

Improved Approximation Bound for Quadratic Optimization Problems with Orthogonality Constraints (Extended Abstract)

Anthony Man–Cho So*

November 21, 2007

Abstract

In this paper we consider approximation algorithms for a class of quadratic optimization problems that contain *orthogonality constraints*, i.e. constraints of the form $X^T X = I$, where $X \in \mathbb{R}^{m \times n}$ is the optimization variable. Such class of problems, which we denote by (QP–OC), is quite general and captures several well–studied problems in the literature as special cases. In a recent work, Nemirovski [*Math. Prog.* 109:283–317, 2007] gave the first non–trivial approximation algorithm for (QP–OC). His algorithm is based on semidefinite programming and has an approximation guarantee of $O((m+n)^{1/3})$. We improve upon this result by providing the first logarithmic approximation guarantee for (QP–OC). Specifically, we show that (QP–OC) can be approximated to within a factor of $O(\ln(\max\{m, n\}))$. The main technical tool used in the analysis is a concentration inequality for the spectral norm of a sum of certain random matrices, which we develop using tools from functional analysis. Such inequality also has ramifications in the design of so–called safe tractable approximations of chance constrained optimization problems. In particular, we use it to improve a recent result of Ben–Tal and Nemirovski [*Manuscript*, 2007] on certain chance constrained linear matrix inequality systems.

*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E–mail: manchoso@se.cuhk.edu.hk

1 Introduction

In recent years, semidefinite programming (SDP) has become an invaluable tool in the design of approximation algorithms. Beginning with the seminal work of Goemans and Williamson [8], who showed how SDP can be used to obtain good approximation algorithms for MAX-CUT and various satisfiability problems, researchers have successfully employed the SDP approach to design approximation algorithms for problems in combinatorial optimization (see, e.g., [7, 4, 1, 3, 2]), telecommunications (see, e.g., [13, 15]) and quadratic optimization (see, e.g., [26, 19, 11, 22]). In fact, for many of those problems, the SDP approach yields the best approximation known to date. In this paper, we consider an SDP-based approximation algorithm for the following class of quadratic optimization problems:

$$\begin{aligned}
 & \text{maximize} && X \bullet \mathcal{A}X \\
 & \text{subject to} && X \bullet \mathcal{B}X \leq 1 && (a) \\
 & && X \bullet \mathcal{B}_i X \leq 1 && \text{for } i = 1, \dots, L && (b) \\
 & \text{(QP-OC)} && \mathcal{C}X = \mathbf{0} && (c) \\
 & && \|X\|_\infty \leq 1 && (d) \\
 & && X \in \mathcal{M}^{m,n} && (e)
 \end{aligned}$$

Here,

- $\mathcal{M}^{m,n}$ is the space of $m \times n$ real matrices equipped with the Frobenius inner product:

$$X \bullet Y = \text{tr}(XY^T) = \sum_{i=1}^m \sum_{j=1}^n X_{ij}Y_{ij} = \text{tr}(X^TY)$$

- $\mathcal{A}, \mathcal{B}, \mathcal{B}_1, \dots, \mathcal{B}_L : \mathcal{M}^{m,n} \rightarrow \mathcal{M}^{m,n}$ are symmetric linear mappings (in particular, they can be represented as symmetric $mn \times mn$ matrices) and \mathcal{A} is not negative semidefinite (so that the optimal value of (QP-OC) is always positive);
- \mathcal{B} is positive semidefinite and has rank at most 1;
- $\mathcal{B}_1, \dots, \mathcal{B}_L$ are positive semidefinite;
- $\mathcal{C} : \mathcal{M}^{m,n} \rightarrow \mathbb{R}^u$ is a linear mapping (in particular, it can be represented as an $u \times mn$ matrix);
- $\|X\|_\infty$ is the spectral norm (i.e. the largest singular value) of X (alternatively, we have $\|X\|_\infty = \max \{\|Xv\|_2 : v \in \mathbb{R}^n, \|v\|_2 = 1\}$ by the Courant–Fischer theorem; see, e.g., [12, Theorem 7.3.10]).

The problem (QP-OC) is quite general and captures several well-studied problems in the literature as special cases. As an illustration, let us consider two such problems, namely the Procrustes Problem and the so-called orthogonal relaxation of the Quadratic Assignment Problem. We remark that both problems contain the orthogonality constraint $X^TX = I$ and at first sight do not seem to fit into the form (QP-OC). However, by exploiting the structure of these problems, we may relax the orthogonality constraint to the norm constraint $\|X\|_\infty \leq 1$ with no loss of generality.

The Procrustes Problem.

