A Framework for Quadratic Form Maximization over Convex Sets through Non-Convex Relaxations

Vijay Bhattiprolu
vspvijay@math.ias.edu
IAS/Princeton

Euiwoong Lee
euiwoong@umich.edu
University of Michigan

Assaf Naor
naor@math.princeton.edu
Princeton University

July 8, 2021

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.

Contents

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

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

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

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

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

$$\mathrm{Q}_K^{\max}(A) = \max_{x \in K} \|A^{\frac{1}{2}} x\|_{\ell_2^n}^2 = \max_{\substack{x \in K \\ y \in \mathrm{Ball}(\ell_2^n)}} \left\langle A^{\frac{1}{2}} x, y \right\rangle^2 = \left(\mathrm{Op}_{K,\mathrm{Ball}(\ell_2^n)}^{\max}(A^{\frac{1}{2}}) \right)^2,$$

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 1.4.

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 [NS09] 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 [NS09] 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 [KNS08] (see [GRSW16] for the corresponding hardness result under a weaker hypothesis than that of [KNS08]). The success of [NS09] was partial, as it pertains only to a certain subclass of convex bodies K that satisfies the following symmetry condition.

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

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 [Nes98, KNS08, NS09], 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, it will be convenient to set some notation and record some basic definitions.

1.1 Notation and Preliminaries

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 Ball(X). By binary search for the smallest $r \geqslant 0$ such that $x \in \mathbb{R}^n$ belongs to rBall(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 [Rud73] for this standard material.

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 [DPR72] 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

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

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

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

These invariants of normed spaces are of immense importance to various areas; see the survey [Mau03] 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 [MS86].

$$T_2(\ell_p^n) \asymp \begin{cases} n^{\frac{1}{p}-\frac{1}{2}} & \text{if } 1 \leqslant p \leqslant 2, \\ \sqrt{\min\{p, \log n\}} & \text{if } 2 \leqslant p \leqslant \infty, \end{cases} \qquad \text{and} \qquad C_2(\ell_p^n) \asymp \begin{cases} 1 & \text{if } 1 \leqslant p \leqslant 2, \\ n^{\frac{1}{2}-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty. \end{cases} \tag{8}$$

[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 $(P \lesssim Q) \wedge (Q \lesssim P)$.

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 [MP76] and the second inequality in (9) is from [Pis80].

1.2 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.

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

For every $w_1,\ldots,w_m \in \mathbb{R}^n$, the random vectors $\sum_{i=1}^m \mathsf{g}_i w_i$ and $\mathbb{W}^{\frac12}\mathsf{g}$, 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,

$$\mathcal{U}(X) = \Big\{\mathbb{W} \in \mathbb{PSD}^n :\ \mathbb{E}\Big[\big\|\mathbb{W}^{\frac12}\mathsf{g}\big\|_X^2\Big] \leqslant 1\Big\}. \tag{11}$$

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 $\mathcal{U}(X)$:

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 $\mathbb{E}\big[\|\mathbb{W}^{\frac12}\mathsf{g}\|_X^2\big] \leqslant 1$, then

$$\langle A\,,\mathbb{W}\rangle = \mathbf{Tr}(A\mathbb{W}) = \mathbf{Tr}\big(\mathbb{W}^{\frac12}A\mathbb{W}^{\frac12}\big) = \mathbb{E}\big[\langle \mathbb{W}^{\frac12}\mathsf{g}, A\mathbb{W}^{\frac12}\mathsf{g}\rangle\big] \leqslant \mathbb{E}\Big[\mathrm{Q}_X^{\max}(A)\big\|\mathbb{W}^{\frac12}\mathsf{g}\big\|_X^2\Big] \leqslant \mathrm{Q}_X^{\max}(A).$$

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 $\langle A\,, ww^*\rangle = \langle w\,, Aw\rangle$.

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 $\mathbb{W}_1,\ldots\mathbb{W}_k \in \mathcal{U}(X)$ and $s_1,\ldots,s_k \in [0,1]$ 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 $j\in\{1,\ldots,k\}$, the assumption $\mathbb{W}_j\in\mathcal{U}(X)$ means that for some $m(j)\in\mathbb{N}$ there are vectors $w_{1,j},\ldots,w_{m(j),j}\in\mathbb{R}^n$ 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,

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

Therefore $T_2(X)^{-2}\sum_{j=1}^k\sum_{i=1}^{m(j)}\big(\sqrt{s_j}w_{ij}\big)\big(\sqrt{s_j}w_{ij}\big)^*=T_2(X)^{-2}\sum_{j=1}^k s_j\mathbb{W}_j$ indeed belongs to $\mathcal{U}(X)$.

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

Definition 1.1. *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 (from Section 9) 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.

Theorem 1.2 (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 1.2 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 1.2, 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 Section 8.

Theorem 1.3 (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 $\mathrm{NP} \not\subseteq \mathrm{P}_{/\mathrm{poly}}$, 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 1.4. *The Small Set Expansion Hypothesis (SSEH) is a commonly used hardness assumption that was formulated in [RS10] and is recalled in Section 8. Of course, the SSEH is less standard than, say, $\mathrm{NP} \not\subseteq \mathrm{P}_{/\mathrm{poly}}$, so one should take Theorem 1.3 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 1.5. *Recalling (8), Theorem 1.3 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 [KNS08], where it was shown that this problem does have a $O_p(1)$ approximation algorithm when $2 \leqslant p < \infty$. In the unpublished manuscript [Alo06] 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 [KN12]. In Section 8 we show that if $X^n = \ell_p^n$ and $1 < p < 2$, then the hardness statement of Theorem 1.3 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 [ABH$^+$05] for hardness when $p = \infty$).*

Approximation Algorithms From Upper Covariance Separation Oracle. Recall that Theorem 1.2 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 1.6. *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 $\mathcal{O}$ 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 $\mathcal{O}$ are as follows.*

- *If the output $\mathcal{O}(x)$ is "Inside," then necessarily $x \in \alpha S$.*
- *If the output $\mathcal{O}(x)$ 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 Section 3).

Theorem 1.7 (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 the introduction, 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 1.7. Another simplifying assumption that we will make throughout this introduction is that the length of the bit description of all inputs (namely matrices) to algorithms is $n^{O(1)}$.

Using Theorem 1.7 and applying it to $\mathcal{U}(X)$, we readily deduce the following approximation algorithm for quadratic maximization (see Section 4.1.1)

Proposition 1.8 (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 [Pis86]), 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 1.3) or for the quotient norm 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 Section 6 for more details). Another advantage which will become apparent in 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 1.9. *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 Section 5 for a detailed introduction to factorization and the relevant theorems we use) as a black box. For the full proof in Section 4, 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 thus caution the reader that the full proof in Section 4 is syntactically different from the ensuing overview. In Section 5 we discuss how the results in Section 4 may be viewed as "dual transpositions" of algorithmic factorization theorems.*

Lower Covariance Region. We define the lower covariance region as follows:

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

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 $s\mathcal{L}(X) \supseteq \mathcal{L}(X)$ for every $0 < s \leqslant 1$.

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 1.1 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 1.10. *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 1.6 is as follows.

Definition 1.11. *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 $\mathcal{O}$ 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 $\mathcal{O}$ are as follows.*

- *If the output $\mathcal{O}(x)$ is "Inside," then necessarily $x \in \frac{1}{\alpha}T$.*
- *If the output $\mathcal{O}(x)$ 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 $\mathcal{O}$ is allowed to either output a hyperplane or output "Inside".

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 1.7. 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 Section 4 for a complete treatment.

Theorem 1.12 (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

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

We will next explain the ingredients that go into (17); the justification of (18) is similar and will be carried out separately in Section 4. 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 Section 4 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 1.12, 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.

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

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 Section 4 we will show that for this it suffices to check that each of the corresponding super-level sets

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

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 1.12 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

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

Note that x_{g} can be found efficiently using polynomially many membership queries to $\mathrm{Ball}(X)$, using the classical theory of convex programming [GLS93]. 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 Section 4.

With the algorithmic groundwork of Theorem 1.12 complete, our final algorithm relies on the analytic inequalities that are contained in the following theorem (see Section 5 for proofs).

Theorem 1.13 (Factorization Inequalities)**.**
For every normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $A \in M_n(\mathbb{R})$ we have

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

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

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

Then,

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

We chose the notation $\gamma_2^{Y\to X^*}(B)$ in (22) purposefully to coincide with the classical functional analytic notation for factorization norms [Pis86], 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 Section 5.2.1. With this identity at hand, the inequality (23) is an application of a deep factorization theorem of Pisier [Pis80]. 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 Section 5.3) and it would be interesting to study its ramifications within functional analysis.

By combining Theorem 1.12 with Theorem 1.13, we get the following algorithmic result.

Theorem 1.14 (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 that is polynomial in n, 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

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

Remark 1.15. *One often wishes not only to approximate efficiently 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 Section 4.2.2. 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 1.14.*

1.3 Examples of Applications

Theorem 1.14 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 (Khintchine-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.

1.3.1 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 Section 6.4 (see the monograph [BL76] 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 1.3.

[2]The real interpolation method (see [BL76]) 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.

In Section 6, 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 1.14. An analogous treatment of the bilinear case is carried out in Section 6 using the second part of Theorem 1.14, 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.

1.3.2 Symmetric Norms

A norm $\|\cdot\|$ on $\mathbb{R}^n$ is said to be a symmetric norm if $\|x\| \asymp \|(\varepsilon_1 x_{\pi(1)}, \dots, \varepsilon_n x_{\pi(n)})\|$ for any $x \in \mathbb{R}^n$, any permutation π of $\{1, \dots, n\}$, and any choice of signs $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$.[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. [LNRW19, ANN+17, BBC+17, ALS+18, SWZ19, SWY+19].

In Section 7.2, we use Theorem 1.14 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 of their dual) is $O(1)$. Combined with Theorem 1.3, 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 $x \in \mathbb{R}^n$,

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

   where $\varphi : [0,\infty) \to [0,\infty)$ is a convex function satisfying $\varphi(0) = 0$ and $\varphi(t) > 0$ for all $t > 0$. Thus, in the special case $\varphi(t) = t^p$ for some $p \geqslant 1$ we have $\ell_\varphi^n = \ell_p^n$. 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 [CW14] as examples of M-estimators with (convex loss functions).

   The class of Orlicz norms with bounded type-2 constant has a complete description [Kat98] as the set of norms ℓ_φ^n where $\varphi : [0,\infty) \to [0,\infty)$ satisfies the following two conditions.

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

(a) There are constants $K, \delta, c > 0$ such that for all $t > 0$, if $\varphi(t) \leqslant \delta$, then $\varphi(2t) \leqslant K\varphi(t) + c$.
(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 $\varphi(K_1 t) \leqslant \psi(t) + c_1$ and $\varphi(t) \leqslant \delta_2$ implies $\psi(K_2 t) \leqslant \varphi(t) + c_2$ 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, [NS09] 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 [BSS18, CS19a, CS19b] and includes e.g. the top-k norm (sum of top k magnitudes of x). The type-2 constant of such norms is bounded whenever a has bounded support.

1.3.3 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

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

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.,

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

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. [Bha97] 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 [GTJ83] asserts that $\|\cdot\|_{\mathsf{S}_E}$ has $O(1)$ type 2 or cotype 2 constant if and only if $\|\cdot\|_E$ does.

In Section 7.3 we use Theorem 1.14 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 [NRV13] (namely, bilinear maximization over the operator norm), albeit with a worse universal constant than in [NRV13]. 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 1.3, we thus obtain a near characterization of unitarily invariant matrix norms over which quadratic maximization admits a constant factor approximation algorithm.

1.3.4 Robust Principle Component Analysis

In [NRV13], 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 [DZHZ06]. 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 set of extreme points of $\mathrm{Ball}(\mathsf{S}_\infty)$ is 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 1.14 may be combined with the lower covariance separation oracles constructed in Section 7 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 1.14 with the separation oracles constructed in Section 7 and using algorithmic closure properties for complex interpolation (Proposition 6.10), 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

$$\|(\Pi_S v_i)_{i=1}^m\|_{X_0} \stackrel{\text{def}}{=} \sum_{i=1}^m \|\Pi_S v_i\|_2 \qquad \text{and} \qquad \|(\Pi_S v_i)_{i=1}^m\|_{X_1} \stackrel{\text{def}}{=} \alpha \sum_{i=1}^m \sum_{j=1}^m \|\Pi_S v_i - \Pi_S v_j\|_2 \,.$$

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.

1.4 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 [HH06, dKLP06], 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 [FL92, BR95, Bri02] 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 [Gro53] and its more common formulation in [LP68], which corresponds to the case $K = \mathrm{Ball}(\ell_\infty^n)$. This leads to a constant factor polynomial time algorithm, as shown in [AN06] (see [BMMN13] 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 [CW04] with application to correlation clustering, and the matching integrality-gap lower bound in this case was obtained in [AMMN06]. Hardness results in these settings (under various complexity assumptions) were obtained in [AN06, ABH$^+$05, KO09, RS09]. The survey [KN12] is devoted to the use of Grothendieck-type inequalities in combinatorial optimization.

Krivine [Kri73] (see also [Pis12]) 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 Section 7.1.1 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 [Nes98], a different proof of this was obtained. The problem of quadratic maximization over exactly 2-convex norms was investigated in [NS09], where a constant factor approximation algorithm was obtained under the additional (necessary) assumption of bounded q-concavity for some finite q (see Section 7.1.1 for definition); this was used in [NS09] 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 [LRV13] is a consequence of the algorithm of [NS09].

Implicit in the non-commutative Khintchine inequality [LPP91] 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 [Gro53] conjectured a noncommutative version of his inequality which was proven in [Pis78] (the sharp constant was obtained in [Haa85]). In [NRV13], 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 [NRV13] to give approximation algorithms for robust principal component analysis and a generalization of the orthogonal procrustes problem. In [RV15], 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 [BRS15] (see also [HV16] for a different proof).

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 ([BBH$^+$12]). Vertex expansion and a related analytic proxy ([LRV13]) can be cast as quadratic maximization, and so can densest-k-subgraph, sparse-PCA, the spread constant of a metric [ABS98], and the poincare 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.

The versatility of quadratic maximization is evident from being able to For appropriate choices of linear maps and convex sets, quadratic maximization also captures (upto constants) the maximization (in 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 [BBH$^+$12], refuting random constraint satisfaction problems [RRS16], statistical physics, tensor principal component analysis and tensor decomposition [BKS15, GM15, MR14, HSS15], game theory, control theory and population dynamics [DK08].

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

2 Detailed Preliminaries

2.1 Vectors and Matrices

All vectors we consider are finite dimensional and real valued (with the exception of Section 6.4 where we consider complex entries). We denote the elementary basis of $\mathbb{R}^n$ by $e_1, e_2 \dots e_n$. Vectors are denoted by lower-case letters $(x, y, z, v, \dots)$ and we often refer to sequences of vectors $x_1, x_2, \dots \in \mathbb{R}^n$ indexed by subscripts. We also use subscripts to denote entries of a vector (e.g. for $x \in \mathbb{R}^n$, x_i denotes its i-th entry), but the distinction between entries of a single vector and a sequence of vectors will be clear in context. For example, $(x_i)_j$ denotes the j-th entry of vector x_i.

Matrices are always finite dimensional and are denoted by upper-case letters $(A, B, C, M, W, \dots)$. For a field $F \in \{\mathbb{R}, \mathbb{C}\}$, let $M_n(F)$ (resp. $M_{n,m}(F)$) be the set of all $n \times n$ (resp. $n \times m$) matrices whose entries are from F. We use $\mathbb{PSD}^n$ to denote the set of symmetric positive semidefinite (henceforth PSD) matrices in $M_n(\mathbb{R})$. We (mostly) use upper case blackboard-bold letters $(\mathbb{X}, \mathbb{Y}, \mathbb{W}, \dots)$ to denote matrices that are the indeterminates in a convex program. Given a vector $v \in \mathbb{R}^n$, let $\mathrm{Diag}(v) \in M_n(\mathbb{R})$ be such that $(\mathrm{Diag}(v))_{i,j} = v_i$ if $i = j$ and 0 otherwise. Similarly, given a matrix $A \in M_n(\mathbb{R})$, $\mathrm{diag}(A)$ is an n-dimensional vector defined as $(\mathrm{diag}(A))_i \stackrel{\text{def}}{=} A_{i,i}$. For a vector or matrix, let $\mathrm{bit}(\cdot)$ denote its bit complexity, which is the number of bits used to represent it.

2.2 Norms

In this paper we consider exclusively finite dimensional norms denoted as $\|\cdot\|_X$. Throughout, the underlying vector space is finite dimensional euclidean space $\mathbb{R}^n$, with the exception of Section 6.4 where we discuss complex interpolation and hence the underlying space is $\mathbb{C}^n$.

We denote the unit ball of a norm $\|\cdot\|_X$ over $\mathbb{R}^n$ by $\mathrm{Ball}(X) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_X \leqslant 1\}$. $\mathrm{Ball}(X)$ is an origin-symmetric (i.e., $\mathrm{Ball}(X) = -\mathrm{Ball}(X)$) convex body. Recall that every origin symmetric convex body $C \subseteq \mathbb{R}^n$ (i.e., $C = -C$) can be realized as the unit ball of a norm defined as $|||x||| \stackrel{\text{def}}{=} \inf_{\lambda > 0}\{x/\lambda \in C\}$ ($|||\cdot|||$ is called the Minkowski functional of C). Thus there is a one-to-one correspondence between norms and origin symmetric convex bodies.

In this work we assume norms are given as input to algorithms via membership oracles, i.e., an oracle that takes $x \in \mathbb{R}^n$ as input and returns "Inside" if $x \in \mathrm{Ball}(X)$ and "Outside" otherwise Equivalently (using binary search) one can assume there is an oracle returning $\|x\|_X$ given input $x \in \mathbb{R}^n$.

Recall the dual norm $\|\cdot\|_{X^*}$ over $\mathbb{R}^n$ is given by

$$\|\xi\|_{X^*} \stackrel{\text{def}}{=} \max_{x \in \mathrm{Ball}(X)} \langle \xi , x \rangle$$

and that $\mathrm{Ball}(X^*) = \mathrm{Ball}(X)^\circ$ where the polar of a set B is defined as $\{\xi \mid \langle x , \xi \rangle \leqslant 1 \forall x \in B\}$. By the bipolar theorem (Theorem 2.4) we also conclude $\mathrm{Ball}(X) = \mathrm{Ball}(X^*)^\circ$ which implies $\|\cdot\|_{(X^*)^*} = \|\cdot\|_X$.

We will repeatedly use the following inequality due to Kahane [Kah64].

Theorem 2.1 (Kahane-Khintchine Inequality). *Consider any $1 \leqslant p < \infty$. Then there are constants*

$C_p, C'_p \lesssim \sqrt{p}$, such that for every finite sequence of vectors $\{x_i\}$ in a normed space X,

$$\frac{1}{C'_p} \leqslant \frac{\mathbb{E}\,[\|\sum_i \mathrm{g}_i x_i\|_{\mathrm{X}}^p]^{1/p}}{\mathbb{E}\,[\|\sum_i \mathrm{g}_i x_i\|_{\mathrm{X}}]} \leqslant C_p\,.$$

Type-2/cotype-2 constants of a norm are powerful classification tools from Banach space theory and informally speaking, provide upper/lower bounds on the expected deviation (measured according to the ambient norm) of a random walk from the origin. Formally,

Definition 2.2 (Type-2/Cotype-2)**.** *The (Gaussian) type-2 constant of a normed space X, denoted by $\widetilde{T}_2(X)$, is the smallest constant C such that for every finite sequence of vectors $\{x_i\}$ in X,*

$$\mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot x_i\|_{\mathrm{X}}^2\right] \leqslant\; C^2 \cdot \sum_i \|x_i\|_{\mathrm{X}}^2$$

where each g_i is an independent standard Gaussian.

The (Gaussian) cotype-2 constant of a normed space X, denoted by $\widetilde{C}_2(X)$, is the smallest constant C such that for every finite sequence of vectors $\{x^i\}$ in X,

$$\mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot x_i\|_{\mathrm{X}}^2\right] \geqslant\; \frac{1}{C^2} \cdot \sum_i \|x_i\|_{\mathrm{X}}^2$$

$\widetilde{C}_2(X^*)$ and $\widetilde{T}_2(X)$ are closely related; in particular $\widetilde{C}_2(X^*) \leqslant \widetilde{T}_2(X)$ and if X is n-dimensional then $\widetilde{T}_2(X) \lesssim \widetilde{C}_2(X^*) \cdot \log n$.

The Gaussian and rademacher type-2 (resp. cotype-2) constants are equivalent within universal constants. More specifically one has $\widetilde{T}_2(X) \leqslant T_2(X) \lesssim \widetilde{T}_2(X)$ and $\widetilde{C}_2(X) \leqslant C_2(X) \lesssim \widetilde{C}_2(X)$ (se [MP76]). In the sequel we work mostly with the Gaussian type-2/cotype-2 constants since in certain places this allows us to provide slightly more precise estimates.

2.3 Polar Operations

In this paper, we use three different notions of polars, and use them to derive equivalences between approximate optimization and approximate separation over various sets. Here we introduce the three notions, beginning with the usual notion of polar.

Definition 2.3 (Standard Polar in $\mathbb{R}^n$)**.** *Let $B \subseteq \mathbb{R}^n$. The* polar *of B is $B^\circ \stackrel{\text{def}}{=} \{\xi \in \mathbb{R}^n \mid \langle \xi, x\rangle \leqslant 1 \;\; \forall x \in B\}$. Note that B° is always convex and moreover $\mathbf{conv}(B)^\circ = B^\circ$.*

This notion has the following nice duality statement.

Theorem 2.4 (Standard Bipolar Theorem)**.** *Let $B \subseteq \mathbb{R}^n$ be a origin-symmetric convex body. Then we have $(B^\circ)^\circ = B$.*

The other two notions of polars are defined with respect to a self-dual cone $K \subseteq \mathbb{R}^n$. In this paper we will consider either the non-negative orthant or the positive semidefinite cone. We first define the notions of upward-closedness and downward-closedness. Given $B \subseteq K$, let $\uparrow B \stackrel{\text{def}}{=} (B+K) \cap K$ and $\downarrow B \stackrel{\text{def}}{=} (B-K) \cap K$ denote upward and downward closures respectively, where $+$ and $-$ denote Minkowski addition and subtraction respectively. B is said to be *upward-closed* when $\uparrow B = B$ and *downward-closed* when $\downarrow B = B$.

We now define a conic version of the polar operation for a subset $B \subseteq K$.

Definition 2.5 (Conic Polars)**.** *Let $K \subseteq \mathbb{R}^n$ be a self-dual cone. We define the* polar *of a set $B \subseteq K$ (with respect to K) as $B^{\circ K} \stackrel{\text{def}}{=} \{\xi \in K \mid \langle \xi, x\rangle \leqslant 1 \ \ \forall x \in B\}$. Note that $B^{\circ K}$ is always convex and moreover $\mathbf{conv}(B)^{\circ K} = B^{\circ K}$.*

For any $A \subseteq K$, $A^{\circ K}$ is downward-closed; indeed if $y \in A^{\circ K}$ and $z = y - w$ for some $w \in K$, then $\langle z, x\rangle \leqslant \langle y, x\rangle \leqslant 1$ for all $x \in A$. We require a conic version of the bipolar theorem. As we could not locate a reference for this, we include a proof here.

Fact 2.6 (Conic Bipolar Theorem)**.** *Let $K \subseteq \mathbb{R}^n$ be a self-dual cone. Let $B \subseteq K$ be a bounded, closed and convex set containing the origin. Then $(B^{\circ K})^{\circ K} = \downarrow B$.*

Proof. By definition, $B \subseteq (B^{\circ K})^{\circ K}$. Since $A^{\circ K}$ is downward-closed for any A, $\downarrow B \subseteq (B^{\circ K})^{\circ K}$. For the other direction, assume towards contradiction that $x \in (B^{\circ K})^{\circ K} \setminus \downarrow B$. Since $x \in K$, it implies $x \notin B - K$. Since $B - K$ is also convex and closed (this follows since B is closed and bounded and K is closed), by Hahn-Banach separation theorem there exists $y \in \mathbb{R}^n$ such that $\langle x, y\rangle > c$ and $\langle z, y\rangle < c$ for all $z \in B - K$. We claim that $y \in K$; otherwise, there exists $w \in K$ with $\langle y, w\rangle < 0$, and $\langle y, -\alpha w\rangle = -\alpha\langle y, w\rangle > c$ for large enough $\alpha > 0$ with $-\alpha w \in B - K$, leading to contradiction. Since B contains 0, $c > 0$. Then $y/c \in B^{\circ K}$, so $\langle x, y/c\rangle > 1$ contradicts that $x \in (B^{\circ K})^{\circ K}$. ■

We define a third notion of polar, namely the *inverse polar*, again with respect to a self-dual cone K, but this time it is the set of elements whose inner product with every element of the original set is at least 1.

Definition 2.7 (Inverse Polar)**.** *Let $K \subseteq \mathbb{R}^n$ be a self-dual cone. For a set $B \subseteq K$ and $c \in \mathbb{R}$, we define $B_c^{\odot} \stackrel{\text{def}}{=} \{\xi \mid \langle \xi, x\rangle \geqslant c \ \ \forall x \in B\}$, and let $B^{\odot} := B_1^{\odot}$. Note that $B^{\odot}$ is always convex and moreover $\mathbf{conv}(B)^{\odot} = B^{\odot}$.*

We will consider B contained in a self-dual cone $K \subseteq \mathbb{R}^n$. By definition, $B \subseteq (B^{\odot})^{\odot}$ for any B. Under the above condition on B, they are indeed equal. We include a proof below.

Fact 2.8 (Inverse Bipolar Theorem)**.** *Let B be a closed, convex, and upward-closed set contained in some self-dual cone $K \subseteq \mathbb{R}^n$. Then (1) $B^{\odot}$ is a closed, convex and upward-closed set contained in K, and (2) $(B^{\odot})^{\odot} = B$.*

Proof. Closure follows straightforwardly from the definition.

We next show that $B^{\odot}$ is also upward-closed. For any $x \in B^{\odot}$, $y \in K$, and $z \in B$, the definition of $B^{\odot}$ implies $\langle x, z\rangle \geqslant 1$, and the self-duality of K implies $\langle y, z\rangle \geqslant 0$. Therefore, $\langle x + y, z\rangle \geqslant 1$, implying $x + y \in B^{\odot}$.

We also show that $B^{\odot} \subseteq K$. More strongly, we show that the set $B_c^{\odot}$ for any $c \in \mathbb{R}$ is contained in K. Assume towards contradiction that there exists $y \in B_c^{\odot} \setminus K$. By the self-duality of K, there exists $z \in K$ with $\langle z, y\rangle < 0$. Let $x \in B$. By the upward-closedness of B, $x + \alpha z \in B$ for any $\alpha \geqslant 0$, but $\langle y, x + \alpha z\rangle = \langle y, x\rangle + \alpha\langle y, z\rangle$ will be strictly less than c for large enough α, achieving the desired contradiction.

Therefore, $B^{\odot}$ is a convex, and upward-closed set contained in K. The same argument implies convexity and upward-closedness of $(B^{\odot})^{\odot}$.

Finally, we prove $(B^{\odot})^{\odot} = B$. Assume towards contradiction that there exists $x \in (B^{\odot})^{\odot} \setminus B$. Since B is closed and convex, there exist $y \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that $\langle x, y\rangle < c$ and $\langle z, y\rangle > c$ for all $z \in B$. In particular, $y \in B_c^{\odot}$, which implies that $y \in K$. Since $x \in (B^{\odot})^{\odot} \subseteq K$, $\langle x, y\rangle \geqslant 0$, so $c > 0$. Then y/c satisfies $\langle z, y/c\rangle > 1$, so $y \in B^{\odot}$, but $\langle x, y/c\rangle < 1$, so it contradicts that $x \in (B^{\odot})^{\odot}$. ■

2.4 Quadratic Maximization and Related Optimization Problems

We are primarily interested in approximation algorithms for maximizing a quadratic form over an origin-symmetric convex body. Formally we define,

Definition 2.9 (Quadratic Maximization)**.**
For an $n \times n$ matrix A and a norm $\|\cdot\|_X$ over $\mathbb{R}^n$, we define quadratic maximization of A over X *as*

$$Q_X^{\max}(A) \stackrel{\text{def}}{=} \max_{x \in \text{Ball}(X)} \langle x \,, Ax \rangle \,. \tag{24}$$

We will also be interested in two special cases of quadratic maximization; namely, quadratic maximization of PSD instances and bilinear form maximization over $C_1 \times C_2$ for origin-symmetric convex bodies C_1, C_2. We define some notation for the bilinear case:

Definition 2.10 (Bilinear Maximization/Operator Norm)**.**
For an $n \times m$ matrix A and norms $\|\cdot\|_X, \|\cdot\|_Y$ over $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, we define bilinear maximization of A over (X, Y), *as*

$$\text{Op}_{X,Y}^{\max}(A) \stackrel{\text{def}}{=} \max_{\substack{x \in \text{Ball}(X) \\ y \in \text{Ball}(Y)}} \langle x \,, Ay \rangle \,. \tag{25}$$

By definition of the dual norm, bilinear maximization can be cast as the operator norm/distortion of A when A is thought of as a linear map from Y to X^*. Formally we have,

$$\|A\|_{Y \to X^*} \stackrel{\text{def}}{=} \max_{y \in \text{Ball}(Y)} \|Ay\|_{X^*} \;=\; \max_{\substack{x \in \text{Ball}(X) \\ y \in \text{Ball}(Y)}} \langle x \,, Ay \rangle \;=\; \max_{\substack{x \in \text{Ball}(X) \\ y \in \text{Ball}(Y)}} \langle A^* x \,, y \rangle \;=\; \|A\|_{X \to Y^*}$$

$$\textit{thus } \|A\|_{Y \to X^*} = \|A\|_{X \to Y^*} = \text{Op}_{X,Y}^{\max}(A) \,. \tag{26}$$

Throughout this document, we will switch notation between operator norms and bilinear maximization based on convenience and context.

Bilinear maximization can be reduced to quadratic maximization using the following identity

$$\text{Op}_{X,Y}^{\max}(A) = Q_{X \oplus_\infty Y}^{\max}(B) \quad \text{where } B \stackrel{\text{def}}{=} \frac{1}{2} \cdot \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$$

where $\|\cdot\|_{X \oplus_\infty Y}$ is a norm over $\mathbb{R}^{n+m}$ defined as $\|(x, y)\|_{X \oplus_\infty Y} \stackrel{\text{def}}{=} \max\{\|x\|_X, \|y\|_Y\}$ and whose unit ball is given simply by $\text{Ball}(X) \times \text{Ball}(Y)$.

We will repeatedly use the following simple property that relates quadratic maximization of a PSD matrix over X to bilinear maximization over (ℓ_2^n, X):

$$Q_X^{\max}(B^* B) = \max_{x \in \text{Ball}(X)} \|Bx\|_2^2 = \|B\|_{X \to \ell_2^n}^2 = \|B^*\|_{\ell_2^n \to X^*}^2 \,. \tag{27}$$

Intuitively, this is useful because it transforms a search problem over $\text{Ball}(X)$ to a search problem over $\text{Ball}(\ell_2^n)$. Quadratic maximization of a PSD matrix $W \succeq 0$ over X is in fact equal to bilinear maximization of W over (X, X). Indeed by Cauchy-Schwarz we have

$$\text{Op}_{X,X}^{\max}(B^* B) = \max_{x,y \in \text{Ball}(X)} \langle x \,, B^* By \rangle = \max_{x,y \in \text{Ball}(X)} \langle Bx \,, By \rangle \leqslant \max_{x,y \in \text{Ball}(X)} \|Bx\|_2 \|By\|_2 = \|B\|_{X \to 2}^2 \,.$$

The inequality in the reverse direction follows by considering the substitution $y = x = x'$ where x' is the vector maximizing $\|B\|_{X\to 2}$.

We will also consider quadratic maximization over sections of convex sets, i.e., maximization over $V \cap \mathrm{Ball}(X)$ where V is a subspace. Upto a factor of $1 + o(1)$, maximization over $V \cap \mathrm{Ball}(X)$ can be captured by the maximization over $\mathrm{Ball}(X)$ by adding a large negative multiple of the projector to the orthogonal complement $V^{\perp}$. This is formalized in Observation 6.13.

The final special case of quadratic maximization we will consider is the minimum factor by which a linear map shrinks a unit vector. (it can be seen as a special case of quadratic maximization by reducing first to subspace quadratic maximization).

Definition 2.11 (Contractivity of a Linear Map)**.**
For an $n \times m$ matrix A and norms $\|\cdot\|_X, \|\cdot\|_Y$ over $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, we define the $Y \to X$ contractivity *of A as*

$$\|A\|^{\min}_{Y\to X} \stackrel{\text{def}}{=} \inf_{\|y\|_Y=1} \|Ay\|_X = \inf_{y\neq 0} \|Ay\|_X / \|y\|_Y \tag{28}$$

We will be concerned with the case of invertible $n \times n$ matrices A, where one has the following useful connection with operator norm (bilinear maximization):

$$\begin{aligned} \|A\|^{\min}_{Y\to X} \;&=\; \inf_{\|y\|_Y=1} \|Ay\|_X \;=\; \inf_{\|Ay\|_X=1} \frac{1}{\|y\|_Y} \\ =\; \inf_{\|x\|_X=1} \frac{1}{\|A^{-1}x\|_Y} \;&=\; \inf_{x\neq 0} \frac{\|x\|_X}{\|A^{-1}x\|_Y} \;=\; \frac{1}{\|A^{-1}\|_{X\to Y}}\,. \end{aligned} \tag{29}$$

2.5 Projective Tensor Norm and Related Measures

Let $\|\cdot\|_X, \|\cdot\|_Y$ be norms over $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. As we make extensive use of duality (and algorithmic versions thereof) in this work, it will be useful for us to investigate the polars of the following level sets:

$$\{A \in M_n(\mathbb{R}) \mid \mathrm{Q}^{\max}_X(A) \leqslant 1\}\,, \{A \in M_{n,m}(\mathbb{R}) \mid \mathrm{Op}^{\max}_{X,Y}(A) \leqslant 1\}\,.$$

2.5.1 Polars of Various Sets of Bounded Forms.

Polar of Bounded Bilinear Forms. We begin by discussing the bilinear case where the polar has been extensively studied in the context of the metric theory of tensor products. $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ is easily checked to be a norm on the space $M_n(\mathbb{R}) \equiv \mathbb{R}^n \otimes \mathbb{R}^n$ and is known as the injective tensor norm. While the injective tensor norm is typically denoted as $\|\cdot\|_{X^*\check{\otimes}Y^*}$, we choose to use the unorthodox notation of $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ for convenience in later sections.

The dual norm of $\|\cdot\|_{X^*\check{\otimes}Y^*}$ is called the projective tensor norm $\|\cdot\|_{X\widehat{\otimes}Y}$ and has the following representation

$$\|\Xi\|_{X\widehat{\otimes}Y} \stackrel{\text{def}}{=} \inf \sum_i \|x_i\|_X \cdot \|y_i\|_Y \;=\; \inf\; (\textstyle\sum_i \|x_i\|_X^2)^{1/2} \cdot (\sum_i \|y_i\|_Y^2)^{1/2} \tag{30}$$

where the infimum runs over all finitary decompositions $\Xi = \sum_i x_i y_i^*$. The unit ball of the projective norm is given simply by $\mathrm{Ball}(X\widehat{\otimes}Y) \stackrel{\text{def}}{=} \mathbf{conv}(\{xy^* \mid x \in \mathrm{Ball}(X)\,, y \in \mathrm{Ball}(Y)\})$. We include a derivation of the injective/projective duality which arises quite naturally. Observe that

$$\mathrm{Op}^{\max}_{X,Y}(A) = \max_{x\in\mathrm{Ball}(X), y\in\mathrm{Ball}(Y)} \langle x\,, Ay\rangle = \max_{x\in\mathrm{Ball}(X), y\in\mathrm{Ball}(Y)} \langle A\,, xy^*\rangle = \max_{\Xi\in\mathrm{Ball}(X\widehat{\otimes}Y)} \langle A\,, \Xi\rangle\,.$$

From the final equality we see that the set of bounded bilinear forms $\{A \in M_{n,m}(\mathbb{R}) \mid \mathrm{Op}^{\max}_{X,Y}(A) \leqslant 1\} = \{A \in M_{n,m}(\mathbb{R}) \mid \max_{\Xi \in \mathrm{Ball}(X \widehat{\otimes} Y)} \langle A, \Xi \rangle \leqslant 1\}$ is precisely $\mathrm{Ball}(X \widehat{\otimes} Y)^{\circ}$. It is then easily checked that $\|\cdot\|_{X \widehat{\otimes} Y}$ as defined above is the Minkowski functional of $\mathbf{conv}(\{xy^* \mid x \in \mathrm{Ball}(X), y \in \mathrm{Ball}(Y)\})$. Thus $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ and $\|\cdot\|_{X \widehat{\otimes} Y}$ are dual norms and therefore also satisfy

$$\|\Xi\|_{X \widehat{\otimes} Y} = \max_{\mathrm{Op}^{\max}_{X,Y}(A) \leqslant 1} \langle A, \Xi \rangle .$$

Polar of Bounded Quadratic Forms. Let $B^{Sym}_{\wedge}(X) \overset{\mathrm{def}}{=} \mathbf{conv}(\{xx^* \mid x \in \mathrm{Ball}(X)\})$. Similar to the bilinear case, we observe that $\mathrm{Q}^{\max}_X(A) = \max_{\mathbb{X} \in B^{Sym}_{\wedge}(X)} \langle A, \mathbb{X} \rangle$ and therefore the set of bounded quadratic forms $\{A \in M_n(\mathbb{R}) \mid \mathrm{Q}^{\max}_X(A) \leqslant 1\}$ is the polar of $B^{Sym}_{\wedge}(X)$. Restricted to the PSD cone $\mathbb{PSD}^n$, we may consider the Minkowski functional of $B^{Sym}_{\wedge}(X)$ which is given by

$$\wedge^{Sym}_X(\mathbb{X}) \overset{\mathrm{def}}{=} \inf \sum_i \|x_i\|_X^2$$

where the infimum runs over all finitary decompositions $\mathbb{X} = \sum_i x_i x_i^*$. Note also that by definition of $\wedge^{Sym}_X(\cdot)$, $B^{Sym}_{\wedge}(X)$ can be alternatively described as $\{\mathbb{X} \in \mathbb{PSD}^X \mid \wedge^{Sym}_X(\mathbb{X}) \leqslant 1\}$ (note that this is not the unit ball of any norm). We caution the reader that even though $\wedge^{Sym}_X(\cdot)$ is homogeneous and satisfies triangle inequality, it is not a norm as it is defined only within the PSD cone. It may be useful to think of $\wedge^{Sym}_X(\mathbb{X})$ as a symmetric version of the projective tensor norm $\|\cdot\|_{X \widehat{\otimes} X}$.

Conic Polar of Bounded PSD Quadratic Forms. Since $B^{Sym}_{\wedge}(X)$ is a subset of the PSD cone which is self-dual, it is possible and beneficial to study its conic polar $B^{Sym}_{\wedge}(X)^{\circ \mathbb{PSD}^n}$ which is in fact the set of bounded PSD quadratic forms, i.e.,

$$B^{Sym}_{\wedge}(X)^{\circ \mathbb{PSD}^n} = \{W \in \mathbb{PSD}^n \mid \mathrm{Q}^{\max}_X(W) \leqslant 1\} .$$

Thus by the conic version of the bipolar theorem (Fact 2.6) we have $(B^{Sym}_{\wedge}(X)^{\circ \mathbb{PSD}^n})^{\circ \mathbb{PSD}^n} = \downarrow B^{Sym}_{\wedge}(X)$. From this we conclude $\downarrow B^{Sym}_{\wedge}(X)$ and the set of bounded PSD quadratic forms are (conically) polar to one another:

$$(\downarrow B^{Sym}_{\wedge}(X))^{\circ \mathbb{PSD}^n} = \{\mathbb{W} \in \mathbb{PSD}^n \mid \mathrm{Q}^{\max}_X(\mathbb{W}) \leqslant 1\}$$
$$\downarrow B^{Sym}_{\wedge}(X) = \{\mathbb{W} \in \mathbb{PSD}^n \mid \mathrm{Q}^{\max}_X(\mathbb{W}) \leqslant 1\}^{\circ \mathbb{PSD}^n} . \tag{31}$$

Restricted to the PSD cone $\mathbb{PSD}^n$, we may describe the Minkowski functional of $\downarrow B^{Sym}_{\wedge}(X)$

$$\wedge^{\downarrow Sym}_X(\mathbb{X}) \overset{\mathrm{def}}{=} \inf_{\mathbb{Y} \succeq \mathbb{X}} \wedge^{Sym}_X(\mathbb{Y})$$

which can be thought of as a symmetric and monotone (in the Loewner ordering) variant of the projective tensor norm $\|\cdot\|_{X \widehat{\otimes} X}$.

Inverse Polar of Non-Contractive Linear Maps. Recall that $B^{\odot} \overset{\mathrm{def}}{=} \{y \mid \langle y, x \rangle \geqslant 1 \ \ \forall x \in B\}$ and that $\|U\|^{\min}_{X^* \to 2} \overset{\mathrm{def}}{=} \inf_{\|\xi\|_{X^*}=1} \|U\xi\|_2$. In this subsection we wish to describe $\{U^*U \mid \|U\|^{\min}_{X^* \to 2} \geqslant 1\}^{\odot}$. Let $B^{\uparrow}_{\wedge}(X) \overset{\mathrm{def}}{=} \mathbf{conv}(\{\xi\xi^* \mid \|\xi\|_{X^*} \geqslant 1\})$. Observe that

$$(\|U\|^{\min}_{X^* \to 2})^2 = \inf_{\|\xi\|_{X^*} \geqslant 1} \|U\xi\|_2^2 = \inf_{\|\xi\|_{X^*} \geqslant 1} \langle U^*U, \xi\xi^* \rangle = \inf_{W \in B^{\uparrow}_{\wedge}(X)} \langle U^*U, W \rangle .$$

Thus we have $\{U^*U \mid \|U\|^{\min}_{X^*\to 2} \geqslant 1\} = B^{\uparrow}_{\wedge}(X)^{\odot}$. Furthermore since $\{U^*U \mid \|U\|^{\min}_{X^*\to 2} \geqslant 1\}$ is a closed, convex and upward-closed (in the Loewner ordering) subset of $\mathbb{PSD}^n$, we may apply a version of the bipolar theorem for upward-closed sets (Fact 2.8) to obtain

$$B^{\uparrow}_{\wedge}(X)^{\odot} = \{U^*U \in M_n(\mathbb{R}) \mid \|U\|^{\min}_{X^*\to 2} \geqslant 1\} \tag{32}$$

and moreover that $B^{\uparrow}_{\wedge}(X)$ is closed, convex and upward-closed.

