Semidefinite Optimization Applications

Anthony Man–Cho So[*]

1 Introduction

Recently, semidefinite programming (SDP) has received a lot of attention in many communities. Such a popularity can partly be attributed to the wide applicability of SDP, as well as recent advances in the design of provably efficient interior–point algorithms and fast heuristics. In this article, we will survey some of the recent applications of SDP and provide further pointers to the literature for the interested readers. Given the versatile modeling power of semidefinite programs and the multitude of work that applies the SDP methodology, we should emphasize that the applications discussed in this article are by no means exhaustive, and they necessarily reflect a certain degree of personal bias.

Before we begin our discussion on SDP applications, let us fix some notation. Let $\mathbb{S}^n$ be the set of $n \times n$ real, symmetric matrices, and let $\mathbb{S}^n_+ \subset \mathbb{S}^n$ be the set of $n \times n$ real, symmetric, positive semidefinite matrices, i.e., if $A \in \mathbb{S}^n_+$, then for any $x \in \mathbb{R}^n$, we have $x^T A x \ge 0$. We shall also write $A \succeq \mathbf{0}$ to denote the fact that $A \in \mathbb{S}^n_+$. For any $A, B \in \mathbb{S}^n$, we define the *trace inner product* $A \bullet B$ via

$$A \bullet B = \mathrm{tr}(A^T B) = \sum_{i,j=1}^{n} A_{ij} B_{ij}.$$

Let $C, A_1, \ldots, A_m \in \mathbb{S}^n$ and $b \in \mathbb{R}^m$. The optimization problem

$$\begin{array}{rll} & \inf & C \bullet X \\ (\mathrm{P}): & \text{subject to} & A_i \bullet X = b_i \quad \text{for } 1 \le i \le m, \\ & & X \succeq \mathbf{0} \end{array}$$

is called a *semidefinite program (SDP) in primal standard form*. Its dual problem, which is given by

$$\begin{array}{rll} & \sup & b^T y \\ (\mathrm{D}): & \text{subject to} & \displaystyle\sum_{i=1}^{m} y_i A_i + Z = C, \\ & & y \in \mathbb{R}^m,\ Z \succeq \mathbf{0}, \end{array}$$

is also an SDP and is called a *semidefinite program (SDP) in dual standard form*. Theoretically, it has been shown that both (P) and (D) can be efficiently solved (up to any desired degree

[*]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`

of accuracy) by the ellipsoid method or interior–point algorithms (see, e.g., [25, 10]). Many techniques have also been recently developed to speed up the solution time of SDPs (see, e.g., [48, 19, 65]).

In the sequel, we shall see how (P) and (D) can be used to model a wide range of problems that arise in practice.

2 Semidefinite Relaxations of Hard Quadratic Optimization Problems

2.1 The Basic Technique

A frequently used and powerful idea for dealing with hard optimization problems is to develop tractable convex relaxations of them. As it turns out, many difficult quadratic optimization problems can be relaxed to SDPs. To illustrate some of the basic ideas, let us consider the following class of homogeneous *quadratically constrained quadratic programs (QCQPs)*:

$$\text{(QCQP)}: \quad \begin{array}{lll} \text{minimize} & x^T A x & \\ \text{subject to} & x^T A_i x \trianglerighteq_i b_i & \text{for } i = 1, \ldots, m, \\ & x \in \mathbb{R}^n. & \end{array}$$

Here, we assume that $A, A_1, \ldots, A_m \in \mathbb{S}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$, and '$\trianglerighteq_i$' can represent either '$\geq$', '$=$', or '$\leq$', where $i = 1, \ldots, m$. Problem (QCQP) is known to be NP–hard, as it captures various NP–hard problems as special cases (see, e.g., [22, 47, 44]). To derive an SDP relaxation of (QCQP), we first observe that

$$x^T A x = A \bullet x x^T \quad \text{and} \quad x^T A_i x = A_i \bullet x x^T \quad \text{for } i = 1, \ldots, m.$$

In particular, both the objective function and constraints in (QCQP) are *linear* in the matrix xx^T. Thus, by introducing a new variable $X = xx^T$ that denotes a rank one symmetric positive semidefinite matrix, we obtain the following *equivalent* formulation of (QCQP):

$$\text{(QCQP–RC)}: \quad \begin{array}{lll} \text{minimize} & A \bullet X & \\ \text{subject to} & A_i \bullet X \trianglerighteq_i b_i & \text{for } i = 1, \ldots, m, \\ & X \succeq \mathbf{0}, \ \ \text{rank}(X) = 1. & \end{array}$$

The upshot of the above equivalent formulation is that it allows us to identify the fundamental difficulty in solving (QCQP). Indeed, in (QCQP–RC), the objective function and all constraints except the $\text{rank}(X) = 1$ one are convex in X. Thus, we may consider dropping the trouble–causing rank constraint to obtain the following so–called *semidefinite relaxation* of (QCQP):

$$\text{(QCQP–SDR)}: \quad \begin{array}{lll} \text{minimize} & A \bullet X & \\ \text{subject to} & A_i \bullet X \trianglerighteq_i b_i & \text{for } i = 1, \ldots, m, \\ & X \succeq \mathbf{0}. & \end{array}$$

It is not hard to see that (QCQP–SDR) is an instance of SDP, and hence can be solved efficiently. However, a fundamental issue that one must address is how to convert a globally optimal solution

X^* to (QCQP–SDR) into a feasible solution $\tilde{x}$ to (QCQP). Note that if X^* is of rank one, then there is nothing to do, for we can write $X^* = x^*(x^*)^T$, and x^* will be a feasible (and in fact optimal) solution to (QCQP). On the other hand, if the rank of X^* is larger than 1, then we must somehow extract from it, in an efficient manner, a vector $\tilde{x}$ that is feasible for (QCQP). There are many ways to do this. However, we must emphasize that even though the extracted solution is feasible for (QCQP), it is in general not an optimal solution (for otherwise we would have solved an NP–hard problem in polynomial time).

We note that the above procedure has been used extensively, e.g., in the signal processing community to tackle various communications problems. We refer the interested readers to [50, 40] for further details and references.

