13 SEMIDEFINITE PROGRAMMING RELAXATIONS OF NONCONVEX QUADRATIC OPTIMIZATION

Yuri Nesterov, Henry Wolkowicz, Yinyu Ye

13.1 INTRODUCTION

Quadratically constrained quadratic programs, denoted Q^2P, are an important modelling tool, e.g.: for hard combinatorial optimization problems, Chapter 12; and SQP methods in nonlinear programming, Chapter 20. These problems are too hard to solve in general. Therefore, relaxations such as the Lagrangian relaxation are used. The dual of the Lagrangian relaxation is the SDP relaxation. Thus SDP has enabled us to efficiently solve the Lagrangian relaxation and find good approximate solutions for these hard, possibly nonconvex, Q^2P. This area has generated a lot of research recently. This has resulted in many strong and elegant theorems that describe the strength/performance of the bounds obtained from solving relaxations of these Q^2P.

For the simple Q^2P case of one quadratic constraint (the trust region subproblem) strong duality holds, even though both the objective function and constraint may be nonconvex, i.e. there is a zero duality gap and the dual is attained. In addition, necessary and sufficient (strengthened) second order optimality conditions and efficient algorithms exist. However, these nice duality results already fail for the two trust region subproblem (CDT problem).

Surprisingly, there are other classes of nonconvex Q^2P where strong duality holds. This includes the special cases of orthogonality type constraints.

Throughout this chapter we emphasize the theme (or open problem) that *Lagrangian duality is best*, i.e. in every case that we have a good (tractable) bound we show that it is equivalent to that obtained from the Lagrangian relaxation of an appropriate problem. Moreover, we include several results on the strength of these bounds. These results follow the pioneering paper [285] and study the theme that a solution of an indefinite quadratic maximization problem with some linear constraints on the squared variables can be approximated with a constant relative accuracy.

In parts 13.2 and 13.3, we present several complexity results on the quality of the SDP relaxations. We present a convex conic relaxation for a problem of maximizing an indefinite quadratic form over a set of convex constraints on the squared variables. We show that for all these problems we get at least $\frac{12}{37}$-relative accuracy of the approximation. In the second part of the paper we derive the conic relaxation by another approach based on the second order optimality conditions. We show that for l_p-balls, $p \geq 2$, intersected by a linear subspace, it is possible to guarantee $(1-\frac{2}{p})$-relative accuracy of the solution. As a consequence, we prove $(1-\frac{1}{e \ln n})$-relative accuracy of the conic relaxation for an indefinite quadratic maximization problem over an n-dimensional unit box with homogeneous linear equality constraints. We discuss the implications of the results for the discussion around the question $P = NP$. We also consider the problem of approximating the global maximum of a quadratic program (QP) subject to bound and (simple) quadratic onstraints. Based on several early results, we show that a 4/7-approximate solution can be obtained in polynomial time.

The rest of the paper is organized as follows. We begin in Section 13.4.1 with the most well known problem in this area, the Max-Cut problem. We present several different relaxations. Surprisingly, following our theme, all these bounds, including the SDP bound, end up being equivalent to the Lagrangian relaxation; see Section 13.4.1. We then present a strengthened SDP bound based on a second lifting procedure.

We discuss the SDP relaxation for general Q^2P in Section 13.4.2. This includes descriptions of the relationships between the SDP relaxation and the Lagrangian relaxation via convex quadratic valid inequalities, following [260, 442]. Several applications, including QAP and GP, are presented in Section 13.4.2.

Occurrences of strong duality for nonconvex quadratic programs is studied in Section 13.4.3. In every instance where one has a tractable bound, we find a Q^2P such that the bound is attained by the Lagrangian relaxation. This follows the work in [41, 37].

13.1.1 Lagrange Multipliers for Q^2P

We now define the (inequality constrained) Q^{2}P in x. (Though our notation does differ slightly in the separate parts (sections) of this chapter.)

$$
(\mathrm{Q}^2\mathrm{P}_{\,x}) \qquad
\begin{array}{rll}
q^* := & \min & q_0(x) := x^T q_0 x + 2g_0^T x + \alpha_0 \\
 & \text{subject to} & q_k(x) := x^T Q_k x + 2g_k^T x + \alpha_k \leq 0 \\
 & & k \in \mathcal{I} := \{1, \ldots, m\} \\
 & & x \in \Re^n,
\end{array}
$$

where the matrices Q_k are symmetric. The Lagrangian of Q^{2}P$_{\,x}$ is

$$L(x, \lambda) := q_0(x) + \sum_{k \in \mathcal{I}} \lambda_k q_k(x),$$

where $\lambda = (\lambda_k) \geq 0$ are nonnegative Lagrange multipliers.