We give a second description of $B^{\uparrow}_{\wedge}(X)$. To this end we define for any $W \in \mathbb{PSD}^n$,

$$\wedge^{\uparrow}_{X}(W) \stackrel{\text{def}}{=} \sup \textstyle\sum_i \|w_i\|^2_{X^*}\,. \tag{33}$$

where the supremum runs over all finitary decompositions $W = \sum_i w_i w_i^*$. It is then easily checked that $B^{\uparrow}_{\wedge}(X) = \{\mathbb{W} \succeq 0 \mid \wedge^{\uparrow}_{X}(\mathbb{W}) \geqslant 1\}$.

2.5.2 Covariance Regions and their Connection to Projective Norm and Related Measures

We define the upper (resp. lower) covariance region which, as discussed in the introduction allow us to formulate a generic approximate convex optimization approach to quadratic (resp. PSD quadratic) maximization in the presence of type-2 (resp. dual cotype-2).

Definition 2.12 (Gaussian Rounding Function)**.**
For a norm X over $\mathbb{R}^n$ and i.i.d. standard Gaussians $\mathsf{g} = (\mathsf{g}_1, \ldots, \mathsf{g}_n)$ we define the Gaussian Rounding Function $\mathcal{N}_X(\cdot) : \mathbb{PSD}^n \to \mathbb{R}_{\geqslant 0}$ *as* $\mathcal{N}_X(\mathbb{X}) \stackrel{\text{def}}{=} \mathbb{E}\,[\|\mathbb{X}^{1/2}\mathsf{g}\|^2_X]$.

Remark 2.13. *Since a Gaussian distribution is uniquely determined by its first two moments, we may alternatively define $\mathcal{N}_X(\mathbb{X})$ as $\mathbb{E}\,[\|\sum_i \mathsf{g}_i \cdot x_i\|^2_X]$ where (x_i) is any finite sequence satisfying $\mathbb{X} = \sum_i x_i x_i^*$. In fact $|||(x_i)||| \stackrel{\text{def}}{=} \mathbb{E}[\|\sum_i \mathsf{g}_i \cdot x_i\|^2_X]^{1/2}$ defines a norm on the space of sequences of a fixed length.*

Definition 2.14 (Upper Covariance Body)**.**
We define the Upper Covariance Body *denoted by $\mathcal{U}(X)$ as $\{\mathbb{X} \succeq 0 \mid \mathcal{N}_X(\mathbb{X}) \leqslant 1\}$.*

Definition 2.15 (Lower Covariance Region)**.**
We define the Lower Covariance Region *denoted by $\mathcal{L}(X)$ as $\{\mathbb{W} \succeq 0 \mid \mathcal{N}_{X^*}(\mathbb{W}) \geqslant 1\}$.*

The lower covariance region is the complement of (the interior of) the upper covariance region of the dual, i.e., $\mathcal{L}(X) = \mathbb{PSD}^n \setminus \mathrm{Int}(\mathcal{U}(X^*))$.

$\mathcal{N}_X(\cdot)$ is known to be non-decreasing (in the Loewner ordering). We include a proof below:

Fact 2.16. *$\mathcal{N}_X(\cdot)$ is non-decreasing in the Loewner ordering.*

Proof. It suffices to show that for any $\mathbb{X}, \mathbb{Y} \in \mathbb{PSD}^n$, we have $\mathcal{N}_X(\mathbb{X}) \leqslant \mathcal{N}_X(\mathbb{X}+\mathbb{Y})$. Let $\mathsf{g}_1, \ldots, \mathsf{g}_n, \mathsf{h}_1, \ldots, \mathsf{h}_n, \mathsf{z}_1, \ldots \mathsf{z}_n$ be i.i.d. standard Gaussians. Observe that the vector $\mathbb{X}^{1/2}\mathsf{g}$ has covariance $\mathbb{X}$ and that the vectors $\mathbb{X}^{1/2}\mathsf{g} \pm \mathbb{Y}^{1/2}\mathsf{h}$, both have covariance $\mathbb{X}+\mathbb{Y}$. Thus $(\mathbb{X}+\mathbb{Y})^{1/2}\mathsf{z}$ has the same distribution as each of the vectors $\mathbb{X}^{1/2}\mathsf{g} \pm \mathbb{Y}^{1/2}\mathsf{h}$. Therefore we have,

$$\begin{aligned}
\mathcal{N}_X(\mathbb{X})^{1/2} &= \mathop{\mathbb{E}}_{\mathsf{g}}\,[\|\mathbb{X}^{1/2}\mathsf{g}\|^2_X]^{1/2} \\
&\leqslant \frac{1}{2}\cdot \mathop{\mathbb{E}}_{\mathsf{g},\mathsf{h}}\,[\|\mathbb{X}^{1/2}\mathsf{g} + \mathbb{Y}^{1/2}\mathsf{h}\|^2_X]^{1/2} \;+\; \frac{1}{2}\cdot \mathop{\mathbb{E}}_{\mathsf{g},\mathsf{h}}\,[\|\mathbb{X}^{1/2}\mathsf{g} - \mathbb{Y}^{1/2}\mathsf{h}\|^2_X]^{1/2} \\
&= \mathop{\mathbb{E}}_{\mathsf{z}}\,[\|(\mathbb{X}+\mathbb{Y})^{1/2}\mathsf{z}\|^2_X]^{1/2}
\end{aligned}$$

$$= \mathcal{N}_X(\mathbb{X} + \mathbb{Y})^{1/2}$$

where the first inequality is an application of the triangle inequality in the space $L_2(\mathbb{R}^n, \gamma)$ where γ denotes the n-variate Gaussian measure. ■

The next observation relates the bodies $\mathcal{U}(X)$, $B_{\wedge}^{Sym}(X)$, $\downarrow B_{\wedge}^{Sym}(X)$, $\mathrm{Ball}(X \widehat{\otimes} X) \cap \mathbb{PSD}^n$.

Observation 2.17 (Relating Projective Measures to Gaussian Rounding Function)**.**
Let X be a finite dimensional normed space. For any $\mathbb{X} \in \mathbb{PSD}^n$, we have,

$$\wedge_X^{\downarrow Sym}(\mathbb{X}) \leqslant \|\mathbb{X}\|_{X \widehat{\otimes} X} \leqslant \wedge_X^{Sym}(\mathbb{X}) \leqslant \mathcal{N}_X(\mathbb{X}) \leqslant \widetilde{T}_2(X)^2 \cdot \wedge_X^{\downarrow Sym}(\mathbb{X}) \,.$$

This implies the inclusions

$$\begin{aligned}
&\widetilde{T}_2(X)^{-2} \cdot \downarrow B_{\wedge}^{Sym}(X) \\
&\subseteq \mathcal{U}(X) \\
&\subseteq B_{\wedge}^{Sym}(X) = \mathbf{conv}(\{xx^* \mid x \in \mathrm{Ball}(X)\}) \\
&\subseteq \mathrm{Ball}(X \widehat{\otimes} X) \cap \mathbb{PSD}^n = \mathbf{conv}(\{xy^* \mid x, y \in \mathrm{Ball}(X)\}) \cap \mathbb{PSD}^n \\
&\subseteq \downarrow B_{\wedge}^{Sym}(X) = \downarrow \mathbf{conv}(\{xx^* \mid x \in \mathrm{Ball}(X)\}) \,.
\end{aligned}$$

Proof. Let $\mathbb{X} \succeq 0$. For the first inequality, assume $\|\mathbb{X}\|_{X \widehat{\otimes} X} \leqslant 1$. Then by definition there exist finite sequences $(u_i), (v_i)$ such that $\mathbb{X} = \sum_i u_i v_i^*$ and $\sum_i \|u_i\|_X^2, \sum_i \|v_i\|_X^2 \leqslant 1$. Also we have $\sum_i (u_i u_i^* + v_i v_i^*)/2 \succeq \sum_i u_i v_i^*/2 + \sum_i v_i u_i^*/2 = \sum_i u_i v_i^* = \mathbb{X}$, where the second last equality follows by symmetry of $\mathbb{X}$. By assumption on $(u_i), (v_i)$, we conclude $\wedge_X^{Sym}(\sum_i (u_i u_i^* + v_i v_i^*)/2) \leqslant 1$. Thus $\wedge_X^{\downarrow Sym}(\mathbb{X}) = \inf_{\mathbb{Y} \succeq \mathbb{X}} \wedge_X^{Sym}(\mathbb{X}) \leqslant 1$.

The second inequality is immediate.

For the third inequality we approximate the Gaussian integral $\mathbb{E}\,[\|\mathbb{X}^{1/2}\mathrm{g}\|_X^2]$ by a sequence of finite sums to obtain a sequence of upper bounds on $\wedge_X^{Sym}(\mathbb{X})$ that converge to $\mathcal{N}_X(\mathbb{X})$.

For the fourth inequality, consider any $\mathbb{Y} \succeq \mathbb{X}$ and any fixed decomposition $\mathbb{Y} = \sum_i y_i y_i^*$. Now since $\mathcal{N}_X(\cdot)$ is non-decreasing in the Loewner ordering (Fact 2.16), we have

$$\mathcal{N}_X(\mathbb{X}) \leqslant \mathcal{N}_X(\mathbb{Y}) \leqslant \widetilde{T}_2(X)^2 \cdot \sum_i \|y_i\|_X^2$$

where the final inequality in the preceding equation follows from the definition of the Gaussian type-2 constant. Taking infimum over all $\mathbb{Y} \succeq \mathbb{X}$ and all decompositions of $\mathbb{Y}$ completes the proof of the claim. ■

Remark 2.18. *It is clear from the proof that $\widetilde{T}_2(X)^2$ is the best possible constant in the fourth inequality.*

Observation 2.17 allows us to prove approximate convexity of the upper covariance body.

Observation 2.19 (Convex Hull of $\mathcal{U}(X)$ and Approximate Convexity)**.**
$\mathbf{conv}(\mathcal{U}(X)) = B_{\wedge}^{Sym}(X) = \mathbf{conv}(\{xx^* \mid x \in \mathrm{Ball}(X)\})$ *and moreover $\mathcal{U}(X)$ is $\widetilde{T}_2(X)^2$-approximately convex.*

Proof. By Observation 2.17 we have $\mathcal{U}(X) \subseteq B_{\wedge}^{Sym}(X)$ and so $\mathbf{conv}(\mathcal{U}(X)) \subseteq B_{\wedge}^{Sym}(X)$. On the other hand $\mathcal{U}(X) \supseteq \{xx^* \mid x \in \mathrm{Ball}(X)\}$ and so $\mathbf{conv}(\mathcal{U}(X)) \supseteq \mathbf{conv}(\{xx^* \mid x \in \mathrm{Ball}(X)\}) = B_{\wedge}^{Sym}(X)$. We conclude that $\mathbf{conv}(\mathcal{U}(X)) = B_{\wedge}^{Sym}(X)$. This yields the first claim.

By Observation 2.17 again, we have $\mathcal{U}(X) \subseteq \mathbf{conv}(\mathcal{U}(X)) \subseteq \widetilde{T}_2(X)^2 \cdot \mathcal{U}(X)$. This yields the second claim. ■

The next observation relates the regions $\mathcal{L}(X)$, $B^{\uparrow}_{\wedge}(X)$.

Observation 2.20. *For any* $\mathbb{W} \in \mathbb{PSD}^n$, *we have* $\widetilde{C}_2(X^*)^{-2} \cdot \wedge^{\uparrow}_X(\mathbb{W}) \leqslant \mathcal{N}_{X^*}(\mathbb{W}) \leqslant \wedge^{\uparrow}_X(\mathbb{W})$. *Equivalently,* $B^{\uparrow}_{\wedge}(X) \supseteq \mathcal{L}(X) \supseteq \widetilde{C}_2(X^*)^2 \cdot B^{\uparrow}_{\wedge}(X)$.

Proof. For the first inequality we approximate the Gaussian integral $\mathbb{E}\,[\|\mathbb{W}^{1/2}g\|^2_{X^*}]$ by a sequence of finite sums to obtain a sequence of lower bounds on $\wedge^{\uparrow}_X(\mathbb{W})$ that converge to $\mathcal{N}_{X^*}(\mathbb{W})$.

For the second inequality, we observe a stronger fact: namely it is true for any fixed decomposition $\mathbb{W} = \sum_i w_i w_i^*$. Indeed $\mathcal{N}_{X^*}(\mathbb{W}) \geqslant \widetilde{C}_2(X^*)^2 \cdot \sum_i \|w_i\|^2_{X^*}$ simply by the definition of the Gaussian cotype-2 constant. The claim follows. ∎

Remark 2.21. *It is clear from this proof that* $\widetilde{C}_2(X^*)^2$ *is the best possible constant for which the first inequality holds.*

Observation 2.20 allows us to prove approximate convexity of the lower covariance region.

Observation 2.22 (Convex Hull of $\mathcal{U}(X)$ and Approximate Convexity)**.**
$\mathbf{conv}(\mathcal{L}(X)) = B^{\uparrow}_{\wedge}(X) = \mathbf{conv}(\{xx^* \mid \|x\|_{X^*} \geqslant 1\})$ *and moreover* $\mathcal{L}(X)$ *is* $\widetilde{C}_2(X^*)^2$*-(inverse) approximately convex.*

Proof. By Observation 2.20 we have $\mathcal{L}(X) \subseteq B^{\uparrow}_{\wedge}(X)$ and so $\mathbf{conv}(\mathcal{U}(X)) \subseteq B^{\uparrow}_{\wedge}(X)$. On the other hand $\mathcal{U}(X) \supseteq \{xx^* \mid \|x\|_{X^*} \geqslant 1\}$ and so $\mathbf{conv}(\mathcal{L}(X)) \supseteq \mathbf{conv}(\{xx^* \mid \|x\|_{X^*} \geqslant 1\}) = B^{\uparrow}_{\wedge}(X)$. We conclude that $\mathbf{conv}(\mathcal{U}(X)) = B^{\uparrow}_{\wedge}(X)$. This yields the first claim.

By Observation 2.20 again, we have $\mathcal{L}(X) \subseteq \mathbf{conv}(\mathcal{L}(X)) \subseteq \widetilde{C}_2(X^*)^{-2} \cdot \mathcal{L}(X)$. This yields the second claim. ∎

Remark 2.23. *While we work predominantly over* $\mathbb{R}^n$, *in Section 6.4 we consider finite dimensional complex normed spaces. Every definition/claim in this Section 2.5 generalizes verbatim to the complex case where we replace* $M_{n,m}(\mathbb{R})$ *by* $M_{n,m}(\mathbb{C})$ *and* $\mathbb{PSD}^n$ *by* $\mathbb{PSD}^{\mathbb{C}^n} \overset{\text{def}}{=} \mathbf{conv}(\{xx^* \mid x \in \mathbb{C}^n\})$. *For* $A, B \in \mathbb{PSD}^{\mathbb{C}^n}$, *we write* $A \succeq B$ *if* $A - B \in \mathbb{PSD}^{\mathbb{C}^n}$.

2.5.3 Verification of Balance

The purpose of this section is to verify that all of the sets/functions to which we apply the approximate ellipsoid method satisfy certain sanity-conditions that are necessary for multiplicative approximation guarantees. We refer to these conditions as "balance", as defined below.

We say a norm $\|\cdot\|_X$ is (R, r)-balanced if $\mathrm{Ball}(X)$ contains a euclidean ball of radius r and is contained in a euclidean ball of radius R. We say a set $B \subseteq \mathbb{PSD}^n$ is $(R, r, \mathbb{PSD}^n)$*-balanced* if $r(\mathbb{PSD}^n \cap \mathrm{Ball}(\ell_2^{n\times n})) \subseteq B \subseteq R(\mathbb{PSD}^n \cap \mathrm{Ball}(\ell_2^{n\times n}))$. We also say that a set $B \subseteq \mathbb{PSD}^n$ is *inverse-*$(R, r, \mathbb{PSD}^n)$*-balanced* if $r \cdot \mathrm{Ball}(\ell_2^{n\times n}) \cap B = \emptyset$ and $\{M \in \mathbb{PSD}^n : \|M\|_{\ell_2^{n\times n}} = R\} \subseteq B$.

For a cone K, we say a function $f : K \to \mathbb{R}$ is (R, r, K)*-balanced* if (1) satisfies $r \cdot f(x) \leqslant \|x\|_2 \leqslant R \cdot f(x)$ for all $x \in K$, and (2) satisfies $\|f(x) - f(y)\| \leqslant R\|x - y\|_2$ for all $x, y \in K$. We say $f : \mathbb{R}^n \to R$ is (R, r)-balanced if it is $(R, r, \mathbb{R}^n)$-balanced.

Throughout this text we make the assumption in our algorithmic results that $\|\cdot\|_X$ is (R, r)-balanced. In this section we verify that such an assumption on $\|\cdot\|_X$ implies that the associated sets/functions in Section 2.5 also satisfy appropriate versions of balance, which is required in order to optimize over them.

Lemma 2.24. *Let $X = (\mathbb{R}^n, \|\cdot\|_X)\,, Y = (\mathbb{R}^m, \|\cdot\|_Y)$ be (R,r)-balanced. Then there exist $R', 1/r' =$ $\mathrm{poly}(n, R, 1/r)$ and $R'', 1/r'' = \mathrm{poly}(n, m, R, 1/r)$ such that*
(1) $\mathrm{Ball}(X \widehat{\otimes} Y)$, $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ and $\mathrm{Ball}(\mathrm{Op}^{\max}_{X,Y}(\cdot))$ are (R'', r'')-balanced.
(2) $\mathcal{U}(X)\,, B^{Sym}_{\wedge}(X)\,, \downarrow B^{Sym}_{\wedge}(X)$ and $\mathrm{Q}^{\max}_X(\cdot)$ are $(R', r', \mathbb{PSD}^n)$-balanced.
(3) $\mathcal{L}(X)\,, B^{\uparrow}_{\wedge}(X)$ are inverse-$(R', r', \mathbb{PSD}^n)$-balanced.

Proof. We begin by showing (1). Recall the $\ell_2^m \to \ell_2^n$-operator norm can be written as

$$\|M\|_{2\to 2} = \max_{\|x\|_2\,,\|y\|_2 \leqslant 1} \langle x\,, My\rangle$$

and combining with the balance condition on X, Y, we obtain

$$r^2 \cdot \|M\|_{2\to 2} \leqslant \mathrm{Op}^{\max}_{X,Y}(M) \leqslant R^2 \cdot \|M\|_{2\to 2}\,.$$

Now since $\|M\|_{2\to 2} \leqslant \|M\|_{\ell_2^{n\times m}} \leqslant \max\{\sqrt{n}, \sqrt{m}\} \cdot \|M\|_{2\to 2}$, it follows that $\{A \in M_{n,m}(\mathbb{R}) \mid \mathrm{Op}^{\max}_{X,Y}(A) \leqslant 1\} = \mathrm{Ball}(\mathrm{Op}^{\max}_{X,Y}(\cdot))$ is $(\max\{\sqrt{n}/r^2, \sqrt{m}/r^2\}, 1/R^2)$-balanced. By polarity, we conclude $\mathrm{Ball}(X \widehat{\otimes} Y)$ is $(R^2, \min\{r^2/\sqrt{n}, r^2/\sqrt{m}\})$-balanced. To show balance of $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ is balanced (as a function), we are left with checking Lipschitzness. By triangle inequality we have $|\mathrm{Op}^{\max}_{X,Y}(A) - \mathrm{Op}^{\max}_{X,Y}(B)| \leqslant \mathrm{Op}^{\max}_{X,Y}(A-B) \leqslant R^2 \cdot \|A-B\|_{\ell_2^{n\times m}}$. Thus $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ is $(\mathrm{poly}(n, R, 1/r), 1/\mathrm{poly}(n, R, 1/r))$-balanced.

We now show (2). From (1) we conclude that $\mathrm{Ball}(X \widehat{\otimes} X) \cap \mathbb{PSD}^n$ is $(R^2, \min\{r^2/\sqrt{n}, r^2/\sqrt{m}\}, \mathbb{PSD}^n)$-balanced. Combining with the equivalences in Observation 2.17 and the fact that $\widetilde{T}_2(X) \leqslant \sqrt{n}$ implies that $\mathcal{U}(X)\,, B^{Sym}_{\wedge}(X)\,, \downarrow B^{Sym}_{\wedge}(X)$ are $(\mathrm{poly}(n, R, 1/r), 1/\mathrm{poly}(n, R, 1/r), \mathbb{PSD}^n)$-balanced. It remains to show Lipschitzness of $\mathrm{Q}^{\max}_X(\cdot)$ restricted to $\mathbb{PSD}^n$, but this follows from Lipschitznes of $\mathrm{Op}^{\max}_{X,X}(\cdot)$ since for any $W \succeq 0$, we have $\mathrm{Q}^{\max}_X(W) = \mathrm{Op}^{\max}_{X,X}(W)$. Thus $\mathrm{Q}^{\max}_X(\cdot)$ is $(\mathrm{poly}(n, R, 1/r), 1/\mathrm{poly}(n, R, 1/r), \mathbb{PSD}^n)$-balanced.

Finally we establish (3). We first show $B^{\uparrow}_{\wedge}(X)$ is inverse-$(\mathrm{poly}(n, R, 1/r), 1/\mathrm{poly}(n, R, 1/r), \mathbb{PSD}^n)$-balanced. Then by the equivalence in Observation 2.20 and the fact that $\widetilde{C}_2(X^*) \leqslant \sqrt{n}$, we also obtain that $\mathcal{L}(X)$ is inverse-$(\mathrm{poly}(n, R, 1/r), 1/\mathrm{poly}(n, R, 1/r), \mathbb{PSD}^n)$-balanced. To this end, observe that for any $W \succeq 0$ and any finite decomposition $W = \sum_i w_i w_i^*$, we have

$$\mathbf{Tr}(W) = \sum_i \|w_i\|_2^2\,.$$

From this as well as the balance of X^* (which follows from balance of X) we conclude that $\wedge^{\uparrow}_X(W)$ and $\mathbf{Tr}(W)$ are equivalent within $\mathrm{poly}(R, 1/r)$. Now since $\mathbf{Tr}(W)$ and $\|W\|_{\ell_2^{n\times n}}$ are equivalent within $\sqrt{n}$ (assuming $W \succeq 0$), we conclude that $\wedge^{\uparrow}_X(W)$ and $\|W\|_{\ell_2^{n\times n}}$ are equivalent within $\mathrm{poly}(n, R, 1/r)$ which immediately implies the desired inverse-balance of $B^{\uparrow}_{\wedge}(X)$. This completes the proof. ■

Remark 2.25. *In Section 3, we use an approximate version of the ellipsoid method to give multiplicative approximation algorithms for various tasks (Proposition 3.11, Proposition 3.12, Theorem 3.14, Theorem 3.16), provided the appropriate balance conditions are satisfied. In all the settings in this text wherein we apply the above theorems, the associated balance conditions are satisfied as a consequence of Lemma 2.24.*

2.6 Oracle Algorithms

In this work, an oracle algorithm denoted $\mathrm{ALG}(\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{O}_1, \mathcal{O}_2, \ldots)$ running in time $T = T(\mathrm{size}(\mathcal{I}_1) + \mathrm{size}(\mathcal{I}_2) + \ldots)$ is an algorithm that takes in a finite set of inputs $\mathcal{I}_1, \mathcal{I}_2, \ldots$ and a finite set of oracles $\mathcal{O}_1, \mathcal{O}_2, \ldots$, and terminates with an output after T steps, where at any given step the algorithm is allowed to make black-box queries to any of the oracles and this counts as a single step in the runtime. For a fixed oracle algorithm, the set of queries made to the oracles is determined completely by the input $\mathcal{I}_1, \mathcal{I}_2, \ldots$ (and in the case of randomized algorithms, the distribution of the set of queries is determined entirely by the input).

We define below some oracles making a repeated appearance in the sequel.

Definition 2.26 (Approximate Optimization Oracles.)**.**

1. *For a norm* $\|\cdot\|_X$ *over* $\mathbb{R}^n$ *an* α-approximate search oracle for quadratic maximization over X *is an oracle that on any input* $A \in M_n(\mathbb{R})$ *outputs a vector* $x \in \mathrm{Ball}(X)$ *satisfying* $\langle x, Ax\rangle \geqslant \mathrm{Q}_X^{\max}(A)/\alpha$.
2. *For a norm* $\|\cdot\|_X$ *over* $\mathbb{R}^n$ *an* α-approximate search oracle for PSD quadratic maximization over X *is an oracle that on any PSD input* $A \succeq 0$ *outputs a vector* $x \in \mathrm{Ball}(X)$ *satisfying* $\langle x, Ax\rangle \geqslant \mathrm{Q}_X^{\max}(A)/\alpha$.
3. *For norms* $(\mathbb{R}^n, \|\cdot\|_X), (\mathbb{R}^m, \|\cdot\|_Y)$*, an* α-approximate search oracle for bilinear maximization over X, Y *is an oracle that on any input* $A \in M_{n,m}(\mathbb{R})$ *outputs vectors* $x \in \mathrm{Ball}(X), y \in \mathrm{Ball}(Y)$ *satisfying* $\langle x, Ay\rangle \geqslant \mathrm{Op}_{X,Y}^{\max}(A)/\alpha$.

3 Approximate Convex Optimization

In this section, we provide optimization tools used in the paper. Intuitively, we show that if a set is "approximately convex" and there is an "approximate separation oracle," the ellipsoid algorithm can be used to approximately minimize a convex function over the set.

3.1 Convex Optimization with an Approximate Separation Oracle

We start with our definition of approximate convexity. Recall that a set $B \subseteq \mathbb{R}^n$ is called *star-shaped with respect to the origin* when $[0,1]B \subseteq B$. We say that $B \subseteq \mathbb{R}^n$ is *inverse star-shaped with respect to the origin* when $[1,\infty)B \subseteq B$, which is equivalent to that $\mathbb{R}^n \setminus B$ is star-shaped with respect to the origin.

Definition 3.1 (Approximately Convex Body)**.**

1. *Consider a body* $B \subseteq \mathbb{R}^n$ *star-shaped with respect to the origin. For* $\alpha \geqslant 1$*, we shall say* B *is* α-approximately convex *if every convex combination of points in* B *is contained in* $\alpha \cdot B$*. Equivalently we have the inclusions* $B \subseteq \mathbf{conv}(B) \subseteq \alpha \cdot B$.

2. *Consider a set* $B \subseteq \mathbb{R}^n$ *inverse star-shaped with respect to the origin. For* $\alpha \geqslant 1$*, we shall say* B *is* α-inverse approximately convex *if every convex combination of points in* B *is contained in* B/α*. Equivalently we have the inclusions* $B \subseteq \mathbf{conv}(B) \subseteq B/\alpha$.

Since $\mathbf{conv}(B)$ is contained in $\alpha \cdot B$ (case 1) or B/α (case 2) in the above definition, the existence of a separating hyperplane for a convex set implies the following.

Definition 3.2 (Approximately Separable Body)**.**

1. *Let $B \subseteq \mathbb{R}^n$ be a body star-shaped with respect to the origin. For $\alpha \geqslant 1$, we shall say B is α-*(forward) separable *if whenever x is outside $\alpha \cdot B$, there exists a hyperplane separating x from B.*

2. *Let $B \subseteq \mathbb{R}^n$ be a set inverse star-shaped with respect to the origin. For $\alpha \geqslant 1$, we shall say B is α-*(inversely) separable *if whenever x is outside B/α, there exists a hyperplane separating x from B.*

   Henceforth we drop the term "forward" or "inversely" whenever it is clear which definition is being referred to.

Remark: α-approximate convexity implies α-separability.

Next, we define *approximate separation oracles* that return approximately separating hyperplanes.

Definition 3.3 (Approximate Separation Oracle)**.**

1. *For $\alpha \geqslant 1$, an α-*approximate separation oracle *for an α-separable body $B \subseteq \mathbb{R}^n$, is an oracle that on any input point $x \in \mathbb{R}^n$, either correctly outputs "Inside" when $x \in \alpha \cdot B$ or outputs a hyperplane separating x from B.*

   Remark: In any ambiguous case, i.e., when $x \notin B$ and $x \in \alpha \cdot B$, the oracle is allowed to either output a hyperplane separating x from B or output "Inside".

2. *For $\alpha \geqslant 1$, an α-*approximate (inverse) separation oracle *for an α-(inversely) separable set $B \subseteq \mathbb{R}^n$, is an oracle that on any input point $x \in \mathbb{R}^n$, either correctly outputs "Inside" when $x \in B/\alpha$ or outputs a hyperplane separating x from B.*

   Remark: In any ambiguous case, i.e., when $x \notin B$ and $x \in B/\alpha$, the oracle is allowed to either output a hyperplane separating x from B or output "Inside".

   Henceforth we drop the term "inverse" whenever it is clear which definition is being referred to.

The following simple observation will be useful for establishing approximate separation oracles in various contexts.

Observation 3.4 (Oracle from Equivalence)**.** *Consider bodies $B_1, B_2 \subseteq \mathbb{R}^n$ satisfying $B_2 \subseteq B_1 \subseteq \alpha \cdot B_2$ and let $\mathcal{SO}$ be a β-approximate separation oracle for B_1. Then $\mathcal{SO}$ is also an $\alpha\beta$-approximate separation oracle for B_2.*

Proof. Indeed if $\mathcal{SO}(x)$ returns "Inside", then $x \in \alpha \cdot B \Rightarrow x \in \alpha\beta \cdot B_2$. On the other hand if $\mathcal{SO}(x)$ returns a hyperplane separating B_1 from x, it of course separates B_2 from x as desired.

The claim for approximate inverse separation oracles follows analogously. ∎

A standard simplification step in convex optimization is to use binary search to reduce the problem to the following decision task: given (exact) separation oracles for two convex bodies, decide whether their intersection is empty. The ellipsoid method is a popular method for solving this subroutine. We use a variant of the ellipsoid method to show that given approximate separation oracles (forward and inverse) for a set, we can efficiently decide whether the intersection of two bodies is close to empty.

Proposition 3.5 (Folklore: Approximate Ellipsoid Method)**.**
Let C_1 have an α_1-approximate separation oracle $\mathcal{SO}_1$, C_2 have an α_2-approximate inverse separation oracle $\mathcal{SO}_2$. Additionally, assume that C_1 is contained in a (Euclidean) ball of radius R. There is an algorithm $\mathrm{ALG}(R, \varepsilon, \mathcal{SO}_1, \mathcal{SO}_2)$ *running in time* $\mathrm{poly}(n, \log R, \log(1/\varepsilon))$ *returning one of the following:*

1. *A point x inside $(\alpha_1 C_1) \cap (C_2/\alpha_2)$.*
2. *$C_1 \cap C_2$ does not contain a (Euclidean) ball of radius ε.*

Proof. We follow the ellipsoid algorithm. Start from an ellipsoid containing the ball of radius R containing $C_1 \cap C_2$. If both oracles return "Inside" when queried on the center x, then we know x is in $(\alpha_1 C_1) \cap (C_2/\alpha_2)$. Otherwise, by definition of the separation oracles, we have for some $i \in [2]$, a hyperplane separating x and C_i. Lemma (3.2.10) of [GLS93] shows that one can compute a new ellipsoid of volume $e^{-1/5n}$ times that of the original ellipsoid, while still maintaining containment of $C_1 \cap C_2$. If we continue this process T times without terminating and T satisfies

$$(\varepsilon/R)^n > e^{-T/5n} \Leftrightarrow T > 5n^2 \log(R/\varepsilon)\,,$$

then the volume of the T-th ellipsoid (and therefore the volume of $C_1 \cap C_2$) is smaller than that of the Euclidean ball of radius ε.

We further note that Lemma 3.2.10 of [GLS93] ensures that each ellipsoid $\{y : \langle y - x\,, A^{-1}(y - x)\rangle \leqslant 1\}$ involved in the algorithm is described by a center x and a PSD matrix A of bit complexities at most $p = \mathrm{poly}(n, \log R, \log(1/\varepsilon))$, where the result of each computation (possibly involving irrational numbers) is rounded down to the closest integer multiple of 2^{-p}. ■

We require some additional notation before proceeding. For a set $K \subseteq \mathbb{R}^n$ and $\varepsilon > 0$, let $K_\varepsilon^+ \stackrel{\mathrm{def}}{=} K + \varepsilon \cdot \mathrm{Ball}(\ell_2^n)$ and $K_\varepsilon^- \stackrel{\mathrm{def}}{=} \{x : x + \varepsilon \cdot \mathrm{Ball}(\ell_2^n) \subseteq K\}$, where $+$ denotes the minkowski sum. We may now combine binary search with Proposition 3.5 to approximately maximize a concave function over a body with a (forward) approximate separation oracle.

Proposition 3.6 (Concave Maximization with an Approximate Separation Oracle)**.**
Let C_1 be a closed set satisfying $x \in C_1 \Rightarrow \alpha x \in C_1$ for all $\alpha \in [0,1]$, with an α_1-approximate separation oracle $\mathcal{SO}_1$. Let f be a homogeneous concave function such that $f_{\geqslant\lambda} \stackrel{\mathrm{def}}{=} \{x : f(x) \geqslant \lambda\}$ has an α_2-approximate inverse separation oracle $\mathcal{SO}_2$ and $f_{\geqslant\lambda} \cap C_1 \subseteq R \cdot \mathrm{Ball}(\ell_2^n)$, for any $\lambda > 0$. Let $\mathsf{OPT} = \sup_{x \in C_1} f(x)$, and assume that $\Lambda > \mathsf{OPT}$ is given. Let $\varepsilon > 0$ and $\mathsf{OPT}_\varepsilon = \sup_{x \in (C_1)_\varepsilon^-} f(x)$. There is an algorithm $\mathrm{ALG}(R, \Lambda, \varepsilon, \mathcal{SO}_1, \mathcal{SO}_2)$ *that, in time* $\mathrm{poly}(n, \log R, \log \Lambda, \log(1/\varepsilon))$*, returns $y \in C_1$ that satisfies*

$$f(y) \geqslant (\mathsf{OPT}_\varepsilon - \varepsilon)/(\alpha_1\alpha_2).$$

Proof. Starting from $\lambda = \Lambda$, we apply Proposition 3.5 ($C_1 \leftarrow C_1$ and $C_2 \leftarrow f_{\geqslant\lambda}$), and perform binary search. In $\mathrm{poly}(\log \Lambda, \log(1/\varepsilon))$ steps, we can find $\lambda \geqslant 0$ such that

1. There is a point $x \in (\alpha_1 \cdot C_1) \cap (f_{\geqslant\lambda}/\alpha_2)$.
2. $C_1 \cap f_{\geqslant\lambda+\varepsilon}$ does not contain a ball of radius ε.

(The first item is satisfied by $\lambda = 0$, and the second item is satisfied by $\lambda = \Lambda$.) The second item implies $\mathsf{OPT}_\varepsilon \leqslant \lambda + \varepsilon$. Therefore, $y = x/\alpha_1$ satisfies $y \in C_1$ and

$$f(y) = f(x)/\alpha_1 \geqslant \lambda/(\alpha_1\alpha_2) \geqslant (\mathsf{OPT}_\varepsilon - \varepsilon)/(\alpha_1\alpha_2),$$

which proves the claim. ■

Similarly we can approximately minimize a concave function over a set with an (inverse) approximate separation oracle.

Proposition 3.7 (Convex Minimization with an Approximate Separation Oracle)**.**
Let C_1 be a closed set with an α_1-approximate inverse separation oracle $\mathcal{SO}_1$. Let f be a nonnegative, homogeneous, and convex function such that $f_{\leqslant\lambda} \overset{\text{def}}{=} \{x : f(x) \leqslant \lambda\}$ has an α_2-approximate separation oracle $\mathcal{SO}_2$ for any $\lambda > 0$. Assume $\Lambda >$ OPT and $R > 0$ are given such that OPT $= \inf_{x\in C_1} f(x) = \inf_{x\in C_1\cap R\cdot \text{Ball}(\ell_2^n)} f(x)$. Let $\varepsilon > 0$ and $\mathsf{OPT}_\varepsilon = \inf_{x\in (C_1\cap R\cdot\text{Ball}(\ell_2^n))_\varepsilon^-} f(x)$. There is an algorithm $\text{ALG}(R,\Lambda,\varepsilon,\mathcal{SO}_1,\mathcal{SO}_2)$ that, in time $\text{poly}(n,\log R,\log\Lambda,\log(1/\varepsilon))$, returns $y \in C_1$ that satisfies

$$f(y) \leqslant (\alpha_1\alpha_2)\cdot(\mathsf{OPT}_\varepsilon + \varepsilon).$$

Proof. Starting from $\lambda = \Lambda$, we apply Proposition 3.5 ($C_1 \cap R\cdot\text{Ball}(\ell_2^n) \leftarrow f_{\leqslant\lambda}$ and $C_2 \leftarrow C_1$), and perform binary search. In $\text{poly}(\log\Lambda, \log(1/\varepsilon))$ steps, we can find $\lambda \geqslant 0$ such that

- There is a point inside $x \in (\alpha_2 \cdot f_{\leqslant\lambda}) \cap (C_1/\alpha_1)$.
- $(f_{\leqslant\lambda-\varepsilon} \cap R\cdot\text{Ball}(\ell_2^n)\cap C_1)$ does not contain a ball of radius ε.

(The first item is satisfied by $\lambda = \Lambda$, and the second item is satisfied by $\lambda = 0$.) The second item implies $\mathsf{OPT}_\varepsilon \geqslant \lambda - \varepsilon$. Therefore, $y = \alpha_1 x$ satisfies $y \in C_1$ and

$$f(y) = \alpha_1 f(x) \leqslant \alpha_1\alpha_2\lambda \leqslant (\alpha_1\alpha_2)\cdot(\mathsf{OPT}_\varepsilon + \varepsilon),$$

which proves the claim. ∎

A subtle issue is that we require versions of Proposition 3.6 and Proposition 3.7 with multiplicative approximation guarantees. For this it is necessary to assume C_1 and f have additional properties, which is in contrast to the exact setting where such assumptions are not required. In our work we only consider sets that are contained in some self-dual cone with the following additional properties. For a vector b, let $\text{bit}(b)$ be its bit complexity.

Definition 3.8 (Tractable Cone)**.** *We will say a closed, self-dual cone $K \subseteq \mathbb{R}^n$ is* tractable *if the following properties hold:*

(T1) Given $y \notin K$, there is a polynomial time algorithm to find a hyperplane separating y and K.

(T2) Given $y \in K$ that admits $x \in K$ with $\langle x, y\rangle = 0$, there is a polynomial time algorithm to find such a vector x.

(T3) If $y \in K$ does not admit such a x, $\min_{x\in K,\|x\|_2=1}\langle x, y\rangle > 2^{-\text{poly}(\text{bit}(y))}$.

(T4) For any $y \in \mathbb{R}^n$, if $\max_{x\in K,\|x\|_2\leqslant 1}\langle x, y\rangle \neq 0$, then it is at least $2^{-\text{poly}(\text{bit}(y))}$.

(T5) For any $\varepsilon > 0$, there exists $y \in K$ with $\|y\|_2 \leqslant \text{poly}(n)\cdot\varepsilon$ such that $y + \varepsilon\cdot\text{Ball}(\ell_2^n) \subseteq K$.

We will only consider cases wherein K is the positive orthant or the positive semidefinite cone.

Definition 3.9 (Balanced Set/Function)**.**
Let K be a tractable cone or $\mathbb{R}^n$. We say a set $B \subseteq \mathbb{R}^n$ is (R,r,K)-balanced *if $(K\cap r\cdot\text{Ball}(\ell_2^n)) \subseteq B \subseteq (K\cap R\cdot\text{Ball}(\ell_2^n))$. When K is a tractable cone, we also say that a set $B \subseteq K$ is* inverse (R,r,K)-balanced *if $r\cdot\text{Ball}(\ell_2^n)\cap B = \emptyset$ and $\{x \in K : \|x\|_2 = R\} \subseteq B$. We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is* (R,r,K)-balanced *if (1) satisfies $r\cdot f(x) \leqslant \|x\|_2 \leqslant R\cdot f(x)$ for all $x \in K$, and (2) satisfies $\|f(x) - f(y)\| \leqslant R\|x-y\|_2$ for all $x, y \in K$.*

We can deduce the following property which states that if C is balanced, C_ε^- is not too much smaller than C.

Lemma 3.10. *Let $K \subseteq \mathbb{R}^n$ be a tractable cone or $\mathbb{R}^n$ itself. Let B be a closed, convex, and (R,r,K)-balanced set. For any $x \in B$ and $\varepsilon \in (0, r/\mathrm{poly}(n))$, there exists $y \in B$ such that $\|x-y\|_2 \leqslant \varepsilon$ and $y + \frac{r\varepsilon}{R\cdot\mathrm{poly}(n)} \cdot \mathrm{Ball}(\ell_2^n) \subseteq B$.*

Proof. By property (T5) of K and the fact that $(K \cap r\cdot\mathrm{Ball}(\ell_2^n)) \subseteq B$, for some $\alpha = r/\mathrm{poly}(n)$, there exists $z \in K$ such that $z + \alpha\cdot\mathrm{Ball}(\ell_2^n) \subseteq B$ and $\|z\|_2 \leqslant r/2$. (If $K = \mathbb{R}^n$, we can take $z = 0$.)