2.2 Quality of Semidefinite Relaxations

Given the SDP relaxations of various QCQP problems, it is natural to ask how good those relaxations are. In their seminal work [22], Goemans and Williamson analyzed an SDP relaxation for the Maximum Cut problem and provided the first provable result concerning the quality of SDP relaxations. Furthermore, it sparked off a great deal of very fruitful research on the quality of various SDP relaxations. Now, let us give an overview of Goemans and Williamson's approach. Such an approach has also been used to establish some of the latter results on the quality of SDP relaxations.

One of the motivations for Goemans and Williamson's work is the Maximum Cut Problem, which is a well–known, NP–hard combinatorial optimization problem and is defined as follows. Suppose that we are given a simple undirected graph $G = (V, E)$ with $V = \{1, \ldots, n\}$, and a function $w : E \to \mathbb{R}_+$ that assigns to each edge $e \in E$ a non–negative weight w_e. The Maximum Cut Problem (Max–Cut) is that of finding a set $S \subset V$ of vertices such that the total weight of the edges in the cut $(S, V \backslash S)$, i.e., sum of the weights of the edges with one endpoint in S and the other in $V \backslash S$, is maximized. By setting $w_{ij} = 0$ if $(i, j) \notin E$, we may denote the weight of a cut $(S, V \backslash S)$ by

$$w(S, V\backslash S) = \sum_{i \in S, j \in V \backslash S} w_{ij}, \tag{1}$$

and our goal is to choose a set $S \subset V$ such that the quantity in (1) is maximized. The Max–Cut problem can be formulated as the following QCQP:

$$\begin{array}{rrll} v^* & = & \text{maximize} & \displaystyle\frac{1}{2} \sum_{(i,j) \in E} w_{ij}(1 - x_i x_j) \\ & & \text{subject to} & x_i \in \{-1, 1\} \qquad\qquad \text{for } i = 1, \ldots, n, \end{array} \tag{2}$$

where the variable x_i indicates which side of the cut vertex i belongs to. Specifically, the cut $(S, V \backslash S)$ is given by $S = \{i \in \{1, \ldots, n\} : x_i = 1\}$. Note that if vertices i and j belong to the same side of a cut, then $x_i = x_j$, whence its contribution to the objective function in (2) is zero. Otherwise, we have $x_i \neq x_j$, whence its contribution to the objective function is $w_{ij}(1 - (-1))/2 = w_{ij}$. Upon applying the techniques described in Section 2.1, we obtain the following SDP relaxation of (2):

$$\begin{array}{rcll} v^*_{sdp} & = & \text{maximize} & \displaystyle\frac{1}{2}\sum_{(i,j)\in E} w_{ij}(1-X_{ij}) \\ & & \text{subject to} & X_{ii} = 1 \qquad\qquad \text{for } i = 1,\ldots,n, \\ & & & X \succeq \mathbf{0}. \end{array} \tag{3}$$

Note that since (3) is a relaxation of (2), we have $v^*_{sdp} \ge v^*$.

Now, let X^* be an optimal solution to (3). In general, the matrix X^* need not be of the form xx^T, and hence it does not immediately yield a feasible solution to (2). However, we can extract from X^* a solution $x' \in \{-1,1\}^n$ to (2) via the following *randomized rounding procedure*:

1. Compute the *Cholesky factorization* $X^* = U^T U$ of X^*, where $U \in \mathbb{R}^{n\times n}$. Let $u_i \in \mathbb{R}^n$ be the i–th column of U. Note that $\|u_i\|_2^2 = 1$ for $i = 1,\ldots,n$.
2. Let $r \in \mathbb{R}^n$ be a vector uniformly distributed on the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.
3. Set $x'_i = \text{sgn}\left(u_i^T r\right)$ for $i = 1,\ldots,n$, where

$$\text{sgn}(z) = \begin{cases} 1 & \text{if } z \ge 0, \\ -1 & \text{otherwise.} \end{cases}$$

In other words, we choose a random hyperplane through the origin (with r as its normal) and partition the vertices according to whether their corresponding vectors lie "above" or "below" the hyperplane.

It is clear that $x' = (x'_1,\ldots,x'_n)$ is a feasible solution to (2), and hence we must have

$$v' \equiv \frac{1}{2}\sum_{(i,j)\in E} w_{ij}(1 - x'_i x'_j) \le v^*.$$

The question now is whether there exists an $\alpha \in (0,1)$ such that $v' \ge \alpha \cdot v^*$. The factor α is commonly known as the *approximation ratio*. More precisely, since the solution $x' \in \{-1,1\}^n$ is produced via a randomized procedure, we are interested in its expected objective value, i.e.,

$$\begin{aligned} \mathbb{E}\left[v'\right] &= \frac{1}{2}\mathbb{E}\left[\sum_{(i,j)\in E} w_{ij}\left(1 - x'_i x'_j\right)\right] \\ &= \frac{1}{2}\sum_{(i,j)\in E} w_{ij}\mathbb{E}\left[1 - x'_i x'_j\right] \\ &= \sum_{(i,j)\in E} w_{ij}\Pr\left[\text{sgn}\left(u_i^T r\right) \ne \text{sgn}\left(u_j^T r\right)\right] \end{aligned} \tag{4}$$

The following theorem provides a lower bound on $\mathbb{E}\left[v'\right]$ and allows us to compare it with v^*_{sdp}:

Table 1: Quality Analyses of Various SDP Relaxations

Type of Problems	**References**
Combinatorial Optimization	[22, 29, 26, 16, 23, 1, 5, 4, 17, 18, 6, 30, 53, 7]
Quadratic Optimization	[47, 12, 44, 67, 49, 68, 8, 41, 43, 63, 27, 62, 3, 57, 2]
Communications	[31, 15, 58, 59, 69]
Geometry	[60, 64, 61, 70]

Theorem 1 *Let $u, v \in S^{n-1}$, and let r be a vector uniformly distributed on S^{n-1}. Then, we have*

$$\Pr\left[\mathrm{sgn}\left(u^T r\right) \neq \mathrm{sgn}\left(v^T r\right)\right] = \frac{1}{\pi} \arccos\left(u^T v\right).$$