Lagrange multipliers can be used in two ways. First, if a constraint qualification holds for Q^{2}P at the optimum $\bar{x}$ (e.g. the Mangasarian-Fromovitz constraint qualification), then the Karush-Kuhn-Tucker necessary conditions for optimality hold, i.e.

$$\nabla L(x, \lambda) = 0, \text{ and } \lambda_k q_k(x) = 0, \forall k \in \mathcal{I}.$$

Therefore, the optimum $\bar{x}$ can be searched among the points satisfying stationarity of the Lagrangian and complementary slackness. Moreover, if the Lagrangian is also convex, then this is a sufficient condition for optimality.

Lagrange multipliers can also be used to derive the Lagrangian dual (or relaxation) of Q^{2}P$_{\,x}$

$$(DQ^2P_x) \qquad q^* \geq d^* := \max_{\lambda \overset{\mu}{\geq} 0} \min_x q_0(x) + \sum_{k \in \mathcal{I}} \lambda_k q_k(x).$$

A zero duality gap holds if $q^* = d^*$. This can fail in the nonconvex case. Strong duality holds if $q^* = d^*$ and also d^* is attained. Moreover, d^* can be efficiently evaluated using SDP.

13.2 GLOBAL QUADRATIC OPTIMIZATION VIA CONIC RELAXATION

Yuri Nesterov

Starting from the pioneering paper [285], there were obtained several results [572, 859, 575], which show that a solution of an indefinite quadratic maximization problem with some linear constraints on the squared variables can be approximated with a constant relative accuracy. In this Section 13.2 we present some improvements and extensions of the results [575].

In Section 13.2.1 we consider a problem of maximizing an indefinite quadratic form subject to arbitrary convex constraints on the squared variables. For convenience of the dual description we use a conic representation of these constraints. We introduce a convex conic relaxation for that problem and prove that it provides us at least with an approximation of $\frac{\pi-2}{6-\pi}$ relative accuracy. In Section 13.2.2 we show how to improve this approximation using the diagonal elements of the quadratic objective function. The relative accuracy, which we can get in this case, is $\frac{12}{37}$. In Section 13.2.3 we extend the results of Section 13.2.1, 13.2.2 onto the case of general convex constraints on squared variables. We conclude the first part of the section with a discussion of the difficulties which arise in the problems with linear equality constraints on the initial variables (Section 13.2.4).

In the second part of the section, which starts from Section 13.2.5, we study another way of deriving the conic relaxation. This approach can be applied only to a small number of sets (l_p-balls, $p \geq 2$), but it allows to treat also the linear equations. We prove that for such problems the conic relaxation gives $(1-\frac{2}{p})$ relative accuracy. In Section 13.2.6 we apply these results to a problem of maximizing a quadratic function over a unit box subject to a system of homogeneous linear equalities. We show that it is possible to compute in polynomial time a $(1-\frac{1}{e\ln n})$-solution of that problem. We conclude the section with a discussion of the results.

We first recall some of the notation we use. For two vectors x, $y \in R^n$ we denote $\langle x, y\rangle$ the standard inner product:

$$\langle x, y\rangle = \sum_{i=1}^{n} x^{(i)} y^{(i)}.$$

Then $\| x \| = \langle x, x\rangle^{1/2}$. Since we work in several finite-dimensional spaces, the meaning of this notation is defined by the spaces of the arguments. For example, for two $(m \times n)$-matrices X and Y we have

$$\langle X, Y\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} Y_{ij}.$$

We use the standard notation for l_p-norms:

$$\| x \|_p = \left[\sum_{i=1}^{n} | x^{(i)} |^p\right]^{1/p}, \quad x \in R^n,\ p \geq 1.$$

Again, the meaning of the notation depends on the dimension of space of the argument. Recall, that for $p = \infty$ we have $\| x \|_\infty = \max\limits_{1 \leq i \leq n} | x^{(i)} |$. The norm dual to $\| \cdot \|_p$ is $\| \cdot \|_{p^*}$ with $p^* = \frac{p}{p-1}$:

$$\| y \|_{p^*} = \max\{\langle y, x\rangle : \| x \|_p \leq 1\}.$$

For a symmetric matrix A we write $A \succeq 0$ if A is positive semidefinite. Notation $B \succeq A$ means that $B - A \succeq 0$. For $x \in R^n$ we denote $\mathrm{Diag}\,(x)$ the diagonal $(n \times n)$-matrix with diagonal entries $x^{(i)}$. Conversely, $\mathrm{diag}(X) \in R^n$ denotes the diagonal of an $(n \times n)$-matrix X. Notation e_i is used for the ith coordinate vector of R^n and $1_n \in R^n$ stands for the vector of all ones. Thus, $I_n = \mathrm{Diag}\,(1_n)$ is a unit matrix. Notation 0_n is used for the zero vector in R^n.