Fix $x \in B$ and let $T := \{t \in z + \alpha\cdot\mathrm{Ball}(\ell_2^n) : \|z-t\|_2 = \alpha, \langle z-x, z-t\rangle = 0\}$, and consider $\mathbf{conv}(x \cup T) \subseteq B$. Let θ be the angle $\angle zxt$ for some $t \in T$, which does not depend on the choice of t. Since $\|t-z\|_2 = \alpha$ and $\|x-t\|_2 \leqslant \|x-z\|_2 + \|z-t\| \leqslant R + r/2 + \alpha \leqslant 2R$, $\sin\theta \geqslant \frac{\alpha}{2R}$. Let y be the point on $\overline{xz}$ which is at distance ε from x. Then the distance from y to the boundary of $\mathbf{conv}(x \cup T)$ is at least $\varepsilon\cdot\sin\theta \geqslant \frac{\varepsilon\alpha}{2R}$. Therefore, $y + \frac{\varepsilon\alpha}{2R}\cdot\mathrm{Ball}(\ell_2^n) = y + \frac{r\varepsilon}{R\cdot\mathrm{poly}(n)}\cdot\mathrm{Ball}(\ell_2^n)$ is contained in $\mathbf{conv}(x \cup T) \subseteq B$. ■

We now prove a multiplicative approximation version of Proposition 3.6.

Proposition 3.11 (Concave Maximization with an Approximate Separation Oracle)**.**
Let K be a tractable cone or $\mathbb{R}^n$ and let C_1 be an (R,r,K)-balanced set with an α_1-approximate separation oracle $\mathcal{SO}_1$. Let f be either a linear function $\langle \xi, x\rangle$ for some $\xi \in \mathbb{R}^n$ (in which case $\alpha_2 = 1$) or an (R,r,K)-balanced, nonnegative, homogeneous, and concave function such that $f_{\geqslant\lambda} \stackrel{\mathrm{def}}{=} \{x : f(x) \geqslant \lambda\}$ has an α_2-approximate inverse separation oracle $\mathcal{SO}_2$ for any $\lambda > 0$. Let $\mathsf{OPT} = \sup_{x\in C_1} f(x)$. There is an algorithm $\mathrm{ALG}(R, r, \mathcal{SO}_1, \mathcal{SO}_2)$ that, in time $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(\xi))$, returns $y \in C_1$ that satisfies

$$f(y) \geqslant \frac{(1-1/n)}{\alpha_1\alpha_2}\cdot\mathsf{OPT}.$$

Proof. If f is a linear function, since we try to get a multiplicative approximation, assume without loss of generality that $f(x) = \langle\xi, x\rangle$ for some ξ with $\|\xi\|_2 = 1$. We first find an upper bound on OPT. By (R,r,K)-balancedness of C_1, whether f is a linear or (R,r,K)-balanced,

$$\mathsf{OPT} \leqslant \sup_{x\in C_1} f(x) \leqslant \max(\|x\|_2/r, \|x\|_2) \leqslant R\max(1, 1/r) =: \mathsf{OPT}_{\max}$$

We also find a lower bound on OPT. If f is (R,r,K)-balanced,

$$\mathsf{OPT} = \sup_{x\in C_1} f(x) \geqslant \sup_{x\in K\cap r\cdot\mathrm{Ball}(\ell_2^n)} f(x) \geqslant \frac{r}{R} =: \mathsf{OPT}_{\min}.$$

If f is linear, by property (T4) of K, unless $\mathsf{OPT} = 0$ (in which case we can just return 0 as the optimal solution),

$$\mathsf{OPT} \geqslant \sup_{x\in K\cap r\cdot\mathrm{Ball}(\ell_2^n)} \langle\xi, x\rangle = r\cdot\sup_{x\in K\cap\mathrm{Ball}(\ell_2^n)}\langle\xi, x\rangle \geqslant r\cdot 2^{-\mathrm{poly}(\mathrm{bit}(\xi))} =: \mathsf{OPT}_{\min}$$

Finally, we bound the difference between OPT_ε and OPT for small enough $\varepsilon > 0$. Let $x^* \in \mathrm{argmax}_{x\in C_1}\langle\xi, x\rangle$. By Lemma 3.10, For any $\varepsilon \in (0, r/\mathrm{poly}(n))$, there exists y^* such that $y \in (C_1)_\varepsilon^-$

and $\|x^* - y^*\|_2 \leqslant \varepsilon \cdot p(n)$ for a fixed polynomial $p(n)$. Since f is Lipschitz with the constant $\max(R, 1)$,

$$\mathsf{OPT}_\varepsilon := \sup_{y \in (C_1)_\varepsilon^-} f(y) \geqslant f(y^*) \geqslant f(x^*) - \varepsilon \cdot p(n) \cdot \max(R, 1) = \mathsf{OPT} - \varepsilon \cdot p(n) \cdot \max(R, 1).$$

We apply Proposition 3.6 with $C_1 \leftarrow C_1$, $f \leftarrow f$, $\Lambda \leftarrow \mathsf{OPT}_{\max}$, $R \leftarrow R$, and $\varepsilon \leftarrow \mathsf{OPT}_{\min}/(2np(n)\max(R,1))$. The running time is $\mathrm{poly}(n, \log R, \log \Lambda, \log(1/\varepsilon)) = \mathrm{poly}(n, \log R, \log(1/r), \mathrm{bit}(\xi))$ and we find $x \in C_1$ that satisfies

$$f(x) \geqslant \alpha_1\alpha_2(\mathsf{OPT}_\varepsilon - \varepsilon) \geqslant \alpha_1\alpha_2\Big(\mathsf{OPT} - \varepsilon \cdot 2p(n)\max(R,1)\Big) \geqslant \alpha_1\alpha_2\Big(1 - \frac{1}{n}\Big)\mathsf{OPT},$$

which proves the claim. ∎

We also give a multiplicative approximation version of Proposition 3.7.

Proposition 3.12 (Convex Minimization with an Approximate Separation Oracle)**.** *Let K be a tractable cone. Let $C_1 \subseteq K$ is a closed, upward-closed, and (R, r, K)-inversed balanced set with an α_1-approximate inverse separation oracle $\mathcal{SO}_1$. Let $f(x) = \langle \xi, x \rangle$ be a linear function for some $\xi \in K$ (in which case $\alpha_2 = 1$) or an (R, r, K)-balanced, nonnegative, homogeneous, and convex function such that $f_{\leqslant \lambda} \stackrel{\mathrm{def}}{=} \{x : f(x) \leqslant \lambda\}$ has an α_2-approximate separation oracle $\mathcal{SO}_2$ for any $\lambda > 0$. Let $\mathsf{OPT} = \inf_{x \in C_1} f(x)$. There is an algorithm $\mathrm{ALG}(R, r, \mathcal{SO}_1, \mathcal{SO}_2)$ that, in time $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(\xi))$, returns $y \in C_1$ that satisfies*

$$f(y) \leqslant \mathsf{OPT} \cdot \alpha_1\alpha_2(1 + 1/n).$$

Proof. We first prove upper and lower bounds on OPT. When f is (R, r, K)-balanced, since C_1 is also (R, r, K)-inversed balanced, $\mathsf{OPT} \in [r^2, R^2]$.

Now we consider linear $f(x) = \langle \xi, x \rangle$. Assume without loss of generality that $\|\xi\|_2 = 1$. The inverse (R, r, K)-balancedness of C_1 implies

$$\mathsf{OPT} := \inf_{x \in C_1} \langle \xi, x \rangle = \min_{x \in C_1 : \|x\| \in [r, R]} \langle \xi, x \rangle.$$

If the optimum is 0, there exists $x \in K$ such that $\langle x, \xi \rangle = 0$. Such x can be found by property (T2) of K, and $Rx/\|x\|_2 \in C_1$ is also an optimal solution. Otherwise, the property (T3) of K implies that

$$\mathsf{OPT} \geqslant \min_{x \in K : \|x\|_2 = r} \langle \xi, x \rangle > \delta = 2^{-\mathrm{poly}(\mathrm{bit}(\xi))} \cdot r.$$

For an upper bound, since C_1 contains a vector of length R, the optimum is also upper bounded by R.

Let $\mathsf{OPT}_{\min} = \min(2^{-\mathrm{poly}(\mathrm{bit}(\xi))} \cdot r, r^2), \mathsf{OPT}_{\max} := \max(R, R^2)$ be the lower and upper bound on OPT whether OPT is linear or (R, r, k)-balanced.

Finally, we bound the difference between OPT_ε and OPT for small enough $\varepsilon > 0$. For any $x^* \in \mathrm{argmin}_{x \in C_1} f(x)$ and $\varepsilon \in (0, R/4)$, the property (T5) of K and the fact that C_1 is upward-closed imply that there exists y^* such that $y^* + \varepsilon \cdot \mathrm{Ball}(\ell_2^n) \in C_1$ and $\|x^* - y^*\|_2 \leqslant \varepsilon \cdot p(n)$ for a fixed polynomial $p(n)$. Furthermore, both x^* and y^* have ℓ_2 norm at most $R + 2\varepsilon < 1.5R$. Since f is Lipschitz with the constant $\max(R, 1)$,

$$\mathsf{OPT}_\varepsilon := \min_{y \in (C_1 \cap 2R \cdot \mathrm{Ball}(\ell_2^n))_\varepsilon^-} f(y) \leqslant f(y^*) \leqslant f(x^*) + \varepsilon \cdot p(n) \cdot \max(R, 1).$$

We apply Proposition 3.7 with with $C_1 \leftarrow C_1$, $f(x) \leftarrow f(x)$, $\Lambda \leftarrow \mathsf{OPT}_{\max}$, $R \leftarrow 2R$, and $\varepsilon \leftarrow \frac{\mathsf{OPT}_{\min}}{2n \cdot p(n) \cdot \max(R,1)}$. The running time is $\mathrm{poly}(n, \log R, \log \Lambda, \log(1/\varepsilon)) = \mathrm{poly}(n,\ \mathrm{bit}(\xi),\ \log R,$ $\log(1/r))$ and we find x that satisfies

$$f(x) \leqslant \alpha_1\alpha_2(\mathsf{OPT}_\varepsilon + \varepsilon) \leqslant \alpha_1\alpha_2(\mathsf{OPT} + \varepsilon \cdot p(n) + \varepsilon) \leqslant \alpha_1\alpha_2(1 + 1/n)\mathsf{OPT},$$

which proves the claim. ∎

3.2 Duality of Approximation Algorithms for Linear Function Optimization

Systematic use of duality is fundamental in the metric theory of tensor products. Since we are interested in approximation algorithms for the injective tensor norm (read. bilinear maximization), we end up making extensive use of algorithmic versions of duality. Such results however, can fail drastically in the approximate optimization setting [SV15]. Nevertheless we show that in the special case of optimization over downward/upward-closed subsets of a self-dual cone, certain approximate algorithmic duality statements hold true.

We establish in this subsection a partial converse of Proposition 3.11 and Proposition 3.12, proving that an approximate linear function maximization algorithm for some downward/upward-closed convex subset B of a self-dual cone K, leads to an approximate separation oracle for B. In particular, we are interested in when K is $\mathbb{R}^n$, the nonnegative orthant, or the positive semidefinite cone.

We first observe that maximization over $B \subseteq K$ leads to separation for the polar of B. Here, two kinds of polars are covered — regular polars 2.3 when the underlying set K is $\mathbb{R}^n$, or conic polars 2.5 when the underlying set is a self-dual cone K.

Observation 3.13 (Maximization Oracle implies Polar Separation Oracle)**.**
Let $K \subseteq \mathbb{R}^n$ be a tractable cone or $\mathbb{R}^n$ itself. Let $B \subseteq K$ be a closed convex body and let $\mathcal{O}$ be an oracle that takes a vector $y \in K$ as input and returns a β-approximately optimal solution to $\sup_{x\in B}\langle y\,, x\rangle$. Then there is an oracle polytime β-approximate separation oracle $\mathcal{SO}$ for $B^{\circ K}$.

Proof. For an input vector $\xi \in \mathbb{R}^n$, if $\xi \notin K$, we use property (T1) of tractability to output a hyperplane (exactly) separating ξ and $B^{\circ K} \subseteq K$. For $\xi \in K$, the description of $\mathcal{SO}$ is as follows: if $\langle \mathcal{O}(\xi)\,, \xi\rangle \leqslant 1$ return "Inside" else return $\{y \mid \langle \mathcal{O}(\xi)\,, y\rangle = 1\}$ as the hyperplane separating ξ from $B^{\circ K}$.

Indeed if $\mathcal{O}$ returns x such that $\langle x\,, \xi\rangle \leqslant 1$, it is valid to return "Inside" since $\sup_{x\in B}\langle x\,, \xi\rangle \leqslant \beta$ and therefore $\xi \in \beta \cdot B^{\circ K}$. On the other hand if $\mathcal{O}$ returns $x \in B$ so that $\langle x\,, \xi\rangle > 1$, we know $\sup_{\overline{\xi}\in B^{\circ K}}\langle x\,, \overline{\xi}\rangle \leqslant 1$ by definition. Thus $\{y \mid \langle x\,, y\rangle = 1\}$ is a hyperplane separating ξ from $B^{\circ K}$. ∎

We now prove a partial converse of Proposition 3.11 for linear maximization over a set with a (forward) approximate separation oracle.

Theorem 3.14 (Duality of Approximate Linear Maximization)**.** *Let $K \subseteq \mathbb{R}^n$ be a tractable cone or $\mathbb{R}^n$ itself. Let $B \subseteq K$ be a closed, convex, and (R, r, K)-balanced set. If $K = \mathbb{R}^n$, additionally assume B is origin-symmetric. Consider any $\beta \geqslant 1$, and let $\mathcal{O}$ be an oracle that $y \in K$ as input and returns a β-approximately optimal solution to $\sup_{x\in B}\langle y, x\rangle$. Then*

(1) There is an algorithm $\mathrm{ALG}(R, r, \mathcal{O})$ that, given $\xi \in K$, returns a $(1 + o(1))\beta$-approximately optimal solution to $\sup_{x\in B^{\circ K}}\langle \xi\,, x\rangle$ in time $\mathrm{poly}(n, \mathrm{bit}(\xi), \log R, \log(1/r))$.

(2) *There is a* $(1+o(1))\beta$*-separation oracle* $\mathcal{SO}(R,r,\mathcal{O})$ *for* B *(if* $K=\mathbb{R}^n$*) or* $\downarrow B$ *(if* K *is a tractable cone) that, on input* ξ*, runs in time* $\mathrm{poly}(n, \mathrm{bit}(\xi), \log R, \log(1/r))$.

Proof. By Observation 3.13 we have a β-approximate separation oracle for $B^{\circ K}$. We use it to get a $(\beta+1/n)$-approximation algorithm for computing $\sup_{x\in B^{\circ K}}\langle \xi, x\rangle$ for any $\xi \in K$. Note that (R,r,K)-balancedness of B implies $(K,1/r,1/R)$-balancedness of $B^{\circ K}$. We apply Proposition 3.11 (with $C_1 \leftarrow B^{\circ K}$ and $f(x) \leftarrow \langle \xi, x\rangle$) to have a $(\beta+1/n)$-approximation algorithm for computing $\sup_{x\in B^{\circ K}}\langle \xi, x\rangle$ for any $\xi \in K$. so the running time becomes $\mathrm{poly}(n, \log \mathrm{bit}(\xi), \log R, \log(1/r))$ instead of polynomially depending on $\mathrm{bit}(\xi)$.

Finally we apply Observation 3.13 to $B^{\circ K}$ (and use Fact 2.6 for $(B^{\circ K})^{\circ K} = \downarrow B$ if K is a tractable cone or $(B^{\circ K})^{\circ K} = B$ if $K = \mathbb{R}^n$) to obtain the separation oracle claimed in (2). ∎

Now we study a similar phenomenon for minimization using inverse polars (Definition 2.7). Our goal is to show that when K is a self-dual cone and $B \subseteq K$ is closed, convex, and upward-closed, an algorithm to approximately minimize a linear function over B implies other oracles. We first check that multiplicative approximation is well defined. If B satisfies the above condition, $B_c^{\odot} \subseteq K$ for any $c \in \mathbb{R}$, implying that the optimal solution $\inf_{x\in B}\langle y, x\rangle = -\infty$ if $y \notin K$ and nonnegative if $y \in K$. Therefore, a β-approximate algorithm is well defined for the problem of computing $\inf_{x\in B}\langle y, x\rangle$; output $-\infty$ if $y \notin K$ and a β-multiplicative approximate solution if $y \in K$.

Observation 3.15 (Minimization Oracle implies Polar Separation Oracle)**.**
Let B *be a closed, convex, and upward-closed set contained in a tractable cone* $K \subseteq \mathbb{R}^n$*. Assume further that* $0 \notin B$*. Let* $\mathcal{O}$ *be an oracle that takes a vector* $\xi \in K$ *as input and returns a* β*-approximately optimal solution to* $\inf_{x\in B}\langle \xi, x\rangle$ *for some* $\beta \geqslant 1$*. Then there is an oracle polytime* β*-approximate inverse separation oracle* $\mathcal{SO}$ *for* $B^{\odot}$.

Proof. For an input vector $\xi \in \mathbb{R}^n$, the description of $\mathcal{SO}$ is as follows. Since $B^{\odot} \subseteq K$, given $\xi \notin K$, it outputs a hyperplane (exactly) separating ξ and K. For $\xi \in K$, the β-approximation algorithm for $\inf_{x\in B}\langle \xi, x\rangle$ returns $\mathcal{O}(\xi)$ such that $\langle \mathcal{O}(\xi), \xi\rangle \geqslant \mathsf{OPT}/\beta$. If $\langle \mathcal{O}(\xi), \xi\rangle \geqslant 1$ return "Inside" else return $\{y \mid \langle \mathcal{O}(\xi), y\rangle = 1\}$ as the hyperplane separating ξ from $B^{\odot}$.

Indeed if $\mathcal{O}$ returns x such that $\langle x, \xi\rangle \geqslant 1$, it is valid to return "Inside" since $\inf_{x\in B}\langle x, \xi\rangle \geqslant 1/\beta$ and therefore $\xi \in B^{\odot}/\beta$.

On the other hand if $\mathcal{O}$ returns $x \in B$ so that $\langle x, \xi\rangle < 1$, we know $\inf_{\bar{\xi}\in B^{\odot}}\langle x, \bar{\xi}\rangle \geqslant 1$ since by the Bipolar theorem, $(B^{\odot})^{\odot} = B$ and hence $x \in (B^{\odot})^{\odot}$. Thus $\{y \mid \langle x, y\rangle = 1\}$ is a hyperplane separating ξ from $B^{\odot}$. ∎

Finally, we prove a partial converse of Proposition 3.12 for linear minimization over a set with an (inverse) approximate separation oracle.

Theorem 3.16 (Duality of Approximate Linear Minimization)**.** *Let* B *be a closed, convex, and upward-closed set contained in a tractable cone* $K \subseteq \mathbb{R}^n$*. Assume that either* B *or* $B^{\odot}$ *is* (R,r,K)*-inverse balanced. Consider any* $\beta \geqslant 1$*, and let* $\mathcal{O}$ *be an oracle that takes* $\xi \in K$ *as input and returns a* β*-approximately optimal solution to* $\inf_{x\in B}\langle \xi, x\rangle$*. Then*

(1) *There is an algorithm* $\mathrm{ALG}(R,r,\mathcal{O})$ *that, given* $\xi \in K$*, returns a* $(1+o(1))\beta$*-approximately optimal solution to* $\inf_{x\in B^{\odot}}\langle \xi, x\rangle$ *in time* $\mathrm{poly}(n, \mathrm{bit}(\xi), \log R, \log(1/r))$.

(2) *There is a* $(1+o(1))\beta$*-separation oracle* $\mathcal{SO}(R,r,\mathcal{O})$ *for* B *that, on input* ξ*, runs in time* $\mathrm{poly}(n, \mathrm{bit}(\xi), \log R, \log(1/r))$.

Proof. By Observation 3.15 we have a β-approximate separation oracle for $B^{\odot}$. We apply Proposition 3.12 ($B \leftarrow B^{\odot}, f \leftarrow \langle \xi, \cdot \rangle$). Note that the (R, r, K)-balancedness of B implies that (1) $(1/R)\mathrm{Ball}(\ell_2^n)$ is disjoint from $B^{\odot}$ and (2) there exists $x \in B^{\odot}$ with $\|x\|_2 \leqslant \mathrm{poly}(n)/r$, and the above two conditions can replace $(\mathrm{poly}(n)/r, 1/R, K)$-balancedness of $B^{\odot}$ in the proof of Proposition 3.12; lower bounding OPT works verbatim as $(1/R)\mathrm{Ball}(\ell_2^n) \cap B^{\odot} = \emptyset$ and upper bounding OPT just requires one point of $x \in B^{\odot}$ with $\|x\|_2 \leqslant \mathrm{poly}(n)/r$. Therefore, we have a $(1 + 1/n)\beta$-approximation algorithm for computing $\inf_{x \in B^{\odot}} \langle \xi, x \rangle$ in time $\mathrm{poly}(n, \mathrm{bit}(\xi), \log R, \log(1/r))$. Finally we apply Observation 3.15 to $B^{\odot}$ (and use the fact that $(B^{\odot})^{\odot} = B$) to obtain the separation oracle claimed in (2). ∎

4 A Generic Framework: Algorithms from Covariance Separation Oracles, and Reductions across Quadratic/Bilinear/PSD Maximization

The main result of this section is a proof of Theorem 1.14 which provides generic framework for quadratic/bilinear maximization under bounded type-2/dual-cotype-2. We do this by developing a theory of polynomial time reductions (with multiplicative loss depending only on $\widetilde{T}_2(X)$ or $\widetilde{C}_2(X^*)$) between the following five oracles

($\mathcal{O}1$) Approximate Search Oracle for Quadratic Maximization.

($\mathcal{O}2$) Approximate Search Oracle for Bilinear Maximization.

($\mathcal{O}3$) Approximate Search Oracle for Quadratic Maximization of PSD instances.

($\mathcal{SO}1$) Approximate Separation Oracle for Upper Covariance Body.

($\mathcal{SO}2$) Approximate Separation Oracle for Lower Covariance Region.

See Fig. 1 for a pictorial depiction of our reductions. While all of our applications in Section 6 and Section 7 may be proved via our framework theorem some proofs do not require its full power. We choose to cast all proofs as an appropriate combination of reductions between the above oracles, which leads to a more distilled and cohesive presentation.

Our framework theorem can be deduced as a consequence of the reductions $(\mathcal{O}1) \to (\mathcal{O}3), (\mathcal{O}2) \to (\mathcal{O}3), (\mathcal{O}3) \to (\mathcal{SO}2)$. The reductions $(\mathcal{O}1) \to (\mathcal{O}3), (\mathcal{O}1) \to (\mathcal{SO}1), (\mathcal{SO}1) \to (\mathcal{O}1)$ will be used in Section 6 to derive closure properties for quadratic maximization under bounded Type-2. The reductions $(\mathcal{O}2) \to (\mathcal{O}3), (\mathcal{O}3) \to (\mathcal{O}2)$ will be used in Section 6 to derive closure properties for bilinear maximization under bounded dual Cotype-2. The reduction $(\mathcal{O}1) \to (\mathcal{SO}1)$ will be used in Section 7 to derive approximation algorithms for quadratic maximization over special families of norms with bounded Type-2. The reduction $(\mathcal{O}2) \to (\mathcal{SO}2)$ will be used in Section 7 to derive approximation algorithms for bilinear maximization over special families of norms with bounded dual Cotype-2.

All our reductions in this section follow from judicious application of one or more of the following three ingredients:

1. Approximate convex optimization under approximate separation oracles (Proposition 3.6, Proposition 3.7).

2. Algorithmic-duality/bipolarity for an appropriate downward/upward-closed subset of a self-dual cone (Theorem 3.14, Theorem 3.16).

3. An appropriate Rounding algorithm based on sampling Gaussians.

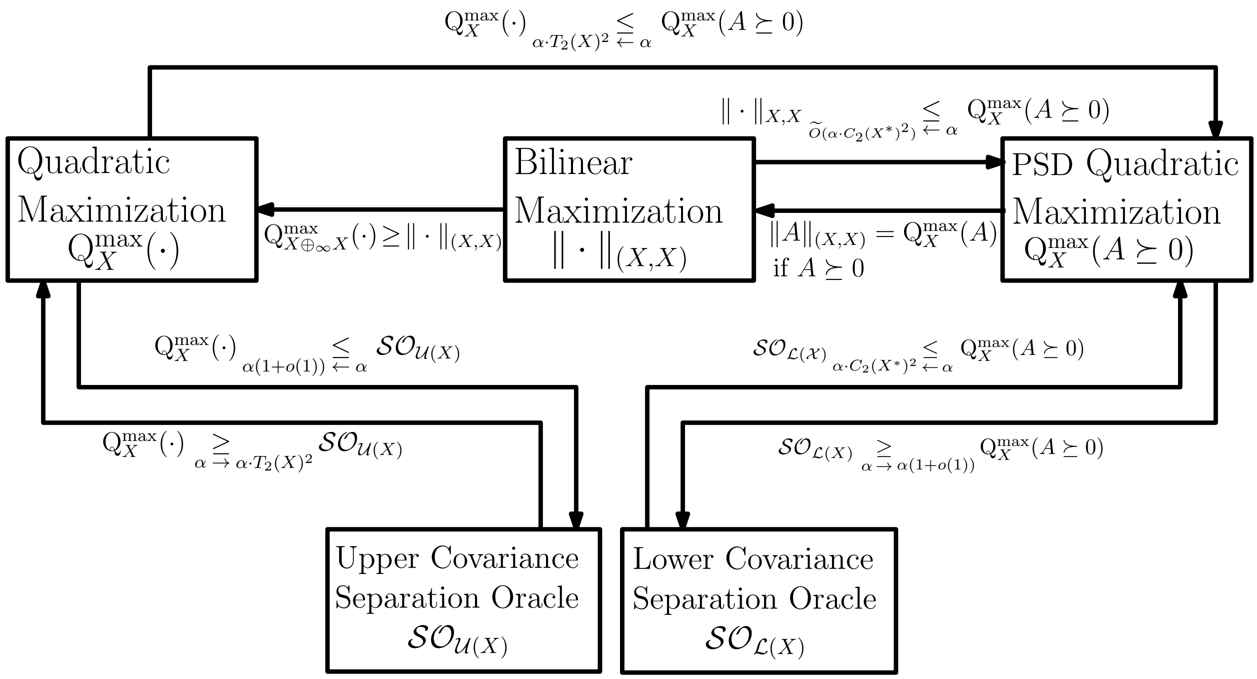

Figure 1: The figure illustrates algorithmic reductions across Quadratic/Bilinear/PSD maximization and Upper/Lower Covariance Separation oracles. For brevity we only depict bilinear maximization over (X, X), however all proofs address bilinear maximization in full generality (i.e., over (X, Y)). The notation $P \underset{\beta(\alpha) \leftarrow \alpha}{\leqslant} Q$ denotes that a polytime α-approximation algorithm for task Q implies a polytime $\beta(\alpha)$-approximation algorithm for task P. The subscript of $\leqslant$ is omitted when the reduction is lossless.

4.1 Approximation Algorithms from Covariance Separation Oracles

In this section we give approximation algorithms for quadratic and PSD-quadratic maximization over X assuming approximate separation oracles for certain bodies associated to X.

An obstacle to convex programming approaches for quadratic maximization is that for norms that are not exactly 2-convex, it is unclear how to choose a computable convex relaxation of the objective. Here we show that simply by abandoning convexity and appealing to the approximate ellipsoid method one can approximate quadratic maximization over X (conditioned on a separation oracle for a body associated to X) without any structural assumptions on X like sign-invariance or 2-convexity. This serves only as the starting point of our approach as designing such a separation oracle can be a non-trivial task.

4.1.1 Quadratic Maximization via Upper Covariance Separation Oracles

Motivated by Gaussian rounding schemes, we consider the following (not convex but approximately convex) relaxation of $\mathrm{Q}_X^{\max}(A)$:

$$\begin{aligned} \text{maximize} \quad & \langle A, \mathbb{X} \rangle \\ \text{subject to} \quad & \mathbb{E}\left[\|\mathbb{X}^{1/2}\mathrm{g}\|_X^2\right] \leqslant 1 \\ & \mathbb{X} \succeq 0 \end{aligned} \tag{34}$$

In fact the relaxation is lossless, i.e.,

Observation 4.1. *(34) is equal to* $Q_X^{\max}(A)$*.*

Proof. To show (34) $\leqslant Q_X^{\max}(A)$, let $\mathbb{X}$ be an optimal solution of (34). We have

$$(34) = \langle A\,, \mathbb{X}\rangle = \mathbb{E}\left[\left\langle \mathbb{X}^{1/2}\mathrm{g}\,, A\mathbb{X}^{1/2}\mathrm{g}\right\rangle\right] \leqslant \mathbb{E}\left[Q_X^{\max}(A)\cdot\|\mathbb{X}^{1/2}\mathrm{g}\|_X^2\right] \leqslant Q_X^{\max}(A)..$$

The direction (34) $\geqslant Q_X^{\max}(A)$ follows from considering the substitution $\mathbb{X} \stackrel{\text{def}}{=} xx^*$ where $x \in \mathrm{Ball}(X)$ is an optimal solution to $Q_X^{\max}(A)$. ■

We show next how applying the approximate ellipsoid method to (34) can provide a conditional approximation algorithm for quadratic maximization.

Proposition 4.2 (Quadratic Maximization Given Separation Oracle for Upper Covariance Body)**.** *There is an algorithm* $\mathrm{ALG}(A, R, r, \mathcal{SO})$ *such that if* $\|\cdot\|_X$ *is an* (R, r)*-balanced norm over* $\mathbb{R}^n$ *and* $\mathcal{SO}$ *is an* α*-approximate separation oracle for* $\mathcal{U}(X)$*, then on any input* $A \in M_n(\mathbb{R})$*,* ALG *runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns a* $(1 + o(1))\alpha$*-approximate solution to* $Q_X^{\max}(A)$ *with probability* $1 - 2^{-\Omega(n)}$*.*

Proof. We use the approximate ellipsoid method (Proposition 3.11 with $C_1 \leftarrow \mathcal{U}(X)$) to compute a $(1 + o(1))\alpha$-approximate solution $\mathbb{X}$ to (34). By Observation 4.1, we know $\mathbb{E}\left[\langle \mathbb{X}^{1/2}\mathrm{g}\,, A\mathbb{X}^{1/2}\mathrm{g}\rangle\right] \geqslant (1 - o(1))Q_X^{\max}(A)/\alpha$. We define a random variable $\mathbf{G} \stackrel{\text{def}}{=} \langle \mathbb{X}^{1/2}\mathrm{g}\,, A\mathbb{X}^{1/2}\mathrm{g}\rangle - Q_X^{\max}(A)\cdot\|\mathbb{X}^{1/2}\mathrm{g}\|_X^2/\tau$ where $\mathrm{g} \in \mathbb{R}^n$ is a vector of independent standard Gaussians and $\tau = \alpha(1 + \varepsilon)$ (where $\varepsilon = o(1)$ is chosen to decay slower than the hidden $o(1)$ factor above). By a routine application of Chebyshev's inequality (and independent resampling) we conclude that polynomially many independent samples of g will guarantee that with probability $1 - 2^{-\Omega(n)}$, $\mathbf{G} \geqslant 0$ for some sample and therefore $\mathbb{X}^{1/2}\mathrm{g}/\|\mathbb{X}^{1/2}\mathrm{g}\|_X$ is the desired approximate solution (the argument with all details is identical to the concentration argument made in the proof of Proposition 4.4, we choose not to duplicate it here for brevity). ■

A necessary condition for an α-separation oracle is α-separability. Interestingly, $\mathcal{U}(X)$ is precisely $\widetilde{T}_2(X)^2$-separable (see Section 4.2.4) and from this the connection to type-2 is evident. Unfortunately, (as we show in Section 9) bounded type-2 is not sufficient in general. However, in Section 6 and Section 7 we exhibit several families of norms wherein one can construct approximate separation oracles for the Gaussian body.

In its present form Proposition 4.2 is not easy to use and in most of our examples it takes significant work to design C-separation oracles for $\mathcal{U}(X)$ with C being a function of $\widetilde{T}_2(X)$. We show in the next section how for the special class of PSD instances one can obtain an analogue of Proposition 4.2 assuming only a separation oracle for the lower covariance region. Designing such separation oracles turns out to be an easier task in many cases.

4.1.2 PSD Quadratic Maximization via Lower Covariance Separation Oracles

Let $A = BB^*$ where B is an $n \times n$ matrix and let X be a norm over $\mathbb{R}^n$. We crucially use the simple fact that $Q_X^{\max}(A) = \mathrm{Op}_{X,X}^{\max}(A) = \|BB^*\|_{X\to X^*} = \|B\|_{2\to X^*}^2$, which informally speaking, allows us to convert a search problem over X into a search problem over ℓ_2^n. We consider the following (not convex but approximately convex) relaxation of $\|B\|_{2\to X^*}^2 = Q_X^{\max}(A)$:

$$\text{maximize} \qquad \mathbb{E}\left[\|B\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2\right]$$

$$\text{subject to} \quad \mathrm{Tr}(\mathbb{W}) \leqslant 1$$
$$\mathbb{W} \in \mathbb{PSD}^n \tag{35}$$

In fact the relaxation is lossless, i.e.,

Observation 4.3. *(35) is equal to $\|B\|_{2\to X^*}^2 = \mathrm{Q}_X^{\max}(A)$.*

Proof. Let $\mathbb{W}$ be the optimal solution of (35).

$$\mathbb{E}\,[\|B\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2] \leqslant \mathbb{E}\,[\|B\|_{2\to X^*}^2 \cdot \|\mathbb{W}^{1/2}\mathrm{g}\|_2^2] = \|B\|_{2\to X^*}^2 \cdot \mathrm{Tr}(\mathbb{W}) \leqslant \|B\|_{2\to X^*}^2$$

which implies one direction of the claim.

The other direction follows from considering the substitution $\mathbb{W} \stackrel{\text{def}}{=} ww^*$ where $w \in \mathrm{Ball}(\ell_2^n)$ is an optimizer of $\|Bw\|_{X^*}$. ∎

We show next how applying the approximate ellipsoid method to (35) can provide an approximation algorithm for PSD quadratic maximization with a weaker assumption than Proposition 4.2.

Proposition 4.4 (PSD Maximization Given Lower Covariance Separation Oracle)**.**
There is an algorithm $\mathrm{ALG}(A, R, r, \mathcal{SO})$ such that if $\|\cdot\|_X$ is an (R, r)-balanced norm over $\mathbb{R}^n$ and $\mathcal{SO}$ is an α-approximate separation oracle for $\mathcal{L}(X)$, then on any input $A \in \mathbb{PSD}^n$, ALG runs in time $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ and returns a $(1+o(1))\alpha$-approximate solution to $\mathrm{Q}_X^{\max}(A)$ with probability $1 - 2^{-\Omega(n)}$.

Proof. Without loss of generality we may assume A is invertible otherwise we may add a small enough multiple of the identity without changing the objective value much (and therefore any $n \times n$ matrix B is invertible if it satisfies $A = BB^*$). By Observation 4.3, to estimate $\mathrm{Q}_X^{\max}(A)$, it suffices to approximately compute (35). By Proposition 3.11 with $C_1 \leftarrow \mathrm{Ball}(\ell_2^n)$ and $f(\mathbb{W}) \leftarrow \mathbb{E}\left[\|B\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2\right]$, it suffices to give for every $\lambda > 0$, an α-approximate separation oracle for the set

$$S_\lambda \stackrel{\text{def}}{=} \left\{\mathbb{W} \succeq 0 \mid \mathbb{E}\left[\|B\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2\right] \geqslant \lambda\right\} = \left\{\mathbb{W} \mid \lambda^{-1} \cdot B\mathbb{W}B^* \in \mathcal{L}(X)\right\}.$$

Since S_λ is a linear transformation of $\mathcal{L}(X)$, it is straightforward to check that $\mathcal{SO}$ can be adapted to provide an α-approximate separation oracle for S_λ. Indeed consider any $\overline{\mathbb{W}} \in M_n(\mathbb{R})$ and run $\mathcal{SO}(\lambda^{-1} \cdot B\overline{\mathbb{W}}B^*)$. If "Inside" is returned, we know $\mathbb{W} \in \alpha \cdot S_\lambda$ and therefore can return "Inside". On the other hand if $\mathcal{SO}$ returns M separating $\mathcal{L}(X)$ from $\lambda^{-1} \cdot B\overline{\mathbb{W}}B^*$ then $\lambda^{-1} \cdot B^*MB$ separates S_λ from $\overline{\mathbb{W}}$ and thus we obtain an α-approximate separation oracle for S_λ.

Our rounding algorithm proceeds by first obtaining w.h.p. a vector $\overline{\mathrm{g}} \in \mathrm{Ball}(\ell_2^n)$ such that $\|B\overline{\mathrm{g}}\|_{X^*}^2 \geqslant (1 - o(1)) \cdot \|B\|_{2\to X^*}^2/\alpha$. Then the desired solution is any vector $\xi \in \mathrm{Ball}(X)$ satisfying $\langle \xi \,, B\overline{\mathrm{g}}\rangle = \|B\overline{\mathrm{g}}\|_{X^*}$ (which can be found using standard convex optimization). This is because

$$\langle A \,, \xi\xi^*\rangle = \|B^*\xi\|_2^2 \geqslant \langle \xi \,, B\overline{\mathrm{g}}\rangle^2 = \|B\overline{\mathrm{g}}\|_{X^*}^2 \geqslant (1 - o(1))\frac{\|B\|_{2\to X^*}^2}{\alpha} = (1 - o(1))\frac{\mathrm{Q}_X^{\max}(A)}{\alpha}.$$

Let $\mathbb{W}$ be a $\alpha/(1-1/n)$-approximately optimal solution to (35). We define a random variable $\mathbf{G} \stackrel{\text{def}}{=} \|B\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2 - \|B\|_{2\to X^*}^2 \cdot \|\mathbb{W}^{1/2}\mathrm{g}\|_2^2/\tau$ where g is an i.i.d. standard Gaussian vector and $\tau > \alpha$ is a parameter to be chosen later. Note that by assumption on $\mathbb{W}$,

$$\frac{1-1/n}{\alpha} \cdot \|B\|_{2\to X^*}^2 \leqslant \mathbb{E}\,[\|B\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2] \leqslant \mathbb{E}\,[\|B\|_{2\to X^*}^2 \cdot \|\mathbb{W}^{1/2}\mathrm{g}\|_2^2] \leqslant \|B\|_{2\to X^*}^2$$

$$\Rightarrow \mathbb{E}[\mathbf{G}]/\|B\|^2_{2\to X^*} \in [(1-1/n)/\alpha - 1/\tau, 1] \tag{36}$$

For $i \in [t]$ let $\mathbf{G_i} \stackrel{\text{def}}{=} \|B\mathbb{W}^{1/2}g^i\|^2_{X^*} - \|B\|^2_{2\to X^*} \cdot \|\mathbb{W}^{1/2}g^i\|_2^2/\tau$ be i.i.d. copies of $\mathbf{G}$ where $t = n^{3C}$ for some constant $C > 1$ and each $g^i \in \mathbb{R}^n$ is a vector of independent standard Gaussians. We set $\overline{g} \stackrel{\text{def}}{=} \mathbb{W}^{1/2}g^j/\|\mathbb{W}^{1/2}g^j\|_2$ for any maximizer $\mathbf{G_j} = \max_{i\in[t]} \mathbf{G_i}$. We argue that $\overline{g}$ is the desired vector w.h.p. through a routine application of Chebyshev's inequality. We start with a bound on the variance:

$$
\begin{aligned}
&\operatorname{Var}\left[\textstyle\sum_{i\in[t]} \mathbf{G_i}\right] \\
&= t \cdot \operatorname{Var}[\mathbf{G}] \\
&\leqslant t \cdot \mathbb{E}[\mathbf{G^2}] \\
&\leqslant t \cdot \left(\mathbb{E}[\|B\mathbb{W}^{1/2}g\|^4_{X^*}] + \mathbb{E}[\|B\|^4_{2\to X^*}\|\mathbb{W}^{1/2}g\|_2^4/\tau^2]\right) \\
&\leqslant C' \cdot t \cdot \left(\mathbb{E}[\|B\mathbb{W}^{1/2}g\|^2_{X^*}]^2 + \mathbb{E}[\|B\|^2_{2\to X^*}\|\mathbb{W}^{1/2}g\|_2^2/\tau]^2\right) \quad &&\text{(Khintchine-Kahane Theorem 2.1)} \\
& &&\text{(where } C' \text{ is an absolute const.)} \\
&\leqslant C'((1-1/n)/\alpha - 1/\tau) \cdot t \cdot \mathbb{E}[\mathbf{G}]^2 &&\text{(by (36))}
\end{aligned}
$$

By Chebyshev's inequality, with probability at least $1 - 1/k^2$ it holds that

$$\sum_{i\in[t]} \mathbf{G_i} \geqslant t \cdot \mathbb{E}[\mathbf{G}] - k \cdot \left(C' \cdot t \cdot \left(\frac{1-1/n}{\alpha} - \frac{1}{\tau}\right)\right)^{1/2} \cdot \mathbb{E}[\mathbf{G}].$$

By averaging, there must exist $i \in [t]$ such that

$$\frac{\|B\mathbb{W}^{1/2}g^i\|^2_{X^*}}{\|\mathbb{W}^{1/2}g^i\|_2^2} \geqslant \mathbb{E}[\mathbf{G}] \cdot \left(1 - O\left(\frac{k}{\sqrt{t}}\left(\frac{1-1/n}{\alpha} - \frac{1}{\tau}\right)\right)\right).$$

Setting $k \stackrel{\text{def}}{=} t^{0.49}$ and $\tau \stackrel{\text{def}}{=} (1 + t^{-0.001})\alpha$, we obtain that with probability at least $1 - n^{-C}$ there exists $i \in [t]$ such that $\mathbf{G_i} \geqslant \mathbb{E}[\mathbf{G}](1 - O(t^{-0.009})) \geqslant 0$. Thus $\|B\mathbb{W}^{1/2}g^i\|^2_{X^*}/\|\mathbb{W}^{1/2}g^i\|_2^2 \geqslant \|B\|^2_{2\to X^*}/\tau$ as desired. We obtain success probability $1 - 2^{-\Omega(n)}$ by repeating the above process independently for polynomially many rounds. This completes the proof. ■

Recall a necessary condition for an α-separation oracle is α-separability. Interestingly, $\mathcal{L}(X)$ is precisely $\widetilde{C}_2(X^*)^2$-separable (see Section 4.2.5) and from this the connection to dual cotype-2 is evident. It turns out that in many cases separation oracles are easier to design for $\mathcal{L}(X)$ than $\mathcal{U}(X)$. In Section 6 and Section 7 we explore several situations wherein one can construct such approximate separation oracles.