In the Procrustes Problem, one is given K collections $\mathcal{P}_1, \dots, \mathcal{P}_K$ of points in \mathbb{R}^n with $|\mathcal{P}_1| = \dots = |\mathcal{P}_K| = m$, and the goal is to find rotations that make these collections as close to each other as possible. More precisely, let A_i be an $n \times m$ matrix whose l -th column represents the l -th point in the i -th collection, where $i = 1, \dots, K$ and $l = 1, \dots, m$. The goal is to find K $n \times n$ orthogonal matrices X_1, \dots, X_K such that the quantity:

$$\sum_{1 \leq i < j \leq K} \sum_{l=1}^m \|X_i A_{il} - X_j A_{jl}\|_2^2$$

is minimized. Here, A_{il} is the l -th column of the matrix A_i , where $i = 1, \dots, K$ and $l = 1, \dots, m$. Note that the quantity $\|X_i A_{il} - X_j A_{jl}\|_2^2$ represents the squared Euclidean distance between the l -th transformed point in the i -th collection and the l -th transformed point in the j -th collection. The Procrustes Problem is first studied in psychometrics and has now found applications in shape and image analyses, market research and biometric identification, just to name a few (see [9] for details). It is not hard to show that the Procrustes Problem as defined above is equivalent to:

$$\text{maximize} \quad \sum_{1 \leq i < j \leq K} \text{tr}(A_i^T X_i^T X_j A_j) \quad \text{subject to} \quad X_i^T X_i = I \text{ for } i = 1, \dots, K \quad (1)$$

Moreover, it can be shown [17] that Problem (1) has the same optimal value as the following problem:

$$\text{maximize} \quad \sum_{1 \leq i < j \leq K} \text{tr}(A_i^T X_i^T X_j A_j) \quad \text{subject to} \quad \|X_i\|_\infty \leq 1 \text{ for } i = 1, \dots, K$$

Thus, after some elementary manipulations, we see that the Procrustes Problem can be cast into the form (QP-OC). □

Orthogonal Relaxation of the Quadratic Assignment Problem.

In the Quadratic Assignment Problem (QAP), one is given a set $\mathcal{N} = \{1, \dots, n\}$ and three $n \times n$ matrices A , B and C , and the goal is to find a permutation π on \mathcal{N} such that the quantity $\sum_{i=1}^n \sum_{j=1}^n A_{\pi(i)\pi(j)} B_{ij} - 2 \sum_{i=1}^n C_{i\pi(i)}$ is maximized. Equivalently, one can formulate the QAP as follows (see [14, 25]):

$$\begin{aligned} &\text{maximize} \quad \text{tr}(AXBX^T - 2CX^T) \\ &\text{subject to} \quad XX^T = I \\ &\quad \quad \quad X_{ij} \in \{0, 1\} \quad \quad \quad \text{for } i = 1, \dots, n; j = 1, \dots, n \end{aligned} \quad (2)$$

The constraints in (2) force the matrix X to be a permutation matrix. Indeed, it is well-known that X satisfies the constraints in (2) iff X is a permutation matrix. The QAP is a classical problem in combinatorial optimization and has found many applications (see, e.g., [20]). However, it is also a notoriously hard computational problem. Therefore, various relaxations have been proposed. One such relaxation, called the *orthogonal relaxation*, is obtained by dropping the binary constraints in (2). In other words, consider the following problem:

$$\text{maximize} \quad \text{tr}(AXBX^T - 2CX^T) \quad \text{subject to} \quad XX^T = I \quad (3)$$

Without loss of generality, we may assume that $A \succ \mathbf{0}$ and $B \succ \mathbf{0}$, for otherwise we may replace A by $A + \theta_A I$ and B by $B + \theta_B I$ with sufficiently large θ_A and θ_B without affecting the optimal solution. Then, it can be shown [17] that Problem (3) is equivalent to the following problem:

$$\text{maximize } \text{tr}(AXBX^T - 2CX^T) \quad \text{subject to } \|X\|_\infty \leq 1$$

which can be cast into the form (QP-OC) after a standard homogenization argument (see, e.g., [17]). \square

The main feature that distinguishes (QP-OC) from the quadratic optimization problems considered in the approximation algorithms literature is the norm constraint (QP-OC(d)). Indeed, if we drop the norm constraint (QP-OC(d)), then (QP-OC) becomes an usual quadratic program, and an $O(\ln L)$ approximation algorithm for it is known [18]. Although (QP-OC) is known to be NP-hard [17], its approximability is not known until only very recently. In a groundbreaking work [17], Nemirovski showed that a natural semidefinite relaxation of (QP-OC) together with a simple rounding scheme yield an $O(\max\{(m+n)^{1/3}, \ln L\})$ approximation algorithm for (QP-OC). The rounding scheme proposed in [17] resembles that of Nemirovski et al. [18]. Roughly speaking, it consists of the following steps:

1. extract from the optimal SDP solution a set $S = \{v_1, \dots, v_{mn}\}$ of vectors and apply a suitable orthogonal transformation to S to obtain vectors v'_1, \dots, v'_{mn}
2. generate a random vector $\xi = (\xi_1, \dots, \xi_{mn})$, where the entries are independent and take on the values ± 1 with equal probability
3. form the (random) vector $\zeta = \sum_{i=1}^{mn} \xi_i v'_i$ and extract from ζ a candidate solution matrix \widehat{X}

In order to analyze the performance of such procedure, one needs to determine the behavior of \widehat{X} with respect to both the objective function and the constraints in (QP-OC). Intuitively, the objective function and the constraints (QP-OC(a))–(QP-OC(c)) pose no difficulty, as one should be able to analyze the behavior of \widehat{X} with respect to those in a manner similar to that in [18]. However, it is more challenging to analyze the behavior of \widehat{X} with respect to the norm constraint (QP-OC(d)). Indeed, as it was shown in [17], the problem boils down to that of estimating the *typical* spectral norm of a sum of certain random matrices whose dimensions are *fixed*. In particular, one cannot utilize the powerful asymptotic results in Random Matrix Theory. Nevertheless, Nemirovski was able to circumvent this difficulty and develop bounds for the norm estimation problem. However, the bounds he obtained are not entirely satisfactory, and consequently he can only obtain a polynomial approximation guarantee for (QP-OC).

