases of PSD quadratic maximization and bilinear maximization (the difference can be as big as $\log n$, as can be seen in the familiar example of $X = \ell_\infty^n$). Below we give a proof sketch for main "framework" theorem, namely an approximation algorithm for quadratic maximization (resp. bilinear maximization) when type-2 (resp. dual cotype-2) is bounded assuming access to only a separation oracle for the lower covariance region.

Remark 3. *In the interest of simplicity, the proof sketch below assumes we only desire to approximate the optimal value (and not produce solution vectors). For this simpler goal it suffices to use certain factorization theorems (see the full version [13] for a detailed introduction to factorization and the relevant theorems we use) as a black box. For the full proof in [13], we give rounding algorithms as well. For technical reasons, it was necessary to "open the factorization black box" and make some parts of the argument constructive, in addition to dualizing the entire argument. We therefore caution the reader that the full proof in [13] is syntactically different from the ensuing overview.*

3.5 Lower Covariance Region

We define the lower covariance region as follows:

$$\begin{aligned}
&\mathcal{L}(X) \stackrel{\text{def}}{=} \\
&\bigcup_{m=1}^{\infty}\Big\{\sum_{i=1}^{m} w_i w_i^* : w_1,\ldots,w_m \in \mathbb{R}^n \text{ and } \mathbb{E}\Big[\Big\|\sum_{i=1}^{m} \mathsf{g}_i w_i\Big\|_{X^*}^2\Big] \geqslant 1\Big\} \\
&= \Big\{\mathbb{W} \in \mathbb{PSD}^n : \mathbb{E}\Big[\big\|\mathbb{W}^{\frac{1}{2}}\mathsf{g}\big\|_{X^*}^2\Big] \geqslant 1\Big\},
\end{aligned} \tag{15}$$

where the second inequality in (15) is justified the same way as (11). Note that because $\mathcal{L}(X)$ is equal to

$$\mathbb{PSD}^n \smallsetminus \{\mathbb{W} \in \mathbb{PSD}^n : \mathbb{E}[\|\mathbb{W}^{\frac{1}{2}}\mathsf{g}\|_{X^*}^2] < 1\},$$

the lower covariance region of X is the complement in $\mathbb{PSD}^n$ of the interior of the upper covariance body of X^*. As such, it is a complement of a set that is star shaped with respect to the origin, and therefore $s \cdot \mathcal{L}(X) \supseteq \mathcal{L}(X)$ for every $0 < s \leqslant 1$.

3.6 Approximate Convexity of Lower Covariance Region

By reasoning analogously to the proof of (13), we see that

$$\mathcal{L}(X) \subseteq \mathbf{conv}\big(\mathcal{L}(X)\big) \subseteq \frac{1}{C_2(X^*)}\mathcal{L}(X). \tag{16}$$

Thus, the lower covariance region of X is $C_2(X^*)^2$-approximately convex in the following sense, which is the natural adaptation of Definition 3.1 to regions that are complements of star shape sets. Recall that by (9) if X has bounded type 2 constant, then X^* has bounded cotype 2 constant.

Definition 3.7. Let $T \subseteq \mathbb{R}^n$ satisfy $[1,\infty)T \subseteq T$ (equivalently, $\mathbb{R}^n \smallsetminus T$ is star shaped with respect to the origin). Given $\alpha \geqslant 1$, we say that T is α-approximately convex if $\mathbf{conv}(T) \subseteq \frac{1}{\alpha}T$.

With this definition at hand, the natural adaptation of Definition 3.4 is as follows.

Definition 3.8. Fix $\alpha \geqslant 1$. Suppose that $T \subseteq \mathbb{R}^n$ satisfies $[1,\infty)T \subseteq T$ and that T is α-approximately convex. An α-approximate separation oracle for T is a function O defined on $\mathbb{R}^n$ that outputs to each input $x \in \mathbb{R}^n$ either "Inside" or an affine hyperplane of $\mathbb{R}^n$. The requirements for O are as follows.

- If the output $O(x)$ is "Inside," then necessarily $x \in \frac{1}{\alpha}T$.
- If the output $O(x)$ is a hyperplane $H \subseteq \mathbb{R}^n$, then H must separate x from T.

If $x \in (\frac{1}{\alpha}T) \smallsetminus \mathbf{conv}(T)$, then O is allowed to either output a hyperplane or output "Inside".

3.7 Approximation Algorithms from Lower Covariance Separation Oracle

With these notions at hand, if the lower covariance region of X has an α-approximate separation oracle for some $\alpha \geqslant C_2(X^*)$, then by analysing a natural approximate version of the ellipsoid method we obtain an (oracle-time) efficient algorithm for approximating certain convex programs up to factor $(1+o(1))\alpha$, in the spirit of Theorem 3.5. For the sake of simplicity, rather than explaining this methodology in the introduction in its full generality, we state the following two consequences of it and refer to the full version [13] for a complete treatment.

Theorem 3.9 (Quadratic/bilinear maximization given separation oracle for lower covariance region). *Suppose that $X = (\mathbb{R}^n, \|\cdot\|_X)$ is a normed space such that $\mathcal{L}(X)$ has an α-approximate separation oracle for some $\alpha \geqslant C_2(X^*)$. Then, there is an algorithm that given an input matrix $A \in M_n(\mathbb{R})$ makes polynomially many oracle calls and runs in time $n^{O(1)}$, and outputs a matrix $\mathbb{W} \in \mathbb{PSD}^n$ with $\mathbb{W} \succcurlyeq A$ that satisfies*

$$\inf\left\{\mathrm{Q}_X^{\max}(\mathbb{M}) : \mathbb{M} \in \mathbb{PSD}^n \text{ and } \mathbb{M} \succcurlyeq A\right\} \gtrsim \frac{\mathrm{Q}_X^{\max}(\mathbb{W})}{\alpha}. \tag{17}$$

