Improved Approximation Bound for Quadratic Optimization Problems with Orthogonality Constraints*

Anthony Man–Cho So†

Abstract

In this paper we consider the problem of approximating a class of quadratic optimization problems that contain *orthogonality constraints*, i.e. constraints of the form $X^TX = I$, where $X \in \mathbb{R}^{m\times n}$ is the optimization variable. This 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 [17] gave the first non–trivial approximation algorithm for (Qp–Oc). His algorithm is based on semidefinite programming and has an approximation guarantee of $O\left((m+n)^{1/3}\right)$. 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 the so–called non–commutative Khintchine inequality, which allows us to prove a concentration inequality for the spectral norm of a Rademacher sum of matrices. As a by–product, we resolve in the affirmative a conjecture of Nemirovski concerning the typical spectral norm of a sum of certain random matrices. The aforementioned concentration inequality also has ramifications in the design of so–called safe tractable approximations of chance constrained optimization problems. In particular, we use it to simplify and improve a recent result of Ben–Tal and Nemirovski [4] concerning certain chance constrained linear matrix inequality systems.

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, 1, 2, 3]), telecommunications (see, e.g., [15, 25, 24]) and quadratic optimization (see, e.g., [21, 27, 11, 26]). 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:

(Qp–Oc)

$$
\begin{array}{lll}
\text{maximize} & X \bullet \mathcal{A}X & \\
\text{subject to} & X \bullet \mathcal{B}X \le 1 & (a) \\
 & X \bullet \mathcal{B}_iX \le 1 \quad \text{for } i=1,\ldots,L & (b) \\
 & \mathcal{C}X = \mathbf{0} & (c) \\
 & \|X\|_\infty \le 1 & (d) \\
 & X \in \mathscr{M}^{m,n} & (e)
\end{array}
$$

Here,

- $\mathscr{M}^{m,n}$ is the space of $m\times n$ real matrices equipped with the Frobenius inner product:
$$X \bullet Y = \operatorname{tr}\left(XY^T\right) = \operatorname{tr}\left(X^TY\right);$$
- $\mathcal{A}$, $\mathcal{B}$, $\mathcal{B}_1,\ldots,\mathcal{B}_L : \mathscr{M}^{m,n} \to \mathscr{M}^{m,n}$ are symmetric linear mappings (in particular, they can be represented as symmetric $mn\times mn$ matrices);
- $\mathcal{B}$ is positive semidefinite and has rank at most 1;
- $\mathcal{B}_1,\ldots,\mathcal{B}_L$ are positive semidefinite;
- $\mathcal{C} : \mathscr{M}^{m,n} \to \mathbb{R}^u$ 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., [13, 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

*This research is supported by CUHK Direct Grant No. 2050401.

†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

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 in these problems, we may relax the orthogonality constraint to the norm constraint $\|X\|_\infty \le 1$ with no loss of generality.

The Procrustes Problem.

In the Procrustes Problem, one is given K collections $\mathcal{P}_1, \ldots, \mathcal{P}_K$ of points in $\mathbb{R}^n$ with $|\mathcal{P}_1| = \cdots = |\mathcal{P}_K| = m$, 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 $i = 1, \ldots, K$ and $l = 1, \ldots, m$. The goal is to find K $n \times n$ orthogonal matrices $X_1, \ldots, X_K$ such that the quantity:

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

is minimized. Here, A_{il} is the l–th column of the matrix A_i, where $i = 1, \ldots, K$ and $l = 1, \ldots, m$. 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:

$$\begin{array}{lll} \text{maximize} & \displaystyle\sum_{1 \le i < j \le K} \operatorname{tr}\left(A_i^T X_i^T X_j A_j\right) & \\ \text{subject to} & X_i^T X_i = I & \text{for } i = 1, \ldots, K \end{array} \tag{1.1}$$

Now, notice that the objective function is linear in each of the X_i's. Thus, we may relax the orthogonality constraint $X_i^T X_i = I$ to the norm constraint $\|X_i\|_\infty \le 1$ for $i = 1, \ldots, K$ without affecting the optimal value of the problem (we refer the reader to [17] for details). In other words, Problem (1.1) has the same optimal value as the following problem:

$$\begin{array}{lll} \text{maximize} & \displaystyle\sum_{1 \le i < j \le K} \operatorname{tr}\left(A_i^T X_i^T X_j A_j\right) & \\ \text{subject to} & \|X_i\|_\infty \le 1 & \text{for } i = 1, \ldots, K \end{array}$$

which, after some elementary manipulations, 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 $\mathcal{N} = \{1, \ldots, n\}$, two $n \times n$ symmetric matrices A and B, and an $n \times n$ matrix 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, e.g., [14, 30]):