From the above discussion, we see that one way of improving the approximation guarantee for (QP-OC) is to obtain better bounds for the norm estimation problem. In this paper, we show that the norm estimation problem is closely related to classical inequalities in functional analysis. Specifically, using a non-commutative version of Khintchine's inequality [16, 21], we are able to obtain optimal bounds for the norm estimation problem. As a corollary, we show that the SDP-based algorithm described in [17] actually yields an $O(\ln(\max\{m, n, L\}))$ approximation for (QP-OC). This significantly improves upon the $O(\max\{(m+n)^{1/3}, \ln L\})$ bound established in [17] and provides the first logarithmic approximation guarantee for (QP-OC). Our bounds for the norm estimation problem also have ramifications in the design of so-called *safe tractable approximations* of chance constrained optimization problems. In particular, our results allow us to simplify and improve a result of Ben-Tal and Nemirovski [5] on certain chance constrained

linear matrix inequality systems. We believe that our techniques are of independent interest and will find further applications. In particular, they can be useful for analyzing norm constraints in other quadratic optimization problems.

The rest of this paper is organized as follows. In Section 2 we derive a natural semidefinite relaxation of (QP–OC). In Section 3, we describe a rounding scheme for the semidefinite relaxation and analyze its performance. The main technical tool used in the analysis is a concentration inequality for the spectral norm of a sum of certain random matrices. This is developed also in Section 3 using a non-commutative version of Khintchine’s inequality. As a further illustration of the power of the aforementioned concentration inequality, we show how it can be used in a chance constrained optimization problem in Section 4. Finally, the concluding remarks will be given in Section 5.

2 A Semidefinite Relaxation of (QP–OC)

We now derive a natural semidefinite relaxation of (QP–OC). The ideas are standard: we first linearize the quadratic terms and then tighten the relaxation with positive semidefinite constraints. To begin, let us identify the mapping \mathcal{A} with an $mn \times mn$ symmetric matrix A whose rows and columns are indexed by pairs (i, j) , where $i = 1, \dots, m$ and $j = 1, \dots, n$, and whose entries are determined by:

$$(\mathcal{A}X)_{ij} = \sum_{k=1}^m \sum_{l=1}^n A_{(i,j)(k,l)} X_{kl} \quad \text{for } i = 1, \dots, m; j = 1, \dots, n$$

In a similar fashion, we identify the mappings \mathcal{B} and \mathcal{B}_i with $mn \times mn$ symmetric positive semidefinite matrices B and B_i , where B has rank 1 and $i = 1, \dots, L$. For the mapping \mathcal{C} , we identify it with an $u \times mn$ matrix C whose entries are determined by:

$$(\mathcal{C}X)_i = \sum_{k=1}^m \sum_{l=1}^n C_{i,(k,l)} X_{kl} \quad \text{for } i = 1, \dots, u$$

Now, for $X \in \mathcal{M}^{m,n}$, let $\text{Vec}(X)$ be the mn -dimensional vector obtained by arranging the columns of X into a single column, and let $\text{Gram}(X)$ be the $mn \times mn$ positive semidefinite matrix $\text{Vec}(X)\text{Vec}(X)^T$. It is then clear that:

$$X \bullet \mathcal{A}X = \sum_{i=1}^m \sum_{j=1}^n X_{ij} (\mathcal{A}X)_{ij} = \sum_{i,k=1}^m \sum_{j,l=1}^n A_{(i,j)(k,l)} X_{ij} X_{kl} = A \bullet \text{Gram}(X)$$

Similarly, we have $X \bullet \mathcal{B}X = B \bullet \text{Gram}(X)$ and $X \bullet \mathcal{B}_i X = B_i \bullet \text{Gram}(X)$ for $i = 1, \dots, L$. Next, observe that for $i = 1, \dots, u$, we have $(\mathcal{C}X)_i = 0$ iff

$$\left(\sum_{k=1}^m \sum_{l=1}^n C_{i,(k,l)} X_{kl} \right)^2 = \sum_{k,k'=1}^m \sum_{l,l'=1}^n C_{i,(k,l)} C_{i,(k',l')} X_{kl} X_{k'l'} = \text{Gram}(C_i) \bullet \text{Gram}(X) = 0$$

where C_i is the $m \times n$ matrix $[C_{i,(k,l)}]_{1 \leq k \leq m, 1 \leq l \leq n}$. Finally, observe that $\|X\|_\infty \leq 1$ iff $XX^T \preceq I$. Now, for $i = 1, \dots, m$ and $j = 1, \dots, n$, the (i, j) -th entry of XX^T is $\sum_{k=1}^n X_{ik} X_{jk}$. It follows that the entries of XX^T are linear combinations of the entries in $\text{Gram}(X)$, which in turn implies

the existence of a linear mapping $\mathcal{S} : \mathcal{S}^{mn} \rightarrow \mathcal{S}^m$ such that $XX^T \preceq I$ iff $\mathcal{S} \text{Gram}(X) \preceq I$ (here, \mathcal{S}^m is the space of $m \times m$ symmetric real matrices). In a similar fashion, we have $\|X\|_\infty \leq 1$ iff $X^T X \preceq I$, and there exists a linear mapping $\mathcal{T} : \mathcal{S}^{mn} \rightarrow \mathcal{S}^n$ such that $X^T X \preceq I$ iff $\mathcal{T} \text{Gram}(X) \preceq I$. Note that both the linear mappings \mathcal{S} and \mathcal{T} can be specified explicitly as matrices of appropriate dimensions in polynomial time. Now, with the above notation and observations, we see that (QP-OC) has the following equivalent formulation:

$$\begin{aligned}
& \text{maximize} && A \bullet \text{Gram}(X) \\
& \text{subject to} && B \bullet \text{Gram}(X) \leq 1 \\
(\text{QP-OC}') & && B_i \bullet \text{Gram}(X) \leq 1 && \text{for } i = 1, \dots, L \\
& && \text{Gram}(C_i) \bullet \text{Gram}(X) = 0 && \text{for } i = 1, \dots, u \\
& && \mathcal{S} \text{Gram}(X) \preceq I, \mathcal{T} \text{Gram}(X) \preceq I \\
& && X \in \mathcal{M}^{m,n}
\end{aligned}$$

Since $\text{Gram}(X) \succeq \mathbf{0}$, we obtain the following straightforward SDP relaxation of (QP-OC'):

$$\begin{aligned}
& \text{maximize} && A \bullet Y \\
& \text{subject to} && B \bullet Y \leq 1 \\
(\text{QP-OC-SDR}) & && B_i \bullet Y \leq 1 && \text{for } i = 1, \dots, L \\
& && \text{Gram}(C_i) \bullet Y = 0 && \text{for } i = 1, \dots, u \\
& && \mathcal{S}Y \preceq I, \mathcal{T}Y \preceq I \\
& && Y \in \mathcal{S}^{mn}, Y \succeq \mathbf{0}
\end{aligned}$$

Although the constraints $\mathcal{S} \text{Gram}(X) \preceq I$ and $\mathcal{T} \text{Gram}(X) \preceq I$ imply each other and are thus redundant in (QP-OC'), the corresponding relaxed constraints $\mathcal{S}Y \preceq I$ and $\mathcal{T}Y \preceq I$ are *not* redundant in (QP-OC-SDR). In fact, as we shall see, they play a crucial role in the quality analysis of (QP-OC-SDR).

Now, using the ellipsoid method [10], the semidefinite program (QP-OC-SDR) can be solved to within an additive error of $\epsilon > 0$ in polynomial time. Specifically, let θ^* be the optimal value of (QP-OC-SDR). Then, for any $\epsilon > 0$, we can compute in polynomial time an $Y' \succeq \mathbf{0}$ that is feasible for (QP-OC-SDR) and satisfies $\theta' \equiv A \bullet Y' \geq \theta^* - \epsilon$.

3 Analysis of the SDP Relaxation

In this section we prove the following theorem, which is the main result of this paper:

Theorem 1 *There exists an efficient randomized algorithm that, given a feasible solution of (QP-OC-SDR) with objective value θ' , produces an $m \times n$ matrix \bar{X} such that:*

- (a) \bar{X} is feasible for (QP-OC)
- (b) $\bar{X} \bullet \mathcal{A}\bar{X} \geq \Omega(1/\ln(\max\{m, n, L\})) \cdot \theta'$

To begin, let us consider the following rounding scheme of Nemirovski [17] that converts a feasible solution Y' of (QP-OC-SDR) into a random $m \times n$ matrix \hat{X} . Since $Y' \succeq \mathbf{0}$, there exists a positive semidefinite matrix $Y'^{1/2} \in \mathcal{S}^{mn}$ such that $Y' = Y'^{1/2} Y'^{1/2}$. Moreover, the matrix $Y'^{1/2} A Y'^{1/2}$

is symmetric, and hence it admits a spectral decomposition $Y^{1/2}AY^{1/2} = U^T\Lambda U$, where Λ is an $mn \times mn$ diagonal matrix and U is an $mn \times mn$ orthogonal matrix. Now, we generate a random mn -dimensional vector $\xi = (\xi_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$, where the entries are independent and take on the values ± 1 with equal probability, and define the random $m \times n$ matrix \widehat{X} via $\text{Vec}(\widehat{X}) = Y^{1/2}U^T\xi$.

Clearly, the above rounding scheme can be implemented in polynomial time. We are now interested in the quality of the solution \widehat{X} . In the sequel, we assume that $\max\{m, n, L\} \geq 7$. The following proposition is straightforward and we omit the proof.

Proposition 1 *The following hold:*

- (a) $\widehat{X} \bullet \mathcal{A}\widehat{X} \equiv \theta'$
- (b) $\mathbb{E} [\widehat{X} \bullet \mathcal{B}\widehat{X}] \leq 1$
- (c) $\mathbb{E} [\widehat{X} \bullet \mathcal{B}_i\widehat{X}] \leq 1$ for $i = 1, \dots, L$
- (d) $\mathcal{C}\widehat{X} \equiv \mathbf{0}$
- (e) $\mathbb{E} [\widehat{X}\widehat{X}^T] \preceq I$ and $\mathbb{E} [\widehat{X}^T\widehat{X}] \preceq I$