Also, if $X = (\mathbb{R}^n, \|\cdot\|_X), Y = (\mathbb{R}^m, \|\cdot\|_Y)$ are normed spaces such that $\mathcal{L}(X), \mathcal{L}(Y)$ have α-approximate separation oracles for $\alpha \geqslant \max\{C_2(X^*), C_2(Y^*)\}$, then there is an algorithm that given an input matrix $B \in M_{n\times m}(\mathbb{R})$ makes polynomially many oracle calls and runs in time that is polynomial in n, m, and outputs a matrices $\mathbb{W} \in \mathbb{PSD}^n, \mathbb{V} \in \mathbb{PSD}^m$ with $\left(\begin{smallmatrix}\mathbb{W} & 0\\ 0 & \mathbb{V}\end{smallmatrix}\right) \succcurlyeq \left(\begin{smallmatrix}0 & B\\ B^* & 0\end{smallmatrix}\right)$ and

$$\begin{aligned}
\inf\Bigg\{&\mathrm{Q}_X^{\max}(\mathbb{M}_1) + \mathrm{Q}_Y^{\max}(\mathbb{M}_2) : (\mathbb{M}_1, \mathbb{M}_2) \in \mathbb{PSD}^n \times \mathbb{PSD}^m \\
&\text{and } \begin{pmatrix}\mathbb{M}_1 & 0\\ 0 & \mathbb{M}_2\end{pmatrix} \succcurlyeq \begin{pmatrix}0 & B\\ B^* & 0\end{pmatrix}\Bigg\} \gtrsim \frac{\mathrm{Q}_X^{\max}(\mathbb{W}) + \mathrm{Q}_Y^{\max}(\mathbb{V})}{\alpha}.
\end{aligned} \tag{18}$$

We will next explain the ingredients that go into (17); the justification of (18) is similar and will be carried out in the full version [13]. The reason why we include (18) here is that it is important for the bilinear variant (2), namely for the question of approximating the operator norm $\|A\|_{Y\to X^*}$.

The goal of (17) is to $O(\alpha)$-approximately minimize the convex function $\mathbb{M} \mapsto \mathrm{Q}_X^{\max}(\mathbb{M})$ over the convex set $\{\mathbb{M} \in \mathbb{PSD}^n : \mathbb{M} \succcurlyeq A\}$. In the full version [13], we will show that in order to efficiently find a $(1+o(1))\alpha$-approximate minimizer, it suffices to show that each of the corresponding sub-level sets $\{\{\mathbb{M} \in \mathbb{PSD}^n : \mathrm{Q}_X^{\max}(\mathbb{M}) \leqslant t\} : t \in \mathbb{R}\}$ has a $(1+o(1))\alpha$-approximate separation oracle. By homogeneity, we therefore need to show that under the assumptions of Theorem 3.9, the convex set $\{\mathbb{M} \in \mathbb{PSD}^n : \mathrm{Q}_X^{\max}(\mathbb{M}) \leqslant 1\}$ has a $(1+o(1))\alpha$-approximate separation oracle.

To this end, fix $\mathbb{M} \in \mathbb{PSD}^n$ and consider the following optimization problem.

$$\max\left\{\mathbb{E}\Big[\|\mathbb{M}^{\frac{1}{2}}\mathbb{V}^{\frac{1}{2}}\mathsf{g}\|_{X^*}^2\Big] : \mathbb{V} \in \mathbb{PSD}^n \text{ and } \mathbf{Tr}(\mathbb{V}) \leqslant 1\right\}. \tag{19}$$

We claim that one can find in polynomial time and with polynomially many oracle calls a matrix $\mathbb{V} \in \mathbb{PSD}^n$ the attains this maximum up to a factor of $(1+o(1))\alpha$. Indeed, in the full version [13] we will show that for this it suffices to check that each of the corresponding super-level sets

$$\left\{\{\mathbf{V} \in \mathbb{PSD}^n : \mathbb{E}\big[\|\mathbb{M}^{\frac{1}{2}}\mathbb{V}^{\frac{1}{2}}\mathsf{g}\|_{X^*}^2\big] \geqslant t\} : t \in \mathbb{R}\right\} \tag{20}$$

has an α-approximate separation oracle. Since each of the sets appearing in (20) is (by definition) a linear transformation of the lower covariance body of X, the assumption of Theorem 3.9 ensures that the desired oracle exists. Therefore, we can find $\mathbb{V} \in \mathbb{PSD}^n$ with $\mathbf{Tr}(\mathbb{V}) \leqslant 1$ at which the maximum in (19) is attained up to a factor of $(1+o(1))\alpha$.

Finally, we can describe what the desired oracle for $\{\mathbb{M}' \in \mathbb{PSD}^n : \mathrm{Q}_X^{\max}(\mathbb{M}') \leqslant 1\}$ will output for the input matrix $\mathbb{M}$. For each realization of the Gaussian vector $\mathrm{g} \in \mathbb{R}^n$, let $x_{\mathrm{g}} \in \mathrm{Ball}(X)$ be the random vector that is given by

$$x_{\mathrm{g}} \stackrel{\mathrm{def}}{=} \underset{x' \in \mathrm{Ball}(X)}{\operatorname{argmax}} \left\langle x', \mathbb{M}^{\frac{1}{2}} \mathbb{V}^{\frac{1}{2}} \mathrm{g} \right\rangle.$$

Note that x_{g} can be found efficiently using polynomially many membership queries to $\mathrm{Ball}(X)$, using the classical theory of convex programming [31]. If

$$\frac{\|\mathbb{M}^{\frac{1}{2}} \mathbb{V}^{\frac{1}{2}} \mathrm{g}\|_{X^*}}{\|\mathbb{V}^{\frac{1}{2}} \mathrm{g}\|_{\ell_2^n}} \leqslant 1,$$

then the oracle outputs "Inside." Otherwise, the oracle outputs the hyperplane

$$\{\mathbb{M}' \in M_n(\mathbb{R}) : \langle \mathbb{M}' x_{\mathrm{g}}, x_g \rangle = 1\}.$$

By tracking the above definitions, one checks that this oracle satisfies the desired properties with positive probability. One gets this to hold with sufficiently high probability (to account for the polynomially many oracle calls) by repeating the above procedure with $n^{O(1)}$ independent samples from g rather than only one such sample; the details appear in the full version [13].

With the algorithmic groundwork of Theorem 3.9 complete, our final algorithm relies on the analytic inequalities that are contained in the following theorem (see the full version [13] for proofs).