Furthermore, for any $z \in [-1, 1]$, we have

$$\frac{1}{\pi} \arccos(z) \geq \alpha \cdot \frac{1}{2}(1 - z) > 0.878 \cdot \frac{1}{2}(1 - z),$$

where

$$\alpha = \min_{0 \leq \theta \leq \pi} \frac{2\theta}{\pi(1 - \cos\theta)}.$$

As a corollary to Theorem 1, we have the following

Corollary 1 *Given an instance (G, w) of the* Max–Cut *problem and an optimal solution to (3), the randomized rounding procedure above will produce a cut $(S', V \backslash S')$ whose expected objective value satisfies $w(S', V \backslash S') \geq 0.878 v^*$.*

Proof Let x' be the solution obtained from the randomized rounding procedure, and let S' be the corresponding cut. By (4) and Theorem 1, we have

$$\begin{aligned} \mathbb{E}\left[w(S', V \backslash S')\right] &= \frac{1}{\pi} \sum_{(i,j) \in E} w_{ij} \cdot \arccos\left(u_i^T u_j\right) \\ &\geq 0.878 \cdot \frac{1}{2} \sum_{(i,j) \in E} w_{ij} \left(1 - u_i^T u_j\right) \\ &= 0.878 v_{sdp}^* \\ &\geq 0.878 v^*, \end{aligned}$$

as desired. □

Note that a crucial step in the above analysis is to study a certain probability question related to the projections of a collection of vectors onto a random vector. This turns out to be a recurring theme in many of the quality analyses of SDP relaxations. Table 1 provides pointers to further theoretical results on the quality of SDP relaxations that arise in various applications.

3 Semidefinite Relaxations of Polynomial Optimization Problems

A fundamental problem in optimization is that of finding the global minimum of a real–valued polynomial $p : \mathbb{R}^n \to \mathbb{R}$ subject to polynomial inequality constraints. Indeed, the class of polynomial functions provides a very powerful modeling tool, and many optimization problems can be formulated as polynomial optimization problems. One example is the MAX–CUT problem in Section 2.2, where the constraint $x_i \in \{-1, 1\}$ can be equivalently formulated as $x_i^2 = 1$. Given its importance, polynomial optimization has been studied intensely in recent years. One approach, which is proposed by Lasserre [34] (see also the excellent survey by Laurent [36]), is to construct a *sequence* of successively tighter SDP relaxations of the given polynomial optimization problem, in such a way that the corresponding optimal values are monotone and converge to the optimal value of the original problem. To fix ideas and simplify notation, let us again focus on the case of *quadratic* optimization problems [33]. To begin, consider the problem

$$\text{(PO)}: \quad \begin{array}{rcll} v^* & = & \text{minimize} & g_0(x) \\ & & \text{subject to} & g_i(x) \geq 0 \qquad \text{for } i = 1, \ldots, m, \end{array}$$

where $g_0, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are quadratic polynomials. Note that we may write the polynomials $g_0, g_1, \ldots, g_m$ in *multi–index notation*, namely,

$$g_i(x) = \sum_{\alpha} (g_i)_\alpha x^\alpha \qquad \text{for } i = 0, 1, \ldots, m, \tag{5}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \geq \mathbf{0}$, $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, and $\sum_{i=1}^n \alpha_i \leq 2$. Hence, we may identify the polynomial g_i with its coefficient vector $((g_i)_\alpha) \in \mathbb{R}^{(n+1)(n+2)/2}$, where $i = 0, 1, \ldots, m$.

Now, observe that each summand in (5) is either a constant or of the form x_i or $x_i x_j$, where $1 \leq i \leq j \leq n$. For terms of the form $x_i x_j$, we may linearize them by introducing the variables y_{ij}. As a result, we have the following *linear programming (LP) relaxation* of (PO):

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{0 \leq i \leq j \leq n} (g_0)_{ij} y_{ij} & \\ \text{subject to} & \displaystyle\sum_{0 \leq i \leq j \leq n} (g_k)_{ij} y_{ij} \geq 0 & \text{for } k = 1, \ldots, m, \end{array}$$

with $y_{00} = 1$ and $y_{0i} = x_i$ for $i = 1, \ldots, n$. To tighten the relaxation, observe that ideally we should have $y_{ij} = y_{0i} y_{0j}$ for $1 \leq i \leq j \leq n$. Hence, we may impose the constraint

$$M_1(\mathbf{y}_1) \equiv \left[\begin{array}{c|cccccc} 1 & y_{01} & y_{02} & \cdots & y_{0i} & \cdots & y_{0n} \\ \hline y_{01} & y_{11} & y_{12} & \cdots & y_{1i} & \cdots & y_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{0j} & y_{1j} & y_{2j} & \cdots & y_{ij} & \cdots & y_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{0n} & y_{1n} & y_{2n} & \cdots & y_{in} & \cdots & y_{nn} \end{array} \right] \succeq \mathbf{0},$$

where $\mathbf{y}_1 = (1, y_{01}, \ldots, y_{0n}, y_{11}, \ldots, y_{nn})$. The first SDP relaxation in the sequence can then be given by

$$\text{(PO–SDR1)}: \quad \begin{array}{rcll} v_1^* & = & \inf & \displaystyle\sum_{0\le i\le j\le n} (g_0)_{ij} y_{ij} \\ & & \text{subject to} & \displaystyle\sum_{0\le i\le j\le n} (g_k)_{ij} y_{ij} \ge 0 \qquad \text{for } k = 1, \ldots, m, \\ & & & M_1(\mathbf{y}_1) \succeq \mathbf{0}. \end{array}$$

The reader may notice that (PO–SDR1) is simply the standard SDP relaxation of a quadratic optimization problem (cf. Section 2.1).