To obtain the results claimed in Theorem 1, we need to analyze the behavior of \widehat{X} with respect to the constraints (QP-OC-(a)), (QP-OC-(b)) and (QP-OC-(d)). Specifically, we would like to show that the following event:

$$\left\{ \left\{ \widehat{X} \bullet \mathcal{B}\widehat{X} \leq 1 \right\} \cap \left\{ \widehat{X} \bullet \mathcal{B}_i\widehat{X} \leq \Gamma^2 \text{ for all } i = 1, \dots, L \right\} \cap \left\{ \|\widehat{X}\|_\infty \leq \Gamma \right\} \right\}$$

occurs with constant probability, where $\Gamma = O\left(\sqrt{\ln(\max\{m, n, L\})}\right)$. Let us first tackle the norm constraint (QP-OC-(d)). One of the contributions of this paper is to show that a measure concentration phenomenon occurs for the spectral norm $\|\widehat{X}\|_\infty$. More precisely, we prove the following:

Theorem 2 *For any $\beta \geq 1/16$, we have:*

$$\Pr\left(\|\widehat{X}\|_\infty \geq e\sqrt{(1+\beta)} \cdot \sqrt{\ln(\max\{m, n\})}\right) \leq (\max\{m, n\})^{-\beta}$$

To begin, let us observe that the matrix \widehat{X} has the form $\sum_{k=1}^m \sum_{l=1}^n \xi_{kl} Q_{kl}$, where each Q_{kl} is an $m \times n$ matrix. Indeed, the (i, j) -th entry of Q_{kl} is simply $(Y^{1/2}U^T)_{(i,j)(k,l)}$, where $i = 1, \dots, m$ and $j = 1, \dots, n$. Moreover, we have:

$$\mathbb{E} [\widehat{X}\widehat{X}^T] = \sum_{i=1}^m \sum_{j=1}^n Q_{ij} Q_{ij}^T \quad \text{and} \quad \mathbb{E} [\widehat{X}^T\widehat{X}] = \sum_{i=1}^m \sum_{j=1}^n Q_{ij}^T Q_{ij}$$

Thus, in order to prove Theorem 2, it suffices to prove the following:

Theorem 2' *Let ξ_1, \dots, ξ_h be independent random variables that take on the values ± 1 with equal probability, and let Q_1, \dots, Q_h be $m \times n$ matrices satisfying the relations $\sum_{i=1}^h Q_i Q_i^T \preceq I$ and $\sum_{i=1}^h Q_i^T Q_i \preceq I$. Set $S = \sum_{i=1}^h \xi_i Q_i$. Then, for any $\beta \geq 1/16$, we have:*

$$\Pr\left(\|S\|_\infty \geq e\sqrt{(1+\beta)} \cdot \sqrt{\ln(\max\{m, n\})}\right) \leq (\max\{m, n\})^{-\beta}$$

The proof of Theorem 2' involves estimating the moments of the random variable $\|S\|_\infty$. Towards that end, let us begin with some notation. For an $m \times n$ matrix X , let $\sigma(X)$ denote the vector of singular values of X . For $1 \leq p < \infty$, define the *Schatten p -norm* $\|X\|_{S_p}$ of the matrix X by $\|X\|_{S_p} = \|\sigma(X)\|_p$, where $\|\cdot\|_p$ is the usual ℓ_p -norm. The following remarkable result is due to Lust-Piquard [16]:

Fact 1 (Non-Commutative Khintchine's Inequality) *Let ξ_1, \dots, ξ_h be independent random variables that take on the values ± 1 with equal probability, and let Q_1, \dots, Q_h be $m \times n$ matrices. Set $S = \sum_{i=1}^h \xi_i Q_i$. Then, for $2 \leq p < \infty$, there exists a constant $\gamma_p > 0$ such that:*

$$\left(\mathbb{E} \left[\|S\|_{S_p}^p \right]\right)^{1/p} \leq \gamma_p \cdot \max \left\{ \left\| \left(\sum_{i=1}^h Q_i Q_i^T \right)^{1/2} \right\|_{S_p}, \left\| \left(\sum_{i=1}^h Q_i^T Q_i \right)^{1/2} \right\|_{S_p} \right\}$$

We remark that Lust-Piquard did not provide an estimate for γ_p . However, it is shown in [21] that $\gamma_p \leq \alpha \sqrt{p}$ for some absolute constant $\alpha > 0$. Using a result of Buchholz [6], it can be shown that $\alpha \leq 2^{-1/4}(\pi/e)^{1/2} < 1$ (see [24]).

Now, it is straightforward to deduce the following proposition from Fact 1:

Proposition 2 *Let $2 \leq p < \infty$. Under the setting of Theorem 2', we have $(\mathbb{E} [\|S\|_\infty^p])^{1/p} \leq \sqrt{p} \cdot (\max\{m, n\})^{1/p}$.*

Proof Under the setting of Theorem 2', all the eigenvalues of $\sum_{i=1}^h Q_i Q_i^T$ and $\sum_{i=1}^h Q_i^T Q_i$ lie in $[0, 1]$. It follows that:

$$\left\| \left(\sum_{i=1}^h Q_i Q_i^T \right)^{1/2} \right\|_{S_p} \leq m^{1/p} \quad \text{and} \quad \left\| \left(\sum_{i=1}^h Q_i^T Q_i \right)^{1/2} \right\|_{S_p} \leq n^{1/p}$$

Using Fact 1, we have $(\mathbb{E} [\|S\|_\infty^p])^{1/p} \leq \left(\mathbb{E} [\|S\|_{S_p}^p]\right)^{1/p} \leq \sqrt{p} \cdot (\max\{m, n\})^{1/p}$ as desired. \square