Theorem 3.10 (Factorization Inequalities). *For every normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $A \in M_n(\mathbb{R})$ we have*

$$\begin{aligned} \mathrm{Q}_X^{\max}(A) &\leqslant \inf\left\{\mathrm{Q}_X^{\max}(\mathbb{W}) : \mathbb{W} \in \mathbb{PSD}^n \text{ and } \mathbb{W} \succcurlyeq A\right\} \\ &\leqslant T_2(X)^2 \cdot \mathrm{Q}_X^{\max}(A). \end{aligned} \tag{21}$$

Also, for every two normed spaces $X = (\mathbb{R}^n, \|\cdot\|_X)$, $Y = (\mathbb{R}^m, \|\cdot\|_Y)$, and every $B \in M_{n\times m}(\mathbb{R})$, denote

$$\begin{aligned} \gamma_2^{Y\to X^*}(B) &\stackrel{\mathrm{def}}{=} \\ &\inf\left\{\frac{\mathrm{Q}_X^{\max}(\mathbb{W}) + \mathrm{Q}_Y^{\max}(\mathbb{V})}{2} : (\mathbb{W}, \mathbb{V}) \in \mathbb{PSD}^n \times \mathbb{PSD}^m \right. \\ &\left. \text{and } \begin{pmatrix} \mathbb{W} & 0 \\ 0 & \mathbb{V} \end{pmatrix} \succcurlyeq \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}\right\}. \end{aligned} \tag{22}$$

Then,

$$\begin{aligned} \|B\|_{Y\to X^*} &\leqslant \gamma_2^{Y\to X^*}(B) \\ &\lesssim C_2(X^*) C_2(Y^*) \log\left(C_2(X^*) C_2(Y^*)\right) \cdot \|B\|_{Y\to X^*}. \end{aligned} \tag{23}$$

We chose the notation $\gamma_2^{Y\to X^*}(B)$ in (22) purposefully to coincide with the classical functional analytic notation for factorization norms [55], namely it is the γ_2 norm of B when it is viewed as an operator from Y to X^*. The equality (22) is therefore a variational characterization of the classical quantity in the left hand side in terms of the infimum on the right hand side; we prove this identity in the full version [13]. With this identity at hand, the inequality (23) is an application of a deep factorization theorem of Pisier [54]. The inequality (21) is inspired by the aforementioned factorization theory, but it seems to be new; it could be viewed as a factorization theorem for quadratic forms (see the full version [13] for details) and it would be interesting to study its ramifications within functional analysis.

By combining Theorem 3.9 with Theorem 3.10, we get the following algorithmic result.

Theorem 3.11 (Generic Framework). *Suppose that $X = (\mathbb{R}^n, \|\cdot\|_X)$ is a normed space such that $\mathcal{L}(X)$ has an α-approximate separation oracle for some $\alpha \geqslant C_2(X^*)$. Then, there is an algorithm that given an input matrix $A \in M_n(\mathbb{R})$ makes polynomially many oracle calls and runs in time $n^{O(1)}$, and outputs a number Alg_1 that is guaranteed to satisfy*

$$\mathrm{Q}_X^{\max}(A) \leqslant \mathrm{Alg}_1 \lesssim \alpha\, T_2(X)^2 \cdot \mathrm{Q}_X^{\max}(A).$$

For the bilinear case, if $X = (\mathbb{R}^n, \|\cdot\|_X)$, $Y = (\mathbb{R}^m, \|\cdot\|_Y)$ are normed spaces such that $\mathcal{L}(X)$, $\mathcal{L}(Y)$ have α-approximate separation oracles for some $\alpha \geqslant \max\{C_2(X^*), C_2(Y^*)\}$, then there is an algorithm that given an input matrix $B \in M_{n\times m}(\mathbb{R})$ makes polynomially many oracle calls and runs in time that is polynomial in n, m, and outputs a number Alg_2 that is guaranteed to satisfy

$$\begin{aligned} &\|B\|_{Y\to X^*} \leqslant \mathrm{Alg}_2 \\ &\lesssim \alpha\, C_2(X^*) C_2(Y^*) \log\left(C_2(X^*) C_2(Y^*)\right) \cdot \|B\|_{Y\to X^*}. \end{aligned}$$

Remark 4. *One often wishes to approximate efficiently not only the values of the quantities $\mathrm{Q}_X^{\max}(A)$ and $\|B\|_{Y\to X^*}$, but also to find efficiently the vector $x \in R^n$ at which $\mathrm{Q}_X^{\max}(A)$ is approximately attained, and correspondingly the vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ at which $\|B\|_{Y\to X^*}$ is approximately attained. For the latter, we need a constructive version of Pisier's factorization theorem that entails several adjustments of its classical proof; the details appear in the full version [13]. For this variant (namely, finding almost maximizing vectors rather than only estimating the quantity $\|B\|_{Y\to X^*}$), we get the slightly worse approximation factor*

$$O(\alpha C_2(X^*) C_2(Y^*) \log(\alpha C_2(X^*) C_2(Y^*)))$$

in the second part of Theorem 3.11.

4 EXAMPLES OF APPLICATIONS

Theorem 3.11 focuses our attention to designing approximate separation oracles for lower covariance bodies. In the specific cases that we examined, it turns out that this task is tractable because it reduces to probabilistic (Khinchine-type) inequalities that are available in the literature. We will examine such applications next. The advantage of the above approach is that it shifts our focus to a new algorithmic task. This task most likely cannot always be achieved due to the aforementioned hardness results, but in specific cases it becomes a concrete new question that lends itself to classical tools that may have not seemed relevant in the initial formulation of the problem. This reframing also allows us to prove various closure properties for the class of convex bodies for which efficient quadratic or bilinear maximization is possible.

4.1 Closure Properties

Given normed spaces $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $Y = (\mathbb{R}^n, \|\cdot\|_Y)$, one can obtain various other normed spaces. The most basic examples are passing to a subspace or a quotient of X. One can also consider the normed spaces $X + Y = (\mathbb{R}^n, \|\cdot\|_{X+Y})$ and $X \cap Y = (\mathbb{R}^n, \|\cdot\|_{X\cap Y})$ whose unit balls are $\mathrm{Ball}(X) + \mathrm{Ball}(Y) = \{x + y : (x, y) \in \mathrm{Ball}(X) \times$

$\mathrm{Ball}(Y)\}$ and $\mathrm{Ball}(X) \cap \mathrm{Ball}(Y)$, respectively; we call the former the Minkowski sum of X and Y and we call the latter the intersection of X and Y. A further operation of great importance is the 1-parameter family of complex[2] interpolation spaces $\{[X, Y]_\theta\}_{\theta \in [0,1]}$ whose definition is recalled in the full version [13] (see the monograph [11] for a thorough account). There are of course more such operations (a notable example is duality), but the above list of constructions is singled out because it always results in a normed space whose type 2 constant does not exceed $O(\max\{T_2(X), T_2(Y)\})$, which is crucial for us due to Theorem 3.3.