(1.2)

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

The constraints in Problem (1.2) force the matrix X to be a permutation matrix. Indeed, it is well–known that X satisfies the constraints in (1.2) iff X is a permutation matrix. The QAP is a classical problem in combinatorial optimization and has found many applications (see, e.g., [22]). 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 (1.2). In other words, consider the following problem:

$$\begin{array}{ll} \text{maximize} & \operatorname{tr}\left(AXBX^T - 2CX^T\right) \\ \text{subject to} & XX^T = I \end{array} \tag{1.3}$$

Suppose now that we have $A, B \succ \mathbf{0}$. Let $A^{1/2}$ and $B^{1/2}$ are $n \times n$ symmetric positive definite matrices such that $A = A^{1/2}A^{1/2}$ and $B = B^{1/2}B^{1/2}$. Then, the objective function in Problem (1.3) can be written as $\operatorname{tr}\left((A^{1/2}XB^{1/2})(A^{1/2}XB^{1/2})^T - 2CX^T\right)$, which is a convex quadratic form in X. Consequently, we may relax the orthogonality constraint $XX^T = I$ to the norm constraint $\|X\|_\infty \le 1$ without affecting the optimal value of the problem (again, we refer the reader to [17] for details). In particular, Problem (1.3) is equivalent to the following problem:

$$\begin{array}{ll} \text{maximize} & \operatorname{tr}\left(AXBX^T - 2CX^T\right) \\ \text{subject to} & \|X\|_\infty \le 1 \end{array}$$

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

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, 26]. 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 $O\left(\max\left\{(m+n)^{1/3}, \ln L\right\}\right)$ 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 $S = \{v_1, \ldots, v_{mn}\}$ of vectors and apply a suitable orthogonal transformation to S to obtain vectors $v_1', \ldots, v_{mn}'$
2. generate a random vector $\xi = (\xi_1, \ldots, \xi_{mn})$, 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, 23], 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 $O\left(\max\left\{(m+n)^{1/3}, \ln L\right\}\right)$ bound established in [17] and provides the first logarithmic approximation guarantee for (QP–OC). Furthermore, our result allows us to settle in the affirmative a conjecture of Nemirovski concerning the norm estimation problem, which in turn allows us to simplify and improve a result of Ben–Tal and Nemirovski [4] concerning the design of so–called *safe tractable approximations* of 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 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 non–commutative version of Khintchine's inequality, which allows us to prove a concentration inequality for the spectral norm of a sum of certain random matrices. As a further illustration of the power of the aforementioned concentration inequality, we show how it can be used to settle a conjecture of Nemirovski and improve a result of Ben–Tal and Nemirovski [4] concerning certain chance constrained optimization problems in Section 4. Finally, we make some concluding remarks 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 $i = 1, \ldots, m$ and $j = 1, \ldots, n$, and whose entries are determined by:

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

where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. 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 $i = 1, \ldots, L$. For the mapping $\mathcal{C}$, we identify it with an $u \times mn$ matrix C whose entries are determined by:

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

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

$$X \bullet \mathcal{A}X = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}(\mathcal{A}X)_{ij} = A \bullet \mathrm{Gram}(X)$$

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

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

where C_i is the $m \times n$ matrix $[C_{i,(k,l)}]_{1\le k\le m,\, 1\le l\le n}$. Finally, observe that $\|X\|_\infty \le 1$ iff $XX^T \preceq I$. Now, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, 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 $\mathrm{Gram}(X)$, which in turn implies the existence of a linear mapping $\mathcal{S} : \mathscr{S}^{mn} \to \mathscr{S}^m$ such that $XX^T \preceq I$ iff $\mathcal{S}\,\mathrm{Gram}(X) \preceq I$ (here, $\mathscr{S}^m$ is the space of $m \times m$ symmetric real matrices). In a similar fashion, we have $\|X\|_\infty \le 1$ iff $X^T X \preceq I$, and there exists a linear mapping $\mathcal{T} : \mathscr{S}^{mn} \to \mathscr{S}^n$ such that $X^T X \preceq I$ iff $\mathcal{T}\,\mathrm{Gram}(X) \preceq I$. Note that both the linear mappings $\mathcal{S}$ and $\mathcal{T}$ can be specified explicitly as matrices of appropriate dimensions in polynomial time. Now, by putting the pieces together and using the fact that $\mathrm{Gram}(X) \succeq \mathbf{0}$, we obtain the following SDP relaxation of (QP–OC):