We use square brackets in order to indicate the component-wise operations with the vectors. For example, notation $[x \cdot y]$ stands for the vector with components $x^{(i)}y^{(i)}$, x, $y \in R^n$. Notation $[x]^2$ is used for the vector with the components $(x^{(i)})^2$. If $f(\tau)$ is a univariate function, we denote $f[x]$ the vector with the components $f(x^{(i)})$. In order to indicate the partial ordering in R^n we use the usual inequality signs. Thus, $x \geq y$ for x and y from R^n means that $x^{(i)} \geq y^{(i)}$, $i = 1, \ldots, n$.

Finally, $[\alpha, \beta]^n$ denotes a continuous box in R^n, that is $\{x \in R^n : \alpha 1_n \leq x \leq \beta 1_n\}$. For a boolean box $\{x \in R^n : x^{(i)} = (\alpha \text{ or } \beta)\}$ we use notation $\{\alpha, \beta\}^n$.

13.2.1 Convex conic constraints on squared variables

Let Q be an arbitrary symmetric $(n \times n)$-matrix. Consider the following pair of optimization problems:

$$\begin{aligned} \text{find } \phi^* = &\quad \max\{\langle Qx, x\rangle : [x]^2 \in \mathcal{F}\}, \\ \text{find } \phi_* = &\quad \min\{\langle Qx, x\rangle : [x]^2 \in \mathcal{F}\}. \end{aligned} \tag{13.2.1}$$

where $\mathcal{F}$ is a closed convex set. Our main assumption on the problem (13.2.1) is as follows.

Assumption 13.2.1 *1). The set $\mathcal{F}$ is bounded. 2). There exists a strictly positive $v \in \mathcal{F}$.*

In order to simplify the dual analysis, in this section we assume that the feasible set $\mathcal{F}$ is presented in a conic form:

$$\mathcal{F} = \{v \in K : Av = b\}, \tag{13.2.2}$$

where K is a convex closed pointed cone in R^n with non-empty interior, A is an $(m \times n)$-matrix and $b \neq 0_m$. Our additional assumption on the set $\mathcal{F}$ is as follows.

Assumption 13.2.2 $\{v \in \mathrm{int}\, K : Av = b\} \neq \emptyset$.

Note that the form (13.2.2) is quite general, since any bounded convex set can be written in this way (see [583] for details). At the same time, in Section

13.2.3 we will show how to transform our result on the case of a general convex feasible set $\mathcal{F}$.

Using the same technique as in [575], we can rewrite the pair of problems (13.2.1) in a trigonometric form.

Lemma 13.2.1

$$\begin{array}{rcl} \phi^* = & \max\limits_{\substack{X \succeq 0,\, \mathrm{diag}(X) = 1_n, \\ d \geq 0,\, [d]^2 \in \mathcal{F},}} & \frac{2}{\pi}\langle Q, \mathrm{Diag}\,(d) \arcsin[X] \mathrm{Diag}\,(d)\rangle, \\ \phi_* = & \min\limits_{\substack{X \succeq 0,\, \mathrm{diag}(X) = 1_n, \\ d \geq 0,\, [d]^2 \in \mathcal{F}.}} & \frac{2}{\pi}\langle Q, \mathrm{Diag}\,(d) \arcsin[X] \mathrm{Diag}\,(d)\rangle, \end{array} \tag{13.2.3}$$

Proof.
Indeed, let us represent a vector $x \in R^n$ as follows:

$$x = [d \cdot \sigma], \quad d \geq 0 \in R^n, \ \sigma \in \{-1, 1\}^n.$$

Note that $[x]^2 = [d]^2$. Therefore $\phi^* = \max\limits_{d}\{\Phi(d) : d \geq 0, \ [d]^2 \in \mathcal{F}\}$ with

$$\Phi(d) = \max\{\langle \mathrm{Diag}\,(d) Q \mathrm{Diag}\,(d)\sigma, \sigma\rangle \ : \ \sigma \in \{-1, 1\}^n\}.$$

Using Theorem 2.3 [575], we can represent $\Phi(d)$ in the following form:

$$\Phi(d) = \max\{\tfrac{2}{\pi}\langle \mathrm{Diag}\,(d) Q \mathrm{Diag}\,(d), \arcsin[X]\rangle \ : \ X \succeq 0, \ \mathrm{diag}(X) = 1_n\}.$$