In the full version [13], we use the above framework to show that the class of normed spaces X with $T_2(X) = O(1)$ for which there exists a polynomial time $O(1)$-approximation algorithm for $\mathrm{Q}_X^{\max}(A)$ is preserved under subspaces, quotients, Minkowski sums, intersection and complex interpolation. Among these operations, passing to subspaces is quite straightforward, but the rest rely on the methodology that is developed here. Beyond the intrinsic interest of such closure properties, we remark that if one starts with the many examples of spaces that belong to the aforementioned class (see below), then these operations produce a rich variety of new examples that were beyond the reach of previous methods. Also, observe that these closure properties do not assume any information whatsoever on the initial algorithms: These algorithms are used as a "black box" to design an approximate separation oracle for the lower covariance body of the resulting normed space, after which one applies the first part of Theorem 3.11. An analogous treatment of the bilinear case is carried out using the second part of Theorem 3.11, where closure under quotients and Minkowski sums is derived under the assumption that cotype 2 constants of the duals of the initial spaces are $O(1)$; we do not treat the rest of the above-listed operations because they do not necessarily preserve this bounded cotype 2 assumption on the dual.

4.2 Symmetric Norms

A norm $\|\cdot\|$ on $\mathbb{R}^n$ is said to be a symmetric norm if $\|x\| \asymp \|(\varepsilon_1 x_{\pi(1)}, \ldots, \varepsilon_n x_{\pi(n)})\|$ for any $x \in \mathbb{R}^n$, any permutation π of $\{1, \ldots, n\}$, and any choice of signs $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$.[3] This is a well studied class of norms occurring frequently in the computer science, learning and optimization literature. Several papers have attempted to characterize the symmetric norms that are appropriate for various algorithmic tasks; see e.g. [5, 6, 14, 42, 62, 63].

In the full version [13], we use Theorem 3.11 to give a constant-factor approximation algorithm for quadratic (respectively bilinear) maximization over unit balls of symmetric norms whose type-2 constant (respectively the cotype-2 constant of their dual) is $O(1)$. Combined with Theorem 3.3, we obtain a near characterization of those symmetric norms for which quadratic maximization admits a constant factor approximation algorithm.

The class of those symmetric norms that have a bounded (or slowly growing) type-2 constant contains many examples that are not covered by the available literature. Below we will list some explicit examples of symmetric norms appearing in the optimization literature for various algorithmic tasks and for which we can conclude either a new quadratic maximization approximation algorithm or a new inapproximability result.

(1) An Orlicz norm ℓ_ϕ^n is defined by setting for every $x \in \mathbb{R}^n$,

$$\|x\|_{\ell_\phi^n} \stackrel{\mathrm{def}}{=} \inf\left\{\lambda > 0 \ : \ \sum_{i=1}^n \phi\Big(\frac{|x_i|}{\lambda}\Big) \leqslant 1\right\},$$

where $\phi : [0, \infty) \to [0, \infty)$ is a convex function satisfying $\phi(0) = 0$ and $\phi(t) > 0$ for all $t > 0$. Thus, in the special case $\phi(t) = t^p$ for some $p \geqslant 1$ we have $\ell_\phi^n = \ell_p^n$. Among the many applications of Orlicz norms, we note that they are important for the study of tail behaviour of random variables and are studied in statistics/machine learning [22] as examples of M-estimators with (convex loss functions).

The class of Orlicz norms with bounded type-2 constant has a complete description [37] as the set of norms ℓ_ϕ^n where $\phi : [0, \infty) \to [0, \infty)$ satisfies the following two conditions.

(a) There are constants $K, \delta, c > 0$ such that for all $t > 0$, if $\phi(t) \leqslant \delta$, then $\phi(2t) \leqslant K\phi(t) + c$.

(b) There is $\psi : [0, \infty) \to [0, \infty)$ such that $t \mapsto \psi(\sqrt{t})$ is convex and ϕ is equivalent to ψ in the sense that there are constants $K_1, K_2, \delta_1, \delta_2, c_1, c_2 > 0$ such that $\psi(t) \leqslant \delta_1$ implies $\phi(K_1 t) \leqslant \psi(t) + c_1$ and $\phi(t) \leqslant \delta_2$ implies $\psi(K_2 t) \leqslant \phi(t) + c_2$ for all $t > 0$.

(2) Norms whose unit balls are of the form $\mathrm{Ball}(\ell_p^n) \cap (\alpha \mathrm{Ball}(\ell_q^n))$ have a $O(1)$ type-2 constant (i.e., independent of n, α) whenever $2 \leqslant p, q < \infty$. Quadratic maximization over such norms is considered in order to capture optimization problems with a sparsity restriction. For instance, the densest k-subgraph and k-sparse principal component analysis, which are extensively studied optimization problems, can be cast as quadratic maximization by taking the underlying norm to be $\mathrm{Ball}(\ell_\infty^n) \cap (k\mathrm{Ball}(\ell_1^n))$ and $\mathrm{Ball}(\ell_2^n) \cap (\sqrt{k}\mathrm{Ball}(\ell_1^n))$, respectively; note that these norms have polynomially large type-2 constant due to the ℓ_1 component, which is consistent with the widespread belief that densest k-subgraph and k-sparse principal component analysis are hard to approximate. The above examples with $2 \leqslant p, q < \infty$ can be viewed as smoothed out versions of these classical algorithmic questions which do admit a polynomial time constant factor approximation algorithm.

(3) Motivated by applications to kernel pattern matching, [51] gave an approximation algorithm for the following symmetric norm that has slowly growing type-2 constant.

$$\|(x_1, \ldots, x_n)\|_{p,\infty} \stackrel{\mathrm{def}}{=} \max_{i \in \{1,\ldots,n\}} i^{\frac{1}{p}} x_i^*,$$

where $p \geqslant 2$ and x_i^* denotes the entry of $(|x_1|, \ldots, |x_n|)$ with the i-th largest magnitude.