(QP–OC–SDR)

$$\begin{array}{lll}
\text{maximize} & A \bullet Y & \\
\text{subject to} & B \bullet Y \le 1 & \\
& B_i \bullet Y \le 1 & \text{for } i = 1, \ldots, L \\
& \mathrm{Gram}(C_i) \bullet Y = 0 & \text{for } i = 1, \ldots, u \\
& \mathcal{S}Y \preceq I,\ \mathcal{T}Y \preceq I & \\
& Y \in \mathscr{S}^{mn},\ Y \succeq \mathbf{0} &
\end{array}$$

Although the constraints $\mathcal{S}\,\mathrm{Gram}(X) \preceq I$ and $\mathcal{T}\,\mathrm{Gram}(X) \preceq I$ 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 $\theta' \equiv A \bullet Y' \ge \theta^* - \epsilon$.

3 Analysis of the SDP Relaxation

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

THEOREM 3.1. *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 $\overline{X}$ such that:*

*(a) $\overline{X}$ is feasible for (*QP–OC*)*

(b) $\overline{X} \bullet \mathcal{A}\overline{X} \ge \Omega\left(1/\ln\left(\max\{m,n,L\}\right)\right) \cdot \theta'$

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 $\widehat{X}$. Since $Y' \succeq \mathbf{0}$, there exists a positive semidefinite matrix $Y'^{1/2} \in \mathscr{S}^{mn}$ such that $Y' = Y'^{1/2}Y'^{1/2}$. Moreover, the matrix $Y'^{1/2}AY'^{1/2}$ 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 $\xi = (\xi_{ij})_{1\le i\le m,\, 1\le j\le n}$, 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 $\mathrm{Vec}(\widehat{X}) = Y'^{1/2}U^T\xi$.

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 $\max\{m,n\} \ge 4$. The following proposition is established in [17]:

PROPOSITION 3.1. *The solution $\widehat{X}$ returned by the rounding procedure satisfies (a) $\widehat{X} \bullet \mathcal{A}\widehat{X} \equiv \theta'$; (b) $\mathbb{E}[\widehat{X} \bullet \mathcal{B}\widehat{X}] \le 1$; (c) $\mathbb{E}[\widehat{X} \bullet \mathcal{B}_i\widehat{X}] \le 1$ for $i = 1, \ldots, L$; (d) $\mathcal{C}\widehat{X} \equiv \mathbf{0}$; and (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 3.1, we need to analyze the behavior of $\widehat{X}$ with respect to the constraints (QP–OC(a)), (QP–OC(b)) and (QP–OC(d)). Towards that end, consider the events:

$$\begin{aligned}
\mathcal{E}_1 &= \left\{\widehat{X} \bullet \mathcal{B}\widehat{X} \le 1\right\} \\
\mathcal{E}_2 &= \left\{\widehat{X} \bullet \mathcal{B}_i\widehat{X} \le \Gamma^2 \text{ for all } i = 1, \ldots, L\right\} \\
\mathcal{E}_3 &= \left\{\|\widehat{X}\|_\infty \le \Gamma\right\}
\end{aligned}$$

where $\Gamma = O\left(\sqrt{\ln(\max\{m,n,L\})}\right)$. We would like to show that the event $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ occurs with constant probability. 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 3.2. *For any $\beta \geq 0$, we have:*

$$\Pr\left(\|\widehat{X}\|_\infty \geq \sqrt{2e(1+\beta)\ln(\max\{m,n\})}\right) \leq (\max\{m,n\})^{-\beta}$$

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 $\left(Y'^{1/2}U^T\right)_{(i,j)(k,l)}$, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Moreover, we have:

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

and

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

Thus, in order to prove Theorem 3.2, it suffices to prove the following:

THEOREM 3.2' *Let $\xi_1, \ldots, \xi_h$ be independent random variables that take on the values ± 1 with equal probability, and let $Q_1, \ldots, Q_h$ 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 $\beta \geq 0$, we have:*

$$\Pr\left(\|S\|_\infty \geq \sqrt{2e(1+\beta)\ln(\max\{m,n\})}\right) \leq (\max\{m,n\})^{-\beta}$$