To obtain the next SDP relaxation in the sequence, consider a particular constraint

$$\sum_{0\le i\le j\le n} (g_k)_{ij} y_{ij} \ge 0,$$

where $k = 1, \ldots, m$. If we multiply the left–hand side by y_{pq} (where $0 \le p \le q \le n$) and linearize, we obtain

$$g_k y_{pq} = \sum_{0\le i\le j\le n} (g_k)_{ij} y_{ijpq}$$

(note that y_{pq} is an entry of $\mathbf{y}_1$). Now, it can be shown [33, 34] that the linear matrix inequality $M_1(g_k\mathbf{y}_1) = [g_k y_{pq}]_{p,q} \succeq \mathbf{0}$ is a valid constraint for the original problem (PO). To further tighten the relaxation, we impose constraints on the terms y_{ijpq} via

$$M_2(\mathbf{y}_2) \equiv \left[\begin{array}{c|cccccc} 1 & y_{01} & \cdots & y_{0n} & y_{11} & \cdots & y_{nn} \\ \hline y_{01} & y_{11} & \cdots & y_{1n} & y_{111} & \cdots & y_{1nn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{0n} & y_{1n} & \cdots & y_{nn} & y_{11n} & \cdots & y_{nnn} \\ y_{11} & y_{111} & \cdots & y_{11n} & y_{1111} & \cdots & y_{11nn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{nn} & y_{1nn} & \cdots & y_{nnn} & y_{11nn} & \cdots & y_{nnnn} \end{array}\right] \succeq \mathbf{0},$$

where $\mathbf{y}_2 = (1, y_{01}, \ldots, y_{nnnn})$. Then, the second SDP relaxation in the sequence is given by

$$\text{(PO–SDR2)}: \quad \begin{array}{rcll} v_2^* & = & \inf & \displaystyle\sum_{0\le i\le j\le n} (g_0)_{ij} y_{ij} \\ & & \text{subject to} & M_1(g_k\mathbf{y}_1) \succeq \mathbf{0} \qquad \text{for } k = 1, \ldots, m, \\ & & & M_2(\mathbf{y}_2) \succeq \mathbf{0}. \end{array}$$

In general, the l–th SDP relaxation in the sequence, which is given by

$$\text{(PO–SDR}l\text{)}: \quad \begin{array}{rcll} v_l^* & = & \inf & \displaystyle\sum_{0\le i\le j\le n} (g_0)_{ij} y_{ij} \\ & & \text{subject to} & M_{l-1}(g_k\mathbf{y}_{l-1}) \succeq \mathbf{0} \qquad \text{for } k = 1, \ldots, m, \\ & & & M_l(\mathbf{y}_l) \succeq \mathbf{0}, \end{array}$$

can be obtained by repeating the above procedure.

The following proposition states that the sequence of optimal values $\{v_l^*\}_l$ is monotone and they all provide lower bounds on the optimal value v^* of the original problem (PO).

Proposition 1 *For $l = 1, 2, \ldots$, we have $v_l^* \le v_{l+1}^* \le v^*$.*

To prove Proposition 1, it suffices to verify that for $l = 1, 2, \ldots$, the feasible region of (PO) is contained in that of (PO–SDR$(l+1)$), which in turn is contained in the feasible region of (PO–SDRl). We refer the reader to [33, 34] for details.

Although Proposition 1 implies that a better lower bound on v^* can be obtained by solving a higher order SDP relaxation (PO–SDRl) (where $l = 1, 2, \ldots$), it is not clear from Proposition 1 whether the sequence of optimal values $\{v_l^*\}_l$ converges to v^*. However, if the feasible region of (PO) is compact, then it can be shown that $v_l^* \nearrow v^*$. The proof of such result relies on representation theorems of polynomials that are positive on a compact set (see, e.g., [51] for a detailed treatment). We remark that the compactness assumption is not as restrictive as it may seem. It can be guaranteed, for instance, if we know *a priori* that there exists an optimal solution x^* to (PO) with $\|x^*\|_2 \le r$ for some $r > 0$, since then we can add the redundant polynomial inequality constraint $g_{m+1}(x) = r^2 - \|x\|_2^2 \ge 0$ to (PO) and obtain a compact feasible region. We summarize as follows:

Theorem 2 *Suppose that the feasible region of (PO) is compact. Then, we have $v_l^* \nearrow v^*$.*

There has been much research on the systematic generation of successively tighter SDP relaxations for various polynomial optimization problems. We refer the interested readers to [39, 37, 36, 35, 42, 52, 28] and the references therein for further details.

4 Safe Tractable Approximations of Chance Constrained Linear Matrix Inequalities

In many practical applications, the data defining an optimization problem may not be known exactly. This can happen, e.g., when we are unable to make precise measurements. Currently, there are several ways to model such data uncertainty. For instance, in the area of robust optimization, the uncertain data are assumed to lie in some bounded set S, and the goal is to find a solution that is feasible for *all* realizations of the uncertain data from the set S. On the other hand, in the area of chance constrained optimization, the uncertain data are assumed to follow certain probability distribution, and it is assumed that certain constraints can be violated *occasionally*. Thus, the goal here is to find a solution whose constraint violation probability is below a certain threshold. It should be noted that both approaches have their own advantages and disadvantages. In this section, we will focus on the chance constrained optimization approach and see how SDP can be used to tackle a large class of chance constrained optimization problems.

To begin, consider the problem

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

where $c \in \mathbb{R}^n$ is a given objective vector; $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_h : \mathbb{R}^n \to \mathscr{S}^m$ 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. Problem (CCP) arises from many engineering applications, such as truss topology design and problems in control theory and communications, and has received much attention lately (see, e.g., [45, 46, 43, 11, 56, 38]). Moreover, it captures chance constrained linear and conic quadratic programming as special cases. In general, the constraint (†) in (CCP) is computationally intractable. In an attempt to circumvent this problem, Ben–Tal and Nemirovski [43, 11] 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

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

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

$$\sum_{i=1}^{h} \left(\mathcal{A}_i'(x)\right)^2 \preceq \gamma^2 I \tag{6}$$

holds, the constraint (†) is satisfied. Then, (6) will be a sufficient condition for (†) to hold. The parameter γ can be viewed as the degree of conservatism of the approximation. In particular, a smaller γ results in a larger "gap" between the safe tractable constraint (6) and the original chance constraint (†). Now, the upshot of (6) is that it can be expressed as a linear matrix inequality using the Schur complement:

$$\begin{bmatrix} \gamma\mathcal{A}_0(x) & \mathcal{A}_1(x) & \cdots & \mathcal{A}_h(x) \\ \mathcal{A}_1(x) & \gamma\mathcal{A}_0(x) & & \\ \vdots & & \ddots & \\ \mathcal{A}_h(x) & & & \gamma\mathcal{A}_0(x) \end{bmatrix} \succeq \mathbf{0}. \tag{7}$$

Thus, by replacing (†) with (7), Problem (CCP) becomes tractable. Moreover, any solution $x \in \mathbb{R}^n$ that satisfies (7) will be feasible for the original chance constrained problem (CCP).