(4) Order statistics norms are defined as the inner product of a non-increasing vector a with the sorted vector x^*. This class is well studied in the clustering literature [18–20] and includes e.g. the top-k norm (sum of top k magnitudes of x).

[2] The real interpolation method (see [11]) furnishes another such 1-parameter family of intermediate norms, but in the present work we will investigate only the complex interpolation method and we expect that it would be mechanical to obtain the analogous results for real interpolation using the same ideas.

[3] One could replace the exact invariance under permutations and signs by the analogous approximate requirement $\|x\| \asymp \|(\varepsilon_1 x_{\pi(1)}, \ldots, \varepsilon_n x_{\pi(n)})\|$. We will no do so here, though our results work under that assumption as well.

The type-2 constant of such norms is bounded whenever a has bounded support.

4.3 Unitarily Invariant Matrix Norms

A norm $\|\cdot\| : M_n(\mathbb{C}) \to [0,\infty)$ on the space $M_n(\mathbb{C})$ of $n\times n$ matrices with complex entries is said to be unitarily invariant if $\|UAV\| = \|A\|$ for any matrix $A \in M_n(\mathbb{C})$ and any two unitary matrices $U, V \in UM_n(\mathbb{C})$; this can be defined analogously for matrices with real entries (using orthogonal matrices), as well as for rectangular matrices, and all of our results hold in these settings. Key examples include the Schatten–von Neumann trace class S_p for $p \in [1,\infty]$, which is defined by

$$\forall A \in M_n(\mathbb{C}), \quad \|A\|_{\mathsf{S}_p} \stackrel{\text{def}}{=} \Big(\mathbf{Tr}\big((AA^*)^{\frac{p}{2}}\big)\Big)^{\frac{1}{p}} = \Big(\mathbf{Tr}\big((A^*A)^{\frac{p}{2}}\big)\Big)^{\frac{1}{p}} = \Big(\sum_{j=1}^{n} \sigma_j(A)^p\Big)^{\frac{1}{p}},$$

where $\sigma_1(A) \geqslant \ldots \geqslant \sigma_n(A) \geqslant 0$ are the singular values of A. Thus, $\|A\|_{\mathsf{S}_\infty} = \|A\|_{\ell_2^n(\mathbb{C})\to\ell_2^n(\mathbb{C})}$ is the usual operator norm of A. Another example is the Ky-Fan k-norm $\|\cdot\|_{(k)}$ for each $k \in \{1,\ldots,n\}$, which is the sum of the top k singular values, i.e.,

$$\forall A \in M_n(\mathbb{C}), \qquad \|A\|_{(k)} \stackrel{\text{def}}{=} \sum_{j=n-k+1}^{n} \sigma_j(A).$$

More generally, if $E = (\mathbb{R}^n, \|\cdot\|_E)$ is a symmetric normed space, then the following norm is unitarily invariant and any unitarily invariant norm is obtained in this way (the fact that this defines a norm in not immediate; see e.g. [12] for a proof).

$$\forall A \in M_n(\mathbb{C}), \qquad \|A\|_{\mathsf{S}_E} \stackrel{\text{def}}{=} \|(\sigma_1(A),\ldots,\sigma_n(A))\|_E.$$

A (substantial) theorem of [28] asserts that $\|\cdot\|_E$ has $O(1)$ type 2 or cotype 2 constant if and only if $\|\cdot\|_E$ does.

In the full version [13] we use Theorem 3.11 to obtain a constant-factor approximation algorithm for quadratic (respectively bilinear) maximization over unitarily invariant norms with bounded type-2 constant (respectively whose dual has cotype-2 constant). In particular, this provides a different rounding algorithm for the non-commutative Grothendieck problem [50] (namely, bilinear maximization over the operator norm), albeit with a worse universal constant than in [50]. As another concrete example, this gives a constant factor approximation algorithm for bilinear maximization over Ky-Fan k-norms when $k = O(1)$. Combined with Theorem 3.3, we thus obtain a near characterization of unitarily invariant matrix norms over which quadratic maximization admits a constant factor approximation algorithm.

4.4 Robust Principle Component Analysis

In [50], efficient bilinear maximization over the operator norm (Schatten-∞) was used to give a constant factor approximation algorithm for the following subspace approximation problem, called R_1-PCA, which was introduced in [25]. Given a set of vectors $v_1,\ldots v_m \in \mathbb{R}^n$ find a k-dimensional subspace $S \subseteq \mathbb{R}^n$ maximizing the sum of the Euclidean lengths of the orthogonal projections $\Pi_S v_1,\ldots,\Pi_S v_m$ of $v_1,\ldots,v_m$ onto S. Thus, the goal of R_1-PCA is to find a k-dimensional subspace $S \subseteq \mathbb{R}^n$ for which the quantity $\sum_{i=1}^m \|\Pi_S v_i\|_{\ell_2^n}$ is (approximately) minimized.

Our framework implies that a more general class of robust PCA variants admits constant factor approximation algorithms. Given a normed space $X = (\mathbb{R}^m, \|\cdot\|_X)$, one can use it to aggregate the length of the projections, thus leading to the following subspace approximation problem.

$$\mathrm{OPT} \stackrel{\text{def}}{=} \max_{\dim(S)=k} \|(\|\Pi_S v_1\|_2,\ldots,\|\Pi_S v_m\|_2)\|_X .$$

Let T denote the linear operator taking an $m\times k$ matrix U with column vectors $u_1,\ldots,u_k \in \mathbb{R}^n$ as input and outputting the vector

$$(\langle u_1, v_1\rangle,\ldots\langle u_k, v_1\rangle) \oplus \cdots \oplus (\langle u_1, v_m\rangle,\ldots\langle u_k, v_m\rangle),$$

where $\oplus$ denotes vector-concatenation. Let $\|\cdot\|_{X(\ell_2^k)}$ be a norm defined over the set of sequences $(a_i)_{i=1}^m \in (\mathbb{R}^k)^m$ of k-dimensional vectors and given by

$$\|(a_i)_{i=1}^m\|_{X(\ell_2^k)} \stackrel{\text{def}}{=} \|(\|a_1\|_2,\ldots,\|a_m\|_2)\|_X .$$

Then, one can cast OPT as a bilinear maximization problem in the following way.

$$\mathrm{OPT} = \max_{U\in O_n} \|T(U)\|_{X(\ell_2^k)} = \max_{\|U\|_{\mathsf{S}_\infty}\leqslant 1} \|T(U)\|_{X(\ell_2^k)} = \|T\|_{\mathsf{S}_\infty\to X(\ell_2^k)},$$

where $O_n \subseteq M_n(\mathbb{R})$ is the set of orthogonal matrices. The second equality above follows since the extreme points of $\mathrm{Ball}(\mathsf{S}_\infty)$ are precisely O_n, and the maximum of a convex function over a convex set occurs at an extreme point.