One of our main technical contributions is using factorization theory to show that quadratic maximization under bounded type-2 (and bilinear maximization under bounded dual cotype-2) reduces (algorithmically) to solving a series of PSD maximization instances. This leads us to several new results for quadratic and bilinear maximization. We now proceed to exhibit the claimed reductions.

4.2 Reductions Across Quadratic/Bilinear/PSD Maximization

In this section we explore in depth polynomial time reductions (with multiplicative loss depending only on $\widetilde{T}_2(X)$ or $\widetilde{C}_2(X^*)$) between the following three oracles

($\mathcal{O}1$) Approximate Search Oracle for Quadratic Maximization.

($\mathcal{O}2$) Approximate Search Oracle for Bilinear Maximization.

($\mathcal{O}3$) Approximate Search Oracle for Quadratic Maximization of PSD instances.

The reductions from quadratic/bilinear maximization to PSD maximization, as we will see in Section 5, are closely related to the theory of factorization of linear operators through ℓ_2.

4.2.1 Reducing Quadratic to PSD under Bounded Type-2

In this section we show how a C-approximate search algorithm for quadratic maximization of PSD instances over X can be used to give a $(1+o(1)) \cdot C \cdot \widetilde{T}_2(X)^2$-approximation algorithm for quadratic maximization of general instances. Our approach is as follows:

1. We will use the PSD oracle and algorithmic-duality for downward-closed sets (Theorem 3.14) to construct a separation oracle for $\downarrow B_{\wedge}^{Sym}(X)$.
2. This allows us to approximately solve the following (not exactly computable) vector relaxation of $\mathrm{Q}_X^{\max}(A)$:
$$\max\{\langle A\,,\mathbb{X}\rangle \;|\; \mathbb{X} \in \downarrow B_{\wedge}^{Sym}(X)\}\,. \tag{37}$$
3. Finally we use Gaussian rounding to show that (37) $\leqslant \widetilde{T}_2(X)^2 \cdot \mathrm{Q}_X^{\max}(A)$.

We proceed with the proof.

Theorem 4.5 (Type-2 Quadratic Maximization Reduces to PSD Quadratic Maximization)**.** *There is an algorithm* $\mathrm{ALG}(A, R, r, \mathcal{O})$ *such that if* $\|\cdot\|_X$ *is an* (R, r)*-balanced norm over* $\mathbb{R}^n$ *and* $\mathcal{O}$ *is an* α*-approximate search oracle for PSD quadratic maximization over* X*, then on any input* $A \in M_n(\mathbb{R})$*,* ALG *runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns a* $(1+o(1)) \cdot \alpha \cdot \widetilde{T}_2(X)^2$*-approximate solution to* $\mathrm{Q}_X^{\max}(A)$ *with probability at least* $1 - 2^{\Omega(n)}$.

Proof. By Proposition 4.2 it suffices to exhibit an approximate separation oracle for $\mathcal{U}(X)$.

To this end, note that $\mathcal{O}$ is an oracle returning an α-approximate solution to $\mathrm{Q}_X^{\max}(\mathbb{W}) = \sup_{\mathbb{X} \in B_{\wedge}^{Sym}(X)} \langle \mathbb{W}\,, \mathbb{X}\rangle$ for any $\mathbb{W} \in \mathbb{PSD}^n$. Therefore we may apply Theorem 3.14 to the set $\downarrow B_{\wedge}^{Sym}(X)$ with $K \stackrel{\text{def}}{=} \mathbb{PSD}^n$ to obtain a $(1+o(1))\alpha$-approximate separation oracle for $\downarrow B_{\wedge}^{Sym}(X)$. (The balancedness of $\downarrow B_{\wedge}^{Sym}(X)$ from Lemma 2.24.)

Lastly Observation 3.4 yields the desired $(1+o(1)) \cdot \alpha \cdot \widetilde{T}_2(X)^2$-approximate separation oracle for $\mathcal{U}(X)$ since by Observation 2.17 we have the equivalence $\widetilde{T}_2(X)^{-2} \cdot \downarrow B_{\wedge}^{Sym}(X) \subseteq \mathcal{U}(X) \subseteq \downarrow B_{\wedge}^{Sym}(X)$. ∎

Remark 4.6. *We make two technical remarks that are required for application of this theorem to algorithmic closure properties for complex interpolation. Note that the proof above can be adapted verbatim for the following additional features:*

1. *We may take* $\|\cdot\|_X$ *to be a norm over* $\mathbb{C}^n$ *and input* $A \in M_n(\mathbb{C})$.
2. *We may replace the PSD maximization search oracle* $\mathcal{O}$ *with a slightly weaker oracle* $\mathcal{O}'$ *that on input* $W \succeq 0$*, returns a witness* $Z \in \mathrm{Ball}(X \widehat{\otimes} X)$ *satisfying* $\langle W\,, Z\rangle \geqslant \mathrm{Q}_X^{\max}(W)/C$*. Such an oracle is weaker as* Z *can be a convex combination of rank-1 matrices of the form* xy^* *(where* $x, y \in \mathrm{Ball}(X)$*) instead of a single such rank-1 matrix.*

4.2.2 Reducing Bilinear to PSD under Bounded Dual Cotype-2

In this section we show how a C-approximate search algorithm for quadratic maximization over PSD instances can be used to give a $C' = C'(C, \widetilde{C}_2(X^*), \widetilde{C}_2(Y^*))$-approximation algorithm for bilinear maximization over X, Y (i.e., $\mathrm{Op}_{X,Y}^{\max}(\cdot)$). Our approach is as follows:

1. We will use the PSD oracles and algorithmic-duality for downward-closed sets (Theorem 3.14) to construct separation oracles for $\downarrow B_{\wedge}^{Sym}(X)$, $\downarrow B_{\wedge}^{Sym}(Y)$.

2. This allows us to approximately solve the following (not exactly computable) convex relaxation $\mathrm{Rlx}(A)$ of $\mathrm{Op}_{X,Y}^{\max}(A)$:

$$\begin{aligned} \mathrm{Rlx}(A) = \quad \max \quad & \langle A\,, Z \rangle \\ \text{s.t.} \quad & \mathbb{X} \in \downarrow B_{\wedge}^{Sym}(X)\,,\ \mathbb{Y} \in \downarrow B_{\wedge}^{Sym}(Y) \\ & \begin{bmatrix} \mathbb{X} & Z \\ Z^* & \mathbb{Y} \end{bmatrix} \succeq 0\,. \end{aligned} \tag{38}$$

3. Finally we need a rounding algorithm taking an optimal solution of $\mathrm{Rlx}(A)$ and producing a good solution to $\mathrm{Op}_{X,Y}^{\max}(A)$. A deep factorization theorem of Pisier [Pis80] states that

$$\|A\|_{Y\to X^*} \leqslant \inf_{BC=A} \|C\|_{Y\to 2} \cdot \|B\|_{2\to X^*} \leqslant C' \cdot \|A\|_{Y\to X^*}\,.$$

   Pisier's proof can be viewed as having two components. The first component is a bound on the intergrality gap of $\mathrm{Rlx}(A)$:

$$\mathrm{Op}_{X,Y}^{\max}(A) \leqslant \mathrm{Rlx}(A) \leqslant C' \cdot \mathrm{Op}_{X,Y}^{\max}(A)\,. \tag{39}$$

   Pisier's proof of (39) is non-constructive and does not furnish a rounding algorithm for bilinear maximization. We give such a rounding algorithm below.

   The second component of the factorization theorem is a dual characterization of $\mathrm{Rlx}(A)$ as a factorization norm that follows from a duality result of Maurey, i.e., $\mathrm{Rlx}(A) = \inf_{BC=A} \|C\|_{Y\to 2} \cdot \|B\|_{2\to X^*}$. We refer the reader to Section 5 for a detailed discussion of factorization theory and the results of Maurey and Pisier.

Pisier's proof exhibits a (randomized) non-constructive map from any pair of sequences $(y_i), (\overline{y}_i)$ to a unit vector $y \in \mathrm{Ball}(Y)$ such that $\mathbb{E}[\|Ay\|_{X^*}]$ is bounded from below by $(\sum_i \|A\,y_i\|_{X^*}^2 / \sum_j \|\overline{y}_j\|_Y^2)/C'$. The analysis of our rounding algorithm borrows heavily from Pisier's proof as we utilize his non-constructive map as a witness of our rounded vector having high expected value.

We proceed with collecting the necessary components of our proof. We require the following fourier analytic decomposition result that is a critical component of Pisier's factorization theorem.

Lemma 4.7 (Pisier [Pis86])**.** *Let* $g_1, \dots, g_m \sim \mathcal{N}(0,1)$ *be i.i.d. standard Gaussians and let* $\varepsilon_1, \dots, \varepsilon_m$ *be i.i.d. rademacher* (± 1) *random variables. If* $\sum_{i\in[m]} y_i y_i^* \in B_{\wedge}^{Sym}(Y)$*, then for any* $0 < \delta < 1$ *there exists a vector valued function* $\varphi : \mathbb{R}^n \times \{\pm 1\}^n \mapsto \mathbb{R}^n$ *such that*

$$\mathop{\mathbb{E}}_{g,\varepsilon}\left[\|\textstyle\sum_i \varepsilon_i g_i y_i + \varphi(g,\varepsilon)\|_Y^2\right]^{1/2} = O(\widetilde{C}_2(Y^*) \log 1/\delta)$$

$$\mathop{\mathbb{E}}_{g,\varepsilon}\left[\|A\varphi(g,\varepsilon)\|_{X^*}^2\right]^{1/2} \leqslant \delta \cdot \mathrm{Rlx}(A) \qquad \textit{for any } A \in M_{n,m}(\mathbb{R})\,.$$

We will also need the following simple fact about scalar valued random variables.

Fact 4.8. *Fix $\alpha, \varepsilon > 0$ and $\beta > 1$. Let x be a random variable supported in the interval $[-\infty, \alpha]$ and let $\mathbb{E}[x] \geqslant \alpha/\beta$. Then*

$$\mathbb{P}\left[x < \frac{\alpha}{(1+\varepsilon)\beta}\right] \leqslant \frac{(1+\varepsilon)(\beta-1)}{\beta(1+\varepsilon)-1} = 1 - \frac{\varepsilon}{\beta+\varepsilon\beta-1}.$$

Proof. If the statement is false,

$$\mathbb{E}[x] < \frac{(1+\varepsilon)(\beta-1)}{\beta(1+\varepsilon)-1} \cdot \frac{\alpha}{(1+\varepsilon)\beta} + \left(1 - \frac{(1+\varepsilon)(\beta-1)}{\beta(1+\varepsilon)-1}\right)\alpha = \alpha/\beta,$$

leading to contradiction. ∎

Finally we need the following simple claim about factorization of a 2×2 positive definite block matrix.

Claim 4.9. *Consider any invertible matrix $\left[\begin{smallmatrix} \mathbb{X} & Z \\ Z^* & \mathbb{Y} \end{smallmatrix}\right] \in \mathbb{PSD}^{n+m}$ and fix any factorization $\mathbb{X} = SS^*$ where $S \in M_{n,s}(\mathbb{R})$. Then there exists $T \in M_{m,s}(\mathbb{R})$ such that $Z = ST^*$ and $\mathbb{Y} \succeq TT^*$.*

Proof. By assumption there exist matrices $U \in M_{n,n+m}(\mathbb{R})$ and $V \in M_{m,n+m}(\mathbb{R})$ such that $\mathbb{X} = UU^*$, $\mathbb{Y} = VV^*$ and $Z = UV^*$. Let $S^\dagger \stackrel{\text{def}}{=} S^*(SS^*)^{-1} = S^*\mathbb{X}^{-1}$ denote the moore-penrose pseudo-inverse of S (recall $S^\dagger$ satisfies the property $SS^\dagger = \mathrm{I}$). Let $T \stackrel{\text{def}}{=} VU^*(S^\dagger)^*$. Then we have $ST^* = SS^\dagger UV^* = UV^* = Z$.

It remains to verify $\mathbb{Y} \succeq TT^*$. To this end note that by definition of pseudo-inverse one has

$$(S^\dagger)^* S^\dagger = (S^\dagger)^* S^* \mathbb{X}^{-1} = \mathbb{X}^{-1}. \tag{40}$$

Let $H \stackrel{\text{def}}{=} U^\dagger U = U^*\mathbb{X}^{-1}U = H^*$. Then $H^2 = U^*\mathbb{X}^{-1}UU^*\mathbb{X}^{-1}U = U^*\mathbb{X}^{-1}U = H$. Thus H is a symmetric orthogonal projector. Now we have

$$TT^* = VU^*(S^\dagger)^*S^\dagger UV^* = VU^*\mathbb{X}^{-1}UV^* = V^*HV \preceq V^*V = \mathbb{Y}$$

where the second equality uses (40), and the (PSD) inequality uses the fact that H is an orthogonal projector. This completes the proof. ∎

We are now ready to prove the main result of this subsection.

Theorem 4.10 (Bilinear Maximization Reduces to PSD Quadratic Maximization)**.**
There is an algorithm $\mathrm{ALG}(A, R, r, \mathcal{O}_X, \mathcal{O}_Y)$ such that if $\|\cdot\|_X$ (resp. $\|\cdot\|_Y$) is an (R, r)-balanced norm over $\mathbb{R}^n$ (resp. $\mathbb{R}^m$) and $\mathcal{O}_X$ (resp. $\mathcal{O}_Y$) is an α-approximate search oracle for PSD quadratic maximization over X (resp. Y), then on any input $A \in M_{n,m}(\mathbb{R})$, ALG runs in time $\mathrm{poly}(n, m, \log R, \log 1/r, \mathrm{bit}(A))$ and returns an $\alpha\beta\log\alpha\beta$-approximate solution to $\mathrm{Op}^{\max}_{X,Y}(A)$ with probability $1 - 2^{-\Omega(n)}$, where $\beta \lesssim C_2(X^*)C_2(Y^*)$.

Proof. Just as in the proof of Theorem 4.5 we obtain $(1 + o(1)) \cdot \alpha$-separation oracles for $\downarrow B_\wedge^{Sym}(X)$, $\downarrow B_\wedge^{Sym}(Y)$. Therefore by Proposition 3.11 we can compute a $(1 + o(1))\alpha$-approximate solution $(\mathbb{X}, \mathbb{Y}, Z)$ to (38) and apply the following rounding algorithm:

1. Let $\mathbb{Y}$ be a $(1+o(1))\alpha$-approximate solution to (38) and fix any decomposition $\mathbb{Y} = \sum_{i\in[m]} y_i y_i^*$.

2. Let $g_1,\ldots,g_m \sim \mathcal{N}(0,1)$ be i.i.d. standard Gaussians and let $\varepsilon_1,\ldots,\varepsilon_m$ be i.i.d. rademacher (± 1) random variables. Using the exact membership oracle of X (and standard convex optimization), compute $x_g \in \mathrm{Ball}(X)$ satisfying $\langle x_g\,, \sum_i \varepsilon_i g_i A y_i\rangle = \|\sum_i \varepsilon_i g_i A y_i\|_{X^*}$.

3. Using the exact membership oracle of Y (and convex optimization), compute $y_g \in \mathrm{Ball}(Y)$ satisfying $\langle A^* x_g\,, y_g\rangle = \|A^* x_g\|_{Y^*}$.

4. Output x_g, y_g.

We now analyze the rounding algorithm. Since $\wedge_X^{\downarrow Sym}(\mathbb{X}) \leqslant 1$, by definition there exists $\mathbb{X}' \succeq \mathbb{X}$ and a finite sequence (x_i) such that $\mathbb{X}' = \sum_i x_i x_i^*$ and $\sum_i \|x_i\|_X^2 \leqslant 1$. Since $\mathbb{X}' \succeq \mathbb{X}$, we have

$$\begin{bmatrix} \mathbb{X}' & Z \\ Z^* & \mathbb{Y} \end{bmatrix} \succeq 0\,.$$

Without loss of generality we assume $\left[\begin{smallmatrix} \mathbb{X}' & Z \\ Z^* & \mathbb{Y} \end{smallmatrix}\right]$ is invertible since otherwise we may add a sufficiently small copy of the identity matrix which affects estimates by only lower order terms. Then by Claim 4.9, there exists a sequence $(\overline{y}_i)$ such that $Z = \sum_i x_i \overline{y}_i^*$ and $\mathbb{Y} \succeq \sum_i \overline{y}_i \overline{y}_i^*$. We have

$$\begin{aligned}
&(1-o(1)) \cdot \mathrm{Rlx}(A)/\alpha \\
&\leqslant \langle A\,, Z\rangle \\
&= \sum_i \langle x_i\,, A\overline{y}_i\rangle \\
&\leqslant \left(\textstyle\sum_i \|x_i\|_X^2\right)^{1/2} \left(\textstyle\sum_i \|A\overline{y}_i\|_{X^*}^2\right)^{1/2} && \text{(by Hölder + Cauchy-Schwarz)} \\
&\leqslant \left(\textstyle\sum_i \|A\overline{y}_i\|_{X^*}^2\right)^{1/2} && \text{(by definition of } (x_i)) \\
&\leqslant \widetilde{C}_2(X^*) \cdot \mathop{\mathbb{E}}_{g}\left[\|\textstyle\sum_i g_i A\overline{y}_i\|_{X^*}^2\right]^{1/2} && \text{(by definition of cotype-2)} \\
&\leqslant \widetilde{C}_2(X^*) \cdot \mathop{\mathbb{E}}_{g}\left[\|\textstyle\sum_i g_i A y_i\|_{X^*}^2\right]^{1/2} && \text{(monotonicity - Fact 2.16)} \qquad (41) \\
&\text{(since } \sum_i (Ay_i)(Ay_i)^* = A\mathbb{Y}A^* \succeq \sum_i (A\overline{y}_i)(A\overline{y}_i)^*)\,.
\end{aligned}$$

We are now equipped to relate $\gamma_2(A)$ to the expected value of the rounding algorithm. By (41), we have

$$\begin{aligned}
&(1-o(1)) \cdot \mathrm{Rlx}(A)/(\alpha \cdot \widetilde{C}_2(X^*)) \\
\leqslant\ & \mathop{\mathbb{E}}_{g}\left[\|\textstyle\sum_i g_i A y_i\|_{X^*}^2\right]^{1/2} \\
\lesssim\ & \mathop{\mathbb{E}}_{g}\left[\|\textstyle\sum_i g_i A y_i\|_{X^*}\right] && \text{(Kahane-Khintchine)} \\
=\ & \mathop{\mathbb{E}}_{g,\varepsilon}\left[\|\textstyle\sum_i \varepsilon_i g_i A y_i\|_{X^*}\right] && \text{(identical distributions)} \\
=\ & \mathop{\mathbb{E}}_{g,\varepsilon}\left[\langle x_g\,, \textstyle\sum_i \varepsilon_i g_i A y_i\rangle\right] \\
=\ & \mathop{\mathbb{E}}_{g,\varepsilon}\left[\langle A^* x_g\,, \textstyle\sum_i \varepsilon_i g_i y_i\rangle\right] \\
\leqslant\ & \mathop{\mathbb{E}}_{g,\varepsilon}\left[\,|\langle A^* x_g\,, \textstyle\sum_i \varepsilon_i g_i y_i + \varphi(g,\varepsilon)\rangle|\,\right] + \mathbb{E}_{g,\varepsilon}\left[\,|\langle A^* x_g\,, \varphi(g,\varepsilon)\rangle|\,\right] && \text{(by triangle inequality)}
\end{aligned}$$

$$\leqslant \mathop{\mathbb{E}}_{g,\varepsilon}[\,|\langle A^* x_g\,, \textstyle\sum_i \varepsilon_i g_i y_i + \varphi(g,\varepsilon)\rangle|\,] + \mathbb{E}_{g,\varepsilon}\left[\langle A^* x_g\,, \varphi(g,\varepsilon)\rangle^2\right]^{1/2} \qquad \text{(by Jensen's inequality)}$$

(where φ is chosen according to Lemma 4.7)

$$\leqslant \mathop{\mathbb{E}}_{g,\varepsilon}[\|A^* x_g\|_Y \cdot \|\textstyle\sum_i \varepsilon_i g_i y_i + \varphi(g,\varepsilon)\|_{Y^*}] + \mathbb{E}_{g,\varepsilon}\left[\|A\varphi(g,\varepsilon)\|_{X^*}^2\right]^{1/2} \qquad \text{(Holder's, } x_g \in \mathrm{Ball}(X))$$

$$\leqslant \mathop{\mathbb{E}}_{g,\varepsilon}[\|A^* x_g\|_Y \cdot \|\textstyle\sum_i \varepsilon_i g_i y_i + \varphi(g,\varepsilon)\|_{Y^*}] + \delta \cdot \mathrm{Rlx}(A) \qquad \text{(Lemma 4.7)}$$

$$\leqslant \mathop{\mathbb{E}}_{g,\varepsilon}\left[\|A^* x_g\|_Y^2\right]^{1/2} \cdot \mathop{\mathbb{E}}_{g,\varepsilon}\left[\|\textstyle\sum_i \varepsilon_i g_i y_i + \varphi(g,\varepsilon)\|_{Y^*}^2\right]^{1/2} + \delta \cdot \mathrm{Rlx}(A) \qquad \text{(Cauchy-Schwarz)}$$

$$\lesssim \mathop{\mathbb{E}}_{g,\varepsilon}\left[\|A^* x_g\|_Y^2\right]^{1/2} \cdot \widetilde{C}_2(Y^*) \cdot \log 1/\delta + \delta \cdot \mathrm{Rlx}(A) \qquad \text{(Lemma 4.7)}$$

$$= \mathop{\mathbb{E}}_{g,\varepsilon}\left[\langle x_g\,, A y_g\rangle^2\right]^{1/2} \cdot \widetilde{C}_2(Y^*) \cdot \log 1/\delta + \delta \cdot \mathrm{Rlx}(A) \qquad \text{(definition of } y) \tag{42}$$

Finally setting $\delta = 1/(C\alpha\beta)$ for a sufficiently large absolute constant C and rearranging (42) implies that

$$\mathop{\mathbb{E}}_{g,\varepsilon}\left[\langle x_g\,, A y_g\rangle^2\right]^{1/2} \geqslant \gamma_2(A)/O(\alpha\beta\log(\alpha\beta)) \geqslant \mathrm{Op}^{\max}_{X,Y}(A)/O(\alpha\beta\log(\alpha\beta))$$

Since the random variable $\langle x_g\,, Ay_g\rangle^2$ lies in the interval $[0, \|A\|^2_{X\to Y}]$, we may apply Fact 4.8 to it and obtain that with probability at least $\varepsilon/\mathrm{poly}(\alpha\beta)$, $\langle x_g\,, Ay_g\rangle \geqslant (1-\varepsilon)\mathrm{Op}^{\max}_{X,Y}(A)/O(\alpha\beta\log(\alpha\beta))$. Finally, sampling independently for polynomially many rounds and returning the best pair x_g, y_g implies the $1 - 2^{-\Omega(n)}$ success probability claimed in (2). ∎

4.2.3 A Generic Framework for Quadratic/Bilinear Maximization

We are finally equipped to prove our claimed framework theorem.

Theorem 4.11 (Framework: Maximization under Lower Covariance Oracle)**.**
There are algorithms $\mathrm{ALG}_1(A_1, R, r, \mathcal{SO}_X)$ and $\mathrm{ALG}_2(A_2, R, r, \mathcal{SO}_X, \mathcal{SO}_Y)$ such that if $\|\cdot\|_X$ (resp. $\|\cdot\|_Y$) is an (R,r)-balanced norm over $\mathbb{R}^n$ (resp. $\mathbb{R}^m$) and $\mathcal{SO}_X$ (resp. $\mathcal{SO}_Y$) is an α-approximate separation oracle for $\mathcal{L}(X)$ (resp. $\mathcal{L}(Y)$) then

- *Quadratic: for any $A_1 \in M_n(\mathbb{R})$, ALG_1 runs in time $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A_1))$ and with probability at least $1 - 2^{-\Omega(n)}$ returns a $(1+o(1)) \cdot \alpha \cdot T_2(X)^2$-approximately optimal solution to $\mathrm{Q}^{\max}_X(A_1)$.*
- *Bilinear: for any $A_2 \in M_{n,m}(\mathbb{R})$, ALG_2 runs in time $\mathrm{poly}(n, m, \log R, \log 1/r, \mathrm{bit}(A_2))$ and with probability at least $1 - 2^{-\Omega(n+m)}$, ALG_2 returns a $\beta\log\beta$-approximately optimal solution to $\mathrm{Op}^{\max}_{X,Y}(A_2)$ where $\beta \lesssim \alpha \cdot \widetilde{C}_2(X^*) \cdot \widetilde{C}_2(Y^*)$.*

Proof. The quadratic (resp. bilinear) claim follows from combining Theorem 4.5 (resp. Theorem 4.10) with Proposition 4.4. ∎

We note the following useful corollary that illustrates the importance of Khintchine-type inequalities for quadratic/bilinear maximization.

Corollary 4.12 (Khintchine-Type Inequalities Yield Approximation Algorithms)**.**

1. *There is an algorithm* $\mathrm{ALG}(A, R, r, \mathcal{O}_f)$ *such that if* $\|\cdot\|_X$ *is an* (R,r)*-balanced norm over* $\mathbb{R}^n$ *and* $\mathcal{O}_f$ *is an exact oracle computing a convex function* $f : \mathbb{PSD}^n \to \mathbb{R}$ *satisfying*

$$\alpha_1^{-1} \cdot f(\mathbb{X}) \leqslant \mathbb{E}\big[\|\mathbb{X}^{1/2}\mathrm{g}\|_X^2\big] \leqslant \alpha_2 \cdot f(\mathbb{X}) \qquad \forall\, \mathbb{X} \in \mathbb{PSD}^n\,,$$

then on any input $A \in M_n(\mathbb{R})$, ALG *runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns an* $\alpha_1\alpha_2\widetilde{T}_2(X)^2$*-approximate solution to* $\mathrm{Q}_X^{\max}(A)$ *with probability* $1 - 2^{-\Omega(n)}$.

2. *There are algorithms* $\mathrm{ALG}_1(A_1, R, r, \mathcal{O}_f)$ *and* $\mathrm{ALG}_2(A_2, R, r, \mathcal{O}_f, \mathcal{O}_g)$ *such that if* $\|\cdot\|_X$ *(resp.* $\|\cdot\|_Y$*) is an* (R,r)*-balanced norm over* $\mathbb{R}^n$ *(resp.* $\mathbb{R}^m$*) and* $\mathcal{O}_f$ *(resp.* $\mathcal{O}_g$*) is an exact oracle computing a concave function* $f : \mathbb{PSD}^n \to \mathbb{R}$ *(resp.* $\mathrm{g} : \mathbb{PSD}^m \to \mathbb{R}$*) satisfying*

$$\alpha_1^{-1} \cdot f(\mathbb{W}) \leqslant \mathbb{E}\big[\|\mathbb{W}^{1/2}\mathrm{g}\|_{X^*}^2\big] \leqslant \alpha_2 \cdot f(\mathbb{W}) \qquad \forall\, \mathbb{W} \in \mathbb{PSD}^n$$
$$\alpha_1^{-1} \cdot g(\mathbb{W}) \leqslant \mathbb{E}\big[\|\mathbb{W}^{1/2}\mathrm{g}\|_{Y^*}^2\big] \leqslant \alpha_2 \cdot g(\mathbb{W}) \qquad \forall\, \mathbb{W} \in \mathbb{PSD}^m\,,$$

then

(A) *on any input* $A_1 \in M_n(\mathbb{R})$, ALG_1 *runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A_1))$ *and returns an* $\alpha_1\alpha_2\widetilde{T}_2(X)^2$*-approximate solution to* $\mathrm{Q}_X^{\max}(A_1)$ *with probability* $1 - 2^{-\Omega(n)}$,

(B) *on any input* $A_2 \in M_{n,m}(\mathbb{R})$, ALG_2 *runs in time* $\mathrm{poly}(n, m, \log R, \log 1/r, \mathrm{bit}(A_2))$ *and returns a* $\beta \log \beta$*-approximate solution to* $\mathrm{Op}_{X,Y}^{\max}(A_2)$ *with probability* $1 - 2^{-\Omega(n+m)}$*, where* $\beta \lesssim \alpha_1\alpha_2 C_2(X^*) C_2(Y^*)$.

Proof. For claim (1.), we know by Proposition 4.2 that it suffices to give a separation oracle for $\mathcal{U}(X)$. Furthermore, the convex set $\{\mathbb{X} \succeq 0 \mid f(\mathbb{X}) \leqslant 1\}$ has an exact membership oracle and therefore also a separation oracle. Applying Observation 3.4 yields the desired approximate separation oracle for $\mathcal{U}(X)$.

For claim (2.), we know by Theorem 4.11 that it suffices to give separation oracles for $\mathcal{L}(X), \mathcal{L}(Y)$. Furthermore, the convex set $\{\mathbb{W} \succeq 0 \mid f(\mathbb{W}) \geqslant 1\}$ has an exact membership oracle and therefore also a separation oracle. Applying Observation 3.4 yields the desired approximate separation oracle for $\mathcal{L}(X)$ and the case for Y is analogous. ■

4.2.4 Type-2 Equivalence of Quadratic Maximization/Upper Covariance Separation

Below we establish a converse of Proposition 4.2 and thus obtain an algorithm vs. covariance separation oracle duality for Type-2 quadratic maximization.

Proposition 4.13 (Quadratic Maximization Algorithm implies Upper Covariance Separation Oracle)**.** *There is an* $(1 + o(1)) \cdot \alpha \cdot \widetilde{T}_2(X)^2$*-approximate separation oracle for* $\mathcal{U}(X)$ *running in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(x))$ *where* x *is the input, assuming access to an* α*-approximate search oracle* $\mathcal{O}$ *for quadratic maximization over an* (R,r)*-balanced norm* $(\mathbb{R}^n, \|\cdot\|_X)$.

Proof. Recall that, by (31), for a PSD matrix A,

$$\mathrm{Q}_X^{\max}(A) = \max_{\mathbb{W} \in B_\wedge^{Sym}(X)} \langle A, \mathbb{W} \rangle = \max_{\mathbb{W} \in \downarrow B_\wedge^{Sym}(X)} \langle A, \mathbb{W} \rangle.$$

Therefore, given an α-approximation search oracle for quadratic maximization, Theorem 3.14 with $K \leftarrow \mathbb{PSD}^n, B \leftarrow \downarrow B_\wedge^{Sym}(X)$ implies that $\downarrow B_\wedge^{Sym}(X)$ has an $(1 + o(1))\alpha$-approximate separation oracle. (The balancedness of $\downarrow B_\wedge^{Sym}(X)$ follows from Lemma 2.24.)

Then the equivalence between $\downarrow B_\wedge^{Sym}(X)$ and $\mathcal{U}(X)$ up to $\widetilde{T}_2(X)^2$ (Observation 2.17 and Observation 3.4) gives a $((1 + o(1)) \cdot \alpha \cdot \widetilde{T}_2(X)^2)$-separation oracle for $\mathcal{U}(X)$ as desired. ■

4.2.5 Cotype-2 Equivalence of PSD Maximization/Lower Covariance Separation

We next give an analogue of Proposition 4.13 for PSD maximization/lower covariance separation in the presence of dual cotype-2.

Proposition 4.14 (PSD Maximization Algorithm implies Lower Covariance Separation Oracle). *There is a $(1+o(1))\cdot\alpha\cdot\widetilde{C}_2(X)^2$-approximate separation oracle for $\mathcal{L}(X)$ running in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(x))$ *where x is the input, assuming access to an α-approximate search oracle $\mathcal{O}$ for PSD quadratic maximization over an (R,r)-balanced norm $(\mathbb{R}^n, \|\cdot\|_X)$.*

Proof. We will construct an α-approximate search oracle for $\inf_{\mathbb{W}\in B^{\uparrow}_{\wedge}(X)}\langle U^*U\,,\mathbb{W}\rangle$ on any input U^*U, and apply Theorem 3.16 with $B \leftarrow B^{\uparrow}_{\wedge}(X)$, $K \leftarrow \mathbb{PSD}^n$ to get a $(1+o(1))\alpha$-separation oracle for $B^{\uparrow}_{\wedge}(X)$. (The inverse balancedness of $B^{\uparrow}_{\wedge}(X)$ follows from Lemma 2.24.) Then the equivalence between $B^{\uparrow}_{\wedge}(X)$ and $\mathcal{L}(X)$ up to $\widetilde{C}_2(X)^2$ (Observation 2.20 and Observation 3.4) gives a $((1+o(1))\cdot\alpha\cdot\widetilde{C}_2(X)^2)$-separation oracle for $\mathcal{L}(X)$ as desired.

Therefore it remains to construct an α-approximate search oracle for $\inf_{\mathbb{W}\in B^{\uparrow}_{\wedge}(X)}\langle U^*U\,,\mathbb{W}\rangle$ on any input U^*U. To this end observe that

$$\inf_{\mathbb{W}\in B^{\uparrow}_{\wedge}(X)}\langle U^*U\,,\mathbb{W}\rangle = (\|U\|^{\min}_{X^*\to 2})^2\,.$$

We conclude that U is invertible otherwise $\|U\|^{\min}_{X^*\to 2}=0$ which contradicts the assumption that $U^*U\in B^{\uparrow}_{\wedge}(X)^{\odot}$. Let v be any vector satisfying $\|v\|_2 \leqslant \alpha^{1/2}/\|U^{-1}\|_{2\to X^*}$ and $\|U^{-1}v\|_{X^*}=1$.

Such a vector can be found by using $\mathcal{O}$ to obtain u such that $\|u\|_2 = 1$ and $\|U^{-1}u\|_{X^*} \geqslant \alpha^{-1/2}\cdot\|U^{-1}\|_{2\to X^*}$. Then setting $v \overset{\text{def}}{=} u/\|U^{-1}u\|_{X^*}$ satisfies the desired properties. Finally taking $y \overset{\text{def}}{=} U^{-1}v$ satisfies the conditions $yy^*\in B^{\uparrow}_{\wedge}(X)$ and

$$\langle U^*U\,,yy^*\rangle = \|Uy\|_2^2 = \|v\|_2^2 \leqslant \alpha\cdot\|U^{-1}\|^{-2}_{2\to X^*} = \alpha\cdot(\|U\|^{\min}_{X^*\to 2})^2\,.$$

This yields the desired minimization oracle and hence completes our proof. ■

4.2.6 Reducing PSD to Bilinear under Bounded Dual Cotype-2

Proposition 4.15 (PSD reduces to Bilinear under Finite Cotype). *Let $(\mathbb{R}^n,\|\cdot\|_{X^n})_{n\in\mathbb{N}}$ and $(\mathbb{R}^n,\|\cdot\|_{Y^n})_{n\in\mathbb{N}}$ be a sequence of (R,r)-balanced norms satisfying $\sup_n \widetilde{C}_2((X^n)^*), \sup_n \widetilde{C}_2((Y^n)^*) < \infty$. If there is a family $(\mathrm{ALG}_{n,m})_{n,m\in\mathbb{N}}$ of $\mathrm{poly}(n,m)$-time search algorithms α-approximating $\mathrm{Op}^{\max}_{X^n,Y^m}(\cdot)$, then there are families of $\mathrm{poly}(n)$-time search algorithms $4\alpha^2$-approximating PSD quadratic maximization over (X^n) and over (Y^n) respectively.*

Proof. We give an algorithm for PSD quadratic maximization over (X^n) and the the algorithm for (Y^n) is analogous. It is known through a classical result of Figiel Lindenstrauss Milman [FLM77] that for any n and $m \gtrsim \widetilde{C}_2(Y^m)^2\cdot n$, there is an $m\times n$ matrix B such that for all $a\in\mathbb{R}^n$,

$$\frac{1}{\sqrt{2}}\cdot\|a\|_{\ell_2^n} \leqslant \|Ba\|_{(Y^m)^*} \leqslant \sqrt{2}\cdot\|a\|_{\ell_2^n}\,. \tag{43}$$

For any matrix B satisfying (43) any $C\in M_n(\mathbb{R})$, we have $\|BC\|_{X^n\to(Y^m)^*}/\sqrt{2} \leqslant \|C\|_{X\to\ell_2^n} \leqslant \sqrt{2}\|BC\|_{X^n\to(Y^m)^*}$.

Consider any instance $A = C^*C \in \mathbb{PSD}^n$ of PSD quadratic maximization over X^n. Our algorithm will be run the (Y^m, X^n) bilinear maximization algorithm on BC to obtain a vector $x \in \mathrm{Ball}(X^n)$ satisfying

$$\|BCx\|_{(Y^m)^*} \geqslant \mathrm{Op}^{\max}_{Y^m,X^n}(BC)/\alpha = \|BC\|_{X^n \to (Y^m)^*}/\alpha \geqslant \|C\|_{X \to \ell_2^n}/(\sqrt{2}\alpha)\,.$$

x is the desired solution since we have

$$\langle x\,, Ax\rangle = \|Cx\|^2_{\ell_2^n} \geqslant \|BCx\|^2_{(Y^m)^*}/2 \geqslant \|C\|^2_{X \to \ell_2^n}/(4\alpha^2) = \mathrm{Q}^{\max}_X(A)/(4\alpha^2)$$

where the first inequality follows again from (43). ■

In the next section we discuss how the reductions from quadratic and bilinear maximization to PSD quadratic maximization may be viewed as dual to (algorithmic) factorization theorems, leading to a streamlined exposition of Pisier's abstract factorization theorem (with a new approximation algorithm to obtain such a factorization), as well as a new factorization theorem for quadratic forms. These results are not necessary to obtain the applications promised in Section 6 and Section 7, and readers only interested in applications of Theorem 4.11 may skip Section 5 and proceed to Section 6.

5 Factorization through ℓ_2 via Gaussian Rounding + Convex Programming Duality

For norms $\|\cdot\|_E$ over $\mathbb{R}^n$, $\|\cdot\|_F$ over $\mathbb{R}^m$ and an linear operator $A : E \to F$, we define the factorization norm

$$\gamma_2(A) := \inf_{BC=A} \|C\|_{E\to 2} \cdot \|B\|_{2\to F}\,. \tag{44}$$

For any factorization $A = BC$ where $C : E \to \ell_2^d$ and $B : \ell_2^d \to F$, the distortion of A (i.e., $\|A\|_{E\to F}$) is trivially upper bounded by the product of distortions (i.e., $\|C\|_{E\to 2} \cdot \|B\|_{2\to F}$). Thus $\gamma_2(A)$ can be thought of as the best upper bound on the distortion of A when it is viewed as a map from $E \to \ell_2$ and then from $\ell_2 \to F$ (there are infinitely many such factorizations). A "factorization theorem" is an inequality of the form

$$\|A\|_{E\to F} \leqslant \gamma_2(A) \leqslant C(E,F)\cdot \|A\|_{E\to F}\,.^{4}$$

i.e., the upper bound on distortion prescribed by the best factorization of A approximates the actual distortion of A. Surprisingly for a wide class of norms, $C(E,F)$ can be taken independent of the dimension. Factorization theory is a fundamental area of research in Banach space theory with many applications.

In this section we discuss how the proofs of powerful theorems on factorization through ℓ_2 (like those of Grothendieck, Maurey and Pisier) can be viewed as having the following components

1. A dual characterization of the factorization norm as a maximization problem (which is a convex but not necessarily computable relaxation of $\|A\|_{E\to F}$).

[4] Note that the first inequality trivially holds for any E, F.

2. A bound on the "integrality gap" of the maximization problem (i.e., a bound on $\sup_A \gamma_2(A)/\|A\|_{E\to F}$). Such a bound can be obtained by analyzing an appropriate Gaussian rounding scheme that maps an optimal solution of the maximization problem to a vector $e \in \mathrm{Ball}(E), f \in \mathrm{Ball}(F)$.

While step (1.) above seems to have been proved using sophisticated Hahn-Banach arguments previously [5], we show how all of these duality results are special cases of conic Lagrangian duality.[6] We then use this perspective to prove factorization theorems for quadratic maximization which to our knowledge were previously overlooked. En route we also show that the optimal factorization can be computed approximately, assuming access to an oracle for PSD quadratic maximization.

We begin with a discussion of sign-invariant norms (lattices) where factorization takes on a particularly simple form, in order to provide a gentle introduction for a reader encountering these notions for the first time. Readers familiar with factorization theory may wish to skip ahead to Section 5.2.

5.1 Warmup: Sign-Invariant Norms

In this section we discuss factorization in sign-invariant norms (lattices) where factorization takes on a simpler form. We begin by discussing the case of ℓ_∞ and the classical results of Grothendieck.