Using some deep results from the functional analysis literature, it can be shown that when the random variables $\xi_1, \ldots, \xi_h$ are "nice", there indeed exists a choice of $\gamma > 0$ such that (6) is a sufficient condition for (†) to hold. Specifically, we have the following

Theorem 3 *Let $\xi_1, \ldots, \xi_h$ be independent mean zero random variables, each of which is either (i) supported on $[-1,1]$, or (ii) normally distributed with unit variance. Consider the chance constrained problem (CCP). Then, for any $\epsilon \in (0,1/2]$, the positive semidefinite constraint (7) with $\gamma \le \gamma(\epsilon) \equiv \left(\sqrt{8e\ln(m/\epsilon)}\right)^{-1}$ is a safe tractable approximation of (†).*

The proof of Theorem 3 can be found in [57]; see also [11].

We refer those readers who are interested in optimization under data uncertainty to the books [55, 9] for further techniques and results.

5 Applications in Data Analysis

Recently, SDP has been used to tackle various problems in data analysis and machine learning. In this section, we will study one such problem, namely sparse principal component analysis, and show how it can be tackled using SDP.

Principal Component Analysis (PCA) (see, e.g, [54]) is a very important tool in data analysis. It provides a way to reduce the dimension of a given data set, thus revealing the sometimes hidden underlying structure of and facilitating further analysis on the data set. To motivate the problem of finding principal components, consider the following scenario. Suppose that we are interested in some attributes $X_1, \ldots, X_n$ of a population. In order to estimate the values of these attributes, one may sample from the population. Specifically, let X_{ij} be the value of the j–th attribute of the i–th individual, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. For $j, k = 1, \ldots, n$, define

$$\bar{X}_j = \frac{1}{m} \sum_{i=1}^{m} X_{ij}, \quad \sigma_{jk} = \frac{1}{m} \sum_{i=1}^{m} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

to be the sample mean of X_j and the sample covariance between X_j and X_k, respectively. Define $\Sigma = [\sigma_{jk}]_{j,k}$ to be the sample covariance matrix. The goal is then to find the *principal components* $u_1, \ldots, u_n$ such that the linear combination $\sum_{j=1}^{n} u_j X_j$ has *maximum* sample variance. In other words, we would like to solve the following problem:

$$\max_{\|u\|_2 \le 1} u^T \Sigma u, \tag{8}$$

where $u = (u_1, \ldots, u_n)$. The model given by (8) is based on the belief that principal components with a large associated variance indicate interesting features in the data. However, one of the shortcomings of (8) is that the linear combination may use a large number of the original variables, i.e., most of the factors $u_1, \ldots, u_n$ are *non–zero*, thus posing a difficulty in interpreting which variables are the most influential. This motivates us to consider the problem of finding *sparse* principal components, i.e., a set $\{u_1, \ldots, u_n\}$ of principal components that maximizes the variance while at the same time having only a few non–zero u_i's. There are many ways to formulate such a problem. For instance, one can take the following *penalty function* approach (see [20]):

$$v(\rho) = \max_{\|u\|_2 \le 1} \left\{ u^T \Sigma u - \rho \|u\|_0 \right\}, \tag{9}$$

where $\|u\|_0 = |\{i \in \{1, \ldots, n\} : u_i \neq 0\}|$ is the number of non–zero elements in the vector $u \in \mathbb{R}^n$, and $\rho \in \mathbb{R}$ is a parameter controlling the level of sparsity. Note that the objective function in (9) is non–convex and the associated optimization problem is hard to solve in general. However, as shown in [20], it can be relaxed to an SDP. Let us now sketch the main ideas in [20].

To begin, observe that $\Sigma \succeq \mathbf{0}$, and we may assume without loss that $\Sigma_{11} \ge \Sigma_{22} \ge \cdots \ge \Sigma_{nn} \ge 0$. Let $\Sigma^{1/2} \succeq \mathbf{0}$ be such that $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$. It is not hard to show that when $\rho > \Sigma_{11}$, the optimal solution to (9) is given by $u = \mathbf{0}$. Thus, we only need to consider the case where $\rho \le \Sigma_{11}$. Then, Problem (9) is equivalent to the problem

$$v(\rho) = \max_{\|u\|_2 = 1} \left\{ u^T \Sigma u - \rho \|u\|_0 \right\}. \tag{10}$$

For a given vector $u \in \mathbb{R}^n$, let $w(u) \in \{0,1\}^n$ be the indicator vector of the *sparsity pattern* of u, i.e., $w(u)_i = 1$ iff $u_i \neq 0$ for $i = 1, \ldots, n$. Observe that

$$u^T \Sigma u = \bar{u}^T \Big(\mathrm{diag}(w(u))\, \Sigma\, \mathrm{diag}(w(u))\Big) \bar{u}$$

for any $\bar{u} \in \mathbb{R}^n$ such that $\bar{u}_i = u_i$ whenever $u_i \neq 0$, where $i = 1, \ldots, n$. On the other hand, by the Courant–Fischer characterization, for a given sparsity pattern vector $w \in \{0,1\}^n$, we have

$$\max_{\|u\|_2=1} u^T \Big(\mathrm{diag}(w)\, \Sigma\, \mathrm{diag}(w)\Big) u = \lambda_{max}\Big(\mathrm{diag}(w)\, \Sigma\, \mathrm{diag}(w)\Big),$$