The proof of Theorem 3.2' involves estimating the moments of the random variable $\|S\|_\infty$. Before we proceed, let us fix 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 3.1. (NON–COMMUTATIVE KHINTCHINE INEQUALITY) *Let $\xi_1, \ldots, \xi_h$ be independent random variables that take on the values ± 1 with equal probability, and let $Q_1, \ldots, Q_h$ be $m \times n$ matrices. Set:*

$$\begin{aligned} S &= \sum_{i=1}^h \xi_i Q_i \\ \Lambda_1 &= \left\| \left(\sum_{i=1}^h Q_i Q_i^T\right)^{1/2} \right\|_{S_p} \\ \Lambda_2 &= \left\| \left(\sum_{i=1}^h Q_i^T Q_i\right)^{1/2} \right\|_{S_p} \end{aligned}$$

Then, for $2 \leq p < \infty$, there exists a constant $\gamma_p > 0$ such that:

$$\left(\mathbb{E}\left[\|S\|_{S_p}^p\right]\right)^{1/p} \leq \gamma_p \cdot \max\{\Lambda_1, \Lambda_2\}$$

We remark that Lust–Piquard did not provide an estimate for γ_p. However, it is shown in [23] that $\gamma_p \leq \alpha\sqrt{p}$ for some absolute constant $\alpha > 0$. Using a result of Buchholz [5], it can be shown that $\alpha \leq 2^{-1/4}(\pi/e)^{1/2} < 1$ (see [29]).

Fact 3.1 immediately leads to the following proposition:

PROPOSITION 3.2. *Let $2 \leq p < \infty$. Under the setting of Theorem 3.2', we have $(\mathbb{E}[\|S\|_\infty^p])^{1/p} \leq \sqrt{p} \cdot (\max\{m,n\})^{1/p}$.*

Proof. Under the setting of Theorem 3.2', 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:

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

and

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

Using Fact 3.1, we have:

$$(\mathbb{E}[\|S\|_\infty^p])^{1/p} \leq \left(\mathbb{E}\left[\|S\|_{S_p}^p\right]\right)^{1/p} \leq \sqrt{p} \cdot (\max\{m,n\})^{1/p}$$

as desired. □

We are now ready to prove Theorem 3.2':

Proof of Theorem 3.2'. By Markov's inequality and Proposition 3.2, for any $t > 0$ and $2 \leq p < \infty$, we have:

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

Upon setting $t = \sqrt{2e(1+\beta)\ln(\max\{m,n\})}$ and $p = t^2/e > 2$ (since $\max\{m,n\} \geq 4$ by assumption), we see that:

$$\Pr\left(\|S\|_\infty \geq \sqrt{2e(1+\beta)\ln(\max\{m,n\})}\right) \leq (\max\{m,n\})^{-\beta}$$

This completes the proof. □

Remarks. By the Central Limit Theorem, we see that Theorem 3.2' also holds in the case where $\xi_1,\ldots,\xi_h$ are i.i.d. standard Gaussian random variables.

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 3.3. *The following hold:*

(a) $\Pr(\widehat{X} \bullet \mathcal{B}\widehat{X} \leq 1) \geq 3/8$

(b) For any $\beta \geq 0$, *we have:*

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

for $i = 1,\ldots,L$.

Proof. Statement (a) follows from a result of Holzman and Kleitman [12], cf. [17, Lemma 4(a)]. To establish (b), we first compute:

$$\begin{aligned}\widehat{X} \bullet \mathcal{B}_i\widehat{X} &= \operatorname{tr}\left(B_i \operatorname{Vec}(\widehat{X}) \operatorname{Vec}(\widehat{X})^T\right) \\ &= \operatorname{tr}\left(UY'^{1/2}B_iY'^{1/2}U^T\xi\xi^T\right) \\ &= \xi^T B_i'\xi\end{aligned}$$