5.1.1 Grothendieck's Inequality/Factorization as a Motivating Example

Recall for an $n \times m$ matrix A,

$$\|A\|_{\infty\to 1} = \mathrm{Op}^{\max}_{\infty,\infty}(A) = \sup_{|x_i|,|y_j|\leqslant 1} \sum_{i,j} A_{i,j} \cdot x_i \cdot y_j \,.$$

Grothendieck's inequality states that $\|A\|_{\infty\to 1} \leqslant \mathrm{SDP}(A) \leqslant K_G \cdot \|A\|_{\infty\to 1}$ for an absolute constant K_G[7] where the semidefinite programming relaxation $\mathrm{SDP}(A)$ is defined as

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j} A_{i,j} \cdot \langle u_i\,, v_j\rangle \quad \text{s.t.} \\ \text{subject to} \quad & \|u_i\|_2, \|v_j\|_2 \leqslant 1 \quad i \in [n],\ j \in [m] \\ & u_1, \ldots, u_n, v_1, \ldots v_m \in \mathbb{R}^{m+n} \end{aligned} \tag{45}$$

and is equivalent to

$$\begin{aligned} \text{maximize} \quad & \langle A\,, Z\rangle \\ \text{subject to} \quad & \mathbb{X}_{i,i},\ \mathbb{Y}_{j,j} \leqslant 1 \quad i \in [n],\ j \in [m] \\ & \begin{bmatrix} \mathbb{X} & Z \\ Z^* & \mathbb{Y} \end{bmatrix} \succeq 0\,. \end{aligned} \tag{46}$$

[5] We make no attempt to study the infinite dimensional case in this work as it takes us too far afield of the discussion of algorithms.

[6] also known in the literature as Lagrangian duality of convex programming with generalized inequalities

[7] We use K_G to denote the best such constant.

Grothendieck [Gro53] observed that the above inequality implies a beautiful factorization result; namely for any matrix A there exist $\alpha \in \text{Ball}(\ell_2^n)$, $\beta \in \text{Ball}(\ell_2^m)$ such that $\text{Op}^{\max}_{\infty,\infty}(A) \leqslant \text{Op}^{\max}_{2,2}(\widetilde{A}) \leqslant K_G \cdot \text{Op}^{\max}_{\infty,\infty}(A)$ where $\widetilde{A}$ is obtained by reweighting the rows (resp. columns) of A by $(1/\alpha_i)$ (resp. $(1/\beta_i)$) (Tropp [Tro09] discovered an algorithmic application of such a factorization– specifically he gave an algorithm for column subset selection by using these reweightings as measures of importance of the rows/columns of A).

Stated in a way that better motivates the term "factorization", $\|A\|_{\infty\to 1} \leqslant \gamma_2^{\text{Diag}}(A) \leqslant K_G \cdot \|A\|_{\infty\to 1}$ where we define the Grothendieck factorization norm as

$$\gamma_2^{\text{Diag}}(A : \ell_\infty^m \to \ell_1^n) := \inf_{D_1 B D_2 = A} \|D_2\|_{\infty\to 2} \cdot \|B\|_{2\to 2} \cdot \|D_1\|_{2\to 1}$$

where the infimum runs over $n \times m$ matrices B and diagonal matrices D_1, D_2. Note that the first inequality above is simply a consequence of sub-multiplicativity of the operator norm under composition ($\|BC\|_{Z\to X} \leqslant \|C\|_{Z\to Y} \cdot \|B\|_{Y\to X}$). Pietsch observed that Grothendieck's constant is tight for this factorization result by showing the equality $\gamma_2^{\text{Diag}}(A) = \text{SDP}(A)$ using an abstract Hahn-Banach argument (see for e.g. Corollary 23.3 in [Pis12] for an exposition) which in this case involves separation arguments on subsets of $\mathbb{R}^{\mathbb{R}^n}$. Tropp [Tro09] proved this equality by observing that the dual semidefinite program of (46) which is given by

$$\begin{aligned} &\inf \quad (\|t\|_1 + \|s\|_1)/2 \quad \text{s.t.} \\ &\begin{bmatrix} \text{Diag}(s) & -A \\ -A^* & \text{Diag}(t) \end{bmatrix} \succeq 0 \quad s \in \mathbb{R}^n,\ t \in \mathbb{R}^m \end{aligned} \tag{47}$$

can be easily massaged into the form of $\gamma_2^{\text{Diag}}(A)$. Thus by now Grothendick's factorization theorem admits a proof involving only objects of routine occurrence in the optimization literature.

5.1.2 Generalizations to 2-Convex Norms.

Krivine [Kri73] observed that Grothendieck's inequality/factorization generalizes to a wide class of normed spaces called 2-convex Banach Lattices. For simplicity we specialize the discussion from lattices to the class of sign-invariant norms over $\mathbb{R}^n$ (i.e., norms invariant to flipping signs of the entries) as this class is rich enough to demonstrate the ideas we aim to communicate.

Before stating the factorization theorem, we require a definition. In what follows, for a scalar function $s : \mathbb{R} \to \mathbb{R}$ and a vector $x \in \mathbb{R}^n$, we use the notation $s(x)$ to denote the vector obtained by entry-wise application of s to x, i.e., $s(x) = (s(x_1), \ldots, s(x_n))$. For e.g., $|x|^p$ denotes the vector $(|x_1|^p, \ldots |x_n|^p)$. This notation appears exclusively in Section 7.1.1 and Section 5.1.

Definition 5.1 (2-convexity)**.** *Let X be a sign-invariant norm over $\mathbb{R}^n$. Then the 2-convexity constant of X, denoted by $M^{(2)}(X)$, is the smallest constant C such that for every finite sequence of vectors (x_i) in X,*

$$\left\| \left(\sum_i |x_i|^2 \right)^{1/2} \right\|_X \leqslant C \cdot \left(\sum_i \|x_i\|_X^2 \right)^{1/2}$$

where $|\cdot|^2$ and $(\cdot)^{1/2}$ are applied to a vector entry-wise in the left hand expression above. We will say X is exactly 2-convex if $M^{(2)}(X) = 1$.

Theorem 5.2 (Krivine/Grothendieck Factorization Theorem)**.**
For sign-invariant norms X over $\mathbb{R}^n$ and Y over $\mathbb{R}^m$ such that X^* and Y are exactly 2-convex and any operator $A : Y \to X$, it holds that

$$\|A\|_{Y\to X} \leqslant \gamma_2^{\text{Diag}}(A) \leqslant K_G \cdot M^{(2)}(X^*) \cdot M^{(2)}(Y) \cdot \|A\|_{Y\to X}$$

where K_G is Grothendieck's constant and

$$\gamma_2^{\mathrm{Diag}}(A : Y \to X) := \inf_{D_1 B D_2 = A} \|D_2\|_{Y\to 2} \cdot \|B\|_{2\to 2} \cdot \|D_1\|_{2\to X}$$

where the infimum runs over diagonal matrices D_1, D_2.
Combining this with the known equivalence of 2-convexity and dual cotype-2 (see for e.g. [Pis86]) for Banach lattices yields

$$\gamma_2^{\mathrm{Diag}}(A) \leqslant O(\widetilde{C}_2(X)\widetilde{C}_2(Y^*) \log \widetilde{C}_2(X) \log \widetilde{C}_2(Y^*)) \cdot \|A\|_{Y\to X}\,.$$

Here again a proof by Lagrangian duality is available (assuming Grothendick's inequality as a blackbox). Indeed Nesterov [Nes98] independently reproved Grothendieck's inequality (and Krivine's 2-convex extension) in 1997. In this case one compares the objective $\mathrm{Op}^{\max}_{X,Y}(A)$ to

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j} A_{i,j} \cdot \langle u_i\,, v_j\rangle \quad \text{s.t.} \\ \text{subject to} \quad & \left\|(\|u_1\|_2, \ldots, \|u_n\|_2)\right\|_X^2 \leqslant 1 \\ & \left\|(\|v_1\|_2, \ldots, \|v_m\|_2)\right\|_Y^2 \leqslant 1 \\ & u_1, \ldots, u_n, v_1, \ldots v_m \in \mathbb{R}^{m+n} \end{aligned} \tag{48}$$

where the above program is convex precisely when X, Y are exactly 2-convex. Without explicitly noting the connection to factorization theory, Nesterov computed the Lagrangian dual of (48) to be

$$\begin{aligned} &\inf \quad (\mathrm{Op}^{\max}_{X,X}(\mathrm{Diag}(s)) + \mathrm{Op}^{\max}_{Y,Y}(\mathrm{Diag}(t)))/2 \quad \text{s.t.} \\ &\begin{bmatrix} \mathrm{Diag}(s) & -A \\ -A^* & \mathrm{Diag}(t) \end{bmatrix} \succeq 0 \quad s \in \mathbb{R}^n,\ t \in \mathbb{R}^m\,. \end{aligned} \tag{49}$$

(49) is readily massaged into the form of $\gamma_2^{\mathrm{Diag}}(A)$ (see lemma A.7. in [BGG+19] for a proof).

Simply applying the approximate ellipsoid method to (49) yields a previously overlooked but somewhat surprising result which states that in the case of sign-invariant norms whose duals have bounded cotype-2 constant, diagonal instances are the bottleneck for bilinear maximization:

Proposition 5.3 (Algorithm for Cotype-2 Lattice Duals assuming Oracle for Diagonal Instances)**.** *There is an algorithm $\mathrm{ALG}(A, \mathcal{O}_X, \mathcal{O}_Y)$ running in time $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ that on any input $A \in M_{n,m}(\mathbb{R})$ returns a C-approximation to $\mathrm{Op}^{\max}_{X,Y}(\cdot)$ assuming access to an α-approximate search oracle $\mathcal{O}_X$ (resp. $\mathcal{O}_Y$) for quadratic maximization of diagonal instances over X (resp. over Y), where*

$$C \stackrel{\text{def}}{=} \alpha\, K_G\, M^{(2)}(X) M^{(2)}(Y) = O(\alpha \cdot C_2(X^*) \log C_2(X^*) \cdot C_2(Y^*) \log C_2(Y^*))\,.$$

Proof. Using the diagonal-instance oracles, we will exhibit α-approximate separation oracles for the sets

$$S_X \stackrel{\text{def}}{=} \{s \in \mathbb{R}^n \mid s \geqslant 0,\ \mathrm{Op}^{\max}_{X,X}(\mathrm{Diag}(s)) \leqslant 1\}\,, \qquad S_Y \stackrel{\text{def}}{=} \{t \in \mathbb{R}^m \mid t \geqslant 0,\ \mathrm{Op}^{\max}_{Y,Y}(\mathrm{Diag}(t)) \leqslant 1\}\,.$$

Then Proposition 3.12 with

$$C_1 \leftarrow \left\{(s,t) : \begin{bmatrix} \mathrm{Diag}(s) & -A \\ -A^* & \mathrm{Diag}(t) \end{bmatrix} \succeq 0\right\} \text{ and } f(s,t) \leftarrow (\mathrm{Op}^{\max}_{X,X}(\mathrm{Diag}(s)) + \mathrm{Op}^{\max}_{Y,Y}(\mathrm{Diag}(t)))/2$$

implies that we can compute an $(1+o(1))\alpha$-approximation to (49), which by Theorem 5.2 implies the desired approximation to $\mathrm{Op}^{\max}_{X,Y}(\cdot)$.

Now we construct α-approximate separation oracles for S_X and S_Y. We will construct an oracle for S_X and the case of S_Y will be analogous. To this end consider any $s \in \mathbb{R}^n$. Let D be the matrix containing s in the diagonal, and let x be the output of $\mathcal{O}_X(D)$. If $\langle D, xx^* \rangle \leqslant 1$ we return "Inside" and if $\langle D, xx^* \rangle > 1$ we return $\{y \mid \langle y, x\rangle^2 = 1\}$ as a hyperplane separating s from S_X. Indeed since $x \in \mathrm{Ball}(X)$, we have $\langle y, x\rangle^2 \leqslant 1$ for any $y \in S_X$. Lastly since $\langle D, xx^* \rangle \leqslant \mathrm{Op}^{\max}_{X,X}(D) \leqslant \alpha \cdot \langle D, xx^* \rangle$, it is easily checked that the above scheme satisfies all conditions of an α-approximate separation oracle for S_X. ∎

It turns out that factorization results with weaker structure (i.e., factorization through ℓ_2) are available for much more general classes of norms (i.e., without any sign-invariance assumptions like in the case of Grothendieck's factorization). We discuss such factorization results in the next section.

5.2 Factorization Through ℓ_2 without Lattice Assumptions

For norms X over $\mathbb{R}^n$, Y over $\mathbb{R}^m$ and an operator $A : Y \to X$, we define the factorization norm

$$\gamma_2(A) := \inf_{BC=A} \|C\|_{Y\to 2} \cdot \|B\|_{2\to X} . \tag{50}$$

Maurey [Mau74] (extending a result of Kwapien) proved a powerful factorization result

Theorem 5.4 (Maurey Factorization Theorem)**.**

$$\|A\|_{Y\to X} \leqslant \gamma_2(A) \leqslant T_2(Y) \cdot C_2(X) \cdot \|A\|_{Y\to X} .$$

However in light of Grothendieck's factorization theorem for $A : \ell_\infty^m \to \ell_1^n$, one expects a dependence on $\widetilde{C}_2(Y^*)$ instead of $\widetilde{T}_2(Y)$ above. An important result of Pisier [Pis80] bridges Grothendieck's and Maurey's factorization theorems:

Theorem 5.5 (Pisier's Factorization Theorem)**.**

$$\|A\|_{Y\to X} \leqslant \gamma_2(A) \leqslant O(C_2(X^*)C_2(Y)\log C_2(X^*)\log C_2(Y)) \cdot \|A\|_{X\to Y} .$$

Previously constant factor approximation algorithms for bilinear maximization were known only for norms admitting an analogue of Grothendieck's inequality; this is the case for exactly 2-convex norms which simply extend the classical Grothendieck inequality and this is also the case for Schatten-∞ (max singular value) where Naor, Regev, and Vidick [NRV13] algorithmicised Haagerup's proof of the celebrated non-commutative Grothendieck inequality. Curiously, both the classical and non-commutative Grothendieck inequalities are equivalent to results about "factorization through ℓ_2" wherein the factorization satisfies additional structural properties.

We obtain constant factor approximation algorithms for a broader class of norms by utilizing Pisier's factorization theorem which applies to a much more general class of norms than those for which Grothendieck-type inequalities are available. This demonstrates that for the purpose of algorithm design, the factorizations need not have additional structure.

We next show how $\gamma_2(\cdot)$ can be cast as a convex program. We then apply conic Lagrangian duality to this formulation to show that the best constant in a factorization theorem is equal to the integrality gap of an appropriate convex relaxation of $\|A\|_{E\to F}$.

5.2.1 Convex Programming Formulation of $\gamma_2(\cdot)$

In this section, we show that $\gamma_2(\cdot)$ can be written as a convex program. As an additional benefit, we can use the approximate ellipsoid method to show that the optimal factorization can be approximately computed using an oracle for quadratic maximization of PSD instances.

Let A be an $n \times m$ matrix. We will show that the following convex program is a reformulation of $\gamma_2(A : Y \to X^*)$:

$$\begin{aligned} &\inf \quad (\mathrm{Q}_X^{\max}(W_1) + \mathrm{Q}_Y^{\max}(W_2))/2 \quad \text{s.t.} \\ &\begin{bmatrix} W_1 & -A \\ -A^* & W_2 \end{bmatrix} \succeq 0 \quad W_1 \in M_n(\mathbb{R}),\ W_2 \in M_m(\mathbb{R}) \end{aligned} \tag{51}$$

which can be equivalently written as

$$\begin{aligned} &\inf \quad (\|U\|^2_{Y\to 2} + \|V^*\|^2_{2\to X^*})/2 \quad \text{s.t.} \\ &\begin{bmatrix} UU^* & -A \\ -A^* & VV^* \end{bmatrix} \succeq 0 \quad U \in M_n(\mathbb{R}),\ V \in M_m(\mathbb{R})\,. \end{aligned} \tag{52}$$

Lemma 5.6. $\gamma_2(A : Y \to X^*) = $ (52).

Proof. We begin with the observation that

$$\inf_{BEC=A} \|C\|_{Y\to 2} \cdot \|E\|_{2\to 2} \cdot \|B\|_{2\to X^*} \;=\; \inf_{BC=A} \|C\|_{Y\to 2} \cdot \|B\|_{2\to X^*}$$

where LHS $\geqslant$ RHS follows since $\|EC\|_{Y\to 2} \leqslant \|C\|_{Y\to 2} \cdot \|E\|_{2\to 2}$ and LHS $\leqslant$ RHS follows by substituting $E = \mathrm{I}$. Consequently we also know

$$\gamma_2(A) = \inf_{\substack{BEC=A \\ \|E\|_{2\to 2}\leqslant 1 \\ \|B\|_{2\to X^*} = \|C\|_{Y\to 2}}} \|C\|_{Y\to 2} \cdot \|B\|_{2\to X^*} \tag{53}$$

since whenever $A = BEC$ we also have $A = B'E'C'$ where

$$E' \stackrel{\text{def}}{=} \frac{E}{\|E\|_{2\to 2}}, \qquad B' \stackrel{\text{def}}{=} \sqrt{\frac{\|E\|_{2\to 2}\|C\|_{Y\to 2}}{\|B\|_{2\to X^*}}} \cdot B\,, \qquad C' \stackrel{\text{def}}{=} \sqrt{\frac{\|E\|_{2\to 2}\|B\|_{2\to X^*}}{\|C\|_{Y\to 2}}} \cdot C$$

By AM-GM inequality we have

$$\gamma_2(A) \;=\; \inf_{\substack{BEC=A \\ \|E\|_{2\to 2}\leqslant 1}} \|C\|_{Y\to 2} \cdot \|B\|_{2\to X^*} \;\leqslant\; \inf_{\substack{BEC=A \\ \|E\|_{2\to 2}\leqslant 1}} (\|C\|^2_{Y\to 2} + \|B\|^2_{2\to X^*})/2\,. \tag{54}$$

Combining (54) with (53) yields

$$\gamma_2(A) \;=\; \inf_{\substack{BEC=A \\ \|E\|_{2\to 2}\leqslant 1}} (\|C\|^2_{Y\to 2} + \|B\|^2_{2\to X^*})/2\,. \tag{55}$$

Consider an optimal solution to (52). We will show (55) $\leqslant$ (52) . We may assume without loss of generality that U and V are invertible since for any $\varepsilon > 0$, adding $(\varepsilon/\max\{\|\mathrm{I}\|_{Y\to 2}, \|\mathrm{I}\|_{2\to X^*}\}) \cdot \mathrm{I}$ to the block matrix in (52) maintains feasibility while increasing the objective value by no more than ε. Thus the infimum in (52) may be taken over invertible U and V without changing the optimum value. We may now make the following substitution: $B \overset{\text{def}}{=} U$, $C \overset{\text{def}}{=} V^*$ and $E \overset{\text{def}}{=} U^{-1}A(V^{-1})^*$.

Since $BCE = A$, it remains to show that $\|E\|_{2\to 2} \leqslant 1$, and indeed we have,

$$\begin{aligned}
&\begin{bmatrix} UU^* & -A \\ -A^* & VV^* \end{bmatrix} \succeq 0 \\
\Leftrightarrow &\begin{bmatrix} U^{-1} & 0 \\ 0 & V^{-1} \end{bmatrix} \begin{bmatrix} UU^* & -A \\ -A^* & VV^* \end{bmatrix} \begin{bmatrix} (U^{-1})^* & 0 \\ 0 & (V^{-1})^* \end{bmatrix} \succeq 0 \\
\Leftrightarrow &\begin{bmatrix} \mathrm{I} & -E \\ -E^* & \mathrm{I} \end{bmatrix} \succeq 0 \\
\Leftrightarrow &\ \|E\|_{2\to 2} \leqslant 1\,.
\end{aligned} \tag{56}$$

Finally, we need to show (52) $\leqslant$ (55). So fix any optimal solution to (55) and by the same reasoning as above, we may assume B and C are invertible. We then make the substitution $U \overset{\text{def}}{=} B$, and $V \overset{\text{def}}{=} C^*$. Since $BEC = A$, we must have $E := U^{-1}A(V^{-1})^*$. Feasibility of U, V then follows from the chain of equivalences in (56). ■

5.2.2 An Alternate Proof of a Dual Characterization of $\gamma_2(A)$

A classical Hahn-Banach result of Lindenstrauss and Pelczynski [LP68] (see for e.g. Theorem 2.4 in [Pis86] or Theorem 7.3.4 in [AK06] for a detailed exposition) gives a dual characterization of $\gamma_2(A : Y \to X^*)$ as the smallest constant C such that for all finite sequences $(y_i), (\overline{y}_j)$ satisfying $\sum_j \overline{y}_j\overline{y}_j^* \succeq \sum_i y_i y_i^*$, it holds that $\sum_i \|A\, y_i\|_{X^*}^2 \leqslant C^2 \cdot \sum_j \|\overline{y}_j\|_Y^2$. A dual reformulation (see chapter 2 in [Pis86]) of this statement is that the dual norm of $\gamma_2(\cdot)$, which we will denote by $\gamma_2^*(\cdot)$, is given by

$$\gamma_2^*(Z) \overset{\text{def}}{=} \inf_{UVW^*=Z\,,\|V\|_{2\to 2}\leqslant 1} \sqrt{\textstyle\sum_i \|u_i\|_X^2} \cdot \sqrt{\textstyle\sum_i \|w_i\|_Y^2} \tag{57}$$

where u_i (resp. w_i) is the i-th column of U (resp. W)

In this section we give an alternate (perhaps more direct) proof of (57) using conic Lagrangian duality.

Lemma 5.7 (Dual Characterization of $\gamma_2(\cdot)$)**.**
For any linear operator $A : Y \to X^*$, $\gamma_2(A) =$ (38) where (38) is defined as

$$\begin{aligned}
\max \quad & \langle A\,, Z\rangle \\
\text{s.t.} \quad & \mathbb{X} \in \downarrow B_\wedge^{Sym}(X)\,,\ \mathbb{Y} \in \downarrow B_\wedge^{Sym}(Y) \\
& \begin{bmatrix} \mathbb{X} & Z \\ Z^* & \mathbb{Y} \end{bmatrix} \succeq 0\,.
\end{aligned}$$

Proof. We will apply Lagrangian duality to the formulation (51) of $\gamma_2(A)$. Indeed by Lagrangian duality for generalized (conic) inequalities (see theorem 2 on page 224 of [Lue97]; see also section 5.9 in [BV04]), we have

$$\inf_{(W_1,W_2)} \left\{ \mathrm{Q}_X^{\max}(W_1) + \mathrm{Q}_Y^{\max}(W_2) \;\middle|\; \begin{bmatrix} W_1 & -A \\ -A^* & W_2 \end{bmatrix} \succeq 0 \right\} \tag{58}$$

$$= \max_{(\mathbb{X},\mathbb{Y},Z)} \left\{ \inf_{W_1,W_2\succeq 0} \{ \mathrm{Q}_X^{\max}(W_1) + \mathrm{Q}_Y^{\max}(W_2) + 2\langle A, Z\rangle - \langle W_1, \mathbb{X}\rangle - \langle W_2, \mathbb{Y}\rangle \} \;\middle|\; \begin{bmatrix} \mathbb{X} & Z \\ Z^* & \mathbb{Y} \end{bmatrix} \succeq 0 \right\}.$$

where slater's condition and therefore strong duality can be verified by the substitution $W_1 = \|A\|_{2\to 2} \cdot \mathrm{I}$, $W_2 = \|A\|_{2\to 2} \cdot \mathrm{I}$. By (31) we have

$$\inf_{W_1\succeq 0} \{ \mathrm{Q}_X^{\max}(W_1) - \langle W_1, \mathbb{X}\rangle \} = \begin{cases} 0 & \text{if } \mathbb{X} \in \downarrow B_\wedge^{Sym}(X) \\ -\infty, & \text{otherwise}. \end{cases}$$

Thus it follows that

$$\inf_{W_1,W_2\succeq 0} \{ \mathrm{Q}_X^{\max}(W_1) + \mathrm{Q}_Y^{\max}(W_2) + 2\langle A, Z\rangle - \langle W_1, \mathbb{X}\rangle - \langle W_2, \mathbb{Y}\rangle \}$$
$$= \begin{cases} 2\langle Z, A\rangle & \text{if } \mathbb{X} \in \downarrow B_\wedge^{Sym}(X),\ \mathbb{Y} \in \downarrow B_\wedge^{Sym}(Y) \\ -\infty, & \text{otherwise}. \end{cases}$$

Combining this with (58) yields the claim:

$$\inf_{(W_1,W_2)} \left\{ (\mathrm{Q}_X^{\max}(W_1) + \mathrm{Q}_Y^{\max}(W_2))/2 \;\middle|\; \begin{bmatrix} W_1 & -A \\ -A^* & W_2 \end{bmatrix} \succeq 0 \right\}$$
$$= \max_{(\mathbb{X},\mathbb{Y},Z)} \left\{ \langle A, Z\rangle \;\middle|\; \mathbb{X} \in \downarrow B_\wedge^{Sym}(X),\ \mathbb{Y} \in \downarrow B_\wedge^{Sym}(Y),\ \begin{bmatrix} \mathbb{X} & Z \\ Z^* & \mathbb{Y} \end{bmatrix} \succeq 0 \right\}.$$

■

Pisier's factorization theorem [Pis80] is now readily deduced from combining Lemma 5.7 with the fact that (38) approximates $\mathrm{Op}_{X,Y}^{\max}(A)$ (which is implicitly shown in Theorem 4.10 via Gaussian rounding).

Theorem 5.8 (Algorithmic Pisier Factorization Theorem)**.**

1. *For any norms* $(\|\cdot\|_X, \mathbb{R}^n)$, $(\|\cdot\|_Y, \mathbb{R}^m)$ *and any linear map* $A : Y \to X^*$*, we have*

$$\mathrm{Op}_{X,Y}^{\max}(A) \leqslant \gamma_2(A : Y \to X^*) \leqslant O(\widetilde{C}_2(X^*)\widetilde{C}_2(Y^*) \log \widetilde{C}_2(X^*)\widetilde{C}_2(Y^*)) \cdot \mathrm{Op}_{X,Y}^{\max}(A).$$

2. *There is an algorithm* $\mathrm{ALG}(A, R, r, \mathcal{O}_X, \mathcal{O}_Y)$ *such that if* $\mathcal{O}_X$ *(resp.* $\mathcal{O}_Y$*) is an* α*-approximate search oracle for PSD quadratic maximization over an* (R, r)*-balanced norm* $(\|\cdot\|_X, \mathbb{R}^n)$ *(resp.* $(\|\cdot\|_Y, \mathbb{R}^m)$*), then on any input* $A \in M_{n,m}(\mathbb{R})$*,* ALG *runs in time* $\mathrm{poly}(m, n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns a factorization* $A = BC$ *satisfying*

$$\gamma_2(A : Y \to X^*) \leqslant \|C\|_{Y\to 2} \cdot \|B\|_{2\to X^*} \leqslant (1 + o(1)) \cdot \alpha \cdot \gamma_2(A : Y \to X^*).$$

Proof. Claim (1.) (i.e., Pisier's factorization theorem) follows from combining Lemma 5.7 (dual characterization of the factorization norm) with the fact that (38) approximates $\mathrm{Op}_{X,Y}^{\max}(A)$ (which is implicitly shown in Theorem 4.10 via Gaussian rounding).

For claim (2.), we will use the search oracles to construct an α-approximate separation oracle for the following sets.

$$S_X \stackrel{\text{def}}{=} \{ W \succeq 0 \mid \mathrm{Q}_X^{\max}(W) \leqslant 1 \}, \qquad S_Y \stackrel{\text{def}}{=} \{ W \succeq 0 \mid \mathrm{Q}_Y^{\max}(W) \leqslant 1 \}.$$

Then Proposition 3.12 with

$$C_1 \leftarrow \left\{ \begin{bmatrix} W_1 & -A \\ -A^* & W_2 \end{bmatrix} : \mathbb{W} \succeq 0 \right\} \text{ and } f\left(\begin{bmatrix} W_1 & -A \\ -A^* & W_2 \end{bmatrix} \right) \leftarrow \mathrm{Q}_X^{\max}(W_1) + \mathrm{Q}_X^{\max}(W_2)$$

implies that we can compute a $(1+o(1))\alpha$-approximation to (52), whose optimal value is equal to $\gamma_2(A : Y \to X^*)$.

It remains to exhibit approximate separation oracles for S_X and S_Y. We will construct an oracle for S_X and the case of S_Y will be analogous. To this end consider any $W \in M_n(\mathbb{R})$. We may assume W is PSD, since otherwise one can use the separation oracle of the cone of PSD matrices. Let x be the output of $\mathcal{O}_X(W)$. If $\langle W, xx^* \rangle \leqslant 1$ we return "Inside" and if $\langle W, xx^* \rangle > 1$ we return $\{B \mid \langle B, xx^* \rangle = 1\}$ as a hyperplane separating W from S_X. Indeed since $x \in \mathrm{Ball}(X)$, we have $\langle B, xx^* \rangle \leqslant 1$ for any $B \in S_X$. Lastly since $\langle W, xx^* \rangle \leqslant \|W\|_{X \to X^*} \leqslant \alpha \cdot \langle W, xx^* \rangle$, it is easily checked that the above scheme satisfies all conditions of an α-approximate separation oracle for S_X. ■

5.3 Factorization Theorem for Quadratic Maximization under Bounded Type-2

In this section we apply the framework of "Gaussian rounding + conic Lagrangian duality" to obtain a new factorization theorem for quadratic maximization under bounded Type-2. Here also we are able to approximately compute the best factorization relative to an oracle for PSD instances.

To do this we first define an appropriate analogue of $\gamma_2(\cdot)$ in the quadratic case and then give a dual characterization of it.

For a norm $(\|\cdot\|_X, \mathbb{R}^n)$, and a symmetric $n \times n$ matrix A, we define the quadratic factorization semi-norm as

$$\begin{aligned} \gamma_2^Q(A) &:= \inf_{B^*CB=A} \|B\|_{X \to \ell_2^n} \cdot \lambda_{\max}(C) \cdot \|B^*\|_{\ell_2^n \to X^*} \\ &= \inf_{B^*CB=A} \|B\|_{X \to \ell_2^n}^2 \cdot \lambda_{\max}(C) \\ &= \inf_W \{\mathrm{Q}_X^{\max}(W) \mid W \succeq A,\ W \succeq 0\} \end{aligned} \tag{59}$$

where the proof of the third equality is a simpler version of the proof of Lemma 5.6. Analogous to the bilinear (operator norm) case, conic Lagrangian duality yields a dual characterization of $\gamma_2^Q(\cdot)$.

Lemma 5.9 (Dual Characterization of $\gamma_2^Q(\cdot)$)**.**
For any norm $(\|\cdot\|_X, \mathbb{R}^n)$ and any symmetric $n \times n$ matrix A, we have

$$\gamma_2^Q(A) = \max_{\mathbb{X}} \{\langle A, \mathbb{X} \rangle \mid \mathbb{X} \in \downarrow B_{\wedge}^{Sym}(X)\}$$

Proof. By Lagrangian duality for generalized (conic) inequalities (see theorem 2 on page 224 of [Lue97]; see also section 5.9 in [BV04]), we have

$$\begin{aligned} &\inf_W \{\mathrm{Q}_X^{\max}(W) \mid W \succeq A,\ W \succeq 0\} \qquad (60) \\ &= \max_{(\mathbb{X},\mathbb{M})} \left\{ \inf_{W \succeq 0} \{\mathrm{Q}_X^{\max}(W) + \langle A, \mathbb{X} \rangle - \langle W, \mathbb{X} \rangle - \langle W, \mathbb{M} \rangle\} \;\middle|\; \mathbb{X}, \mathbb{M} \succeq 0 \right\}. \end{aligned}$$

where slater's condition and therefore strong duality can be verified by the substitution $W = \lambda_{\max}(A) \cdot \mathrm{I}$. By (31) we have

$$\inf_{W \succeq 0} \{ \mathrm{Q}^{\max}_X(W) - \langle W, \mathbb{X} + \mathbb{M} \rangle + \langle A, \mathbb{X} \rangle \} = \begin{cases} \langle A, \mathbb{X} \rangle & \text{if } (\mathbb{X} + \mathbb{M}) \in \downarrow B^{Sym}_{\wedge}(X) \\ -\infty, & \text{otherwise}. \end{cases}$$

Substituting this in (60) yields

$$\begin{aligned} &\max_{(\mathbb{X},\mathbb{M})} \left\{ \inf_{W \succeq 0} \{ \mathrm{Q}^{\max}_X(W) + \langle A, \mathbb{X} \rangle - \langle W, \mathbb{X} + \mathbb{M} \rangle \} \;\middle|\; \mathbb{X}, \mathbb{M} \succeq 0 \right\} \\ &= \max_{\mathbb{X},\mathbb{M} \succeq 0} \{ \langle A, \mathbb{X} \rangle \mid \mathbb{X} + \mathbb{M} \in \downarrow B^{Sym}_{\wedge}(X) \} \\ &= \max_{\mathbb{X} \succeq 0} \{ \langle A, \mathbb{X} \rangle \mid \mathbb{X} \in \downarrow B^{Sym}_{\wedge}(X) \} \end{aligned}$$

where the final equality follows since $\wedge^{\downarrow Sym}_X(\cdot)$ is monotone in the Loewner ordering (equivalently $\downarrow B^{Sym}_{\wedge}(X)$ is downward-closed w.r.t. $\mathbb{PSD}^n$). This completes the proof. ∎

Combining Lemma 5.9 with Theorem 4.5 yields

Theorem 5.10. *For any norm $(\|\cdot\|_X, \mathbb{R}^n)$ we have*

1. *For any symmetric $n \times n$ matrix A,*

$$\mathrm{Q}^{\max}_X(A) \leqslant \gamma^{Q}_2(A) \leqslant \widetilde{T}_2(X)^2 \cdot \mathrm{Q}^{\max}_X(A).$$

2. *There is an algorithm $\mathrm{ALG}(A, R, r, \mathcal{O})$ such that if $\mathcal{O}$ is an α-approximate search oracle for PSD quadratic maximization over an (R, r)-balanced norm $(\|\cdot\|_X, \mathbb{R}^n)$, then on any input $A \in M_n(\mathbb{R})$, ALG runs in time $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ and returns a factorization $A = B^*CB$ satisfying*

$$\gamma^{Q}_2(A) \leqslant \|B\|^2_{X \to 2} \cdot \lambda_{\max}(C) \leqslant (1 + o(1)) \cdot \alpha \cdot \gamma^{Q}_2(A).$$

Proof. The first claim follows from combining (60) (dual characterization of the factorization norm) with the fact that (37) is a $\widetilde{T}_2(X)^2$-approximation to $\mathrm{Q}^{\max}_X(A)$. The latter fact can be proved via Gaussian rounding (i.e., combining Observation 2.17 and Observation 4.1).

For the second claim, we will use the oracle $\mathcal{O}$ to construct an α-approximate separation oracle for $S_X = \{ W \succeq 0 \mid \mathrm{Q}^{\max}_X(W) \leqslant 1 \}$ (just as in the proof of Theorem 5.8). Then Proposition 3.12 with

$$C_1 \leftarrow \{ W \mid W \succeq A, W \succeq 0 \} \text{ and } f(W) = \mathrm{Q}^{\max}_X(W)$$

implies that we can compute a $(1 + o(1))\alpha$-approximation to (59). ∎

6 Algorithmic Closure Properties of Quadratic/Bilinear Maximization

In this section we show that under bounded type-2 (resp. bounded dual cotype-2), constant factor approximability of quadratic (resp. bilinear) maximization over X (resp. X, Y) is preserved under the following operations: (a) Minkowski Sum (b) Intersection (c) Subspaces (d) Quotients (e) Interpolation (resp. (a) Minkowski Sum (b) Quotients). Formally, we define T_2-Quad-Apx (resp. C_2-Bi-Apx) as the set of norm sequences $(\mathbb{R}^n, \|\cdot\|_{X^n})$ (resp. $(\mathbb{R}^n, \|\cdot\|_{X^n}), (\mathbb{R}^m, \|\cdot\|_{Y^m})$) satisfying the following properties

(C1) $\sup_n \widetilde{T}_2(X^n) < \infty$ (resp. $\sup_n \widetilde{C}_2((X^n)^*), \sup_m \widetilde{C}_2((Y^m)^*) < \infty$).

(C2) X^n (resp. X^n, Y^n) is $(n^{O(1)}, 1/n^{O(1)})$-balanced.

(C3) There is a poly(n) time algorithm computing $\|\cdot\|_{X^n}$ (resp. there is a family of poly(n)-time algorithms $\|\cdot\|_{X^n}, \|\cdot\|_{Y^n}$) within a factor of $1 + 1/\text{poly}(n)$.

(C4) For an absolute constant $C \geqslant 1$, there is a poly(n, bit$A$) time randomized algorithm that on any $n \times n$ input matrix A, with probability at least $1 - n^{-\omega(1)}$ returns a C-approximate solution $x \in \text{Ball}(X^n)$ to $\text{Q}^{\max}_{X^n}(A)$ (resp. there is a family $(\text{ALG}_{n,m})_{n,m\in\mathbb{N}}$ of poly(n, m, bit$A$)-time search algorithms returning a C-approximate solution $(x, y) \in \text{Ball}(X^n) \times \text{Ball}(Y^m)$ to $\text{Op}^{\max}_{X^n,Y^m}(A)$ given an $n \times m$ input matrix A).

The main result of this section is the following theorem (notation and definitions are given in the sequel):

Theorem 6.1 (Algorithmic Closure Properties)**.**
T_2-Quad-Apx is closed under the following operations:
(1a) Minkowski Sum: $(X^n), (\overline{X}^n) \mapsto (X^n + \overline{X}^n)$.
(1b) Intersection: $(X^n), (\overline{X}^n) \mapsto (X^n \vee \overline{X}^n)$.
(1c) Subspaces: $(X^n), (E_n) \mapsto (E_n, \|\cdot\|_{X^n})$.
(1d) Quotients: $(X^n), (E_n) \mapsto (X^n/E_n)$.
(1e) Complex Interpolation: $(\mathbb{C}^n, \|\cdot\|_{X^n}), (\mathbb{C}^n, \|\cdot\|_{\overline{X}^n}) \mapsto ([X^n, \overline{X}^n]_\theta)$.

C_2-Bi-Apx is closed under the following operations:
(2a) Minkowski Sum: $((X^n), (Y^m)), ((\overline{X}^n), (\overline{Y}^m)) \mapsto (X^n + \overline{X}^n), (Y^m + \overline{Y}^m)$.
(2b) Quotients: $((X^n), (Y^m)), (E_n), (F_m) \mapsto (X^n/E_n), (Y^m/F_m)$.

Remark 6.2.

- *Above we assume we are given a polytime algorithm returning a generator matrix for any subspace in the sequence (E_n) (resp. (F_m)).*
- *An interesting point to note above is that while bilinear maximization admits closure properties under the milder assumption that $\widetilde{C}_2(X^*)$ is bounded, quadratic maximization enjoys more closure properties albeit under the stronger assumption of bounded $\widetilde{T}_2(X)$. Indeed bounded $\widetilde{C}_2(X^*)$ bilinear maximization is only closed under Minkowski-sum/quotients whereas bounded $\widetilde{T}_2(X)$ quadratic maximization is closed under Minkowski-sum/quotients as well as the duals of these operations (namely intersection/subspaces). Thus the bounded type-2 assumption is more robust for closure properties.*
- *(1c) and (2b) above are easy to establish and in fact hold without any type-2/cotype-2 assumptions.*

Closure of (C1), (C2) and (C3) under the above operations is immediate (except in the case of complex interpolation where closure of (C2) requires substantial work and was shown in [ANN+18]; closure of (C1) under complex interpolation is well known and follows from Riesz-Thorin interpolation). Thus we need only be concerned with closure of (C4) under all of the above operations. We proceed to prove Theorem 6.1 one operation at a time.

6.1 Minkowski Sums

For norms $\|\cdot\|_{X_1}, \|\cdot\|_{X_2}$ over $\mathbb{R}^n$, we denote by $X_1 + X_2$ and $X_1 \vee X_2$ the norms

$$\|x\|_{X_1+X_2} \stackrel{\text{def}}{=} \inf_{y+z=x} \|y\|_{X_1} + \|z\|_{X_2}, \qquad \|x\|_{X_1\vee X_2} \stackrel{\text{def}}{=} \max\{\|x\|_{X_1}, \|x\|_{X_2}\}$$

which respectively correspond to the Minkowski sum and intersection of $\mathrm{Ball}(X_1)$, $\mathrm{Ball}(X_2)$. It is straightforward to check the duality relation $(X_1 + X_2)^* = X_1^* \vee X_2^*$.

It is easily verified that $\widetilde{T}_2(X_1 + X_2) \leqslant \sqrt{2} \cdot \max\{\widetilde{T}_2(X_1), \widetilde{T}_2(X_2)\}$ and so boundedness of type-2 is preserved under Minkowski sum. Similarly $\widetilde{C}_2((X_1 + X_2)^*) = \widetilde{C}_2(X_1^* \vee X_2^*) \leqslant \sqrt{2} \cdot \max\{\widetilde{C}_2(X_1^*), \widetilde{C}_2(X_2^*)\}$ and so boundedness of dual cotype-2 is preserved under Minkowski sum. In this section we show that approximation algorithms for quadratic (resp. bilinear) maximization are also preserved under Minkowski sum when type-2 (resp. dual cotype-2) is bounded.