We are now ready to prove Theorem 2'. By Markov's inequality, for any $t > 0$ and $2 \leq p < \infty$, we have:

$$\Pr(\|S\|_\infty \geq t) \leq \frac{\mathbb{E} [\|S\|_\infty^p]}{t^p} \leq \frac{p^{p/2} \cdot \max\{m, n\}}{t^p} \quad (4)$$

Upon setting $t = e\sqrt{(1+\beta)} \cdot \sqrt{\ln(\max\{m, n\})}$ and $p = (t/e)^2 \geq 2$, we see that:

$$\Pr\left(\|S\|_\infty \geq e\sqrt{(1+\beta)} \cdot \sqrt{\ln(\max\{m, n\})}\right) \leq (\max\{m, n\})^{-\beta}$$

This completes the proof of Theorem 2'. \square

Remarks. *By the Central Limit Theorem, we see that Theorem 2' also holds in the case where ξ_1, \dots, ξ_h are i.i.d. standard Gaussian random variables.*

For the purpose of determining the approximation ratio of the outlined rounding scheme, we need the following proposition, whose proof is essentially the same as that of Theorem 2:

Proposition 3 *For any $\beta \geq 1/16$, we have:*

$$\Pr\left(\|\widehat{X}\|_\infty \geq e\sqrt{(1+\beta)} \cdot \sqrt{\ln(\max\{m, n, L\})}\right) \leq (\max\{m, n, L\})^{-\beta}$$

Now, it remains to analyze the behavior of \widehat{X} with respect to the constraints (QP-OC-(a)) and (QP-OC-(b)). This is done in the following proposition:

Proposition 4 *The following hold:*

(a) $\Pr(\widehat{X} \bullet \mathcal{B}\widehat{X} \leq 1) \geq 1/3$

(b) *For any $\beta \geq 1/16$, we have:*

$$\Pr\left(\widehat{X} \bullet \mathcal{B}_i \widehat{X} \geq e^2(1 + \beta) \ln(\max\{m, n, L\})\right) \leq (\max\{m, n, L\})^{-(1+\beta)}$$

for $i = 1, \dots, L$

Proof Statement (a) is established in [17, Lemma 4(a)]. To establish (b), we first compute:

$$\widehat{X} \bullet \mathcal{B}_i \widehat{X} = \text{tr}\left(B_i \text{Vec}(\widehat{X}) \text{Vec}(\widehat{X})^T\right) = \text{tr}\left(UY^{\prime 1/2} B_i Y^{\prime 1/2} U^T \xi \xi^T\right) = \xi^T B_i' \xi$$

where $B_i' = UY^{\prime 1/2} B_i Y^{\prime 1/2} U^T$ for $i = 1, \dots, L$. Since $B_i' \succeq \mathbf{0}$, we have $\widehat{X} \bullet \mathcal{B}_i \widehat{X} = \|(B_i')^{1/2} \xi\|_2^2$. Moreover, note that:

$$1 \geq \mathbb{E}\left[\widehat{X} \bullet \mathcal{B}_i \widehat{X}\right] = \sum_{k=1}^m \sum_{l=1}^n (B_i')_{(k,l)(k,l)} = \sum_{k=1}^m \sum_{l=1}^n \|(B_i')_{(\cdot, \cdot)(k,l)}\|_2^2 \quad (5)$$

where $(B_i')_{(\cdot, \cdot)(k,l)}$ is the (k, l) -th column of B_i' . Hence, it suffices to prove the following:

$$\Pr\left(\left\|\sum_{k=1}^m \sum_{l=1}^n \xi_{kl} (B_i')_{(\cdot, \cdot)(k,l)}\right\|_2 \geq e\sqrt{(1 + \beta)} \cdot \sqrt{\ln(\max\{m, n, L\})}\right) \leq (\max\{m, n, L\})^{-(1+\beta)} \quad (6)$$

Towards that end, we first recall the following result of Tomczak–Jaegermann [23]:

Fact 2 *Let ξ_1, \dots, ξ_h be independent random variables that take on the values ± 1 with equal probability, and let Q_1, \dots, Q_h be $m \times n$ matrices. Then, for $2 \leq p < \infty$, we have:*

$$\left(\mathbb{E}\left[\left\|\sum_{i=1}^h \xi_i Q_i\right\|_{S_p}^p\right]\right)^{1/p} \leq \sqrt{p} \left(\sum_{i=1}^h \|Q_i\|_{S_p}^2\right)^{1/2}$$

Now, note that for a vector v , we have $\|v\|_{S_p} = \|v\|_2$ for $1 \leq p < \infty$. Hence, by (5) and Fact 2, we have:

$$\left(\mathbb{E}\left[\left\|\sum_{k=1}^m \sum_{l=1}^n \xi_{kl} (B_i')_{(\cdot, \cdot)(k,l)}\right\|_2^p\right]\right)^{1/p} \leq \sqrt{p} \left(\sum_{k=1}^m \sum_{l=1}^n \|(B_i')_{(\cdot, \cdot)(k,l)}\|_2^2\right)^{1/2} \leq \sqrt{p}$$