Inserting this representation in the above expression for ϕ^* we get the first statement of the lemma. The second one can be obtained in a similar way. ■

Note that in general none of the problems (13.2.1) is convex in x. Therefore, in order to estimate their optimal values, we need to use a kind of convex relaxation. Let us define the *conic relaxation* of problems (13.2.1):

$$\begin{array}{rl} \psi^* = & \max\{\langle Q, X\rangle \ : \ \mathrm{diag}(X) \in \mathcal{F}, \ X \succeq 0\}, \\ \psi_* = & \min\{\langle Q, X\rangle \ : \ \mathrm{diag}(X) \in \mathcal{F}, \ X \succeq 0\}. \end{array} \tag{13.2.4}$$

Sometimes it is convenient to use a dual form of these relaxations. Recall that for a convex cone $K \subseteq R^n$ the dual cone K^* is defined as follows:

$$K^* = \{u \in R^n \ : \ \langle u, v\rangle \geq 0, \ \forall v \in K\}.$$

Lemma 13.2.2

$$\begin{array}{rl} \psi^* = & \min\limits_{y \in R^m, u \in R^n} \{\langle b, y\rangle \ : \ Q + \mathrm{Diag}\,(u) \preceq \mathrm{Diag}\,(A^T y), \ u \in K^*\}, \\ \psi_* = & \max\limits_{y \in R^m, u \in R^n} \{\langle b, y\rangle \ : \ Q \succeq \mathrm{Diag}\,(u) + \mathrm{Diag}\,(A^T y), \ u \in K^*\}. \end{array} \tag{13.2.5}$$

Proof.
In view of Assumptions 13.2.1, 13.2.2, we can get a dual representation of the upper relaxation ψ^* as follows:

$$\psi^* = \max_{X,v}\{\langle Q, X\rangle : \ A\mathrm{diag}(X) = b, \ \mathrm{diag}(X) = v, \ X \succeq 0, \ v \in K\}$$

$$= \max_{X \succeq 0, v \in K} \ \min_{y \in R^m, u \in R^n} \{\langle Q, X\rangle + \langle y, b - A\mathrm{diag}(X)\rangle + \langle u, \mathrm{diag}(X) - v\rangle\}$$

$$= \min_{y \in R^m, u \in R^n} \ \max_{X,v}\{\langle Q + \mathrm{Diag}\,(u - A^T y), X\rangle + \langle b, y\rangle - \langle u, v\rangle : ||X \succeq 0, \ v \in K\}$$

$$= \min_{y \in R^m, u \in R^n}\{\langle b, y\rangle : \ Q + \mathrm{Diag}\,(u) \preceq \mathrm{Diag}\,(A^T y), \ u \in K^*\}.$$

Similarly, for the lower relaxation we get the following:

$$\psi\min_{X,v}\{\langle Q, X\rangle : \ A\mathrm{diag}(X) = b, \ \mathrm{diag}(X) = v, \ X \succeq 0, \ v \in K\}$$

$$= \min_{X \succeq 0, v \in K} \ \max_{y \in R^m, u \in R^n} \{\langle Q, X\rangle + \langle y, b - A\mathrm{diag}(X)\rangle + \langle u, v - \mathrm{diag}(X)\rangle\}$$

$$= \max_{y \in R^m, u \in R^n} \ \min_{X,v}\{\langle Q - \mathrm{Diag}\,(u + A^T y), X\rangle + \langle b, y\rangle + \langle u, v\rangle : ||X \succeq 0, \ v \in K\}$$

$$= \max_{y \in R^m, u \in R^n}\{\langle b, y\rangle : \ Q \succeq \mathrm{Diag}\,(u) + \mathrm{Diag}\,(A^T y), \ u \in K^*\}.$$

■

Let us establish some relations between the relaxations (13.2.4) and the optimal values of the problems (13.2.1). Denote

$$\psi(\alpha) = \alpha\psi^* + (1 - \alpha)\psi_*. \tag{13.2.6}$$

The proof of the following theorem is similar to that of Theorem 3.3 [575].

Theorem 13.2.1

$$\psi_* \le \phi_* \le \psi(1 - \tfrac{2}{\pi}) \le \psi(\tfrac{2}{\pi}) \le \phi^* \le \psi^*. \qquad \textit{(13.2.7)}$$

Proof.
Note that $\psi_* \le \psi^*$ by definition. So, the middle inequality in (13.2.7) is correct. Further, if $[x]^2 \in \mathcal{F}$ then the matrix $X = xx^T$ is feasible for both relaxation problems (13.2.4) since $\mathrm{diag}(X) = [x]^2$. Moreover, $\langle Q, X\rangle = \langle Qx, x\rangle$. Thus, both bounding inequalities in the chain (13.2.7) are valid. Let us prove now two remaining inequalities.