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

$$\begin{aligned}1 &\geq \mathbb{E}\left[\widehat{X} \bullet \mathcal{B}_i\widehat{X}\right] \\ &= \sum_{k=1}^m\sum_{l=1}^n (B_i')_{(k,l)(k,l)} \\ &= \sum_{k=1}^m\sum_{l=1}^n \left\|(B_i')^{1/2}_{(\cdot,\cdot)(k,l)}\right\|_2^2 \qquad (3.4)\end{aligned}$$

where $(B_i')^{1/2}_{(\cdot,\cdot)(k,l)}$ is the (k,l)–th column of $(B_i')^{1/2}$. Hence, it suffices to prove the following:

$$\Pr\left(\left\|\sum_{k=1}^m\sum_{l=1}^n \xi_{kl}(B_i')^{1/2}_{(\cdot,\cdot)(k,l)}\right\|_2 \geq \Gamma\right) \leq (\max\{m,n,L\})^{-(1+\beta)} \qquad (3.5)$$

where $\Gamma = \sqrt{2e(1+\beta)\ln(\max\{m,n,L\})}$. Towards that end, we first recall the following result of Tomczak–Jaegermann [28]:

FACT 3.2. *Let* $\xi_1,\ldots,\xi_h$ *be independent random variables that take on the values* ±1 *with equal probability, and let* $Q_1,\ldots,Q_h$ *be* $m \times n$ *matrices. Then, for* $2 \leq p < \infty$, *we have:*

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

Now, note that for a vector v, we have $\|v\|_{S_p} = \|v\|_2$ for $1 \leq p < \infty$. Hence, by (3.4) and Fact 3.2, we have:

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

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

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

Upon setting $t = e\sqrt{(1+\beta)}\cdot\sqrt{\ln(\max\{m,n,L\})}$ and $p = (t/e)^2 \geq 2$, we obtain (3.5) and complete the proof. □

We are now ready to finish the proof of Theorem 3.1:

Proof of Theorem 3.1. Let $\beta = 1$ in Theorem 3.2' and Proposition 3.3(b). Since $\max\{m,n\} \geq 4$ by assumption, we see from Theorem 3.2' and Propositions 3.1(a,d) and 3.3 that with probability at least $3/8 - (1/4 + 1/16) = 1/16$, the solution $\widehat{X}$ 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) $\widehat{X} \bullet \mathcal{B}_i\widehat{X} \leq 4e\ln(\max\{m,n,L\})$ for $i = 1,\ldots,L$;

(d) $\mathcal{C}\widehat{X} \equiv \mathbf{0}$; and

(e) $\|\widehat{X}\|_\infty \le \sqrt{4e\ln(\max\{m,n,L\})}$.

It follows that the matrix:

$$\overline{X} = \widehat{X}/\sqrt{4e\ln(\max\{m,n,L\})}$$

has the required properties. □

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

One of Nemirovski's original motivations for studying concentration inequalities of the form given in Theorem 3.2' is to develop a so–called safe tractable approximation of the following chance constrained optimization problem:

(4.6)

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & F(x) \le \mathbf{0} \\ & \Pr\left(\mathcal{A}_0(x) - \sum_{i=1}^h \xi_i \mathcal{A}_i(x) \succeq \mathbf{0} \right) \ge 1-\epsilon \ (\dagger) \\ & x \in \mathbb{R}^n \end{array}$$

Here, $c \in \mathbb{R}^n$ is a given objective vector; $F : \mathbb{R}^n \to \mathbb{R}^m$ is an efficiently computable vector–valued function with convex components; $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_h : \mathbb{R}^n \to \mathscr{S}^m$ are affine functions in x with $\mathcal{A}_0(x) \succ \mathbf{0}$ for all $x \in \mathbb{R}^n$; $\xi_1, \ldots, \xi_h$ are independent (but not necessarily identical) mean zero random variables; and $\epsilon \in (0,1)$ is the error tolerance parameter. The chance constrained problem (4.6) arises from many engineering applications, such as truss topology design and problems in control theory, and has received much attention lately (see, e.g., [19, 20, 17, 4]). In general, the constraint (†) in (4.6) is computationally intractable. In an attempt to circumvent this problem, Ben–Tal and Nemirovski [17, 4] proposed a *safe tractable approximation* of (†) — that is, a system of constraints $\mathcal{H}$ such that (i) x is feasible for (†) whenever it is feasible for $\mathcal{H}$, and (ii) the constraints in $\mathcal{H}$ are efficiently computable. Specifically, their strategy is as follows. First, observe that:

$$\begin{aligned} & \Pr\left(\mathcal{A}_0(x) - \sum_{i=1}^h \xi_i \mathcal{A}_i(x) \succeq \mathbf{0} \right) \\ = \ & \Pr\left(\sum_{i=1}^h \xi_i \mathcal{A}_i'(x) \preceq I \right) \end{aligned}$$