By Markov's inequality, for any $t > 0$ and $2 \leq p < \infty$, we have:

$$\Pr\left(\left\|\sum_{k=1}^m \sum_{l=1}^n \xi_{kl} (B_i')_{(\cdot, \cdot)(k,l)}\right\|_2 \geq t\right) \leq t^{-p} \cdot \mathbb{E}\left[\left\|\sum_{k=1}^m \sum_{l=1}^n \xi_{kl} (B_i')_{(\cdot, \cdot)(k,l)}\right\|_2^p\right] \leq \frac{p^{p/2}}{t^p}$$

Upon setting $t = e\sqrt{(1+\beta)} \cdot \sqrt{\ln(\max\{m, n, L\})}$ and $p = (t/e)^2 \geq 2$, we obtain (6) and complete the proof. \square

We are now ready to finish the proof of Theorem 1:

Proof of Theorem 1 Let $\beta = 1$ in Propositions 3 and 4(b). Then, by Propositions 1(a,d), 3 and 4, we see that with probability at least $1/3 - 2(\max\{m, n, L\})^{-1} \geq 5/21$, the matrix \widehat{X} returned by the rounding scheme satisfies:

- (a) $\widehat{X} \bullet \mathcal{A}\widehat{X} = \theta'$
- (b) $\widehat{X} \bullet \mathcal{B}\widehat{X} \leq 1$
- (c) $\widehat{X} \bullet \mathcal{B}_i\widehat{X} \leq 2e^2 \ln(\max\{m, n, L\})$ for $i = 1, \dots, L$
- (d) $\mathcal{C}\widehat{X} \equiv \mathbf{0}$
- (e) $\|\widehat{X}\|_\infty \leq \sqrt{2e^2 \ln(\max\{m, n, L\})}$

It follows that the matrix $\overline{X} = \widehat{X} / \sqrt{2e^2 \ln(\max\{m, n, L\})}$ has the required properties. \square

4 Concentration of Spectral Norm Revisited: Application to Chance Constrained Linear Matrix Inequality Systems

As pointed out in [5, 17], Theorem 2' can also be used to design a so-called safe tractable approximation of chance constrained optimization problems of the form:

$$\begin{aligned}
 & \text{minimize} && c^T x \\
 & \text{subject to} && F(x) \leq \mathbf{0} \\
 & && \Pr \left(\mathcal{A}_0(x) + \sum_{i=1}^h \xi_i \mathcal{A}_i(x) \succeq \mathbf{0} \right) \geq 1 - \epsilon \quad (\dagger) \\
 & && x \in \mathbb{R}^n
 \end{aligned} \tag{7}$$

Here,

- F is an efficiently computable vector-valued function with convex components;
- $\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_h : \mathbb{R}^n \rightarrow \mathcal{S}^m$ are affine functions, with $\mathcal{A}_0(x) \succ \mathbf{0}$ for all $x \in \mathbb{R}^n$;
- ξ_1, \dots, ξ_h are i.i.d. mean zero random variables with a symmetric distribution;
- $\epsilon \in (0, 1)$ is the error tolerance parameter.

The above chance constrained problem arises from many engineering applications, such as truss topology design and problems in control theory. In general, the constraint (\dagger) in (7) is computationally intractable. In order to obtain a more tractable problem, we could develop a *safe tractable approximation* of (\dagger) — that is, a system of constraints \mathcal{H} such that (i) x is feasible for (\dagger) whenever it is feasible for \mathcal{H} , and (ii) the constraints in \mathcal{H} are efficiently computable — as follows. First, from the assumptions on the random perturbations ξ_1, \dots, ξ_h , we have:

$$\Pr \left(\mathcal{A}_0(x) + \sum_{i=1}^h \xi_i \mathcal{A}_i(x) \succeq \mathbf{0} \right) = \Pr \left(-I \preceq \sum_{i=1}^h \xi_i \mathcal{A}'_i(x) \preceq I \right) = \Pr \left(\left\| \sum_{i=1}^h \xi_i \mathcal{A}'_i(x) \right\|_\infty \leq 1 \right)$$

where $\mathcal{A}'_i(x) = \mathcal{A}_0(x)^{-1/2} \mathcal{A}_i(x) \mathcal{A}_0^{-1/2}$. Now, suppose that ξ_i takes on the values ± 1 with equal probability or is a standard Gaussian. Then, Theorem 2' implies that for any $\beta \geq \ln(1/\epsilon)/\ln m \geq 1/16$, if:

$$\sum_{i=1}^h (\mathcal{A}'_i(x))^2 \preceq \frac{1}{(1+\beta)e^2 \ln m} \cdot I \preceq \frac{1}{e^2 \ln(1/\epsilon)} \cdot I \tag{8}$$

then we have:

$$\Pr \left(\left\| \sum_{i=1}^h \xi_i \mathcal{A}'_i(x) \right\|_{\infty} \leq 1 \right) \geq 1 - \frac{1}{m^{\beta}} \geq 1 - \epsilon$$

Observe that by the Schur complement, (8) is equivalent to:

$$\begin{bmatrix} \gamma \mathcal{A}_0(x) & \mathcal{A}_1(x) & \cdots & \mathcal{A}_h(x) \\ \mathcal{A}_1(x) & \gamma \mathcal{A}_0(x) & & \\ \vdots & & \ddots & \\ \mathcal{A}_h(x) & & & \gamma \mathcal{A}_0(x) \end{bmatrix} \succeq \mathbf{0} \tag{9}$$

where $\gamma = \left(e \cdot \sqrt{\ln(1/\epsilon)} \right)^{-1}$. Thus, whenever $\ln(1/\epsilon) \geq \frac{1}{16} \ln m$, the positive semidefinite constraint (9) is a safe tractable approximation of (†). We remark that this is an improvement over Ben-Tal and Nemirovski's result [5], which requires that $\ln(1/\epsilon) \geq \Omega(m^{1/3})$. Now, by replacing (†) with (9), Problem (7) becomes tractable. Moreover, any solution x that satisfies $F(x) \leq \mathbf{0}$ and (9) will be feasible for the original chance constrained problem.