where $\lambda_{max}(M)$ is the largest eigenvalue of the matrix M. Thus, upon letting $\sigma_i \in \mathbb{R}^n$ to be the i–th column of $\Sigma^{1/2}$ (where $i = 1, \ldots, n$), we may write (10) as

$$\begin{aligned}
v(\rho) &= \max_{w \in \{0,1\}^n} \Big\{ \lambda_{max}\Big(\mathrm{diag}(w)\, \Sigma\, \mathrm{diag}(w)\Big) - \rho e^T w \Big\} \\
&= \max_{w \in \{0,1\}^n} \Big\{ \lambda_{max}\Big(\mathrm{diag}(w)\, \Sigma^{1/2}\Sigma^{1/2}\, \mathrm{diag}(w)\Big) - \rho e^T w \Big\} \\
&= \max_{w \in \{0,1\}^n} \Big\{ \lambda_{max}\Big(\Sigma^{1/2}\, \mathrm{diag}(w)\, \Sigma^{1/2}\Big) - \rho e^T w \Big\} \qquad (11) \\
&= \max_{\|u\|_2=1} \max_{w \in \{0,1\}^n} \Big\{ u^T \Big(\Sigma^{1/2}\, \mathrm{diag}(w)\, \Sigma^{1/2}\Big) u - \rho e^T w \Big\} \\
&= \max_{\|u\|_2=1} \max_{w \in \{0,1\}^n} \sum_{i=1}^n w_i \Big[\left(\sigma_i^T u\right)^2 - \rho \Big] \qquad (12) \\
&= \max_{\|u\|_2=1} \sum_{i=1}^n \Big[\left(\sigma_i^T u\right)^2 - \rho \Big]_+, \qquad (13)
\end{aligned}$$

where (11) follows from the fact that $\lambda_{max}\left(M^T M\right) = \lambda_{max}\left(M M^T\right)$ for any matrix M and $\mathrm{diag}(w)^2 = \mathrm{diag}(w)$ for any $w \in \{0,1\}^n$, and (12) follows from the fact that

$$u^T \left(\Sigma^{1/2}\, \mathrm{diag}(w)\, \Sigma^{1/2}\right) u = \left\| \mathrm{diag}(w)\, \Sigma^{1/2} u \right\|_2^2 = \sum_{i=1}^n w_i^2 \left(\sigma_i^T u\right)^2 = \sum_{i=1}^n w_i \left(\sigma_i^T u\right)^2.$$

Note that the objective function in (13) is still non–convex. However, the upshot of (13) is that the objective function is linear in $u_i u_j$, where $i, j = 1, \ldots, n$. In particular, let $U = uu^T \in \mathbb{R}^{n \times n}$. Then, it is easy to verify that $\left(\sigma_i^T u\right)^2 = \sigma_i^T U \sigma_i$ for $i = 1, \ldots, n$. Moreover, for a symmetric matrix U, we have $U = uu^T$ and $\|u\|_2 = 1$ iff $\mathrm{tr}(U) = 1$, $\mathrm{rank}(U) = 1$ and $U \succeq \mathbf{0}$. Thus, we see that (13) is equivalent to the following problem:

$$\begin{aligned}
v(\rho) = \ &\text{maximize} && \sum_{i=1}^n \left(\sigma_i^T U \sigma_i - \rho\right)_+ \\
&\text{subject to} && \mathrm{tr}(U) = 1, \\
& && U \succeq \mathbf{0},\ \mathrm{rank}(U) = 1.
\end{aligned} \qquad (14)$$

It is not hard to see that the function $U \mapsto \sum_{i=1}^{n} \left(\sigma_i^T U \sigma_i - \rho\right)_+$ is *convex*. Unfortunately, Problem (14) is still hard to solve in general, since we are maximizing a convex function, and the constraint $\text{rank}(U) = 1$ is not convex. However, not all is lost, as we may proceed as follows. First, let us introduce a notation. Given an $n \times n$ symmetric matrix M, let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Define

$$\text{tr}(M)_+ = \sum_{i=1}^{n} \max\{\lambda_i, 0\}.$$

Now, observe that if $U = uu^T$ and $\|u\|_2 = 1$, then $U^{1/2} = U = uu^T$. Moreover, for any $\alpha \in \mathbb{R}$, the matrix αuu^T has one eigenvalue equal to α and $n-1$ eigenvalues equal to 0, which implies that $\text{tr}\left(\alpha uu^T\right)_+ = \alpha_+$. It follows that

$$\begin{aligned} \left(\sigma_i^T U \sigma_i - \rho\right)_+ &= \text{tr}\Big(\left(\sigma_i^T uu^T \sigma_i - \rho\right) uu^T \Big)_+ \\ &= \text{tr}\Big(u \left(u^T \sigma_i \sigma_i^T u - \rho\right) u^T \Big)_+ \\ &= \text{tr}\left(U^{1/2} \sigma_i \sigma_i^T U^{1/2} - \rho U \right)_+ . \end{aligned}$$

The point of the above manipulation is that the function $U \mapsto \text{tr}\left(U^{1/2}\sigma_i\sigma_i^T U^{1/2} - \rho U\right)_+$ is *concave*. Specifically, one can prove the following

Theorem 4 *Let $M \in \mathbb{S}^n$ and $X \in \mathbb{S}_+^n$. Then, we have*

$$tr\left(X^{1/2} M X^{1/2}\right)_+ = \max_{X \succeq P \succeq \mathbf{0}} tr(PM) = \min_{Y \succeq M,\, Y \succeq \mathbf{0}} tr(YX) \tag{15}$$

In particular, the function $X \mapsto tr\left(X^{1/2} M X^{1/2}\right)_+$ is concave on $\mathbb{S}_+^n$.