where $\mathcal{A}_i'(x) = \mathcal{A}_0^{-1/2}(x)\mathcal{A}_i(x)\mathcal{A}_0^{-1/2}(x)$. Now, suppose that one can choose $\gamma = \gamma(\epsilon) > 0$ such that whenever

$$\sum_{i=1}^h \left(\mathcal{A}_i'(x)\right)^2 \preceq \gamma^2 I \tag{4.7}$$

holds, the constraint (†) is satisfied. Then, (4.7) will be a sufficient condition for (†) to hold. The upshot of (4.7) is that it can be expressed as a linear matrix inequality using the Schur complement:

$$\begin{bmatrix} \gamma\mathcal{A}_0(x) & \mathcal{A}_1(x) & \cdots & \mathcal{A}_h(x) \\ \mathcal{A}_1(x) & \gamma\mathcal{A}_0(x) & & \\ \vdots & & \ddots & \\ \mathcal{A}_h(x) & & & \gamma\mathcal{A}_0(x) \end{bmatrix} \succeq \mathbf{0} \tag{4.8}$$

Thus, by replacing (†) with (4.8), Problem (4.6) becomes tractable. Moreover, any solution $x \in \mathbb{R}^n$ that satisfies $F(x) \le \mathbf{0}$ and (4.8) will be feasible for the original chance constrained problem (4.6).

Now, it is not hard to see that if each of the ξ_i takes on the values ± 1 with equal probability or is a standard Gaussian, then Theorem 3.2' implies that for any $\epsilon \in (0, 1/m]$, we have:

$$\Pr\left(\left\| \sum_{i=1}^h \xi_i \mathcal{A}_i'(x) \right\|_\infty \le 1 \right) \ge 1-\epsilon$$

provided that $\gamma \ge \gamma(\epsilon) \equiv \left(\sqrt{4e\ln(1/\epsilon)}\right)^{-1}$. In other words, for any $\epsilon \in (0,1/m]$, the positive semidefinite constraint (4.8) with $\gamma \ge \gamma(\epsilon)$ is a safe tractable approximation of (†). Indeed, upon setting $\beta = \ln(1/\epsilon)/\ln m > 0$ in Theorem 3.2' and noting that $(1+\beta) \le 2\beta$ whenever $\epsilon \le 1/m$, we have:

$$\begin{aligned} & \Pr\left(\left\| \sum_{i=1}^h \xi_i \mathcal{A}_i'(x) \right\|_\infty \le 1 \right) \\ \ge \ & \Pr\left(\left\| \sum_{i=1}^h \xi_i \mathcal{A}_i'(x) \right\|_\infty \le \sqrt{\frac{2e(1+\beta)\ln m}{4e\ln(1/\epsilon)}} \right) \\ = \ & \Pr\left(\left\| \sum_{i=1}^h \xi_i \left(\frac{1}{\gamma}\mathcal{A}_i'(x) \right) \right\|_\infty \le \sqrt{2e(1+\beta)\ln m} \right) \\ \ge \ & 1 - \frac{1}{m^\beta} = 1-\epsilon \end{aligned}$$

as claimed.

Our result in the preceding paragraph applies only to the case where ξ_i takes on the values ± 1 with equal probability or is a standard Gaussian. Now, following [4, 17], let us consider the case where $\xi_1, \ldots, \xi_h$ are independent mean zero random variables that are supported on $[-1, 1]$. To handle this case, we need the following extension of Theorem 3.2':

THEOREM 4.1. *Let $\xi_1, \ldots, \xi_h$ be independent mean zero random variables that are supported on $[-1, 1]$, and let $Q_1, \ldots, Q_h$ 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 $\beta \geq 0$, we have:*

$$\begin{aligned} &\Pr\left(\|S\|_\infty \geq \sqrt{8e(1+\beta)\ln\left(\max\{m,n\}\right)}\right) \\ \leq\ & \left(\max\{m,n\}\right)^{-\beta} \end{aligned}$$

Proof. Let $\epsilon_1, \ldots, \epsilon_h$ be independent random variables that are independent of the ξ_i's and take on the values ± 1 with equal probability. Set:

$$\begin{aligned} \Lambda_1 &= \left\| \left(\sum_{i=1}^h \xi_i^2 Q_i Q_i^T \right)^{1/2} \right\|_{S_p}^p \\ \Lambda_2 &= \left\| \left(\sum_{i=1}^h \xi_i^2 Q_i^T Q_i \right)^{1/2} \right\|_{S_p}^p \end{aligned}$$