5 Conclusion

In this paper we studied a class of quadratic optimization problems with orthogonality constraints and showed that an SDP-based algorithm yields a solution whose objective value is within a logarithmic factor of the optimum. Our proof relies on a concentration inequality for the spectral norm of a sum of certain random matrices. Such inequality is developed using a non-commutative version of Khintchine's inequality. We also showed how such a concentration inequality can be used to develop improved safe tractable approximations of certain chance constrained linear matrix inequality systems. An interesting future direction would be to find other applications for which our techniques apply.

References

- [1] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ Approximation Algorithms for Min UnCut, Min 2CNF Deletion, and Directed Cut Problems. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC '05)*, pages 573–581, 2005.
- [2] N. Alon, K. Makarychev, Y. Makarychev, and A. Naor. Quadratic Forms on Graphs. *Inventiones Mathematicae*, 163(3):499–522, 2006.
- [3] S. Arora, J. R. Lee, and A. Naor. Euclidean Distortion and the Sparsest Cut. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC '05)*, pages 553–562, 2005.
- [4] S. Arora, S. Rao, and U. Vazirani. Expander Flows, Geometric Embeddings and Graph Partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC '04)*, pages 222–231, 2004.
- [5] A. Ben-Tal and A. Nemirovski. On Safe Tractable Approximations of Chance Constrained Linear Matrix Inequalities. Manuscript, 2007.
- [6] A. Buchholz. Operator Khintchine Inequality in Non-Commutative Probability. *Mathematische Annalen*, 319:1–16, 2001.
- [7] M. X. Goemans. Semidefinite Programming in Combinatorial Optimization. *Mathematical Programming*, 79:143–161, 1997.
- [8] M. X. Goemans and D. P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, 42(6):1115–1145, 1995.
- [9] J. C. Gower and G. B. Dijkstra. *Procrustes Problems*, volume 30 of *Oxford Statistical Science Series*. Oxford University Press, 2004.
- [10] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, second corrected edition, 1993.
- [11] S. He, Z.-Q. Luo, J. Nie, and S. Zhang. Semidefinite Relaxation Bounds for Indefinite Homogeneous Quadratic Optimization. Technical Report SEEM2007–01, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong, 2007.
- [12] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [13] M. Kisiailiou and Z.-Q. Luo. Performance Analysis of Quasi-Maximum-Likelihood Detector Based on Semidefinite Programming. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, volume 3, pages 433–436, 2005.
- [14] T. C. Koopmans and M. Beckmann. Assignment Problems and the Location of Economic Activities. *Econometrica*, 25(1):53–76, 1957.

- [15] Z.-Q. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang. Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints. *SIAM Journal on Optimization*, 18(1):1–28, 2007.
- [16] F. Lust-Piquard. Inégalités de Khintchine dans C_p ($1 < p < \infty$). *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, 303(7):289–292, 1986.
- [17] A. Nemirovski. Sums of Random Symmetric Matrices and Quadratic Optimization under Orthogonality Constraints. *Mathematical Programming, Series B*, 109(2–3):283–317, 2007.
- [18] A. Nemirovski, C. Roos, and T. Terlaky. On Maximization of Quadratic Form over Intersection of Ellipsoids with Common Center. *Mathematical Programming, Series A*, 86:463–473, 1999.
- [19] Y. Nesterov, H. Wolkowicz, and Y. Ye. Semidefinite Programming Relaxations of Nonconvex Quadratic Optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, International Series in Operations Research and Management Science, pages 361–419. Kluwer Academic Publishers, 2000.
- [20] P. M. Pardalos and H. Wolkowicz, editors. *Quadratic Assignment and Related Problems*, volume 16 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, 1994.
- [21] G. Pisier. Non-Commutative Vector Valued L_p -Spaces and Completely p -Summing Maps. *Astérisque*, 247, 1998.
- [22] A. M.-C. So, Y. Ye, and J. Zhang. A Unified Theorem on SDP Rank Reduction. Manuscript, 2007.
- [23] N. Tomczak-Jaegermann. The Moduli of Smoothness and Convexity and the Rademacher Averages of Trace Classes S_p ($1 \leq p < \infty$). *Studia Mathematica*, 50:163–182, 1974.
- [24] J. A. Tropp. The Random Paving Property for Uniformly Bounded Matrices. Manuscript, 2007.
- [25] H. Wolkowicz. Semidefinite Programming Approaches to the Quadratic Assignment Problem. In P. M. Pardalos and L. S. Pitsoulis, editors, *Nonlinear Assignment Problems: Algorithms and Applications*, pages 143–174. Kluwer Academic Publishers, 2000.
- [26] Y. Ye. Approximating Global Quadratic Optimization with Convex Quadratic Constraints. *Journal of Global Optimization*, 15(1):1–17, 1999.