Upon applying Theorem 4 and dropping the problematic rank constraint $\text{rank}(U) = 1$ in (14), we obtain the following *convex relaxation*:

$$\begin{aligned} v'(\rho) \quad = \quad \text{maximize} \quad & \sum_{i=1}^{n} \text{tr}\left(U^{1/2} \sigma_i \sigma_i^T U^{1/2} - \rho U \right)_+ \\ \text{subject to} \quad & \text{tr}(U) = 1, \\ & U \succeq \mathbf{0}. \end{aligned} \tag{16}$$

Upon letting $M_i(\rho) = \sigma_i \sigma_i^T - \rho I$ for $i = 1, \ldots, n$ and using (15), we see that (16) is equivalent to

$$\begin{aligned} v'(\rho) \quad = \quad \text{maximize} \quad & \sum_{i=1}^{n} \text{tr}(P_i M_i(\rho)) \\ \text{subject to} \quad & \text{tr}(U) = 1, \\ & U \succeq P_i \succeq \mathbf{0} \qquad \text{for } i = 1, \ldots, n, \\ & U \succeq \mathbf{0}, \end{aligned}$$

which is an SDP.

It turns out that SDP is a very powerful tool for addressing a host of data analysis and machine learning problems. We refer the interested readers to [32, 21, 14, 24, 13, 66] for further details.

Acknowledgments

The author would like to thank the reviewers for their detailed comments and suggestions. This research is supported by Hong Kong Research Grants Council (RGC) General Research Fund (GRF) Project No. CUHK 416908.

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 2005)*, pages 573–581, 2005.

[2] W. Ai, Y. Huang, and S. Zhang. New Results on Hermitian Matrix Rank–One Decomposition. To appear in *Mathematical Programming*, 2010.

[3] W. Ai and S. Zhang. Strong Duality for the CDT Subproblem: A Necessary and Sufficient Condition. *SIAM Journal on Optimization*, 19(4):1735–1756, 2009.

[4] N. Alon, K. Makarychev, Y. Makarychev, and A. Naor. Quadratic Forms on Graphs. *Inventiones Mathematicae*, 163(3):499–522, 2006.

[5] N. Alon and A. Naor. Approximating the Cut–Norm via Grothendieck's Inequality. *SIAM Journal on Computing*, 35(4):787–803, 2006.

[6] 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.

[7] S. Arora, S. Rao, and U. Vazirani. Expander Flows, Geometric Embeddings and Graph Partitioning. *Journal of the ACM*, 56(2):Article 5, 2009.

[8] A. Beck and Y. C. Eldar. Strong Duality in Nonconvex Quadratic Optimization with Two Quadratic Constraints. *SIAM Journal on Optimization*, 17(3):844–860, 2006.

[9] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009.

[10] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. MPS–SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2001.

[11] A. Ben-Tal and A. Nemirovski. On Safe Tractable Approximations of Chance–Constrained Linear Matrix Inequalities. *Mathematics of Operations Research*, 34(1):1–25, 2009.

[12] D. Bertsimas and Y. Ye. Semidefinite Relaxations, Multivariate Normal Distributions, and Order Statistics. In D.-Z. Du and P. M. Pardalos, editors, *Handbook of Combinatorial Optimization*, volume 3, pages 1–19. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[13] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? Technical Report 2009–13, Department of Statistics, Stanford University, Stanford, CA 94305, 2009.

[14] E. J. Candès and B. Recht. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[15] T.-H. Chang, Z.-Q. Luo, and C.-Y. Chi. Approximation Bounds for Semidefinite Relaxation of Max–Min–Fair Multicast Transmit Beamforming Problem. *IEEE Transactions on Signal Processing*, 56(8):3932–3943, 2008.

[16] M. Charikar. On Semidefinite Programming Relaxations for Graph Coloring and Vertex Cover. In *Proceedings of the 13th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2002)*, pages 616–620, 2002.

[17] M. Charikar, K. Makarychev, and Y. Makarychev. Near–Optimal Algorithms for Unique Games. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 2006)*, pages 205–214, 2006.

[18] E. Chlamtac, K. Makarychev, and Y. Makarychev. How to Play Unique Games Using Embeddings. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 687–696, 2006.

[19] A. d'Aspremont. Subsampling Algorithms for Semidefinite Programming. Preprint, 2009.

[20] A. d'Aspremont, F. Bach, and L. El Ghaoui. Optimal Solutions for Sparse Principal Component Analysis. *The Journal of Machine Learning Research*, 9:1269–1294, 2008.

[21] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. G. Lanckriet. A Direct Formulation for Sparse PCA Using Semidefinite Programming. *SIAM Review*, 49(3):434–448, 2007.

[22] 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.

[23] M. X. Goemans and D. P. Williamson. Approximation Algorithms for Max–3–Cut and Other Problems via Complex Semidefinite Programming. *Journal of Computer and System Sciences*, 68(2):442–470, 2004.

[24] D. Gross. Recovering Low–Rank Matrices from Few Coefficients in Any Basis. Preprint, 2009.

[25] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer–Verlag, Berlin Heidelberg, second corrected edition, 1993.

[26] E. Halperin and U. Zwick. A Unified Framework for Obtaining Improved Approximation Algorithms for Maximum Graph Bisection Problems. In K. Aardal and B. Gerards, editors, *Proceedings of the 8th Conference on Integer Programming and Combinatorial Optimization (IPCO VIII)*, volume 2081 of *Lecture Notes in Computer Science*, pages 210–225, 2001.

[27] 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.

[28] J. W. Helton and J. Nie. Semidefinite Representation of Convex Sets. *Mathematical Programming, Series A*, 122(1):21–64, 2010.

[29] D. Karger, R. Motwani, and M. Sudan. Approximate Graph Coloring by Semidefinite Programming. *Journal of the ACM*, 45(2):246–265, 1998.

[30] S. Khot and A. Naor. Linear Equations Modulo 2 and the L_1 Diameter of Convex Bodies. *SIAM Journal on Computing*, 38(4):1448–1463, 2008.

[31] M. Kisialiou 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, 2005 (ICASSP 2005)*, volume 3, pages III–433—III–436, 2005.

[32] G. R. G. Lanckriet, N. Cristianini, P. Barlett, L. El Ghaoui, and M. I. Jordan. Learning the Kernel Matrix with Semidefinite Programming. *The Journal of Machine Learning Research*, 5:27–72, 2004.

[33] J. B. Lasserre. Convergent LMI Relaxations for Nonconvex Quadratic Programs. In *Proceedings of the 39th IEEE Conference on Decision and Control*, volume 5, pages 5041–5046, 2000.