Our approach is to show the claim for PSD maximization first (which is immediate) and then appeal to our reduction theorem from quadratic (resp. bilinear) to PSD maximization to obtain the general case. Indeed the following fact is easy to check.

$$\mathrm{Q}^{\max}_{X_1+X_2}(BB^*) = \|B\|^2_{2\to X_1^*\vee X_2^*} = \max\{\|B\|^2_{2\to X_1^*}, \|B\|^2_{2\to X_2}\} = \max\{\mathrm{Q}^{\max}_{X_1}(BB^*), \mathrm{Q}^{\max}_{X_2}(BB^*)\}\,. \quad (61)$$

Combining (61) and Theorem 4.5 immediately yields

Proposition 6.3 (Closure of Type-2 Quadratic Maximization under Minkowski Sum)**.** *There is an algorithm* $\mathrm{ALG}(A, R, r, \mathcal{O}_{X_1}, \mathcal{O}_{X_2})$ *such that if* $\mathcal{O}_{X_1}$ *(resp.* $\mathcal{O}_{X_2}$*) is an* α*-approximate search oracle for quadratic maximization over an* (R, r)*-balanced norm* $(\mathbb{R}^n, \|\cdot\|_{X_1})$ *(resp.* $(\mathbb{R}^n, \|\cdot\|_{X_2})$*), then on any input* $A \in M_n(\mathbb{R})$*,* ALG *runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns a* β*-approximate solution to* $\mathrm{Q}^{\max}_{X_1+X_2}(A)$ *with probability* $1 - 2^{-\Omega(n)}$*, where* $\beta \stackrel{\text{def}}{=} 2\alpha \cdot \max\{\widetilde{T}_2(X_1)^2, \widetilde{T}_2(X_2)^2\}$*.*

The bilinear version follows from combining Proposition 4.15, Theorem 4.11, and (61):

Proposition 6.4 (Closure of Dual Cotype-2 Bilinear Maximization under Minkowski Sum)**.** *Consider any pairs of norm sequences* $((X^n), (Y^m)) \in C_2$*-Bi-Apx and* $((\overline{X}^n), (\overline{Y}^m)) \in C_2$*-Bi-Apx (see Section 6 for the definition of* C_2*-Bi-Apx). Then* $((X^n + \overline{X}^n), (Y^m + \overline{Y}^m)) \in C_2$*-Bi-Apx.*

6.2 Intersection

Recall for norms $\|\cdot\|_{X_1}, \|\cdot\|_{X_2}$ over $\mathbb{R}^n$, the intersection $\mathrm{Ball}(X_1) \cap \mathrm{Ball}(X_2)$ is the unit ball of the norm $X_1 \vee X_2$ defined as

$$\|x\|_{X_1 \vee X_2} \stackrel{\text{def}}{=} \max\{\|x\|_{X_1}, \|x\|_{X_2}\}\,.$$

It is easily verified that $\widetilde{T}_2(X_1 \vee X_2) \leqslant \sqrt{\widetilde{T}_2(X_1)^2 + \widetilde{T}_2(X_2)^2}$ and so boundedness of type-2 is preserved under intersection. We show that approximability of quadratic maximization is closed under intersection by simply observing that the upper covariance body $\mathcal{U}(X_1 \vee X_2)$ is equivalent to the intersection of the upper covariance bodies $\mathcal{U}(X_1)$, $\mathcal{U}(X_2)$. Formally it is easily checked that for any $\mathbb{X} \in \mathbb{PSD}^n$,

$$\max\{\mathcal{N}_{X_1}(\mathbb{X}), \mathcal{N}_{X_1}(\mathbb{X})\} \leqslant \mathcal{N}_{X_1 \vee X_2}(\mathbb{X}) \leqslant \mathcal{N}_{X_1}(\mathbb{X}) + \mathcal{N}_{X_1}(\mathbb{X}) \leqslant 2 \cdot \max\{\mathcal{N}_{X_1}(\mathbb{X}), \mathcal{N}_{X_1}(\mathbb{X})\} \quad (62)$$

where the first inequality follows immediately from Jensen's inequality and the remaining inequalities are straightforward. Thus we have

$$\frac{1}{2} \cdot \mathcal{U}(X_1) \cap \mathcal{U}(X_2) \subseteq \mathcal{U}(X_1 \vee X_2) \subseteq \mathcal{U}(X_1) \cap \mathcal{U}(X_2)\,. \quad (63)$$

We then obtain

Proposition 6.5 (Closure of Type-2 Quadratic Maximization under Intersection).
There is an algorithm $\mathrm{ALG}(A, R, r, \mathcal{O}_{X_1}, \mathcal{O}_{X_2})$ *such that if* $\mathcal{O}_{X_1}$ *(resp.* $\mathcal{O}_{X_2}$*) is an* α*-approximate search oracle for quadratic maximization over an* (R,r)*-balanced norm* $(\mathbb{R}^n, \|\cdot\|_{X_1})$ *(resp.* $(\mathbb{R}^n, \|\cdot\|_{X_2})$*), then on any input* $A \in M_n(\mathbb{R})$*, ALG runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns a* β*-approximate solution to* $\mathrm{Q}^{\max}_{X_1 \vee X_2}(A)$ *with probability* $1 - 2^{-\Omega(n)}$*, where* $\beta \stackrel{\text{def}}{=} 2\alpha \cdot \max\{\widetilde{T}_2(X_1)^2, \widetilde{T}_2(X_2)^2\}$.

Proof. By Proposition 4.13 we obtain $\widetilde{T}_2(X_1)^2$-approximate and $\widetilde{T}_2(X_2)^2$-approximate separation oracles for $\mathcal{U}(X_1)$ and $\mathcal{U}(X_2)$ respectively. Therefore we obtain a $\max\{\widetilde{T}_2(X_1)^2, \widetilde{T}_2(X_2)^2\}$-approximate separation oracle for $\mathcal{U}(X_1) \cap \mathcal{U}(X_2)$. Combining this fact with Observation 3.4 and (63) implies a $2 \cdot \max\{\widetilde{T}_2(X_1)^2, \widetilde{T}_2(X_2)^2\}$-approximate separation oracle for $\mathcal{U}(X_1 \vee X_2)$. Finally applying Proposition 4.2 completes the proof. ■

6.3 Quotients

Let $\|\cdot\|_X$ be a norm over $\mathbb{R}^n$. For a subspace E of $\mathbb{R}^n$, the quotient norm $\|X/E\|_\cdot$ is a norm defined on the space X/E (which can be identified with the orthogonal complement $E^\perp$) and is given by

$$\|x\|_{X/E} \stackrel{\text{def}}{=} \min_{y \in E} \|x - y\|_X$$

i.e., the distance of x to the subspace E. The dual is the norm $\|\cdot\|_{X^*}$ restricted to the subspace $E^\perp$, i.e.,

$$\text{for any } x \in E^\perp, \quad \sup_{\substack{\xi \in E^\perp \\ \|\xi\|_{X^*} \leqslant 1}} \langle \xi, x \rangle = \|x\|_{X/E} \qquad \text{and for any } \xi \in E^\perp, \quad \sup_{\substack{x \in E^\perp \\ \|x\|_{X/E} \leqslant 1}} \langle x, \xi \rangle = \|\xi\|_{X^*}.$$

More generally we say a surjective linear map $T : \mathbb{R}^n \to \mathbb{R}^k$ (where $k < n$) is an (a,b)-quotient map from X to Q if $\|T\|_{X \to Q} \leqslant a$ and $\mathrm{Ball}(Q) \subseteq b \cdot T\mathrm{Ball}(X)$. It is easily checked that for any $x \in \mathbb{R}^n$, $a^{-1} \cdot \|T(x)\|_Q \leqslant \|x\|_{X/\ker T} \leqslant b \cdot \|T(x)\|_Q$. An obvious dual transposition implies that for any $x \in \ker T^\perp$,

$$b^{-1} \cdot \|T(x)\|_{Q^*} \leqslant \|x\|_{X^*} \leqslant a \cdot \|T(x)\|_{Q^*}. \tag{64}$$

It is easily verified that $\widetilde{T}_2(Q) \leqslant ab \cdot \widetilde{T}_2(X)$ and so quotients of X inherit boundedness of type-2. Similarly $\widetilde{C}_2(Q^*) \leqslant ab \cdot \widetilde{C}_2(X*)$ and so quotients of X inherit boundedness of dual cotype-2. In this section we show that approximation algorithms for quadratic (resp. bilinear) maximization are also preserved under quotienting when type-2 (resp. dual cotype-2) is bounded.

Our approach again is to show the claim for PSD maximization first (which is immediate) and then obtain the general case by appealing to our reduction theorem from quadratic (resp. bilinear) to PSD maximization.

Given access to an oracle for T one can compute the linear map $T^\dagger : \mathbb{R}^k \to \ker T^\perp$ corresponding to the inverse of $T|_{\ker T^\perp}$.

Fact 6.6. *Consider any PSD matrix* $A = BB^* \in \mathbb{PSD}^k$*. Then we have*

$$\begin{aligned} &\frac{1}{b} \cdot \|B\|_{2 \to Q^*} \leqslant \|T^\dagger B\|_{2 \to X^*} \leqslant a \cdot \|B\|_{2 \to Q^*} && \textit{(by (64))} \\ \Rightarrow\ &\frac{1}{b^2} \cdot \mathrm{Q}^{\max}_Q(A) \leqslant \mathrm{Q}^{\max}_X(T^\dagger A (T^\dagger)^*) \leqslant a^2 \cdot \mathrm{Q}^{\max}_Q(A). \end{aligned}$$

Thus if x is an α-approximate solution to $\mathrm{Q}_X^{\max}(T^\dagger A(T^\dagger)^*)$, then $T(x)$ is an $\alpha\cdot a^2b^2$-approximate solution to $\mathrm{Q}_Q^{\max}(A)$.

Combining Fact 6.6 and Theorem 4.5 immediately yields

Proposition 6.7 (Closure of Type-2 Quadratic Maximization under Quotienting)**.**
There is an algorithm $\mathrm{ALG}(A,R,r,\mathcal{O}_X,\mathcal{O}_T)$ such that if $\mathcal{O}_X$ is an α-approximate search oracle for quadratic maximization over a norm $(\mathbb{R}^n,\|\cdot\|_X)$, $\|\cdot\|_Q$ is an (R,r)-balanced norm over $\mathbb{R}^k$ and $\mathcal{O}_T$ is an oracle computing an (a,b)-quotient map $T: X\to Q$, then on any input $A\in M_k(\mathbb{R})$, ALG runs in time $\mathrm{poly}(n,\log R,\log 1/r,\mathrm{bit}(A))$ and returns a β-approximate solution to $\mathrm{Q}_Q^{\max}(A)$ with probability $1-2^{-\Omega(n)}$, where $\beta=\alpha\cdot a^2b^2\cdot\widetilde{T}_2(Q)^2\leqslant\alpha\cdot a^4b^4\cdot\widetilde{T}_2(X)^2$.

The bilinear case follows (without any type-2/cotype-2 assumptions) from the following straightforward generalization of Fact 6.6.

Fact 6.8 (Closure of Bilinear Maximization under Quotienting)**.**
Consider norms $(\mathbb{R}^n,\|\cdot\|_X),(\mathbb{R}^m,\|\cdot\|_Y),(\mathbb{R}^k,\|\cdot\|_P),(\mathbb{R}^\ell,\|\cdot\|_Q)$ admitting (a,b)-quotient maps $S:\mathbb{R}^n\to\mathbb{R}^k$ from X to P and $T:\mathbb{R}^m\to\mathbb{R}^\ell$ from Y to Q. Consider any $k\times\ell$ matrix A. Then for $S^\dagger,T^\dagger$ defined as above, we have

$$\frac{1}{b}\cdot\|A^*\|_{P\to Q^*}\leqslant\|T^\dagger A^*\|_{P\to Y^*}\leqslant a\cdot\|A^*\|_{P\to Q^*}\qquad\text{(by (64))}$$

and

$$\frac{1}{b}\cdot\|A(T^\dagger)^*\|_{Y\to P^*}\leqslant\|S^\dagger A(T^\dagger)^*\|_{Y\to X^*}\leqslant a\cdot\|A(T^\dagger)^*\|_{Y\to P^*}\qquad\text{(by (64))}$$

Combining the above four inequalities (repeatedly) with the fact that $\mathrm{Op}_{E,F}^{\max}(M)=\|M\|_{F\to E^*}=\|M^*\|_{E\to F^*}$, yields

$$\frac{1}{b^2}\cdot\mathrm{Op}_{P,Q}^{\max}(A)\leqslant\mathrm{Op}_{X,Y}^{\max}(S^\dagger A(T^\dagger)^*)\leqslant a^2\cdot\mathrm{Op}_{P,Q}^{\max}(A).$$

Thus if (x,y) is an α-approximate solution to $\mathrm{Op}_{X,Y}^{\max}(S^\dagger A(T^\dagger)^*)$, then $(S(x),T(y))$ is an $\alpha\cdot a^2b^2$-approximate solution to $\mathrm{Q}_Q^{\max}(A)$.

6.4 Complex Interpolation

6.4.1 Interpolation Preliminaries

Let $\mathcal{S}\stackrel{\text{def}}{=}\{z\in\mathbb{C}\mid\mathrm{Re}(z)\in(0,1)\}$ be the complex unit open strip, let $\partial\mathcal{S}$ denote its boundary and lastly let $\overline{\mathcal{S}}\stackrel{\text{def}}{=}\mathcal{S}\cup\partial\mathcal{S}$ denote the closed unit strip.

Let $\mathcal{F}$ be the space of bounded continuous functions $f:\overline{\mathcal{S}}\to\mathbb{C}^n$ that are holomorphic in $\mathcal{S}$. Given complex norms $(\mathbb{C}^n,\|\cdot\|_{X_0})$ and $(\mathbb{C}^n,\|\cdot\|_{X_1})$, and a parameter $0<\theta<1$, the complex interpolant of X_0 and X_1 is defined to be the norm $(\mathbb{C}^n,\|\cdot\|_{[X_0,X_1]_\theta})$ given by

$$\|x\|_{[X_0,X_1]_\theta}\stackrel{\text{def}}{=}\inf_{\substack{f\in\mathcal{F}\\ f(\theta)=x}}\max\left\{\sup_{\mathrm{Re}(z)=0}\|f(z)\|_{X_0},\ \sup_{\mathrm{Re}(z)=1}\|f(z)\|_{X_1}\right\}.\qquad(65)$$

Whenever the source and destination are labeled as X_0 and X_1 we will use the shorthand X_θ to denote the interpolant $[X_0,X_1]_\theta$.

In this section we will be interested in quadratic and bilinear maximization over interpolants X_θ. We obtain C-approximation algorithms for a constant C depending on $T_2(X_0)$ and $T_2(X_1)$. We use the shorthand $(X,Y)_\theta$ to denote the interpolant $[(X_0,Y_0),(X_1,Y_1)]_\theta$.

6.4.2 Approximation Algorithms for Interpolants of Type-2 Norms

The main result of this section is that if quadratic maximization is computationally approximable over type-2 norms $(\mathbb{C}^n, \|\cdot\|_{X_0})$ and $(\mathbb{C}^n, \|\cdot\|_{X_1})$, then it is also approximable over any interpolant of X_0 and X_1:

Theorem 6.9 (Quadratic Maximization over Interpolants of Type-2 Norms.)**.**
Fix any $\theta \in (0,1)$. There is an algorithm $\mathrm{ALG}(A, R, r, \mathcal{O}_{X_0}, \mathcal{O}_{X_1})$ such that if $\mathcal{O}_{X_0}, \mathcal{O}_{X_1}$ are α-approximate search oracles for quadratic maximization over (R,r)-balanced norms $(\mathbb{C}^n, \|\cdot\|_{X_0}), (\mathbb{C}^n, \|\cdot\|_{X_1})$ respectively, then ALG runs in time $\mathrm{poly}(n, R, 1/r, \mathrm{bit}(A))$ and returns a C-approximate solution to $\mathrm{Q}^{\max}_{X_\theta}(A)$ with probability $1 - 2^{-\Omega(n)}$, where $C = (1 + o(1)) \cdot \alpha \cdot \max_{i\in\{0,1\}}\{\widetilde{T}_2(X_i)^4\}$.

Here again our approach is to reduce quadratic maximization to bilinear maximization and then appeal to a factorization result of Kouba in order to solve the bilinear problem. Indeed Theorem 6.9 follows from combining Theorem 4.5 with the following theorem

Proposition 6.10 (Bilinear Maximization over Interpolants of Type-2 Norms.)**.**
Fix any $\theta \in (0,1)$. There is an algorithm $\mathrm{ALG}(A, \mathcal{O}_{(X_0,Y_0)}, \mathcal{O}_{(X_1,Y_1)})$ such that if $\mathcal{O}_{(X_0,Y_0)}$ (resp. $\mathcal{O}_{(X_1,Y_1)}$) is and α-approximate search oracle for bilinear maximization over the pair of (R,r)-balanced norms $(\mathbb{C}^n, \|\cdot\|_{X_0})$, $(\mathbb{C}^m, \|\cdot\|_{Y_0})$ (resp. $(\mathbb{C}^n, \|\cdot\|_{X_1}), (\mathbb{C}^m, \|\cdot\|_{Y_1})$), then ALG runs in time $\mathrm{poly}(n, m, R, 1/r, \mathrm{bit}(A))$ and returns a C-approximation to $\mathrm{Op}^{\max}_{X_\theta,Y_\theta}(A)$, where $C = (1 + o(1)) \cdot \alpha \cdot \max_{i\in\{0,1\}}\{\widetilde{T}_2(X_i)^2 \cdot \widetilde{T}_2(Y_i)^2\}$.
Moreover ALG returns a witness $Z \in \mathrm{Ball}(X_\theta \widehat{\otimes} Y_\theta)$ satisfying $\langle A, Z\rangle \geqslant \mathrm{Op}^{\max}_{X_\theta,Y_\theta}(A)/C$.

6.4.3 Proof of Proposition 6.10

Our approach is to use a result of Kouba [Kou91] who extended a factorization result of Pisier for matrix valued analytic functions in order to obtain conditions under which $\|\cdot\|_{[(X_0,Y_0),(X_1,Y_1)]_\theta}$ and $\mathrm{Op}^{\max}_{X_\theta,Y_\theta}(\cdot)$ are equivalent.

Theorem 6.11 (Kouba [Kou91])**.** *Let X_0, X_1 (resp. Y_0, Y_1) be norms over $\mathbb{C}^n$ (resp. $\mathbb{C}^m$). Then for any $A \in M_{n,m}(\mathbb{C})$ we have*

$$\|A\|_{(X,Y)_\theta} \leqslant \mathrm{Op}^{\max}_{X_\theta,Y_\theta}(A) \leqslant \max_{i\in\{0,1\}}\{\widetilde{T}_2(X_i)^2 \cdot \widetilde{T}_2(Y_i)^2\} \cdot \|A\|_{(X,Y)_\theta}$$

$$\|A\|_{(X\widehat{\otimes}Y)_\theta} \geqslant \|A\|_{X_\theta\widehat{\otimes}Y_\theta} \geqslant \min_{i\in\{0,1\}}\{\widetilde{T}_2(X_i)^{-2} \cdot \widetilde{T}_2(Y_i)^{-2}\} \cdot \|A\|_{(X\widehat{\otimes}Y)_\theta}$$

Armed with this equivalence, we need only approximate $(X,Y)_\theta \overset{\text{def}}{=} [(X_0,Y_0),(X_1,Y_1)]_\theta$ in order to prove Proposition 6.10. Andoni et al. [ANN+18] show that given membership oracles for norms W_0, W_1, W_θ can be computed using the ellipsoid method. We use their result with the substitution $W_0 \overset{\text{def}}{=} X_0\widehat{\otimes}Y_0$ and $W_1 \overset{\text{def}}{=} X_1\widehat{\otimes}Y_1$[8] with the caveat that we only have approximate separation oracles for W_0, W_1 and so we instead use the approximate ellipsoid method.

To this end note that $\|\cdot\|_{(X\widehat{\otimes}Y)_\theta} \overset{\text{def}}{=} \|\cdot\|_{[X_0\widehat{\otimes}Y_0, X_1\widehat{\otimes}Y_1]_\theta}$ is the dual norm of $\|\cdot\|_{(X,Y)_\theta}$ and so

$\|A\|_{(X,Y)_\theta}$

[8] We interpolate the projective norms along with the injective norms as the reduction from quadratic maximization to psd maximization (Theorem 4.5) requires a 'witness' in order to apply.

$$= \sup_{\|Z\|_{(X\widehat{\otimes}Y)_\theta}\leqslant 1} \langle A\,,Z\rangle$$

$$= \sup\left\{\langle A\,,Z\rangle \mid \exists f\in\mathcal{F}\ s.t.\ f(\theta)=Z,\ \ \forall t\in\mathbb{R},\ \|f(i\cdot t)\|_{X_0\widehat{\otimes}Y_0}, \|f(1+i\cdot t)\|_{X_1\widehat{\otimes}Y_1}\leqslant 1\right\} \tag{66}$$

where $\mathcal{F}$ is the set of matrix valued functions holomorphic on the strip $\overline{\mathcal{S}}$.

The following result of [ANN+18] states that $\mathcal{F}$ can be replaced with a set of functions having only polynomially many non-zero fourier coefficients, thereby enabling us to formulate (66) (approximately) as a convex program in polynomially many variables.

Lemma 6.12 (Lemma 5.11 + Claim 5.15 in Andoni et al.)**.**
Fix any $C\geqslant 1$. Then there exists $C'\geqslant 1$ and $M,N,R\leqslant n^{C'}$ such that for any $Z\in \mathrm{Ball}((X\widehat{\otimes}Y)_\theta)$, there is a sequence of matrices $(v_q)_{q\in\mathbb{Q}_M}$ in $M_{n,m}(\mathbb{C})$ such that

$$\begin{aligned}
&i.\ \ \forall z\in\mathbb{D}_N^{(0)}\ \ \|f_V(z)\|_{X_0\widehat{\otimes}Y_0}\leqslant 1+1/n^C\\
&ii.\ \forall z\in\mathbb{D}_N^{(1)}\ \ \|f_V(z)\|_{X_1\widehat{\otimes}Y_1}\leqslant 1+1/n^C\\
&iii.\ \forall q\in\mathbb{Q}_M\ \ \|v_q\|_H\cdot\max\{e^q,1\}\leqslant n^{C'}\\
&iv.\ \ \|f_V(\theta)-Z\|_H\leqslant 1/n^C
\end{aligned}$$

$$\text{where } f_V(z)\stackrel{\text{def}}{=} e^{z^2/M}\cdot\sum_{q\in\mathbb{Q}_M} v_q\cdot e^{qz}\,. \tag{67}$$

Furthermore if for some $Z\in M_n(\mathbb{C})$, a sequence (v_q) satisfying (67) exists, then $\|Z\|_{(X\widehat{\otimes}Y)_\theta}\leqslant 1+1/n$.

We thus apply the approximate ellipsoid method to the following program

$$\begin{aligned}
\text{maximize}\quad &\langle A\,,Z\rangle\quad\text{such that}\\
&i.\ \ \forall z\in\mathbb{D}_N^{(0)}\ \ \|f_V(z)\|_{W_0}\leqslant 1+1/n^C\\
&ii.\ \forall z\in\mathbb{D}_N^{(1)}\ \ \|f_V(z)\|_{W_1}\leqslant 1+1/n^C\\
&iii.\ \forall q\in\mathbb{Q}_M\ \ \|v_q\|_H\cdot\max\{e^q,1\}\leqslant n^{C'}\\
&iv.\ \ \|f_V(\theta)-w\|_H\leqslant 1/n^C\\
\text{where } f_V(z)\stackrel{\text{def}}{=}\ &e^{z^2/M}\cdot\sum_{q\in\mathbb{Q}_M} v_q\cdot e^{qz}\\
&Z\,,v_q\in M_{n,m}(\mathbb{C})\,.
\end{aligned} \tag{68}$$

We are ready to prove our bilinear maximization result.

Proof of Proposition 6.10. By Theorem 3.14, $\mathcal{O}_{(X_0,Y_0)}$ (resp. $\mathcal{O}_{(X_1,Y_1)}$) can be adapted to obtain a poly-time α-approximate separation oracle for $\mathrm{Ball}(X_0\widehat{\otimes}Y_0)$ (resp. $\mathrm{Ball}(X_1\widehat{\otimes}Y_1)$). Therefore by applying Proposition 3.11 to the program (68) (where $B\in M_{n,m}(\mathbb{C})$ is treated as an object in $(\mathbb{R}^2)^{n\times m}$), we obtain $Z\in M_{n,m}(\mathbb{C})$ such that $\langle A\,,Z\rangle=(1-o(1))\cdot\|A\|_{(X,Y)_\theta}$ and $(1-o(1))Z/\alpha$ satisfies the conditions of (67). Thus by Lemma 6.12 $\|Z\|_{(X\widehat{\otimes}Y)_\theta}\leqslant\alpha(1+o(1))$ and furthermore by Theorem 6.11 (the trivial direction), $\|Z\|_{X_\theta\widehat{\otimes}Y_\theta}\leqslant\alpha(1+o(1))$. Again by Theorem 6.11 (the non-trivial direction), $\langle A\,,Z\rangle\geqslant(1-o(1))\cdot\min_{i\in\{0,1\}}\{\widetilde{T}_2(X_i)^{-2}\cdot\widetilde{T}_2(Y_i)^{-2}\}\cdot\mathrm{Op}^{\max}_{X_\theta,Y_\theta}(A)$. This completes the proof since $(1-o(1))Z/\alpha$ is the desired witness. ∎

6.5 Sections (Subspaces)

A section of a origin-symmetric convex body C is defined as the intersection of C with a subspace V. Let X be the norm whose ball is C and let E denote the norm on the space V whose unit ball is $C \cap V$. The corresponding quadratic maximization problem over E is to compute given a linear map $A : V \to V$, the following quantity

$$\mathrm{Q}_E^{\max}(A) \stackrel{\text{def}}{=} \sup_{x \in C \cap V} \langle x, A(x) \rangle .$$

It is easy to reduce $\mathrm{Q}_E^{\max}(\cdot)$ to $\mathrm{Q}_X^{\max}(\cdot)$. In fact, this holds without any type-2 assumption on X – a fact that will come in handy for hardness reductions in Section 8. Indeed we have

Observation 6.13 (Enforcing a Subspace Constraint)**.**
Let A be an $n \times n$ matrix, $V \subseteq \mathbb{R}^n$ be a subspace and X be a norm over $\mathbb{R}^n$ such that $r \cdot \mathrm{Ball}(\ell_2^n) \subseteq \mathrm{Ball}(X)$. Let Π be the projector to $V^\perp$ and let $\alpha \stackrel{\text{def}}{=} \mathrm{Q}_X^{\max}(A')/(n\|A'\|_{X \to X^*})$. Let $A' \stackrel{\text{def}}{=} A - (r\alpha)^{-2} \cdot \mathrm{Q}_X^{\max}(A) \cdot \Pi$. Then we have

$$\sup_{x \in \mathrm{Ball}(X) \cap V} \langle x, Ax \rangle \leqslant \mathrm{Q}_X^{\max}(A') \leqslant (1 + O(1/n)) \cdot \sup_{x \in \mathrm{Ball}(X) \cap V} \langle x, Ax \rangle$$

Proof. The first inequality is immediate. For the second inequality, consider any $x \in \mathrm{Ball}(X)$ that maximizes $\langle x, A'x \rangle$. Decompose x as $x^{\|} + x^\perp$ where $x^\perp$ is the component of x in $V^\perp$ and $x^{\|}$ is the component in V. We may assume $\|x^\perp\|_2 \leqslant r\alpha$ since otherwise $\langle x, A'x \rangle < 0$ which would contradict optimality of x. This implies that $\|x^\perp\|_X \leqslant \alpha$ and therefore $\|x^{\|}\|_X \leqslant 1 + \alpha \leqslant 1 + 1/n$. Finally we have

$$\begin{aligned}
\mathrm{Q}_X^{\max}(A') = \langle x, A'x \rangle &= \langle x^{\|}, A'x^{\|} \rangle + \langle x^{\|}, A'x^\perp \rangle + \langle x^\perp, A'x^{\|} \rangle \langle x^\perp, A'x^\perp \rangle \\
&\leqslant \langle x^{\|}, A'x^{\|} \rangle + (2(1+\alpha)\alpha + \alpha^2) \cdot \|A'\|_{X \to X^*} \\
&\leqslant \langle x^{\|}, A'x^{\|} \rangle + 4 \cdot \mathrm{Q}_X^{\max}(A')/n \\
&= (1 + O(1/n)) \langle x^{\|}, A'x^{\|} \rangle / \|x^{\|}\|_X^2 + 4 \cdot \mathrm{Q}_X^{\max}(A')/n \\
&= (1 + O(1/n)) \langle x^{\|}, Ax^{\|} \rangle / \|x^{\|}\|_X^2 + 4 \cdot \mathrm{Q}_X^{\max}(A')/n
\end{aligned}$$

as desired. ∎

7 Unconditional Algorithms for Special Families

In this section we study various special families of norms for which one can design separation oracles for the lower covariance region $\mathcal{L}(X)$ using only a membership oracle for $\mathrm{Ball}(X)$. Thus using Theorem 4.11 we obtain approximation algorithms for quadratic/bilinear maximization that require only a membership oracle for $\mathrm{Ball}(X)$. In doing so we recover constant factor approximation algorithms for the cases studied previously in the literature and also obtain new results. We obtain constant factor algorithms for the following norm families (see Section 7.1.1 and Section 7.3.1 for definitions):

1. Quadratic maximization over exactly 2-convex sign-invariant norms with bounded q-concavity for finite q. This recovers a result of Naor and Schechtman [NS09].

2. Bilinear maximization over exactly 2-convex sign-invariant norms. This recovers (with a worse constant) a result implicit in the work of Krivine [Kri73] and independently rediscovered by Nesterov [Nes98].

3. Quadratic (resp. bilinear) maximization over symmetric norms (i.e., invariant to permutation and sign changes) that have bounded type-2 (resp. bounded dual cotype-2). This includes many new examples not covered in the exactly 2-convex case.

4. Quadratic (resp. bilinear) maximization over unitarily invariant matrix norms that have bounded type-2 (resp. bounded dual cotype-2). This recovers as a special case (with a worse constant) a result of Naor, Regev and Vidick [NRV13] for bilinear maximization over Schatten-∞.

7.1 Approximation Algorithms for Sign Invariant Norms

We require some preliminaries.

7.1.1 p-convexity and q-concavity Preliminaries

The notions of p-convexity and q-concavity are well defined for a wide class of normed spaces known as Banach lattices. In this document we only define these notions for finite dimensional norms that are 1-unconditional in the elementary basis (i.e., those norms $\|\cdot\|_X$ for which flipping the sign of an entry of x does not change the norm. We shall refer to such norms as *sign-invariant norms*). Most of the statements we make in this context can be readily extended to the case of norms admitting some 1-unconditional basis, but we choose to fix the elementary basis in the interest of simplicity.

In what follows, for a scalar function $s : \mathbb{R} \to \mathbb{R}$ and a vector $x \in \mathbb{R}^n$, we use the notation $s(x)$ to denote the vector obtained by entry-wise application of s to x, i.e., $s(x) = (s(x_1), \ldots, s(x_n))$. For e.g., $|x|^p$ denotes the vector $(|x_1|^p, \ldots |x_n|^p)$. This notation appears exclusively in Section 7.1.1 and Section 5.

Definition 7.1 (p-convexity/q-concavity)**.** *Let X be a sign-invariant norm over $\mathbb{R}^n$. Then for $1 \leqslant p \leqslant \infty$ the p-convexity constant of X, denoted by $M^{(p)}(X)$, is the smallest constant C such that for every finite sequence of vectors (x_i) in X,*

$$\left\| \left(\sum_i |x_i|^p \right)^{1/p} \right\|_X \leqslant C \cdot \left(\sum_i \|x_i\|_X^p \right)^{1/p}$$

We will say X is exactly p-convex if $M^{(p)}(X) = 1$.

For $1 \leqslant q \leqslant \infty$, the q-concavity constant of X, denoted by $M_{(q)}(X)$, is the smallest constant C such that for every finite sequence of vectors $\{x_i\}$ in X,

$$\left\| \left(\sum_i |x_i|^q \right)^{1/q} \right\|_X \geqslant \frac{1}{C} \cdot \left(\sum_i \|x_i\|_X^q \right)^{1/q}.$$

We will say X is exactly q-concave if $M_{(q)}(X) = 1$.

Remark 7.2. *Every sign-invariant norm is exactly 1-convex and exactly ∞-concave. It is also known (see Lindenstrauss-Tzafriri) that any sign-invariant norm X is C-equivalent to an exactly p-convex norm (resp. an exactly q-concave norm) for C depending only on $M^{(p)}(X)$ (resp. $M_{(q)}(X)$).*

Fact 7.3. *Let X be a sign-invariant norm over* $\mathbb{R}^n$*. Then whenever* $|[x]| \leqslant |[y]|$ *(entry-wise), it must be that* $\|x\|_X \leqslant \|y\|_X$*.*

For a sign-invariant norm $\|\cdot\|_X$ over $\mathbb{R}^n$, and any $0 < p < \infty$ let $\|\cdot\|_{X^{(p)}}$ denote the function $\||x|^p\|_X^{1/p}$. $X^{(p)}$ is referred to as the p-convexification of X. It is easily verified that $M^{(p)}(X^{(p)}) = M^{(1)}(X)$ and further that $\|\cdot\|_{X^{(p)}}$ is an exactly p-convex sign-invariant norm if and only if $\|\cdot\|_X$ is an exactly 1-convex sign-invariant norm.

We will appeal to a known equivalence between (2-convexity + $(q<\infty)$-concavity) and Type-2 for Banach lattices that is implicit in the following Khintchine-type inequality.

Theorem 7.4 (Banach Lattice Khintchine [Mau73])**.**
Let X be a sign-invariant norm over R^n*. Then for any* $1 \leqslant p \leqslant 2 \leqslant q < \infty$ *and any finite sequence* (x_i) *and i.i.d. standard Gaussians* (g_i)*,*

$$\frac{\gamma_p}{M^{(p)}(X)} \cdot \left\| \left(\sum_i |x_i|^2 \right)^{1/2} \right\|_X \leqslant \sqrt{\mathbb{E}\left[\|\sum_i \mathrm{g}_i \cdot x_i\|_X^2 \right]} \leqslant \gamma_q \cdot M_{(q)}(X) \cdot \left\| \left(\sum_i |x_i|^2 \right)^{1/2} \right\|_X$$

where $\gamma_p \stackrel{\text{def}}{=} \mathbb{E}_{\mathrm{g}\sim\mathcal{N}(0,1)}\left[|g|^p\right]^{1/p}$*.*

Remark 7.5.

1. *If X is exactly 2-convex then the constant in the left inequality can be taken as 1 since in this case* $\gamma_2 / M^{(2)}(X) = 1$*.*
2. *If X is exactly 2-concave then the constant in the right inequality can be taken as 1 since in this case* $\gamma_2 \cdot M_{(2)}(X) = 1$*.*
3. *Since X is a sign-invariant norm,* $M^{(1)}(X) = 1$ *and so the constant in the left inequality can always be taken as* $\gamma_1 = \sqrt{2/\pi}$*.*
4. *Since X is a norm,* $M_{(\infty)}(X) = 1$ *and so* $M_{(\log n)}(X) \leqslant e$*. Thus the constant in the right inequality can always be taken as* $e \cdot \gamma_{\log n} = (1+o(1))\sqrt{e \log n}$*.*

Finally, we record the well known equivalence between 2-convexity and dual cotype-2 (see for e.g. [Pis86]) for Banach lattices:

$$\widetilde{C}_2(X^*) \lesssim M^{(2)}(X) \lesssim \widetilde{C}_2(X^*) \log \widetilde{C}_2(X^*) \,.$$

7.1.2 Approximation Algorithms for Maximization over Exactly 2-Convex Norms

Naor and Schectman [NS09] gave an $M_{(q)}(X)^2 \cdot \gamma_q^2$-approximation algorithm for quadratic maximization over an exactly 2-convex norm X. We show how our framework for quadratic maximization (Proposition 4.2) recovers their result.

Theorem 7.6 (Quadratic Maximization over Exactly 2-convex Norms)**.**
There is an algorithm $\mathrm{ALG}(A, R, r, \mathcal{O}_X)$ *such that if* $\mathcal{O}_X$ *is an oracle for exactly computing a sign-invariant, exactly 2-convex,* (R, r)*-balanced norm* $(\mathbb{R}^n, \|\cdot\|_{X_1})$*, then on any input* $A \in M_n(\mathbb{R})$*, ALG runs in time* $\mathrm{poly}(n, \log R, \log 1/r, \mathrm{bit}(A))$ *and returns a* β*-approximate solution to* $\mathrm{Q}_X^{\max}(A)$ *with probability* $1 - 2^{-\Omega(n)}$*, where* $\beta \stackrel{\text{def}}{=} \inf_{2 \leqslant q < \infty} M_{(q)}(X)^2 \cdot \gamma_q^2$*.*

Proof. For the domain $\mathbb{X} \in M_n(\mathbb{R})$, $\mathbb{X} \succeq 0$, we define a computable convex function f as

$$f(\mathbb{X}) \stackrel{\text{def}}{=} \|\text{diag}(\mathbb{X})\|_{X^{(1/2)}} = \|\text{diag}(\mathbb{X})^{1/2}\|_X^2 = \left\| \left(\sum_i |x_i|^2 \right)^{1/2} \right\|_X^2 \quad (69)$$
$$\text{for any decomposition } \mathbb{X} = \sum_{k \in [K]} x_i (x_i)^* .$$

Convexity is evident from the first equality above since exact 2-convexity of X implies $X^{(1/2)}$ is a norm. Computability (using $\mathcal{O}_X$) is evident from the second equality above. Theorem 7.4 implies that $\mathcal{N}_X(\mathbb{X})$ and $f(\mathbb{X})$ are β-equivalent for the claimed β. Thus combining part (1.) of Corollary 4.12 and Theorem 7.4 yields the claim. ∎

Krivine [Kri73] observed that Grothendieck's inequality extends (with the same constant K_G) to the case of bilinear maximization over exactly 2-convex lattices. Nesterov [Nes98] independently rediscovered Grothendieck's inequality (with constant ~ 2.29) and also noted its extension to the exactly 2-convex case.

We give an alternate proof of this result (with a worse constant) using our framework (Theorem 4.11).

Theorem 7.7 (Bilinear Maximization over Exactly 2-convex Norms)**.**
There is an absolute constant $\beta > 1$ and an algorithm $\text{ALG}(A, R, r, \mathcal{O}_X, \mathcal{O}_Y)$ *such that if $\mathcal{O}_X$ (resp. $\mathcal{O}_Y$) is an oracle for exactly computing a sign-invariant, exactly* 2*-convex,* (R, r)*-balanced norm* $(\mathbb{R}^n, \|\cdot\|_X)$ *(resp.* $(\mathbb{R}^m, \|\cdot\|_Y)$*), then on any input $A \in M_{n,m}(\mathbb{R})$,* ALG *runs in time* $\text{poly}(n, m, \log R, \log 1/r, \text{bit}(A))$ *and returns a β-approximate solution to* $\text{Op}^{\max}_{X,Y}(A)$ *with probability* $1 - 2^{-\Omega(n+m)}$.

Proof. For the domain $\mathbb{W} \in M_n(\mathbb{R})$, $\mathbb{W} \succeq 0$, we define a computable concave function f as

$$f(\mathbb{W}) \stackrel{\text{def}}{=} \|\text{diag}(\mathbb{W})\|_{(X^*)^{(1/2)}} = \|\text{diag}(\mathbb{W})^{1/2}\|_{X^*}^2 = \left\| \left(\sum_i |x_i|^2 \right)^{1/2} \right\|_{X^*}^2 \quad (70)$$
$$\text{for any decomposition } \mathbb{W} = \sum_{k \in [K]} x_i (x_i)^* .$$

Concavity is evident from the first equality above since exact 2-convexity of X implies $(X^*)^{(1/2)}$ is concave. Computability (using $\mathcal{O}_X$) is evident from the second equality above. Theorem 7.4 implies that $\mathcal{N}_{X^*}(\mathbb{W})$ and $f(\mathbb{W})$ are equivalent within a universal constant. We proceed similarly for Y. Thus combining Theorem 7.4 and part (B) of part (2.) of Corollary 4.12 yields the claim. ∎

7.2 Approximation Algorithms for Symmetric Norms

In this section we will use Theorem 4.11 to give constant factor approximation algorithms for quadratic (resp. bilinear) maximization over symmetric norms with bounded type-2 (resp. dual cotype-2) constant (assuming only oracles computing the norms being optimized over).

To do so, we will design a separation oracle for the lower covariance region. We begin with a technical ingredient common to all our proofs - namely that linear optimization over symmetric downward/upward-closed subsets of the non-negative orthant can be done efficiently (assuming only an approximate membership oracle for the set).

7.2.1 Preliminaries: Linear Optimization over Symmetric Upward/Downward-closed Sets

The following lemma shows that one can efficiently minimize linear functions over a symmetric upward closed subset of the non-negative orthant. This will be used in constructing a separation oracle for the lower covariance region of a symmetric norm.

Lemma 7.8. *Let $K = (\mathbb{R}^n_{\geqslant 0})$ be the positive orthant, and let $B \subseteq K$ be a, permutation symmetric, upward-closed, inverse (R, r, K)-balanced, and exactly convex body with a C-approximate membership oracle $\mathcal{O}$. Then for any $\varepsilon > 0$, there is a $\mathrm{poly}(n, R/r, 1/\log \varepsilon, \mathrm{bit}(z))$-time algorithm* ALG *that takes a vector $z \in \mathbb{R}^n_{\geqslant 0}$ as input and outputs y' with $y' \in B$ such that*

$$\inf_{y \in B} \langle z\,, y\rangle \leqslant \langle z, y' \rangle \leqslant (C \cdot (1+\varepsilon)) \inf_{y \in B} \langle z\,, y\rangle$$

Proof. Without loss of generality, assume $r = 1$. If $z_i = 0$ for some $i \in [n]$, the optimum is 0 since $R \cdot e_i \in B$, and we are done. Otherwise, by sign-invariance, and scaling, we may assume z and the optimal vector $y^* \in B$ satisfy $1 = z_1 \leqslant \ldots \leqslant z_n \leqslant n^c$, and $y^*_1 \geqslant \ldots \geqslant y^*_{n'} = 1 > y^*_{n'+1} = \cdots = y^*_n = 0$. (I.e., the smallest nonzero entry of y^* is 1.) Then $\mathsf{OPT} = \langle z, y^* \rangle \geqslant 1$, and since $R \cdot e_1 \in B$, $\mathsf{OPT} \leqslant R$. This implies that $y^*_1 \leqslant R$ and $z_{n'} \leqslant R$. Let z and y^* be their projections to the first n' coordinates. (We do not know n' in advance, but one can guess it with n tries.) We perform the following operations to covert z and y^* to nicer forms while approximately preserving $\langle z, y^* \rangle$ and $y^* \in B$.