Now, a standard symmetrization argument (see, e.g., [6, Lemma 1.2.6]) together with Fubini's theorem and Fact 3.1 imply that:

$$\begin{aligned} \mathbb{E}\left[\left\| \sum_{i=1}^h \xi_i Q_i \right\|_{S_p}^p \right] &\leq 2^p \cdot \mathbb{E}_\xi \mathbb{E}_\epsilon \left[\left\| \sum_{i=1}^h \epsilon_i \xi_i Q_i \right\|_{S_p}^p \right] \\ &\leq 2^p \cdot p^{p/2} \cdot \mathbb{E}_\xi \left[\max\left\{\Lambda_1, \Lambda_2\right\} \right] \\ &\leq 2^p \cdot p^{p/2} \cdot \max\{m,n\} \end{aligned}$$

Here, $\mathbb{E}_\xi$ (resp. $\mathbb{E}_\epsilon$) denotes the mathematical expectation with respect to the random variables $\xi_1, \ldots, \xi_h$ (resp. $\epsilon_1, \ldots, \epsilon_h$). The desired result then follows from an application of Markov's inequality. □

Theorem 4.1 settles in the affirmative a conjecture of Nemirovski ([17, Conjecture 1]) regarding the typical spectral norm of a sum of certain random matrices. Furthermore, it leads to the following result concerning the chance constrained problem (4.6):

PROPOSITION 4.1. *Let $\xi_1, \ldots, \xi_h$ be as in Theorem 4.1, and consider the chance constrained problem (4.6). Then, for any $\epsilon \in (0, 1/m]$, the positive semidefinite constraint (4.8) with $\gamma \geq \gamma(\epsilon) \equiv \left(\sqrt{16e\ln(1/\epsilon)}\right)^{-1}$ is a safe tractable approximation of (†).*

We remark that Ben–Tal and Nemirovski established a similar result in [17, 4], but only under the additional assumptions that the random variables $\xi_1, \ldots, \xi_h$ have zero third moments, and that $\epsilon \leq O(\exp(-m^{1/3}))$. Thus, Proposition 4.1 yields an improvement over Ben–Tal and Nemirovski's result.

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. *Journal of the American Mathematical Society*, 21(1):1–21, 2008.

[4] A. Ben-Tal and A. Nemirovski. On Safe Tractable Approximations of Chance Constrained Linear Matrix Inequalities. Manuscript, 2007.

[5] A. Buchholz. Operator Khintchine Inequality in Non–Commutative Probability. *Mathematische Annalen*, 319:1–16, 2001.

[6] V. H. de la Peña and E. Giné. *Decoupling: From Dependence to Independence.* Probability and Its Applications. Springer–Verlag, 1999.

[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. Dijksterhuis. *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. *SIAM Journal on Optimization*, 19(2):503–523, 2008.

[12] R. Holzman and D. J. Kleitman. On the Product of Sign Vectors and Unit Vectors. *Combinatorica*, 12(3):303–316, 1992.

[13] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[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] A. Nemirovski and A. Shapiro. Convex Approximations of Chance Constrained Programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[20] A. Nemirovski and A. Shapiro. Scenario Approximations of Chance Constraints. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 3–47. Springer–Verlag, London, 2006.

[21] 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.

[22] 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.

[23] G. Pisier. Non–Commutative Vector Valued L_p–Spaces and Completely p–Summing Maps. *Astérisque*, 247, 1998.

[24] A. M.-C. So. On the Performance of Semidefinite Relaxation MIMO Detectors for QAM Constellations. Manuscript, 2008.

[25] A. M.-C. So. Probabilistic Analysis of the Semidefinite Relaxation Detector in Digital Communications. Manuscript, 2008.

[26] A. M.-C. So, Y. Ye, and J. Zhang. A Unified Theorem on SDP Rank Reduction. To appear in *Mathematics of Operations Research*, 2008.

[27] A. M.-C. So, J. Zhang, and Y. Ye. On Approximating Complex Quadratic Optimization Problems via Semidefinite Programming Relaxations. *Mathematical Programming, Series B*, 110(1):93–110, 2007.

[28] 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.

[29] J. A. Tropp. The Random Paving Property for Uniformly Bounded Matrices. *Studia Mathematica*, 185(1):67–82, 2008.

[30] 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.