Thanks to this bilinear maximization formulation, Theorem 3.11 may be combined with the lower covariance separation oracles constructed in the full version [13], to provide good approximation algorithms for a variety of norms $\|\cdot\|_X$, like constant approximations for sign-invariant norms with 2-concavity constant 1 or symmetric norms with bounded cotype-2 constant. We illustrate the versatility of our framework by providing a more intricate example; by combining Theorem 3.11 with the separation oracles constructed in the full version [13] and using algorithmic closure properties for complex interpolation (see the full version [13]), we obtain a $(\log n)^{O(1)}$-factor approximation algorithm for the following refinement of robust-PCA: Find a k-dimensional subspace $S \subseteq \mathbb{R}^n$ (approximately) maximizing

$$\|(\Pi_S v_i)_{i=1}^m\|_{[X_0,X_1]_\theta},$$

where $[\cdot,\cdot]_\theta$ denotes complex interpolation, $\alpha \geqslant 0$ is a parameter, and

$$\|(\Pi_S v_i)_{i=1}^m\|_{X_0} \stackrel{\text{def}}{=} \sum_{i=1}^m \|\Pi_S v_i\|_2 \qquad \text{and}$$

$$\|(\Pi_S v_i)_{i=1}^m\|_{X_1} \stackrel{\text{def}}{=} \alpha \cdot \sum_{i=1}^m \sum_{j=1}^m \|\Pi_S v_i - \Pi_S v_j\|_2 .$$

As defined above, X_1 is a semi-norm but can be made into a norm by adding a sufficiently small multiple of ℓ_2^n which would cause negligible change to the objective value. By tuning the parameters $\alpha \geqslant 0$ and $\theta \in [0,1]$, the above optimization problem intuitively asks for a subspace maximizing its correlation with the given vectors $\{v_i\}_{i=1}^m$, while also requiring that the orthogonal projections onto S of these vectors are not clustered together much on average.

4.5 Brief Summary of the Literature and Problems Captured by Quardatic Maximization

Here we will mention some of what is known about the quadratic and bilinear optimization problem over convex bodies. Quadratic/bilinear maximization over $\mathrm{Ball}(\ell_2^n)$ correspond to the familiar linear-algebraic quantities maximum eigenvalue/maximum singular value. The (non-origin-symmetric) case of (1) when K is a simplex has been investigated in [24, 34], partly in connection to problems in computational biology. The case when K is a polytope with polynomially many facets is classical. It is among the most important non-linear optimization problems, with a wide range of applications in operations research, computational biology and economics. See [10, 16, 27] for more information on the computational complexity of such problems.

Perhaps the first nontrivial and most influential case of bilinear maximization is Grothendieck's classical inequality [30] and its more common formulation in [43], which corresponds to the case $K = \mathrm{Ball}(\ell_\infty^n)$. This leads to a constant factor polynomial time algorithm, as shown in [4] (see [15] for the best known approximation factor), with a variety of applications to combinatorial optimization. The quadratic maximization problem over $\mathrm{Ball}(\ell_\infty)$ was studied in [21] with application to correlation clustering, and the matching integrality-gap lower bound in this case was obtained in [3]. Hardness results in these settings (under various complexity assumptions) were obtained in [4, 7, 39, 58]. The survey [38] is devoted to the use of Grothendieck-type inequalities in combinatorial optimization.

Krivine [41] (see also [56]) observed that Grothendieck's inequality generalizes (with the same constant) to the class of norms of the form

$$|||(x_1, \ldots, x_n)||| \stackrel{\text{def}}{=} \|(x_1^2, \ldots, x_n^2)\|_Y^{\frac{1}{2}},$$

where $\|\cdot\|_Y$ is a norm on $\mathbb{R}^n$ that satisfies the symmetry condition (3). Such norms are clearly invariant to flipping signs of the entries and are precisely those norms having a 2-convexity constant of 1 (see the full version [13] for definitions). Hereafter, we shall refer to them as exactly 2-convex norms. Note in particular that the above class includes the norm ℓ_p^n whenever $p \geqslant 2$. Underlying Krivine's observation is a constant factor bound on the integrality gap of the bilinear analogue of the convex programming relaxation (4) over exactly 2-convex norms; in [52], a different proof of this was obtained. The problem of quadratic maximization over exactly 2-convex norms was investigated in [51], where a constant factor approximation algorithm was obtained under the additional (necessary) assumption of bounded q-concavity for some finite q (see the full version [13] for the definition); this was used in [51] to obtain a $(\log\log n)^{O(1)}$-approximation algorithm for a special case of the quadratic assignment problem. It can also be shown that the $(\log n)$-approximation algorithm for vertex expansion of a graph due to [44] is a consequence of the algorithm of [51].

Implicit in the non-commutative Khintchine inequality [45] is a constant factor convex programming algorithm for Quadratic Maximization over Schatten-p when $2 \leqslant p < \infty$ (and a $\log n$-approximation when $p = \infty$). In the bilinear Schatten-∞ case, Grothendieck [30] conjectured a noncommutative version of his inequality which was proven in [53] (the sharp constant was obtained in [33]). In [50], algorithmic proofs of the non-commutative Grothendieck inequality were derived, thereby obtaining efficient constant factor rounding algorithms for bilinear maximization over Schatten-∞. This was used in [50] to give approximation algorithms for robust principal component analysis and a generalization of the orthogonal Procrustes problem. In [60], it was shown how this can be used to bound the power of entanglement in quantum XOR games. A corresponding (sharp) hardness result was obtained in [17] (see also [35] for a different proof).