- Round up each entry of z_i to the smallest $(1+\varepsilon)^t$ for some integer $t \geqslant 0$. The value $\langle z, y^* \rangle$ increases multiplicatively by a factor at most $(1+\varepsilon)$. Let $L = \lceil \log_{1+\varepsilon} R \rceil$ and Let $S_0, \ldots, S_L \subseteq [n']$ be the set of coordinates such that $S_i := \{j \in [n'] : z_j = (1+\varepsilon)^i\}$. Note that $S_0, \ldots, S_L$ partition $[n']$ and they are "consecutive" in the sense that S_0 contains the first $|S_0|$ coordinates, S_1 contains the next $|S_1|$ coordinates, $\ldots$, and S_L contains the last $|S_L|$ coordinates.
- For $i \in \{0, \ldots, L\}$, let $a_i := (\sum_{j \in S_i} y^*_j)/|S_i|$, and change y^* by setting $y^*_j = a_i$ for all $j \in S_i$ for all $i \in \{0, \ldots, L\}$. Since the new y^* can be written as a convex combination of coordinate permutations of the old y^*, the convexity of B implies the new y^* satisfies $y^* \in B$. The value $\langle z, y^* \rangle$ does not change.
- Similarly to z, round up each entry of y^*_i to the largest power of $(1+\varepsilon)^t$ for some $t \geqslant 0$. The value $\langle z, y^* \rangle$ increases multiplicatively by a factor at most $(1+\varepsilon)$. Since we only increased coordinates of y^*, by upward-closedness of B, y^* is still in B.

Therefore, for both z and y^*, the coordinates in the same S_i have the same value, which is of the form $(1+\varepsilon)^t$ for some $t \in \{1, \ldots, L\}$. The objective function $\langle z, y^* \rangle$ is at most $(1+\varepsilon)^2 OPT$, and $y \in B$. Note that we do not know y^*, but we can compute the new z and $S_0, \ldots, S_L$ given the original z and correctly guessed n'.

Finally, we exhaustively search the coordinate values for y^*. Since we search for at most $L+1$ different values and the number of possible values is also at most $L+1$, the number of possible choices is bounded by $2^{O(L)} = \mathrm{poly}(R)$. Once we find the correct y^*, the C-approximate membership oracle will at least correctly certify that $Cy^* \in B$. The returned value is $\langle Cy^*, z \rangle \leqslant C(1+O(\varepsilon))OPT$. ∎

The next lemma states that one can efficiently maximize a linear function over a downward-closed subset of the non-negative orthant.

Lemma 7.9. *Let Y be a sign-invariant, symmetric (under permutation), quasi-norm over $\mathbb{R}^n$ with quasi-triangle constant C and an exact membership oracle $\mathcal{O}$. Then for any $\varepsilon > 0$, there is an* $\text{poly}(n^{O(1/\varepsilon)}, \text{bit}(z))$*-time algorithm* ALG *that takes a vector $z \in \mathbb{R}^n$ as input and outputs $y' \in \text{Ball}(Y)$ such that*

$$\frac{\sup_{y\in\text{Ball}(Y)}\langle z\,,y\rangle}{(1+o(1))\cdot C\cdot(1+\varepsilon)} \leqslant \langle z, y'\rangle \leqslant \sup_{y\in\text{Ball}(Y)} \langle z\,,y\rangle$$

Proof. By symmetry and sign-invariance we may assume z as well as the optimal vector y^* have non-negative entries sorted in descending order. Without loss of generality we may assume $z_1 = 1$ and $\|e_1\|_Y = 1$, which implies that $y_1^* \leqslant 1$. Let $OPT := \langle y^*, z\rangle$. Since $y = e_1$ is a feasible solution with $\langle y, z\rangle = 1$, it also implies that $OPT \geqslant 1$. We perform the following operations to covert z and y^* to nicer forms while approximately preserving $\langle z, y^*\rangle$.

- For any $i \in [n]$, if z_i or y_i^* is less than $1/n^3$, make it 0. The value $\langle z, y^*\rangle$ decreases additively by at most $n^2/n^3 = 1/n \leqslant OPT/n$.
- Round down each entry of z_i to the largest power of $(1-\varepsilon)^t$ for some $t \geqslant 0$ while keeping 0 as 0. The value $\langle z, y^*\rangle$ decreases multiplicatively by a factor at most $1/(1-\varepsilon)$. Let $L = \lceil \log_{1/(1-\varepsilon)} n^3 \rceil$ and Let $S_0, \ldots, S_L \subseteq [n]$ be the set of coordinates such that $S_i := \{j \in [n] : z_j = (1-\varepsilon)^i\}$. Let $S_z := \{j \in [n] : z_j = 0\}$. Note that $S_0, \ldots, S_L, S_z$ partition $[n]$ and they are "consecutive" in the sense that S_0 contains the first $|S_0|$ coordinates, S_1 contains the next $|S_1|$ coordinates, …, and S_z contains the last $|S_z|$ coordinates.
- For $i \in \{0, \ldots, L\}$, let $a_i := (\sum_{j\in S_i} y_j^*)/|S_i|$, and change y^* by setting $y_j^* = a_i$ for all $j \in S_i$ for all $i \in \{0, \ldots, L\}$. Since the new y^* can be written as a convex combination of coordinate permutations of the old y^*, the quasi-triangle inequality implies that $\|y^*\|_Y$ is increased by a factor at most C. The value $\langle z, y^*\rangle$ does not change.
- Similarly to z, round down each entry of y_i^* to the largest power of $(1-\varepsilon)^t$ for some $t \geqslant 0$ while keeping 0 as 0. The value $\langle z, y^*\rangle$ decreases multiplicatively by a factor at most $1/(1-\varepsilon)$, and $\|y^*\|_Y$ does not increase.

Therefore, for both z and y^*, the coordinates in the same S_i have the same value, which is of the form $(1-\varepsilon)^t$ for some $t \in \{0, \ldots, L\}$ or 0. The objective function $\langle z, y^*\rangle$ is at least $(1 - 1/n)(1-\varepsilon)^2 OPT$, and $\|y^*\|_Y \leqslant C$. Note that we do not know y^*, but we can compute the new z and $S_0, \ldots, S_L, S_z$ given the original z.

Finally, we exhaustively search the coordinate values for y^*. Since we search for at most $L+2$ different values and the number of possible values is also at most $L+2$, the number of possible choices is bounded by $2^{O(L)} = n^{O(1/\varepsilon)}$. Once we fine the correct y^*, we can divide by C so that the $\|y^*\|_Y \leqslant 1$. The returned value of $\langle y^*, z\rangle \geqslant \frac{(1-1/n)(1-2\varepsilon)}{C} OPT$. ■

The following lemma shows that one can reduce linear function maximization over a unitarily invariant subset of the positive semidefinite cone to an instance of linear function maximization over a symmetric subset of the non-negative orthant. This will be used in constructing a separation oracle for the lower covariance region of a unitarily invariant norm.

Lemma 7.10. *Let $\Lambda \in (\mathbb{R}_{\geqslant 0})^n$ be a permutation symmetric set, and let $\mathcal{B}$ be the set of PSD matrices whose eigenvalue sequence $(y_1, \ldots, y_n)$ belongs to Λ. Let A be an $n \times n$ PSD matrix, and let $z_1 \geqslant \ldots \geqslant z_n \geqslant 0$*

be its eigenvalues. Then

$$\inf_{B\in\mathcal{B}}\langle A,B\rangle = \inf_{y\in\Lambda}\langle z,y\rangle, \quad \textit{and} \quad \sup_{B\in\mathcal{B}}\langle A,B\rangle = \sup_{y\in\Lambda}\langle z,y\rangle,$$

and the optimal $B \in \mathcal{B}$ can be efficiently computed given the optimal solution $y \in \Lambda$.

Proof. Let $A = \sum z_i u_i u_i^T$ be its eigendecomposition for some orthonormal basis $\{u_i\}$. For any $y \in \Lambda$, $B = \sum_i y_i u_i u_i^T \in \mathcal{B}$ satisfies $\langle A,B\rangle = \sum z_i y_i$, so $\inf_B\langle A,B\rangle \leqslant \inf_y\langle z,y\rangle$ and $\sup_B\langle A,B\rangle \geqslant \sup_y\langle z,y\rangle$. To prove the other direction, take any $B \in \mathcal{B}$ and let $B = \sum_i y_i u_i u_i^T$ be its eigendecomposition with $y \in \Lambda$ with $y_1 \geqslant \ldots \geqslant y_n \geqslant 0$. Since $\langle A,B\rangle = \sum_{i,j} z_i y_j \langle v_i,u_j\rangle^2$ and both $\{u_i\}$ and $\{v_i\}$ are orthonormal bases, $\sum_j\langle v_i,u_j\rangle^2 = 1$ and $\sum_j\langle v_j,u_i\rangle^2 = 1$. Therefore, $\langle A,B\rangle$ is maximized when $v_i = u_i$ and minimized $v_i = u_{n+1-i}$. Therefore, $\langle A,B\rangle \in [\sum_i z_i y_{n+1-i}, \sum_i z_i y_i] \in [\inf_y\langle z,y\rangle, \sup_y\langle z,y\rangle]$. ■

7.2.2 Symmetric Type-2 Norms

In this section we will give an unconditional constant factor approximation algorithm for quadratic (resp. bilinear) maximization over symmetric norms with bounded type-2 (resp. dual cotype-2) constant.

We begin by constructing a separation oracle for the lower covariance region $\mathcal{L}(X)$ of a symmetric norm $\|\cdot\|_X$.

Lemma 7.11 (Lower Covariance Separation Oracle for a Symmetric Norm with Bounded Dual Cotype-2)**.** *For any (R,r)-balanced symmetric norm $\|\cdot\|_X$ over $\mathbb{R}^n$ the lower covariance region $\mathcal{L}(\mathsf{S}_X)$ has a* poly$(n, \log R, \log 1/r, x)$ *time α-approximate separation oracle with input x, assuming access to an oracle $\mathcal{O}_X$ computing $\|\cdot\|_X$, where $\alpha \lesssim M^{(2)}(X)^4 = O(\widetilde{C}_2(X^*)^4\log^4(\widetilde{C}_2(X^*))$.*

Proof. Let F be the renorming of X that is symmetric, exactly 2-convex and satisfies $\|x\|_F \leqslant \|x\|_X \leqslant M^{(2)}(X)\cdot\|x\|_X$. We know F^* is exactly 2-concave and therefore $\|\cdot\|_{(F^*)^{(1/2)}}$ is a concave function on $\mathbb{R}^n_{\geqslant 0}$ (we abuse notation and use $\|\cdot\|_{(F^*)^{(1/2)}}$ even though it is not a norm). Similar to the proof of Theorem 7.7 we define a concave function $g : \mathbb{PSD}^n \to \mathbb{R}_{\geqslant 0}$ and its associated convex level set L_g as follows

$$\begin{aligned} g(\mathbb{W}) &\stackrel{\text{def}}{=} \|[\mathrm{diag}(\mathbb{W})^{1/2}\|_{F^*}^2 = \|\mathrm{diag}(\mathbb{W})\|_{(F^*)^{(1/2)}} \\ L_g &\stackrel{\text{def}}{=} \{\mathbb{W} \succeq 0 \mid g(\mathbb{W}) \geqslant 1\}. \end{aligned}$$

Theorem 7.4 implies that $\mathcal{N}_{F^*}(\mathbb{W})$ and $g(\mathbb{W})$ are equivalent within a factor of $\pi/2$. Thus by Observation 3.4 it suffices to give a C-separation oracle for L_g since this would imply a $C\cdot\pi/2$-approximate separation oracle for $\mathcal{L}(F)$ and therefore a $C\cdot M^{(2)}(X)^2\cdot\pi/2$-approximate separation oracle for $\mathcal{L}(X)$.

Observe further that it suffices to design a separation oracle for the convex set $L_{(F^*)^{(1/2)}} \stackrel{\text{def}}{=} \{\xi \in \mathbb{R}^n_{\geqslant 0} \mid \|\xi\|_{(F^*)^{(1/2)}} \geqslant 1\}$. Indeed consider any $\mathbb{W} \in \mathbb{PSD}^n \setminus L_g/C$ (if $\mathbb{W} \not\succeq 0$ then we may use the separation oracle of $\mathbb{PSD}^n$). By definition $\mathrm{diag}(\mathbb{W}) \in \mathbb{R}^n_{\geqslant 0} \setminus L_{(F^*)^{(1/2)}}/C$ and so there exists a hyperplane $\{\xi \mid \langle h, \xi\rangle = 1\}$ that separates $\mathrm{diag}(\mathbb{W})$ from $L_{(F^*)^{(1/2)}}$. Thus the hyperplane $\{M \mid \langle \mathrm{Diag}(h), M\rangle = 1\}$ separates $\mathbb{W}$ from L_g.

To obtain a separation oracle for $L_{(F^*)^{(1/2)}}$, we combine Lemma 7.8 (i.e., $(1+\varepsilon)M^{(2)}(X)^2$-approximate linear minimization can be done over $L_{(F^*)^{(1/2)}}$) with Theorem 3.16 (i.e., linear minimization over an upward-closed set implies a separation oracle). In order to ensure linear minimization runs in $\text{poly}(n)$ time, we need to ensure that $L_{(F^*)^{(1/2)}}$ is inverse $(R, r, \mathbb{R}^n_{\geqslant 0})$-balanced for $R/r = \text{poly}(n)$. This can be verified from the following simple inequalities:

$$\|\xi\|_\infty \leqslant \frac{\|\xi\|_{(F^*)^{(1/2)}}}{\|e_1\|_{F^*}} \leqslant n^2 \cdot \|\xi\|_\infty \quad \forall \xi \in \mathbb{R}^n_{\geqslant 0}$$

where the first inequality uses monotonicity of $\|\cdot\|_{F^*}$ (in the entry-wise ordering) and the second inequality follows from triangle inequality. Thus we obtain an $O(M^{(2)}(X)^2)$-approximate separation oracle for $L_{(F^*)^{(1/2)}}$ running in time $\text{poly}(n)$. ∎

Finally we obtain

Theorem 7.12 (Maximization over Symmetric Norms under Type-2/Dual Cotype-2)**.**
There are algorithms $\text{ALG}_1(A_1, \mathcal{O}_X)$ *and* $\text{ALG}_2(A_2, \mathcal{O}_X, \mathcal{O}_Y)$ *such that if* $\mathcal{O}_X$ *(resp.* $\mathcal{O}_Y$*) is an oracle computing a symmetric norm* $(\mathbb{R}^n, \|\cdot\|_X)$ *(resp.* $(\mathbb{R}^m, \|\cdot\|_Y)$*), then*

- *Quadratic: on any input* $A_1 \in M_n(\mathbb{R})$, ALG_1 *runs in time* $\text{poly}(n, \text{bit}(A_1))$ *and returns an* α*-approximate solution to* $\text{Q}^{\max}_{\mathsf{S}_X}(A_1)$ *with probability* $1 - 2^{-\Omega(n)}$*, where*

$$\alpha \lesssim T_2(X)^2 \cdot \widetilde{C}_2(X^*)^4 \cdot \log^4 \widetilde{C}_2(X^*) \leqslant \widetilde{T}_2(X)^6 \log^4 \widetilde{T}_2(X)\,.$$

- *Bilinear: on any input* $A_2 \in M_{nd,mh}(\mathbb{R})$, ALG_2 *runs in time* $\text{poly}(n, m, \text{bit}(A_2))$ *and returns a* β*-approximate solution to* $\text{Op}^{\max}_{\mathsf{S}_X,\mathsf{S}_Y}(A_2)$ *with probability* $1 - 2^{-\Omega(n)}$*, where*

$$\beta \lesssim \max\{\widetilde{C}_2(X^*)^6 \cdot \log^5 \widetilde{C}_2(X^*)\ ,\ \widetilde{C}_2(Y^*)^6 \cdot \log^5 \widetilde{C}_2(Y^*)\}\,.$$

Proof. Without loss of generality we may assume $\|e_1\|_X = \|e_1\|_Y = 1$ since otherwise we can just renormalize. $\|\cdot\|_X$ (resp. $\|\cdot\|_Y$) is always $(n^{O(1)}, n^{-\Omega(1)})$-balanced (resp. $(m^{O(1)}, m^{-\Omega(1)})$-balanced) since

$$\|x\|_\infty \leqslant \frac{\|x\|_X}{\|e_1\|_X} \leqslant n \cdot \|x\|_\infty \quad \forall x \in \mathbb{R}^n$$

and $\|e_1\|_X = 1$ by assumption.

Combining Theorem 4.11 with Lemma 7.11 yields the desired theorem. ∎

Remark 7.13. *Combining Proposition 5.3 and Lemma 7.9 yields an alternate proof of Theorem 7.12, achieving the approximation factor* $(1+\varepsilon) \cdot K_G \cdot M^{(2)}(X) \cdot M^{(2)}(Y) \cdot \max\{M^{(2)}(X)^2, M^{(2)}(Y)^2\}$*. In fact the factor can be improved to* $(1+\varepsilon) \cdot K_G \cdot M^{(2)}(X) \cdot M^{(2)}(Y)$ *if one allows non-uniform algorithms with quasipolynomial runtime.*

We chose to present the above proof via Theorem 4.11 to highlight the generality of our approach and furthermore because it generalizes well to the unitarily invariant case.

7.3 Approximation Algorithms for Unitarily Invariant Norms

In this section we will use Theorem 4.11 to give constant factor approximation algorithms for quadratic (resp. bilinear) maximization over unitarily-invariant norms with bounded type-2 (resp. dual cotype-2) constant (assuming only oracles computing the norms being optimized over). To do so, we will design a separation oracle for the lower covariance region, for which we first require some additional preliminaries.

7.3.1 Preliminaries: Non-Commutative Khintchine Inequality

Let E be a symmetric norm over $\mathbb{R}^n$ (i.e., invariant to permutation and flipping signs of entries). For a matrix $M \in M_{n,d}(\mathbb{R})$ we define the norm $\|M\|_{\mathsf{S}_E} \stackrel{\text{def}}{=} \|\sigma(M)\|_E$ where $\sigma(M)$ denotes the vector of singular values of M (when $d \geqslant n$ we assume $\sigma(M)$ returns only the n largest singular values of M – the rest are of course 0's). We use the shorthand C_p for $C_{\ell_p^n}$.

For a finite sequence of matrices $(M_k)_{k=1}^K$ in $M_{n,d}(\mathbb{R})$, we define the norms (we refer the reader to [LPP91] for a more detailed discussion of the these norms and their associated properties quoted here):

$$\begin{aligned}
\|(M_k)\|_{\mathsf{S}_E(\ell_R^2)} &\stackrel{\text{def}}{=} \left\|\sqrt{\textstyle\sum_k M_k^* M_k}\right\|_{\mathsf{S}_E} \\
\|(M_k)\|_{\mathsf{S}_E(\ell_L^2)} &\stackrel{\text{def}}{=} \left\|\sqrt{\textstyle\sum_k M_k M_k^*}\right\|_{\mathsf{S}_E} \\
\|(M_k)\|_{\mathsf{S}_E(\ell_R^2)\vee\mathsf{S}_E(\ell_L^2)} &\stackrel{\text{def}}{=} \max\{\|(M_k')\|_{\mathsf{S}_E(\ell_R^2)}, \|(M_k'')\|_{\mathsf{S}_E(\ell_L^2)}\} \\
\|(M_k)\|_{\mathsf{S}_E(\ell_R^2)+\mathsf{S}_E(\ell_L^2)} &\stackrel{\text{def}}{=} \inf_{(M_k)=(M_k')+(M_k'')} \|(M_k')\|_{\mathsf{S}_E(\ell_R^2)} + \|(M_k'')\|_{\mathsf{S}_E(\ell_L^2)}\,.
\end{aligned}$$

For a finite sequence of matrices $(\overline{M}_k)_{k=1}^K$ in $M_{n,d}(\mathbb{R})$, we define the inner product $\langle (M_k)\,,(\overline{M}_k)\rangle \stackrel{\text{def}}{=} \sum_k \langle M_k\,,\overline{M}_k\rangle = \sum_k \mathrm{Tr}(M_k\overline{M}_k^*)$. It is a standard fact that $\mathsf{S}_E(\ell_R^2)+\mathsf{S}_E(\ell_L^2)$ and $(\mathsf{S}_{E^*}(\ell_R^2)\vee\mathsf{S}_{E^*}(\ell_L^2)$ are dual to each other, i.e.,

$$\|(\overline{M}_k)\|_{(\mathsf{S}_E(\ell_R^2)+\mathsf{S}_E(\ell_L^2))^*} = \sup_{\|(M_k)\|_{\mathsf{S}_E(\ell_R^2)+\mathsf{S}_E(\ell_L^2)}\leqslant 1} \langle (M_k)\,,(\overline{M}_k)\rangle = \|(\overline{M}_k)\|_{\mathsf{S}_{E^*}(\ell_R^2)\vee\mathsf{S}_{E^*}(\ell_L^2)}\,. \tag{71}$$

We require the following matrix khintchine inequality due to Lust-Piquard and Xu [LPX07]

Theorem 7.14 (2-concave Matrix Khintchine Inequality)**.**
There is an absolute constant $C > 1$ such that for any exactly 2-concave symmetric norm E and any finite sequence of matrices $(M_k)_{k=1}^K$ in $M_{n,d}(\mathbb{R})$ we have

$$\frac{1}{C}\cdot\|(M_k)\|_{\mathsf{S}_E(\ell_R^2)+\mathsf{S}_E(\ell_L^2)} \leqslant \sqrt{\mathbb{E}\left[\|\textstyle\sum_k \mathrm{g}_k\cdot M_k\|_{\mathsf{S}_E}^2\right]} \leqslant \|(M_k)\|_{\mathsf{S}_E(\ell_R^2)+\mathsf{S}_E(\ell_L^2)}\,.$$

where $(\mathrm{g}_k)_{k=1}^K$ is a sequence of i.i.d. standard Gaussians.

We define linear maps $T_L : \mathbb{PSD}^{n\cdot d} \to M_n(\mathbb{R})$ (resp. $T_R : \mathbb{PSD}^{n\cdot d} \to M_d(\mathbb{R})$) as

$$\begin{aligned}
(T_R(\mathbb{X}))[i,j] &\stackrel{\text{def}}{=} \sum_{\ell\in[n]} \mathbb{X}[(\ell,i),(\ell,j)] \\
(T_L(\mathbb{X}))[i,j] &\stackrel{\text{def}}{=} \sum_{\ell\in[n]} \mathbb{X}[(i,\ell),(j,\ell)]
\end{aligned}$$

Note that for any decomposition $\mathbb{X} = \sum_{k\in[K]} \mathrm{vec}(M_k)\,\mathrm{vec}(M_k)^*$ one has

$$\begin{aligned}
T_R(\mathbb{X}) &= \sum_{k\in[K]} M_k^* M_k \\
T_L(\mathbb{X}) &= \sum_{k\in[K]} M_k M_k^*
\end{aligned}$$

where $\mathrm{vec}(H)$ denotes the (row-wise) nd-dimensional vector associated with the $n\times d$ matrix H.

7.3.2 Unitarily Invariant Type-2 Matrix Norms

In this section we will give an unconditional constant factor approximation algorithm for quadratic (resp. bilinear) maximization over unitarily invariant matrix norms with bounded type-2 (resp. dual cotype-2) constant.

We first construct a separation oracle for the lower covariance region of a unitarily invariant norm.

Lemma 7.15 (Separation Oracle for $\mathcal{L}(\mathsf{S}_E)$)**.**
For any (R,r)-balanced unitarily invariant matrix norm $\|\cdot\|_{\mathsf{S}_E}$ over $M_{n,d}(\mathbb{R})$ the lower covariance region $\mathcal{L}(\mathsf{S}_E)$ has a $\mathrm{poly}(n,d,\log R,\log 1/r,\mathrm{bit}(x))$ time α-approximate separation oracle on input x, assuming access to an oracle $\mathcal{O}_{\mathsf{S}_E}$ computing $\|\cdot\|_{\mathsf{S}_E}$, where $\alpha = O(M^{(2)}(E)^6) = O(\widetilde{C}_2(E^*)^6\log^6(\widetilde{C}_2(E^*))$.

Proof. Let F be the renorming of E that is symmetric, exactly 2-convex and satisfies $\|x\|_F \leqslant \|x\|_E \leqslant M^{(2)}(E)\cdot\|x\|_F$. For $\mathbb{W}\in\mathbb{PSD}^{n\cdot d}$, we define a concave function g non-decreasing in the Loewner ordering as follows

$$g(\mathbb{W}) \overset{\text{def}}{=} \sup_{(M_k)} \|(M_k)\|_{\mathsf{S}_{F^*}(\ell_R^2)+\mathsf{S}_{F^*}(\ell_L^2)} \tag{72}$$

where the supremum runs over all decompositions

$$\mathbb{W}\succeq \sum_{k\in[2n^2]} \mathrm{vec}(M_k)\,\mathrm{vec}(M_k)^*\,.$$

By Theorem 7.14 and monotonicity in the Loewner ordering of $\mathcal{N}_{\mathsf{S}_{F^*}}(\cdot)$, we conclude $\mathcal{N}_{\mathsf{S}_{F^*}}(\mathbb{W})^{1/2}$ and $g(\mathbb{W})$ are equal within an absolute constant. $g(\mathbb{W})$ is alternatively given by the following (not exactly computable) convex program

$$\begin{aligned}
&\sup\ \mathrm{Tr}(Z)\quad \text{s.t.}\\
&\|T_L(\mathbb{X})\|_{\mathsf{S}_{F^{(1/2)}}},\|T_R(\mathbb{X})\|_{\mathsf{S}_{F^{(1/2)}}}\leqslant 1\\
&\begin{bmatrix}\mathbb{X} & Z\\ Z^* & \mathbb{W}'\end{bmatrix}\succeq 0\\
&\mathbb{W}\succeq\mathbb{W}'\\
&\mathbb{X},Z\in M_{n\cdot d}(\mathbb{R})\,.
\end{aligned}\tag{73}$$

Before verifying equivalence of the two definitions, we note that concavity of g follows easily from (73) since whenever $\mathbb{X}_1,Z_1$ (resp. $\mathbb{X}_2,Z_2$) are feasible for $\mathbb{W}_1$ (resp. $\mathbb{W}_2$), $\lambda\mathbb{X}_1+(1-\lambda)\mathbb{X}_2,\lambda Z_1+(1-\lambda)Z_2$ are feasible for $\lambda\mathbb{W}_1+(1-\lambda)\mathbb{W}_2$. It follows that $g(\lambda\mathbb{W}_1+(1-\lambda)\mathbb{W}_2)\geqslant\lambda g(\mathbb{W}_1)+(1-\lambda)g(\mathbb{W}_2)$.

Claim 7.16. *(72) = (73).*

Proof. To show (72) $\leqslant$ (73), consider any sequence (M_k) and let $(\overline{M}_k)$ be any sequence satisfying (such a sequence exists by (71)),

$$\langle(\overline{M}_k),(M_k)\rangle = \|(M_k)\|_{\mathsf{S}_{F^*}(\ell_R^2)+\mathsf{S}_{F^*}(\ell_L^2)}\qquad\text{where}\ \ \|(\overline{M}_k)\|_{\mathsf{S}_F(\ell_L^2)},\|(\overline{M}_k)\|_{\mathsf{S}_F(\ell_R^2)}\leqslant 1\,.$$

The claim then follows by considering the substitution

$$\mathbb{X}\overset{\text{def}}{=}\sum_k \mathrm{vec}(\overline{M}_k)\,\mathrm{vec}(\overline{M}_k)^*,\ \ Z\overset{\text{def}}{=}\sum_k \mathrm{vec}(\overline{M}_k)\,\mathrm{vec}(M_k)^*,\ \ \mathbb{W}'\overset{\text{def}}{=}\sum_k \mathrm{vec}(M_k)\,\mathrm{vec}(M_k)^*\,.$$

For the other direction, consider any $\mathbb{X}, Z, \mathbb{W}'$ feasible for $\mathbb{W}$. By the spectral theorem there exist sequences $(M_k)_{k\in[K]}, (\overline{M}_k)_{k\in[K]}$ for $K = 2n^2$, such that

$$\mathbb{X} = \sum_{k\in[K]} \operatorname{vec}(\overline{M}_k)\operatorname{vec}(\overline{M}_k)^*\,,\ \ Z = \sum_{k\in[K]} \operatorname{vec}(\overline{M}_k)\operatorname{vec}(M_k)^*\,,\ \ \mathbb{W}' = \sum_{k\in[K]} \operatorname{vec}(M_k)\operatorname{vec}(M_k)^*\,.$$

Since $\|(\overline{M}_k)\|_{\mathsf{S}_F(\ell_L^2)}, \|(\overline{M}_k)\|_{\mathsf{S}_F(\ell_R^2)} \leqslant 1$, we have

$$\operatorname{Tr}(Z) = \langle (\overline{M}_k), (M_k)\rangle \leqslant \|(M_k)\|_{\mathsf{S}_{F^*}(\ell_R^2)+\mathsf{S}_{F^*}(\ell_L^2)}\,.$$

Taking (M_k) as the sequence in (72) then implies (72) $\geqslant$ (73). ∎

We define an upwards closed convex set $B \overset{\text{def}}{=} \{\mathbb{W} \succeq 0 \mid g(\mathbb{W}) \geqslant 1\}$ where convexity of B follows from concavity of g and upward-closure follows from monotonicity of g (in the Loewner ordering). By Claim 7.16, we know that $\inf_{\mathbb{W}\in B}\langle A, \mathbb{W}\rangle$ is alternatively given by

$$\begin{aligned} &\inf\ \langle A, \mathbb{W}\rangle \quad \text{s.t.}\\ &\operatorname{Tr}(Z) \geqslant 1\\ &\|T_L(\mathbb{X})\|_{\mathsf{S}_{F(1/2)}}, \|T_R(\mathbb{X})\|_{\mathsf{S}_{F(1/2)}} \leqslant 1\\ &\begin{bmatrix} \mathbb{X} & Z\\ Z^* & \mathbb{W}'\end{bmatrix} \succeq 0\\ &\mathbb{W} \succeq \mathbb{W}'\\ &\mathbb{X}, Z \in M_{n\cdot d}(\mathbb{R})\,. \end{aligned} \tag{74}$$

By Proposition 3.12, it suffices to give an approximate separation oracle for

$$\{\mathbb{X} \in \mathbb{PSD}^{n\cdot d} \mid \|T_L(\mathbb{X})\|_{\mathsf{S}_{F(1/2)}} \leqslant 1\} \quad (\text{resp. } \|T_R(\mathbb{X})\|_{\mathsf{S}_{F(1/2)}})\,.$$

For this it suffices to give a separation oracle for the set $S \overset{\text{def}}{=} \{H \in \mathbb{PSD}^n \mid H \in \operatorname{Ball}(\mathsf{S}_{F(1/2)})\}$. By Lemma 7.10 and Lemma 7.9 for any fixed $\varepsilon > 0$, one can compute in oracle-polytime an $M^{(2)}(E)^4 \cdot (1+\varepsilon)$-approximate solution to $\sup_{H\in S}\langle M, H\rangle$ for any $M \in \mathbb{PSD}^n$. So by linear function maximization duality for downward closed sets (Theorem 3.14) there is an oracle-polytime algorithm to find an $M^{(2)}(E)^4 \cdot (1+\varepsilon)$-approximate separation oracle for S. Then by Proposition 3.12, we obtain an oracle-polytime algorithm to find an $O(M^{(2)}(E)^4)$-approximate solution to (74) (i.e., linear function minimization over B).

Thus by linear function minimization duality for upward closed sets (Theorem 3.16), B admits an $M^{(2)}(E)^4 \cdot (1+\varepsilon)$-approximate separation oracle. Recalling that by Theorem 7.14 $\mathcal{L}(\mathsf{S}_E)$ and B are equivalent within $O(M^{(2)}(E))^2$ and applying Observation 3.4 yields the claimed separation oracle for $\mathcal{L}(\mathsf{S}_E)$. ∎

We are now ready to prove the main result of this section.

Theorem 7.17 (Maximization over Unitarily Invariant Norms under Type-2/Dual Cotype-2). *There are algorithms* $\mathrm{ALG}_1(A_1, \mathcal{O}_{\mathsf{S}_X})$ *and* $\mathrm{ALG}_2(A_2, \mathcal{O}_{\mathsf{S}_X}, \mathcal{O}_{\mathsf{S}_Y})$ *such that if* $\mathcal{O}_{\mathsf{S}_X}$ *(resp.* $\mathcal{O}_{\mathsf{S}_Y}$*) is an oracle computing a unitarily invariant matrix norm* $(M_{n,d}(\mathbb{R}), \|\cdot\|_{\mathsf{S}_X})$ *(resp.* $(M_{m,h}(\mathbb{R}), \|\cdot\|_{\mathsf{S}_Y})$*), then*

- *Quadratic: on any input* $A_1 \in M_{nd}(\mathbb{R})$, ALG_1 *runs in time* $\operatorname{poly}(n, d, \operatorname{bit}(A_1))$ *and returns an* α*-approximate solution to* $\mathrm{Q}^{\max}_{\mathsf{S}_X}(A_1)$ *with probability* $1 - 2^{-\Omega(n)}$*, where*

$$\alpha \lesssim T_2(X)^2 \cdot \widetilde{C}_2(X^*)^6 \cdot \log^6 \widetilde{C}_2(X^*) \leqslant \widetilde{T}_2(X)^8 \log^6 \widetilde{T}_2(X)\,.$$

- *Bilinear: on any input* $A_2 \in M_{nd,mh}(\mathbb{R})$, ALG_2 *runs in time* $\mathrm{poly}(n,d,m,h,\mathrm{bit}(A_2))$ *and returns a* β*-approximate solution to* $\mathrm{Op}^{\max}_{\mathsf{S}_X,\mathsf{S}_Y}(A_2)$ *with probability* $1-2^{-\Omega(n)}$*, where*

$$\beta \lesssim \max\{\widetilde{C}_2(X^*)^8 \cdot \log^7 \widetilde{C}_2(X^*)\,,\, \widetilde{C}_2(Y^*)^8 \cdot \log^7 \widetilde{C}_2(Y^*)\}\,.$$

Proof. Without loss of generality we may assume $\|e_1e_1^*\|_{\mathsf{S}_X} = \|e_1e_1^*\|_{\mathsf{S}_Y} = 1$ since otherwise we can just renormalize. $\|\cdot\|_{\mathsf{S}_X}$ (resp. $\|\cdot\|_{\mathsf{S}_Y}$) is always $(n^{O(1)}, n^{-\Omega(1)})$-balanced (resp. $(m^{O(1)}, m^{-\Omega(1)})$-balanced) since

$$\|M\|_{C_\infty} \leqslant \frac{\|M\|_{\mathsf{S}_E}}{\|e_1e_1^*\|_{\mathsf{S}_E}} \leqslant \max\{n,d\}\cdot\|M\|_{C_\infty} \quad \forall M \in M_{n,d}(\mathbb{R})$$

and $\|e_1e_1^*\|_{\mathsf{S}_E} = 1$ by assumption.

Combining Theorem 4.11 with Lemma 7.15 yields the desired theorem. ∎

8 Hardness in the Absence of Type-2

In this section, we prove that if a sequence of norms $(\|\cdot\|_{X^n}, \mathbb{R}^n)_n$ has $\widetilde{T}_2(X^n) = n^{\Omega(1)}$, then a constant-factor approximation algorithm for $\mathrm{Q}^{\max}_{X^n}(\cdot)$ will refute the (randomized) Small Set Expansion Hypothesis (SSEH) defined below. Given a d-regular graph $G=(V,E)$ with n vertices and $\delta \in [0,1]$, let

$$\Phi(\delta) := \min_{S\subseteq V, |S|=\delta n} \frac{|E(S, V\setminus S)|}{d|S|}.$$

Definition 8.1 (Gap Small Set Expansion (Gap-SSE) [RS10])**.** *Given a regular graph* $G=(V,E)$ *and* $\varepsilon,\delta > 0$*, the Gap Small Set Expansion problem Gap-SSE*(ε,δ) *asks to distinguish between*

- *YES:* $\Phi(\delta) \leqslant \varepsilon$.
- *NO:* $\Phi(\delta) \geqslant 1-\varepsilon$.

Hypothesis 8.2 (Small Set Expansion Hypothesis [RS10])**.** *For any* $\varepsilon > 0$*, there exists* $\delta > 0$ *such that Gap-SSE*(ε,δ) *is NP-hard.*

Similarly, in the sequel we will say "Gap-SSE does not admit an algorithm of runtime $T(n)$" to refer to the assumption that for every constant $\varepsilon > 0$, there exists $\delta > 0$ such that no algorithm of runtime $T(n)$ solves the distinguishing problem Gap-SSE(ε,δ).

8.1 SSE-Hardness of ℓ_p-Quadratic Maximization when $p < 2$

Barak et al. [BBH+12] proved hardness of approximating $2\to 4$ norm assuming the SSEH.

We will require a hardness result for $2\to q$ norm when q approaches 2 sufficiently slowly (i.e., $q-2$ is subconstant in n). To do this we exploit the fact that in the reduction of [BBH+12] the optimal vector in the completeness case does not depend on q. We then use interpolation to argue that the soundness decays sufficiently slowly – here we make use of another special property of the reduction, namely the instance is a projector and therefore the $2\to 2$ norm is bounded. We begin by stating the result of [BBH+12] (also see theorem 21 in [BBB+19] for details) with the desired additional properties.

Theorem 8.3 ([BBH+12]). *Assuming SSEH, for any sufficiently small $\delta > 0$, no polynomial time algorithm can distinguish between the following two cases given an input matrix $P \in M_n(\mathbb{R})$ satisfying $\|P\|_{2\to 2} \leqslant 1$:*

YES: $\|P\|_{2\to q} \geqslant (n/(10\cdot\delta))^{1/2-1/q} \quad \forall q \in (2,\infty)$.

NO: $\|P\|_{2\to 4} \leqslant 2\cdot n^{1/4}/\delta^{1/8}$.

By interpolation (Holder's inequality in this case) we obtain the following hardness result for q approaching 2.

Corollary 8.4. *Fix any $\theta \in (0,1)$ and let $q_\theta \stackrel{\text{def}}{=} ((1-\theta)/2+\theta/4)^{-1} = 4/(2-\theta)$. Assuming SSEH, for any $C > 1$, no polynomial time algorithm can approximate $\|\cdot\|_{2\to q_\theta}$ within a factor better than $1 + C\theta$.*

Proof. By Theorem 8.3 in the YES case we have

$$\|P\|_{2\to q_\theta} \geqslant n^{1/2-1/q_\theta}/(10\cdot\delta^{1/2-1/q_\theta}) = (n/(10\cdot\delta))^{\theta/4}.$$

In the NO case by Holder's inequality we have

$$\|P\|_{2\to q_\theta} \leqslant \|P\|_{2\to 2}^{1-\theta}\cdot\|P\|_{2\to 4}^{\theta} \leqslant (2n)^{\theta/4}/\delta^{\theta/8}.$$

Thus we obtain a gap of $1/(20\cdot\delta)^{\theta/8} \geqslant 1 + (1/100)\cdot\theta\cdot\log 1/\delta$. Taking δ sufficiently small yields the claim. ∎

We next amplify the gap in the previous hardness result, and also adapt it to obtain hardness results for ℓ_p-quadratic maximization when $1 < p < 2$ (even for the case when $p = p(n)$ approaches 2 as $n \to \infty$).

Proposition 8.5 (ℓ_p-Quadratic Maximization Hardness for $p < 2$)**.**

(1) *Consider any fixed $q > 2$. Assuming the SSEH and that* P $\neq$ NP*, there is no polynomial time constant factor approximation algorithm for $\|\cdot\|_{2\to q}$.*

(2) *Consider any fixed $q > 2, \varepsilon > 0$. Assuming that Gap-SSE does not admit a quasi-polynomial time algorithm, there is no polynomial time $2^{\log^{1-\varepsilon} n}$-approximation algorithm for $\|\cdot\|_{2\to q}$.*

(3) *Consider any increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $f(n) \leqslant n$, and let $g \stackrel{\text{def}}{=} f^{-1}$. For any constant $C > 1$ there is a reduction running in time* $\operatorname{poly}(g(\overline{m}^{O(1)}))$ *from a size $\overline{m}$ instance of Gap-SSE to the problem of obtaining a C-approximation to $\|\cdot\|_{2\to 2+\log f(n)/\log n}$.*

(4) *All results above extend to the case of $q^* \to q$ norm.*

(5) *All results above extend to the case of $\mathrm{Q}^{\max}_{q^*}(B)$ (even for instances B with 0's on the diagonal).*

Proof. Since for any fixed $q > 2$ arbitrary constant SSEH-hardness of $2 \to q$ norm was shown in [BBH+12], (1) and (2) follow from amplifying the gap using the fact that $\|A^{\otimes t}\|_{2\to q} = \|A\|_{2\to q}^t$ (see [BGG+19]).

The proof of (3) proceeds again by amplifying the gap obtained in Corollary 8.4, the caveat being that q now depends on the size of the matrix and so care must be taken in the calculations. To this end, consider any constant $C > 1$. Let $q(n) \stackrel{\text{def}}{=} 2 + \log f(n)/\log n$, $\overline{q}(m) \stackrel{\text{def}}{=} q(g(m)) =$

$2 + \log m / \log g(m)$ and let $t \overset{\text{def}}{=} \log C \cdot \log g(m) / \log m$. We set $m = m(n)$ so that $m^t = n$. By Corollary 8.4 there is a reduction from Gap-SSE of size $\overline{m} = m^{\Omega(1)}$ to an instance A of $\ell_2^m \to \ell_{\overline{q}(m)}^m$ with gap $1 + 10 \log m / \log g(m)$. We then reduce to $\ell_2^n \to \ell_{\overline{q}(m)}^n$ by considering $A^{\otimes t}$ and as claimed we obtain a gap of at least

$$(1 + 10 \log m / \log g(m))^t \geqslant e^{t \log m / \log g(m)} = C\,.$$

Since $m^t = n$, we have $n = \mathrm{poly}(g(m))$. Therefore the runtime of the reduction is $\mathrm{poly}(g(\overline{m}^{O(1)}))$ as desired (where $\overline{m}$ is the size of the Gap-SSE instance).