[34] J. B. Lasserre. Global Optimization with Polynomials and the Problem of Moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[35] J. B. Lasserre. Convex Sets with Semidefinite Representation. *Mathematical Programming, Series A*, 120(2):457–477, 2009.

[36] M. Laurent. Sums of Squares, Moment Matrices and Optimization over Polynomials. In M. Putinar and S. Sullivant, editors, *Emerging Applications of Algebraic Geometry*, volume 149 of *The IMA Volumes in Mathematics and Its Applications*, pages 157–270. Springer Science+Business Media, LLC, New York, 2009.

[37] M. Laurent and F. Rendl. Semidefinite Programming and Integer Programming. In K. Aardal, G. L. Nemhauser, and R. Weismantel, editors, *Discrete Optimization*, volume 12 of *Handbooks in Operations Research and Management Science*, pages 393–514. Elsevier B. V., Amsterdam, The Netherlands, 2005.

[38] W. L. Li, Y. J. Zhang, A. M.-C. So, and M. Z. Win. Slow Adaptive OFDMA through Chance Constrained Programming. To appear in *IEEE Transactions on Signal Processing*, 2009.

[39] L. Lovász and A. Schrijver. Cones of Matrices and Set–Functions and 0–1 Optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

[40] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang. Semidefinite Relaxation of Quadratic Optimization Problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

[41] 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.

[42] M. Marshall. Representations of Non–Negative Polynomials, Degree Bounds and Applications to Optimization. *Canadian Journal of Mathematics*, 61(1):205–221, 2009.

[43] A. Nemirovski. Sums of Random Symmetric Matrices and Quadratic Optimization under Orthogonality Constraints. *Mathematical Programming, Series B*, 109(2–3):283–317, 2007.

[44] A. Nemirovski, C. Roos, and T. Terlaky. On Maximization of Quadratic Form over Intersection of Ellipsoids with Common Center. *Mathematical Programming, Series A*, 86(3):463–473, 1999.

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

[46] 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.

[47] Yu. Nesterov. Quality of Semidefinite Relaxation for Nonconvex Quadratic Optimization. CORE Discussion Paper 9719, Université Catholique de Louvain, Belgium, 1997.

[48] Yu. Nesterov. Smoothing Technique and Its Applications in Semidefinite Optimization. *Mathematical Programming, Series A*, 110(2):245–259, 2007.

[49] Yu. 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*, volume 27 of *International Series in Operations Research and Management Science*, pages 361–419. Kluwer Academic Publishers, Boston, Massachusetts, 2000.

[50] D. P. Palomar and Y. C. Eldar, editors. *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, New York, 2010.

[51] A. Prestel and C. N. Delzell. *Positive Polynomials: From Hilbert's 17th Problem to Real Algebra*. Springer Monographs in Mathematics. Springer–Verlag, Berlin/Heidelberg, 2001.

[52] M. Putinar and S. Sullivant, editors. *Emerging Applications of Algebraic Geometry*, volume 149 of *The IMA Volumes in Mathematics and Its Applications*. Springer Science+Business Media, LLC, New York, 2009.

[53] P. Raghavendra. Optimal Algorithms and Inapproximability Results for Every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC 2008)*, pages 245–254, 2008.

[54] A. Shapiro. Rank–Reducibility of a Symmetric Matrix and Sampling Theory of Minimum Trace Factor Analysis. *Psychometrika*, 47(2):187–199, 1982.

[55] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MPS–SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2009.

[56] M. B. Shenouda and T. N. Davidson. Outage–Based Designs for Multi–User Transceivers. In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, pages 2389–2392, 2009.

[57] A. M.-C. So. Moment Inequalities for Sums of Random Matrices and Their Applications in Optimization. To appear in *Mathematical Programming*, 2009.

[58] A. M.-C. So. On the Performance of Semidefinite Relaxation MIMO Detectors for QAM Constellations. In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, pages 2449–2452, 2009.

[59] A. M.-C. So. Probabilistic Analysis of the Semidefinite Relaxation Detector in Digital Communications. In *Proceedings of the 21st Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2010)*, pages 698–711, 2010.

[60] A. M.-C. So and Y. Ye. A Semidefinite Programming Approach to Tensegrity Theory and Realizability of Graphs. In *Proceedings of the 17th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2006)*, pages 766–775, 2006.

[61] A. M.-C. So and Y. Ye. Theory of Semidefinite Programming for Sensor Network Localization. *Mathematical Programming, Series B*, 109(2):367–384, 2007.

[62] A. M.-C. So, Y. Ye, and J. Zhang. A Unified Theorem on SDP Rank Reduction. *Mathematics of Operations Research*, 33(4):910–920, 2008.

[63] 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.

[64] K. Varadarajan, S. Venkatesh, Y. Ye, and J. Zhang. Approximating the Radii of Point Sets. *SIAM Journal on Computing*, 36(6):1764–1776, 2007.

[65] Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg. Row by Row Methods for Semidefinite Programming. Technical report, Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, 2009.

[66] X.-M. Wu, A. M.-C. So, Z. Li, and S.-Y. R. Li. Fast Graph Laplacian Regularized Kernel Learning via Semidefinite–Quadratic–Linear Programming. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22: Proceedings of the 2009 Conference*, pages 1964–1972, Cambridge, Massachusetts, 2009. The MIT Press.

[67] Y. Ye. Approximating Quadratic Programming with Bound and Quadratic Constraints. *Mathematical Programming*, 84(2):219–226, 1999.

[68] Y. Ye and S. Zhang. New Results on Quadratic Minimization. *SIAM Journal on Optimization*, 14(1):245–267, 2003.

[69] Y. J. Zhang and A. M.-C. So. Optimal Spectrum Sharing in MIMO Cognitive Radio Networks via Semidefinite Programming. Preprint, 2009.

[70] Z. Zhu, A. M.-C. So, and Y. Ye. Universal Rigidity: Towards Accurate and Efficient Localization of Wireless Networks. In *Proceedings of the 29th Conference on Computer Communications (INFOCOM 2010)*, 2010.