4.6 Other Problems in the Literature Captured by Quadratic Maximization

The bilinear ℓ_p case captures the problem of certifying hypercontractivity which in turn has connections to small set expansion and quantum separability ([8]). Vertex expansion and a related analytic proxy ([44]) can be cast as quadratic maximization, and so can densest-k-subgraph, sparse-PCA, the spread constant of a metric [2], and the Poincaré constant (in discrete domains). Approximability/inapproximability aspects of these expansion-type problems have been the subject of a large body of work. Expansion-type problems are of interest in part due to their connection to the unique games conjecture, and also due to their relevance to hardness results for optimization over pseudo-random instances.

For appropriate choices of linear maps and convex sets, quadratic maximization also captures (upto constants) the maximization of the absolute value of homogeneous polynomials of any constant degree. Homogeneous polynomial maximization is a very expressive class of problems in its own right, and has connections to quantum information theory [8], refuting random constraint satisfaction problems [57], statistical physics, tensor principal component analysis and tensor decomposition [9, 29, 36, 49], game theory, control theory and population dynamics [23].

Quadratic maximization also captures problems of interest in compressed sensing and coding theory, like subspace distortion, or the sparsest vector in a subspace.

ACKNOWLEDGMENTS

The research that is presented here was conducted under the auspices of the Simons Algorithms and Geometry (A& G) Think Tank. The first author was supported indirectly by the National Science Foundation Grant No. CCF-1900460. The third author was supported in part by the BSF, the Packard Foundation and the Simons Foundation.

The first and second authors would like to thank Venkatesan Guruswami and Madhur Tulsiani for useful discussions on these topics over the last few years.

The authors would like to thank the anonymous STOC'21 referees for useful comments on the writing of the manuscript.

REFERENCES

[1] Noga Alon. 2006. Maximizing a quadratic form on the ℓ_1^n unit ball. (2006). Unpublished manuscript.