(4) follows from combining (1), (2) and (3) with the observation that $\|AA^*\|_{q^* \to q} = \|A\|_{2 \to q}^2$. Lastly, (5) follows from combining (4) with the following claim:

Claim 8.6. *For any $p \in [1, \infty]$, there is a polytime computable matrix $B = B(A)$ (with 0's on the diagonal) such that $\mathrm{Q}_p^{\max}(B) \leqslant \|A\|_{p \to p^*} \leqslant 2 \cdot \mathrm{Q}_p^{\max}(B)$.*

Proof. Consider any $A \in M_{m,n}(\mathbb{R})$ and let $B \in M_{m+n}(\mathbb{R})$ be given by

$$B \overset{\text{def}}{=} \frac{1}{2} \cdot \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$$

Let $p \oplus_\infty p$ be the norm over $\mathbb{R}^{m+n}$ defined as $\|x \oplus y\|_{p \oplus_\infty p} \overset{\text{def}}{=} \max\{\|x\|_p, \|y\|_p\}$. Observe that $\mathrm{Q}_{p \oplus_\infty p}^{\max}(B) = \|A\|_{p \to p^*}$. Further, it is easily checked that $\|\cdot\|_{p \oplus_\infty p}$ and ℓ_p^{m+n} are equivalent within a factor of 2. Thus $\mathrm{Q}_p^{\max}(B)$ and $\|A\|_{p \to p^*}$ are within a factor of 2 as desired. ∎

This completes the proof of Proposition 8.5. ∎

Remark 8.7. *Observe that if $q(n) = 2 + O(1/\log n)$, then $\|\cdot\|_{2 \to q(n)}$ is equivalent to $\|\cdot\|_{2 \to 2}$ within a constant and therefore admits a polytime constant factor approximation algorithm. Part (2) of Proposition 8.5 provides an almost matching hardness result, i.e., assuming SSEH and ETH, for any $q(n) = 2 + \omega(\log \log n / \log n)$, there is no polytime constant factor approximation algorithm for $\|\cdot\|_{2 \to q(n)}$. We speculate that one can bridge this small gap with a hardness result for the case of $q(n) = 2 + \omega(1/\log n)$. One approach to establishing such a result would be to employ a more size-efficient gap-amplification procedure than tensoring (for example using random walks).*

8.2 SSE-Hardness of Approximation when Type-2 Fails

We begin with some preliminaries.

8.2.1 Preliminaries: Embedding Copies of ℓ_p^k in X

To obtain more general hardness results we will require the notion of *stable type*. For $1 \leqslant p \leqslant 2$, the stable type-$p$ constant $ST_p(X)$ is defined analogously to $T_p(X)$ with p-stable random variables replacing rademachers. We also define a constant c_p which will appear in the sequel. Let $(\mathbb{E}_i)_{i \in \mathbb{N}}$ be an independent sequence of exponential random variables and let $\Gamma_j \overset{\text{def}}{=} \sum_{i=1}^{j} E_i$. Then we define

$$c_p \overset{\text{def}}{=} \mathbb{E}\,[|\sum_{j \in \mathbb{N}} \Gamma_j^{-1/p} \cdot \varepsilon_j||]$$

where $(\varepsilon_j)_{j\in\mathbb{N}}$ is an independent sequence of rademacher random variables. It's not hard to check that $c_p \gtrsim 1$ (a fact we will use later).

We now state an important result of Pisier [Pis83] which states that for any $1 < p \leqslant 2$, norms with large stable type-p constant contain large-dimensional isomorphic copies of ℓ_p^k.

Theorem 8.8 (Pisier's Stable Type Theorem [Pis83])**.** *Let X be a norm over $\mathbb{R}^n$. Fix any $1 < p < 2$ and $\varepsilon > 0$, and let $\delta(\varepsilon, p) \stackrel{\text{def}}{=} \frac{2-p}{p}\left(\frac{\varepsilon c_p}{2p^*+2}\right)^{p^*}$. Then for any $k < \delta(\varepsilon,p)\cdot(\mathrm{ST}_p(X))^{1/p^*}$, there is an $n \times k$ matrix B such that for all $a \in \mathbb{R}^k$,*

$$(1-\varepsilon)\cdot\|a\|_p \leqslant \|Ba\|_X \leqslant (1+\varepsilon)\cdot\|a\|_p\,.$$

The existence of the matrix B above is shown by a probabilistic construction. Below we check that it can be efficiently sampled.

Corollary 8.9 (Efficient Sampling of [Pis83])**.**
Fix any $1 < p < 2$ and $\varepsilon > 0$, and let $\delta(\varepsilon, p) \stackrel{\text{def}}{=} \frac{2-p}{p}\left(\frac{\varepsilon c_p}{2p^*+2}\right)^{p^*}$. Let $X = (\mathbb{R}^n, \|\cdot\|_X)$ be a normed space and assume we are given a sequence of vectors $x_1, \ldots x_t \in \mathrm{Ball}(X)$ with $t = n^{O(1)}$, and such that $\mathbb{E}[\|\sum_i \varepsilon_i x_i\|_X^2]^{1/2} \geqslant f(n)\cdot\sqrt{t}$. Then for $k(n) = \frac{2-p}{100p}\left(\frac{c_p\cdot f(n)}{(p^*+1)\,n^{1/p-1/2}}\right)^{p^*}$, there is an $n \times k(n)$ random matrix B that can be sampled in time $n^{O(1/(2-p))}$, and a scalar C, such that with probability $1 - 1/n^4$ it holds that for all $a \in \mathbb{R}^k$,

$$0.9\cdot\mathsf{C}\cdot\|a\|_p \leqslant \|\mathsf{B}a\|_X \leqslant 1.1\cdot\mathsf{C}\cdot\|a\|_p\,.$$

Proof. We describe the distribution of an $n \times k$ random matrix B' as specified in a presentation of Theorem 8.8 in [MS86]. Let v be the $\mathbb{R}^n$ valued random variable given by choosing uniformly at random from the set $\{x_1, \ldots x_t\}$. Let $(\mathsf{v}_{i,j})_{i\in[k],j\in\mathbb{N}}$ be a double sequence of i.i.d. $\mathbb{R}^n$-valued random variables having the same distribution as v and similarly let $(\varepsilon_{i,j})_{i\in[k],j\in\mathbb{N}}$ be a double sequence of i.i.d. Rademacher random variables. Let B'_i denote the i-th column of B'. Then each B'_i (for $i \in [k]$) is given by

$$\mathsf{B}'_i \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{-1/p}\cdot\varepsilon_{i,j}\cdot\mathsf{v}_{i,j}\,.$$

Let $\tau \stackrel{\text{def}}{=} n^{50(2-p)}$ and let B be an $n \times k$ random matrix whose i-th column B_i is given by

$$\mathsf{B}_i \stackrel{\text{def}}{=} \sum_{1\leqslant j\leqslant\tau} j^{-1/p}\cdot\varepsilon_{i,j}\cdot\mathsf{v}_{i,j}\,.$$

It is shown in [MS86] (see Theorem 13.12 and Proposition 13.14) that with probability $1 - 2^{-\Omega(k)}$, it holds that for all $a \in \mathbb{R}^k$,

$$0.99\cdot\mathsf{C}\cdot\|a\|_p \leqslant \|\mathsf{B}'a\|_X \leqslant 1.01\cdot\mathsf{C}\cdot\|a\|_p\,.$$

where $\mathsf{C} \stackrel{\text{def}}{=} \mathbb{E}\,[\|\mathsf{B}'_1\|_X]$. We will show that w.h.p. $\mathsf{B}'' \stackrel{\text{def}}{=} \mathsf{B}' - \mathsf{B}$ has $p \to X$ operator norm bounded by C/n. Since $\|\mathsf{B}'a\|_X - \|\mathsf{B}''\|_{p\to X}\cdot\|a\|_p \leqslant \|\mathsf{B}a\| \leqslant \|\mathsf{B}'a\|_X + \|\mathsf{B}''\|_{p\to X}\cdot\|a\|_p$, this would imply (for sufficiently large n) that

$$0.9\cdot C\cdot\|a\|_p \leqslant \|\mathsf{B}a\|_X \leqslant 1.1\cdot C\cdot\|a\|_p\,.$$

In what follows we assume WLOG that $k \leqslant n \leqslant t$. We proceed with showing $\|\mathsf{B}''\|_{p\to X} \leqslant C/t^{3/2}$ with probability at least $1 - \Omega(1/t^4)$. It suffices to show that $\|\mathsf{B}''_i\|_X \leqslant C/t^{5/2}$ with probability $1 - \Omega(1/t^5)$, since by triangle inequality we have $\|\mathsf{B}''a\|_X \leqslant \|a\|_1 \cdot \max_{i\in[k]}\|\mathsf{B}''_i\|_X \leqslant k^{1-1/p} \cdot \|a\|_p \cdot \max_{i\in[k]}\|\mathsf{B}''_i\|_X$. To this end we have,

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{(\mathsf{v}_{i,j})}\Big[\mathop{\mathbb{E}}_{(\varepsilon_{i,j})}[\|\mathsf{B}''_i\|_X^2]\Big] \\
&= \mathop{\mathbb{E}}_{(\mathsf{v}_{i,j})}\Big[\mathop{\mathbb{E}}_{(\varepsilon_{i,j})}[\|\textstyle\sum_{1\leqslant j\leqslant \tau} j^{-1/p}\cdot \varepsilon_{i,j}\cdot \mathsf{v}_{i,j}\|_X^2]\Big] \\
&\leqslant n \cdot \mathop{\mathbb{E}}_{(\mathsf{v}_{i,j})}\Big[\textstyle\sum_{\tau\leqslant j\leqslant\infty} j^{-2/p}\cdot\|\mathsf{v}_{i,j}\|_X^2\Big] && \text{(since } \widetilde{T}_2(X)\leqslant\sqrt{n}\text{)} \\
&= n\cdot \sum_{\tau\leqslant j\leqslant\infty} j^{-2/p}\cdot \mathop{\mathbb{E}}_{(\mathsf{v}_{i,j})}[\|\mathsf{v}_{i,j}\|_X^2] \\
&\leqslant n\cdot \big(\textstyle\sum_{\ell\in[t]}\|x_\ell\|_X^2/t\big)\cdot \textstyle\sum_{\tau\leqslant j\leqslant\infty} j^{-2/p} && \text{(by definition of } \mathsf{v}\text{)} \\
&\leqslant n\cdot \max_{\ell\in[t]}\|x_\ell\|_X \cdot \sum_{\tau\leqslant j\leqslant\infty} j^{-2/p} \\
&\lesssim \frac{n}{\tau^{2/p-1}}\cdot \max_{\ell\in[t]}\|x_i\|_X && \text{(integration)}\,.
\end{aligned}
$$

Taking $\tau \overset{\text{def}}{=} t^{20/(2-p)}$ and applying Markov's inequality implies that with probability $1-\Omega(1/t^5)$ we have $\|\mathsf{B}''_i\|_X \leqslant \max_{\ell\in[t]}\|x_i\|_X/t^4$.

By an analogous argument to above (this time using the cotype inequality $\widetilde{C}_2(X) \leqslant \sqrt{n}$), we conclude that

$$
C = \mathbb{E}\,[\|\mathsf{B}'_1\|_X] \gtrsim n^{-1/2}\cdot \mathop{\mathbb{E}}_{(\mathsf{v})}[\|\mathsf{v}\|_X^2]^{1/2} \geqslant \max_{i\in[t]}\|x_i\|_X/(\sqrt{n}t)\,.
$$

This implies $\|\mathsf{B}''_i\|_X \leqslant C/t^{5/2}$ as desired.

The truncation B of B' can be computed in time $t^{O(1/(2-p))} = n^{O(1/(2-p))}$ and this completes the proof. ∎

8.2.2 The Final Reduction

We are finally ready to prove our main hardness result. The core idea is to use Pisier's stable type theorem in order to encode an ℓ_p-quadratic maximization instance (for $p < 2$) as an instance of quadratic maximization over a general norm X with large type-2. The proof is then completed by appealing to hardness results for ℓ_p-quadratic maximization.

Theorem 8.10 (Hardness of Quadratic Maximization when Type-2 Grows Sufficiently Fast)**.** *Let $(\|\cdot\|_{X^n}, \mathbb{R}^n)_{n\in\mathbb{N}}$ be a sequence of normed spaces and let $f(n) \overset{\text{def}}{=} \widetilde{T}_2(X^n)$. Assume there is a* $\mathrm{poly}(n)$ *time algorithm computing $\|\cdot\|_{X^n}$, and an algorithm that on input $n\in\mathbb{N}$ returns in* $\mathrm{poly}(n)$ *time a sequence of vectors $x_1,\ldots,x_t \in \mathrm{Ball}(X^n)$ satisfying*

$$
\mathbb{E}\,[\|\sum_{i\in[m]} \varepsilon_i\cdot x_i\|_{X^n}^2] \gtrsim f(n)\cdot\sqrt{t}\,. \tag{75}
$$

Then we have

(1) If $f(n) = n^{\Omega(1)}$, then assuming the SSEH and that NP $\not\subseteq$ BPP, *there is no polynomial time algorithm approximating $\mathrm{Q}^{\max}_{X^n}(\cdot)$ within a constant factor.*

(2) *If $f(n) = n^{\Omega(1)}$, then assuming Gap-SSE cannot be solved in randomized quasi-polynomial time, there is no polynomial time algorithm approximating $\mathrm{Q}^{\max}_{X^n}(\cdot)$ within a factor of $2^{\log^{1-\varepsilon} n}$.*

(3) *If for every constant $C > 1$, $f(f^{1/C}(n^{1/C})) = \log^{\omega(1)} n$, then assuming Gap-SSE $\notin \bigcap_{\varepsilon>0} \mathrm{RTIME}(2^{n^{\varepsilon}})$, there is no polynomial time algorithm approximating $\mathrm{Q}^{\max}_{X^n}(\cdot)$ within a constant factor.*

Proof. We proceed by reducing $\mathrm{Q}^{\max}_{\ell_p^k}(\cdot)$ to quadratic maximization over a subspace of X^n. Applying Observation 6.13 then completes the proof. We conclude (1) and (2) by appropriately choosing $p = p(n)$ and $k = k(n)$ at the end.

By Corollary 8.9, setting $k(n) \stackrel{\mathrm{def}}{=} \frac{2-p}{100p}\left(\frac{c_p \cdot f(n)}{(p^*+1)\, n^{1/p-1/2}}\right)^{p^*}$, there is an $n \times k(n)$ random matrix B that can be sampled in time $n^{O(1/(2-p))}$, and a scalar C such that with probability $1 - o(1)$, it holds that for all $a \in \mathbb{R}^k$,

$$0.9 \cdot C \cdot \|a\|_p \leqslant \|\mathsf{B}a\|_{X^n} \leqslant 1.1 \cdot C \cdot \|a\|_p \,. \tag{76}$$

Let $\mathsf{B}^\dagger$ be the $k \times n$ matrix defined as the linear map that maps $x \in \mathbb{R}^n$ to $a \in \mathbb{R}^k$ where a is such that $\mathsf{B}a$ is the projection of x onto the column span of B (note that a is unique since the columns of B must be linearly independent in order to satisfy (76)). Let E be the column span of B. Then we have for any $H \in M_k(\mathbb{R})$

$$\sup_{x \in E \setminus \{0\}} \frac{\langle x\,, (\mathsf{B}^\dagger)^* H \mathsf{B}^\dagger x\rangle}{\|x\|^2_{X^n}} = \sup_{a \in \mathbb{R}^k \setminus \{0\}} \frac{\langle a\,, Ha\rangle}{\|\mathsf{B}a\|^2_{X^n}} \asymp \frac{1}{C^2} \cdot \sup_{a \in \mathbb{R}^k \setminus \{0\}} \frac{\langle a\,, Ha\rangle}{\|a\|^2_p}$$

where the final step follows from (76). Thus $H \mapsto (\mathsf{B}^\dagger)^* H \mathsf{B}^\dagger$ is a valid reduction from quadratic maximization over ℓ_p^k to quadratic maximization over the subspace E of X^n.

We are now ready to set parameters. Let $p(n) \stackrel{\mathrm{def}}{=} 2(1 + \log f(n)/(2\log n))^{-1}$ (so that $f(n)/n^{1/p-1/2} = \sqrt{f(n)}$). We then have the lower estimate

$$k(n) \geqslant f^{1/5}(n)/\log n \tag{77}$$

where the choice of $1/5$ above is not special - any sufficiently small constant will do.

For (1), it follows from (77) that if $f(n) = n^{\Omega(1)}$, then $k(n) = n^{\Omega(1)}$ and thus composing with Proposition 8.5 part (1), we obtain a reduction from a size m instance of Gap-SSE to $\mathrm{Q}^{\max}_{X^n}(\cdot)$ where $n = m^{O(1)}$. The runtime of the reduction is $n^{O(1/(2-p(n)))} = n^{O(1)} = m^{O(1)}$. This completes the proof of (1).

For (2), it follows from (77) that if $f(n) = n^{\Omega(1)}$, then $k(n) = n^{\Omega(1)}$ and thus composing with Proposition 8.5 part (2), we obtain a reduction from a size m instance of Gap-SSE to $\mathrm{Q}^{\max}_{X^n}(\cdot)$ where $n = m^{\log^{O(1)} m}$. The runtime of the reduction is $n^{O(1/(2-p(n)))} = n^{O(1)} = m^{\log^{O(1)} m}$. This completes the proof of (2).

For (3), it follows from (77) that if for every constant $C > 1$, $f(f^{1/C}(n^{1/C})) = \log^{\omega(1)} n$, then $k(n) \geqslant f^{1/5}(n)/\log n \geqslant f^{1/6}(n)$ (where we use the fact that $c_p \gtrsim 1$) and thus composing with Proposition 8.5 part (3), we obtain a reduction from a size m instance of Gap-SSE to $\mathrm{Q}^{\max}_{X^n}(\cdot)$ where $n = g(g(m^{O(1)})^{O(1)})$ and $g = f^{-1}$. Finally since for every $C > 1$, $f(f^{1/C}(n^{1/C})) = \log^{\omega(1)} n$, it follows that $n = g(g(m^{O(1)})^{O(1)}) \leqslant 2^{m^{o(1)}}$. The runtime of the reduction is $n^{O(1/(2-p(n)))} = n^{O(1/\log n)} = (2^{m^{o(1)}})^{m^{o(1)}} = 2^{m^{o(1)}}$, which completes the proof of (3). ∎

Remark 8.11.

1. *Theorem 1.3 in the introduction states that if $\widetilde{T}_2(X) = n^{\Omega(1)}$ then $Q_X^{\max}(\cdot)$ cannot be approximated within a constant in polynomial time assuming the (randomized) small set expansion hypothesis. This is implied by the above proof (with parameters set differently). Above, we chose to highlight two settings of parameters we believed to be most interesting.*

2. *If $Q_{X^n}^{\max}(\cdot)$ admits a polytime constant factor approximation, then (3) above necessitates in particular that $\widetilde{T}_2(X^n)$ grows slower than every function in the following series: $2^{\log^{\varepsilon} n}, 2^{2^{\log\log^{\varepsilon} n}}, \cdots$.*

3. *The assumption (75) of Theorem 8.10 may be removed in any of (1), (2), (3) provided one replaces the assumption "there is no algorithm for Gap-SSE with runtime $T(n)$" by the stronger hypothesis that "Gap-SSE cannot be solved by circuits of size $T(n)$". In (1) for instance, this is equivalent to assuming the SSEH and that $\mathrm{NP} \not\subseteq \mathrm{P}_{/\mathrm{poly}}$.*

4. *By [FLM77] the type-2 constant of an n-dimensional normed space is attained up to a universal constant factor for $t = O(n^2)$ vectors (see also [JN09] for the proof of this). In fact, by [TJ79] we can take $t = O(n)$. The assumption (75) of Theorem 8.10 states that such vectors can be found efficiently. It is open whether or not it is possible to find such vectors using only polynomially many calls to the assumed membership oracle for X; this would amount to making the proofs in [FLM77] or [TJ79] algorithmic. If this were possible, then we could remove assumption (75). This nuance is secondary to the main content of Theorem 8.10, namely to demonstrate that type-2 is inherently linked to the computational complexity of quadratic maximization. We note that if one were only interested in removing the assumption that vectors satisfying (75) could be found efficiently, then it would suffice to do so using polynomially many calls to an oracle that approximates $Q_{X^n}^{\max}(\cdot)$. But, it is an independently interesting question to understand when it is possible to approximate type and cotype constants efficiently.*

9 Oracle Lower Bound for General Type-2 Norms

Our construction is identical to that of Brieden et al. [BGK+01] who gave a query lower bound for approximating the ℓ_2-diameter of a convex body in the membership oracle model with the caveat that a constant bound on Type-2 needs to be verified. We include the full analysis for completeness.

9.1 The Construction

Fix any constant $\delta < 1/2$ and let $r \stackrel{\text{def}}{=} \mathrm{g}/n^{1/2-\delta} \in \mathbb{R}^n$ where g is a standard Gaussian random vector (so that r has length n^{δ} with probability $1 - o(1)$). Let $\|\cdot\|_X$ be the norm whose unit ball is $\mathbf{conv}(\mathbb{S}^{n-1} \cup \{\pm r\})$.

9.2 Oracle Lower Bound

We now show that PSD quadratic maximization over general norms with bounded type-2 (and therefore also bilinear maximization) cannot be approximated (within even $n^{o(1)}$).

Theorem 9.1 (Bounded Type-2 Hardness)**.**
Consider constants $\delta \in (0, 1/2)$ and $\varepsilon < 1 - 2\delta$. Then for the random norm X over $\mathbb{R}^n$ defined as above

(1) $\widetilde{T}_2(\mathsf{X}) \lesssim 1$.

(2) For any fixed set $S \subseteq \mathbb{R}^n$ *of size at most* 2^{n^ε}, $\mathrm{Ball}(\mathsf{X}) \cap S = \mathrm{Ball}(\ell_2^n) \cap S$ *with probability* $1 - o(1)$.

(3) $\mathrm{Op}^{\max}_{\mathsf{X},\mathsf{X}}(\mathrm{I}) = \mathrm{Q}^{\max}_{\mathsf{X}}(\mathrm{I}) = \Omega(n^{2\delta})$ *with probability* $1 - o(1)$*, while on the other hand* $\mathrm{Op}^{\max}_{\ell_2^n,\ell_2^n}(\mathrm{I}) = \mathrm{Q}^{\max}_{\ell_2^n}(\mathrm{I}) = 1$.

Proof. The squared ℓ_2-radius of $\mathrm{Ball}(\mathsf{X})$ is $\Omega(n^{2\delta})$ w.h.p. by construction. (3) then follows from observing that $\mathrm{Q}^{\max}_{\mathsf{X}}(\mathrm{I})$ is precisely the squared ℓ_2-radius of $\mathrm{Ball}(\mathsf{X})$.

We next establish (1). Let $x_1, \dots, x_m \in \mathbb{R}^n$ be a sequence of vectors. We may write $x_i / \|x_i\|_{\mathsf{X}}$ as a convex combination $x_i/\|x_i\|_{\mathsf{X}} = \lambda_i \cdot r + (1 - \lambda_i) \cdot y_i$ where $y_i \in \mathrm{Ball}(\ell_2^n)$. Then for a sequence $(\mathrm{g}_i)_{i \in [m]}$ of i.i.d. standard Gaussians we have

$$\begin{aligned}
&\mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i x_i\|_{\mathsf{X}}\right] \\
&\leqslant \mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot (1-\lambda_i) \cdot y_i\|_{\mathsf{X}}\right] + \mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot \lambda_i \cdot r\|_{\mathsf{X}}\right] \\
&\leqslant \mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot (1-\lambda_i) \cdot y_i\|_2\right] + \mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot \lambda_i \cdot r\|_{\mathsf{X}}\right] \\
&\leqslant \sqrt{\textstyle\sum_i \|x_i\|_{\mathsf{X}}^2} + \mathbb{E}\left[\|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot \lambda_i \cdot r\|_{\mathsf{X}}\right] \\
&= \sqrt{\textstyle\sum_i \|x_i\|_{\mathsf{X}}^2} + \|r\|_{\mathsf{X}} \cdot \mathbb{E}\left[|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot \lambda_i|\right] \\
&= \sqrt{\textstyle\sum_i \|x_i\|_{\mathsf{X}}^2} + \mathbb{E}\left[|\textstyle\sum_i \mathrm{g}_i \cdot \|x_i\|_{\mathsf{X}} \cdot \lambda_i|\right] \\
&\leqslant 2 \cdot \sqrt{\textstyle\sum_i \|x_i\|_{\mathsf{X}}^2}\,.
\end{aligned}$$

Lastly by Kahane-Khintchine inequality Theorem 2.1 we have $\mathbb{E}\left[\|\sum_i \mathrm{g}_i x_i\|_{\mathsf{X}}\right] \gtrsim \mathbb{E}\left[\|\sum_i \mathrm{g}_i x_i\|_{\mathsf{X}}^2\right]^{1/2}$. This completes the proof of (1).

Finally we shall establish (2). Consider any any fixed set $S \subseteq \mathbb{R}^n$ of size at most 2^{n^ε} (ε to be chosen later) and let $S' \stackrel{\text{def}}{=} \{v \in S \mid \|v\|_2 > 1\}$. Since $\mathrm{Ball}(\ell_2^n) \subseteq \mathrm{Ball}(\mathsf{X})$, we need only show that $S' \cap \mathrm{Ball}(\mathsf{X}) = \varnothing$ with probability $1 - o(1)$. We do so by exhibiting dual witnesses. Note that for any $v \in S'$, we have $\langle v, v/\|v\|_2 \rangle > 1$. So if $v/\|v\|_2 \in \mathrm{Ball}(\mathsf{X}^*)$, then we conclude $v \notin \mathrm{Ball}(\mathsf{X})$. Thus in order to prove (2), it suffices to show that with probability $1 - o(1)$, the set $\widehat{S} \stackrel{\text{def}}{=} \{v/\|v\|_2 \mid v \in S\}$ is contained in $\mathrm{Ball}(\mathsf{X}^*)$.

To this end note that $\mathrm{Ball}(\mathsf{X}^*)$ is simply given by

$$\{\xi \in \mathrm{Ball}(\ell_2^n) \mid |\langle \xi, r \rangle| \leqslant 1\} = \{\xi \in \mathrm{Ball}(\ell_2^n) \mid |\langle \xi, \mathrm{g} \rangle| \leqslant n^{1/2-\delta}\}\,.$$

Thus showing that $\widehat{S} \subseteq \mathrm{Ball}(\mathsf{X}^*)$ is equivalent to showing that $\sup_{\xi \in \widehat{S}} |\langle \mathrm{g}, \xi \rangle| \leqslant n^{1/2-\delta}$. Since for each $\xi \in \widehat{S}$, $\langle \mathrm{g}, \xi \rangle$ is a standard Gaussian, a standard union bound argument yields that with probability $1 - 1/|\widehat{S}|$,

$$\sup_{\xi \in \widehat{S}} |\langle \mathrm{g}, \xi \rangle| \lesssim \sqrt{\log |\widehat{S}|} \lesssim n^{\varepsilon/2}\,.$$

Thus it suffices to take $\varepsilon < 1 - 2\delta$. This completes the proof. ∎

Remark 9.2. *Note that the above theorem rules out the possibility of any approximation algorithm for PSD quadratic maximization assuming only membership access to* $\mathrm{Ball}(\mathsf{X})$*. This follows from taking the set* S *to be the set of queries made by the algorithm.*

We conclude by noting that the above hardness result applies even to quadratic maximization of matrices with 0's on the diagonal (this is a special case considered frequently in the literature). Let

$$A \stackrel{\text{def}}{=} \frac{1}{2} \cdot \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

Further let $\|\cdot\|_{\mathsf{X}\oplus_\infty \mathsf{X}}$ and $\|\cdot\|_{\ell_2^n\oplus_\infty \ell_2^n}$ be the norms over $\mathbb{R}^{2n}$ given by $\|(x,y)\|_{\mathsf{X}\oplus_\infty\mathsf{X}} = \max\{\|x\|_{\mathsf{X}}, \|y\|_{\mathsf{X}}\}$ and $\|(x,y)\|_{\ell_2^n\oplus_\infty \ell_2^n} = \max\{\|x\|_2, \|y\|_2\}$. From the simple observation that $\sup_{x,y\in\mathrm{Ball}(\mathsf{X})}\langle x\,, y\rangle = \mathrm{Q}^{\max}_{\mathsf{X}\oplus_\infty\mathsf{X}}(A)$, we obtain the following corollary.

Corollary 9.3 (Bounded Type-2 Hardness even with 0's on the Diagonal)**.**
Consider constants $\delta \in (0, 1/2)$ *and* $\varepsilon < 1 - 2\delta$*. Then for the matrix* A *and the norms* $\mathsf{X}\oplus_\infty \mathsf{X}\,, \ell_2^n \oplus_\infty \ell_2^n$ *over* $\mathbb{R}^{2n}$ *defined as above, we have*

(1) All diagonal entries of A *are* 0*.*

(2) $\widetilde{T}_2(\mathsf{X}\oplus_\infty\mathsf{X}), \widetilde{T}_2(\ell_2^n\oplus_\infty \ell_2^n) \lesssim 1$*.*

(3) For any fixed set $S \subseteq \mathbb{R}^n$ *of size at most* 2^{n^ε}, $\mathrm{Ball}(\mathsf{X}\oplus_\infty\mathsf{X}) \cap S = \mathrm{Ball}(\ell_2^n\oplus_\infty \ell_2^n)\cap S$ *with probability* $1 - o(1)$*.*

(4) $\mathrm{Q}^{\max}_{\mathsf{X}\oplus_\infty\mathsf{X}}(A) = \Omega(n^{2\delta})$ *w.h.p., while on the other hand* $\mathrm{Q}^{\max}_{\ell_2^n\oplus_\infty\ell_2^n}(A) = 1$*.*

References

[ABH+05] Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 206–215. IEEE, 2005. 6, 15

[ABS98] Noga Alon, Ravi Boppana, and Joel Spencer. An asymptotic isoperimetric inequality. *Geometric & Functional Analysis GAFA*, 8(3):411–436, 1998. 16

[AK06] Fernando Albiac and Nigel J Kalton. *Topics in Banach space theory, volume 233 of Graduate Texts in Mathematics*. Springer, New York, 2006. 53

[Alo06] Noga Alon. Maximizing a quadratic form on the ℓ_1^n unit ball. Unpublished manuscript, 2006. 6

[ALS+18] Alexandr Andoni, Chengyu Lin, Ying Sheng, Peilin Zhong, and Ruiqi Zhong. Subspace embedding and linear regression with orlicz norm. *arXiv preprint arXiv:1806.06430*, 2018. 12

[AMMN06] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *Inventiones mathematicae*, 163(3):499–522, 2006. Conference version in STOC '05. 15

[AN06] Noga Alon and Assaf Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.*, 35(4):787–803, 2006. 15

[ANN+17] Alexandr Andoni, Huy L Nguyen, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. Approximate near neighbors for general symmetric norms. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 902–913, 2017. 12

[ANN+18] Alexandr Andoni, Assaf Naor, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. Hölder homeomorphisms and approximate nearest neighbors. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 159–169. IEEE, 2018. 57, 61, 62

[BBB+19] Frank Ban, Vijay Bhattiprolu, Karl Bringmann, Pavel Kolev, Euiwoong Lee, and David P Woodruff. A ptas for lp-low rank approximation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 747–766. SIAM, 2019. 74

[BBC+17] Jarosław Błasiok, Vladimir Braverman, Stephen R Chestnut, Robert Krauthgamer, and Lin F Yang. Streaming symmetric norms via measure concentration. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 716–729, 2017. 12

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 16, 74, 75

[BGG+19] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Approximability of $p \to q$ matrix norms: generalized krivine rounding and hypercontractive hardness. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1358–1368. SIAM, 2019. 50, 75

[BGK+01] Andreas Brieden, Peter Gritzmann, Ravindran Kannan, Victor Klee, László Lovász, and Miklós Simonovits. Deterministic and randomized polynomial-time approximation of radii. *Mathematika*, 48(1-2):63–105, 2001. 80

[Bha97] Rajendra Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. 14

[BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. 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*, pages 143–151. ACM, 2015. 16

[BL76] Jöran Bergh and Jörgen Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223. 11

[BMMN13] Mark Braverman, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. The Grothendieck constant is strictly smaller than Krivine's bound. In *Forum of Mathematics, Pi*, volume 1. Cambridge University Press, 2013. Conference version in FOCS '11. 15

[BR95] Mihir Bellare and Phillip Rogaway. The complexity of approximating a nonlinear program. *Math. Programming*, 69(3, Ser. A):429–441, 1995. 15

[Bri02] Andreas Brieden. Geometric optimization problems likely not contained in $\mathbb{APX}$. *Discrete Comput. Geom.*, 28(2):201–209, 2002. 15

[BRS15] Jop Briët, Oded Regev, and Rishi Saket. Tight hardness of the non-commutative Grothendieck problem. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1108–1122. IEEE, 2015. 16

[BSS18] Jarosław Byrka, Krzysztof Sornat, and Joachim Spoerhase. Constant-factor approximation for ordered k-median. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 620–631, 2018. 13

[BV04] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 53, 55

[CS19a] Deeparnab Chakrabarty and Chaitanya Swamy. Approximation algorithms for minimum norm and ordered optimization problems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 126–137, 2019. 13

[CS19b] Deeparnab Chakrabarty and Chaitanya Swamy. Simpler and better algorithms for minimum-norm load balancing. *arXiv preprint arXiv:1905.00044*, 2019. 13

[CW04] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: Extending grothendieck's inequality. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 54–60. IEEE Computer Society, 2004. 15

[CW14] Kenneth L Clarkson and David P Woodruff. Sketching for m-estimators: A unified approach to robust regression. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 921–939. SIAM, 2014. 12

[DK08] Etienne De Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research*, 16(2):111–125, 2008. 16

[dKLP06] Etienne de Klerk, Monique Laurent, and Pablo A. Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoret. Comput. Sci.*, 361(2-3):210–225, 2006. 15

[DPR72] Ed Dubinsky, A. Pełczyński, and H. P. Rosenthal. On Banach spaces X for which $\Pi_2(\mathcal{L}_\infty, X) = B(\mathcal{L}_\infty, X)$. *Studia Math.*, 44:617–648, 1972. 3

[DZHZ06] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha. 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)*, pages 281—-288, 2006. 14

[FL92] U. Feige and L. Lovász. 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*, pages 733–744, New York, 1992. ACM. 15

[FLM77] Tadeusz Figiel, Joram Lindenstrauss, and Vitali D Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139(1):53–94, 1977. 46, 80

[GLS93] M Grotschel, L Lovász, and A Schrijver. Geometric algorithms and combinatorial optimization, 1993. 10, 29

[GM15] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 829, 2015. 16

[Gro53] Alexander Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo* ***8***, 1953. 15, 16, 49

[GRSW16] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing UGC from some optimal geometric inapproximability results. *ACM Transactions on Algorithms (TALG)*, 12(1):6, 2016. Conference version in SODA '12. 2

[GTJ83] D. J. H. Garling and N. Tomczak-Jaegermann. The cotype and uniform convexity of unitary ideals. *Israel J. Math.*, 45(2-3):175–197, 1983. 14

[Haa85] Uffe Haagerup. The grothendieck inequality for bilinear forms on C*-algebras. *Advances in Mathematics*, 56(2):93–116, 1985. 16

[HH06] Eran Halperin and Elad Hazan. HAPLOFREQ—estimating haplotype frequencies efficiently. *J. Comput. Biol.*, 13(2):481–500 (electronic), 2006. 15

[HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory*, pages 956–1006, 2015. 16

[HV16] Steven Heilman and Thomas Vidick. A moment majorization principle for random matrix ensembles with applications to hardness of the noncommutative grothendieck problem. Manuscript available at `https://arxiv.org/abs/1603.05620`, 2016. 16

[JN09] William B. Johnson and Assaf Naor. The Johnson-Lindenstrauss lemma almost characterizes Hilbert space, but not quite. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 885–891. SIAM, Philadelphia, PA, 2009. 80

[Kah64] Jean-Pierre Kahane. Sur les sommes vectorielles sigma plus minus un. *COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES*, 259(16):2577, 1964. 17

[Kat98] Eleni Katirtzoglou. Type and cotype of musielak–orlicz sequence spaces. *Journal of mathematical analysis and applications*, 226(2):431–455, 1998. 12

[KN12] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *Communications on Pure and Applied Mathematics*, 65(7):992–1035, 2012. 6, 15

[KNS08] Guy Kindler, Assaf Naor, and Gideon Schechtman. The UGC hardness threshold of the l_p Grothendieck problem. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 64–73. ACM, New York, 2008. 2, 6

[KO09] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory OF Computing*, 5:83–117, 2009. 15

[Kou91] Omran Kouba. On the interpolation of injective or projective tensor products of banach spaces. *Journal of Functional Analysis*, 96(1):38–61, 1991. 61

[Kri73] JL Krivine. Théoremes de factorisation dans les espaces réticulés. *Séminaire Analyse fonctionnelle (dit" Maurey-Schwartz")*, pages 1–22, 1973. 15, 49, 64, 66

[LNRW19] Jerry Li, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. On mean estimation for general norms with statistical queries. *arXiv preprint arXiv:1902.02459*, 2019. 12

[LP68] Joram Lindenstrauss and Aleksander Pełczyński. Absolutely summing operators in $\mathcal{L}_p$-spaces and their applications. *Studia Mathematica*, 29(3):275–326, 1968. 15, 53

[LPP91] Françoise Lust-Piquard and Gilles Pisier. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.*, 29(2):241–260, 1991. 16, 71

[LPX07] Françoise Lust-Piquard and Quanhua Xu. The little grothendieck theorem and khintchine inequalities for symmetric spaces of measurable operators. *Journal of Functional Analysis*, 244(2):488–503, 2007. 71

[LRV13] Anand Louis, Prasad Raghavendra, and Santosh Vempala. The complexity of approximating vertex expansion. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 360–369. IEEE, 2013. 16

[Lue97] David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1997. 53, 55

[Mau73] B Maurey. Type et cotype dans les espaces munis de structures locales inconditionnelles. *Séminaire Analyse fonctionnelle (dit" Maurey-Schwartz")*, pages 1–25, 1973. 65

[Mau74] Bernard Maurey. Théorèmes de factorisation pour les opérateurs à valeurs dans un espace l_p. *Séminaire Analyse fonctionnelle (dit)*, pages 1–5, 1974. 51

[Mau03] Bernard Maurey. Type, cotype and K-convexity. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1299–1332. North-Holland, Amsterdam, 2003. 3

[MP76] Bernard Maurey and Gilles Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.*, 58(1):45–90, 1976. 4, 18

[MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014. 16

[MS86] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov. 3, 77

[Nes98] Yurii Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software*, 9(1-3):141–160, 1998. 2, 16, 50, 64, 66

[NRV13] Assaf Naor, Oded Regev, and Thomas Vidick. Efficient rounding for the noncommutative grothendieck inequality. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 71–80. ACM, 2013. 14, 16, 51, 64

[NS09] Assaf Naor and Gideon Schechtman. An approximation scheme for quadratic form maximization on convex bodies. 2009. 2, 13, 16, 63, 65

[Pis78] Gilles Pisier. Grothendieck's theorem for noncommutative C*-algebras, with an appendix on grothendieck's constants. *Journal of Functional Analysis*, 29(3):397–415, 1978. 16

[Pis80] Gilles Pisier. 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):23–43, 1980. 4, 10, 41, 51, 54

[Pis83] Gilles Pisier. On the dimension of the ℓ_p^n-subspaces of banach spaces, for $1 \leqslant p < 2$. *Transactions of the American Mathematical Society*, 276(1):201–211, 1983. 77

[Pis86] Gilles Pisier. *Factorization of linear operators and geometry of Banach spaces*. American Mathematical Soc., 1986. 7, 10, 41, 50, 53, 65

[Pis12] Gilles Pisier. Grothendieck's theorem, past and present. *Bulletin of the American Mathematical Society*, 49(2):237–323, 2012. 15, 49

[RRS16] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold. *arXiv preprint arXiv:1605.00058*, 2016. 16

[RS09] Prasad Raghavendra and David Steurer. Towards computing the Grothendieck constant. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 525–534. Society for Industrial and Applied Mathematics, 2009. 15

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd annual ACM Symposium on Theory of Computing*, pages 755–764. ACM, 2010. 6, 74

[Rud73] Walter Rudin. *Functional analysis*. McGraw-Hill Book Co., New York-Düsseldorf-Johannesburg, 1973. McGraw-Hill Series in Higher Mathematics. 3

[RV15] Oded Regev and Thomas Vidick. Quantum xor games. *ACM Transactions on Computation Theory (ToCT)*, 7(4):15, 2015. 16

[SV15] Yaron Singer and Jan Vondrák. Information-theoretic lower bounds for convex optimization with erroneous oracles. In *Advances in Neural Information Processing Systems*, pages 3204–3212, 2015. 33

[SWY+19] Zhao Song, Ruosong Wang, Lin Yang, Hongyang Zhang, and Peilin Zhong. Efficient symmetric norm regression via linear sketching. In *Advances in Neural Information Processing Systems*, pages 830–840, 2019. 12

[SWZ19] Zhao Song, David Woodruff, and Peilin Zhong. Towards a zero-one law for column subset selection. In *Advances in Neural Information Processing Systems*, pages 6123–6134, 2019. 12

[TJ79] Nicole Tomczak-Jaegermann. Computing 2-summing norm with few vectors. *Ark. Mat.*, 17(2):273–277, 1979. 80

[Tro09] Joel A Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 978–986. Society for Industrial and Applied Mathematics, 2009. 49