[2] Noga Alon, Ravi Boppana, and Joel Spencer. 1998. An asymptotic isoperimetric inequality. *Geometric & Functional Analysis GAFA* 8, 3 (1998), 411–436.
[3] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. 2006. Quadratic forms on graphs. *Inventiones mathematicae* 163, 3 (2006), 499–522. Conference version in STOC '05.
[4] Noga Alon and Assaf Naor. 2006. Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.* 35, 4 (2006), 787–803. https://doi.org/10.1137/S0097539704441629
[5] Alexandr Andoni, Chengyu Lin, Ying Sheng, Peilin Zhong, and Ruiqi Zhong. 2018. Subspace embedding and linear regression with Orlicz norm. *arXiv preprint arXiv:1806.06430* (2018).
[6] Alexandr Andoni, Huy L Nguyen, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. 2017. Approximate near neighbors for general symmetric norms. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 902–913.
[7] Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, and Muli Safra. 2005. On non-approximability for quadratic programs. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*. IEEE, 206–215.
[8] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. 2012. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. ACM, 307–326.
[9] Boaz Barak, Jonathan A Kelner, and David Steurer. 2015. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*. ACM, 143–151.
[10] Mihir Bellare and Phillip Rogaway. 1995. The complexity of approximating a nonlinear program. *Math. Programming* 69, 3, Ser. A (1995), 429–441.
[11] Jöran Bergh and Jörgen Löfström. 1976. *Interpolation spaces. An introduction.* Springer-Verlag, Berlin-New York. x+207 pages. Grundlehren der Mathematischen Wissenschaften, No. 223.
[12] Rajendra Bhatia. 1997. *Matrix analysis.* Graduate Texts in Mathematics, Vol. 169. Springer-Verlag, New York. xii+347 pages. https://doi.org/10.1007/978-1-4612-0653-8
[13] Vijay Bhattiprolu, Euiwoong Lee, and Assaf Naor. 2021. A framework for quadratic form maximization over convex sets through nonconvex relaxations. *Soon to appear on arxiv* (2021).
[14] Jarosław Błasiok, Vladimir Braverman, Stephen R Chestnut, Robert Krauthgamer, and Lin F Yang. 2017. Streaming symmetric norms via measure concentration. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 716–729.
[15] Mark Braverman, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. 2013. The Grothendieck constant is strictly smaller than Krivine's bound. In *Forum of Mathematics, Pi*, Vol. 1. Cambridge University Press. Conference version in FOCS '11.
[16] Andreas Brieden. 2002. Geometric optimization problems likely not contained in $\mathbb{APX}$. *Discrete Comput. Geom.* 28, 2 (2002), 201–209.
[17] Jop Briët, Oded Regev, and Rishi Saket. 2015. Tight hardness of the non-commutative Grothendieck problem. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE, 1108–1122.
[18] Jarosław Byrka, Krzysztof Sornat, and Joachim Spoerhase. 2018. Constant-factor approximation for ordered k-median. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 620–631.
[19] Deeparnab Chakrabarty and Chaitanya Swamy. 2019. Approximation algorithms for minimum norm and ordered optimization problems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 126–137.
[20] Deeparnab Chakrabarty and Chaitanya Swamy. 2019. Simpler and Better Algorithms for Minimum-Norm Load Balancing. *arXiv preprint arXiv:1905.00044* (2019).
[21] Moses Charikar and Anthony Wirth. 2004. Maximizing Quadratic Programs: Extending Grothendieck's Inequality. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*. IEEE Computer Society, 54–60. https://doi.org/10.1109/FOCS.2004.39
[22] Kenneth L Clarkson and David P Woodruff. 2014. Sketching for M-estimators: A unified approach to robust regression. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 921–939.
[23] Etienne De Klerk. 2008. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research* 16, 2 (2008), 111–125.
[24] Etienne de Klerk, Monique Laurent, and Pablo A. Parrilo. 2006. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoret. Comput. Sci.* 361, 2-3 (2006), 210–225.
[25] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha. 2006. R_1-PCA: Rotational invariant L_1-norm principal component analysis for robust subspace factorization. In *Proceedings of the 23rd International Conference on Machine Learning (ICML'06)*. 281—-288.
[26] Ed Dubinsky, A. Pełczyński, and H. P. Rosenthal. 1972. On Banach spaces X for which $\Pi_2(L_\infty, X) = B(L_\infty, X)$. *Studia Math.* 44 (1972), 617–648. https://doi.org/10.4064/sm-44-6-617-648
[27] U. Feige and L. Lovász. 1992. Two-prover one-round proof systems, their power and their problems. In *STOC'92: Proceedings of the 24th Annual ACM Symposium on Theory of Computing*. ACM, New York, 733–744.
[28] D. J. H. Garling and N. Tomczak-Jaegermann. 1983. The cotype and uniform convexity of unitary ideals. *Israel J. Math.* 45, 2-3 (1983), 175–197. https://doi.org/10.1007/BF02774015
[29] Rong Ge and Tengyu Ma. 2015. Decomposing Overcomplete 3rd Order Tensors using Sum-of-Squares Algorithms. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (2015), 829.
[30] Alexander Grothendieck. 1953. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo* ***8*** (1953).
[31] M Grotschel, L Lovász, and A Schrijver. 1993. Geometric Algorithms and Combinatorial Optimization.
[32] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. 2016. Bypassing UGC from Some Optimal Geometric Inapproximability Results. *ACM Transactions on Algorithms (TALG)* 12, 1 (2016), 6. Conference version in SODA '12.
[33] Uffe Haagerup. 1985. The Grothendieck inequality for bilinear forms on C*-algebras. *Advances in Mathematics* 56, 2 (1985), 93–116.
[34] Eran Halperin and Elad Hazan. 2006. HAPLOFREQ—estimating haplotype frequencies efficiently. *J. Comput. Biol.* 13, 2 (2006), 481–500 (electronic).
[35] Steven Heilman and Thomas Vidick. 2016. A Moment Majorization principle for random matrix ensembles with applications to hardness of the noncommutative Grothendieck problem. (2016). Manuscript available at https://arxiv.org/abs/1603.05620.
[36] Samuel B Hopkins, Jonathan Shi, and David Steurer. 2015. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory*. 956–1006.
[37] Eleni Katirtzoglou. 1998. Type and cotype of Musielak–Orlicz sequence spaces. *Journal of mathematical analysis and applications* 226, 2 (1998), 431–455.
[38] Subhash Khot and Assaf Naor. 2012. Grothendieck-Type Inequalities in Combinatorial Optimization. *Communications on Pure and Applied Mathematics* 65, 7 (2012), 992–1035.
[39] Subhash Khot and Ryan O'Donnell. 2009. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory OF Computing* 5 (2009), 83–117.
[40] Guy Kindler, Assaf Naor, and Gideon Schechtman. 2008. The UGC hardness threshold of the l_p Grothendieck problem. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York, 64–73.
[41] JL Krivine. 1973. Théoremes de factorisation dans les espaces réticulés. *Séminaire Analyse fonctionnelle (dit" Maurey-Schwartz")* (1973), 1–22.
[42] Jerry Li, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. 2019. On mean estimation for general norms with statistical queries. *arXiv preprint arXiv:1902.02459* (2019).
[43] Joram Lindenstrauss and Aleksander Pełczyński. 1968. Absolutely summing operators in L_p-spaces and their applications. *Studia Mathematica* 29, 3 (1968), 275–326.
[44] Anand Louis, Prasad Raghavendra, and Santosh Vempala. 2013. The complexity of approximating vertex expansion. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 360–369.
[45] Françoise Lust-Piquard and Gilles Pisier. 1991. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.* 29, 2 (1991), 241–260. https://doi.org/10.1007/BF02384340
[46] Bernard Maurey. 2003. Type, cotype and K-convexity. In *Handbook of the geometry of Banach spaces, Vol. 2*. North-Holland, Amsterdam, 1299–1332. https://doi.org/10.1016/S1874-5849(03)80037-2
[47] Bernard Maurey and Gilles Pisier. 1976. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.* 58, 1 (1976), 45–90. https://doi.org/10.4064/sm-58-1-45-90
[48] Vitali D. Milman and Gideon Schechtman. 1986. *Asymptotic theory of finite-dimensional normed spaces*. Lecture Notes in Mathematics, Vol. 1200. Springer-Verlag, Berlin. viii+156 pages. With an appendix by M. Gromov.
[49] Andrea Montanari and Emile Richard. 2014. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*. 2897–2905.
[50] Assaf Naor, Oded Regev, and Thomas Vidick. 2013. Efficient rounding for the noncommutative Grothendieck inequality. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 71–80.
[51] Assaf Naor and Gideon Schechtman. 2009. An Approximation Scheme for Quadratic Form Maximization on Convex Bodies. (2009).
[52] Yurii Nesterov. 1998. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software* 9, 1-3 (1998), 141–160.
[53] Gilles Pisier. 1978. Grothendieck's theorem for noncommutative C*-algebras, with an appendix on Grothendieck's constants. *Journal of Functional Analysis* 29, 3 (1978), 397–415.
[54] Gilles Pisier. 1980. Un théorème sur les opérateurs linéaires entre espaces de Banach qui se factorisent par un espace de Hilbert. *Ann. Sci. École Norm. Sup. (4)* 13, 1 (1980), 23–43. http://www.numdam.org/item?id=ASENS_1980_4_13_1_23_0

[55] Gilles Pisier. 1986. *Factorization of linear operators and geometry of Banach spaces.* American Mathematical Soc.

[56] Gilles Pisier. 2012. Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc.* 49, 2 (2012), 237–323.

[57] Prasad Raghavendra, Satish Rao, and Tselil Schramm. 2016. Strongly Refuting Random CSPs Below the Spectral Threshold. *arXiv preprint arXiv:1605.00058* (2016).

[58] Prasad Raghavendra and David Steurer. 2009. Towards computing the Grothendieck constant. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms.* Society for Industrial and Applied Mathematics, 525–534.

[59] Prasad Raghavendra and David Steurer. 2010. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd annual ACM Symposium on Theory of Computing.* ACM, 755–764.

[60] Oded Regev and Thomas Vidick. 2015. Quantum XOR games. *ACM Transactions on Computation Theory (ToCT)* 7, 4 (2015), 15.

[61] Walter Rudin. 1973. *Functional analysis.* McGraw-Hill Book Co., New York-Düsseldorf-Johannesburg. xiii+397 pages. McGraw-Hill Series in Higher Mathematics.

[62] Zhao Song, Ruosong Wang, Lin Yang, Hongyang Zhang, and Peilin Zhong. 2019. Efficient symmetric norm regression via linear sketching. In *Advances in Neural Information Processing Systems.* 830–840.

[63] Zhao Song, David Woodruff, and Peilin Zhong. 2019. Towards a zero-one law for column subset selection. In *Advances in Neural Information Processing Systems.* 6123–6134.