Let us choose arbitrary $u \in K$ and $y \in R^m$, which satisfy the constraints of the dual form (13.2.5) of the lower relaxation ψ_*:

$$(u, y) \in \mathcal{F}_d = \{(u, y) \in K^* \times R^m : \; Q \succeq \mathrm{Diag}\,(u) + \mathrm{Diag}\,(A^T y)\}. \qquad (13.2.8)$$

Consider a pair (X, d), which satisfies the constraints of the trigonometric representation (13.2.3) for ϕ^*:

$$X \succeq 0, \quad \mathrm{diag}(X) = 1_n, \quad d \geq 0, \quad A[d]^2 = b, \quad [d]^2 \in K. \qquad (13.2.9)$$

Since $X \succeq 0$ and $\mid X_{ij} \mid \leq 1$ we have $\arcsin[X] \succeq X$ in view of Corollary 3.2 [575]. Therefore, using Lemma 13.2.1 we get the following:

$$\begin{aligned}\phi^* &\geq \tfrac{2}{\pi}\langle \mathrm{Diag}\,(d) Q \mathrm{Diag}\,(d), \arcsin[X]\rangle \\ &= \tfrac{2}{\pi}\langle \mathrm{Diag}\,(d)(Q - \mathrm{Diag}\,(u) - \mathrm{Diag}\,(A^T y))\mathrm{Diag}\,(d), \arcsin[X]\rangle \\ &\qquad\qquad + \langle u + A^T y, [d]^2\rangle \\ &\geq \tfrac{2}{\pi}\langle \mathrm{Diag}\,(d)(Q - \mathrm{Diag}\,(u) - \mathrm{Diag}\,(A^T y))\mathrm{Diag}\,(d), X\rangle + \langle u + A^T y, [d]^2\rangle \\ &= \tfrac{2}{\pi}\langle Q, \mathrm{Diag}\,(d) X \mathrm{Diag}\,(d)\rangle + (1 - \tfrac{2}{\pi})\langle u + A^T y, [d]^2\rangle.\end{aligned}$$

Note that $u \in K^*$ and $[d]^2 \in K$. Therefore $\langle u, [d]^2\rangle \geq 0$. In view of (13.2.9) we have

$$\langle A^T y, [d]^2\rangle = \langle A[d]^2, y\rangle = \langle b, y\rangle.$$

Finally, for any pair (X, d), which satisfy (13.2.9) we have $Y = \mathrm{Diag}\,(d) X \mathrm{Diag}\,(d)$ feasible for the primal relaxation problems:

$$Y \in \mathcal{F}_p = \{Y : \; Y \succeq 0, \; A\mathrm{diag}(Y) = b, \; \mathrm{diag}(Y) \in K\}.$$

On the other hand, any $Y \in \mathcal{F}_p$ can be represented as $Y = \mathrm{Diag}\,(d) X \mathrm{Diag}\,(d)$ with X and d, which satisfy (13.2.9). Therefore, we conclude that

$$\phi^* \geq \tfrac{2}{\pi}\langle Q, Y\rangle + (1 - \tfrac{2}{\pi})\langle b, y\rangle, \quad \forall Y \in \mathcal{F}_p, \; (u, y) \in \mathcal{F}_d.$$

This proves the forth inequality in the chain (13.2.7). The remaining inequality can be proved in a similar way. ■

Definition 13.2.1 *We say that the value ψ approximates ϕ^* with a relative accuracy $\mu \in [0, 1]$ if $\mid \psi - \phi^* \mid \leq \mu(\phi^* - \phi_*)$. We call this approximation implementable if $\psi \leq \phi^*$.*

Corollary 13.2.1 *1. Let $\alpha = \frac{2}{\pi}$. Then the value $\psi(\alpha)$ is an implementable approximation of ϕ^* with the relative accuracy $\mu = \frac{\pi}{2} - 1 < \frac{4}{7}$.*

2. Let $\beta = \frac{(1+\alpha)^2 - 2}{3\alpha - 1}$. Then the value $\psi(\beta)$ approximates ϕ^* with the relative accuracy $\mu = \frac{\pi - 2}{6 - \pi} < \frac{2}{5}$.

The proof of that statement is exactly the same as that of Corollary 3.4 in [575].

13.2.2 Using additional information

In this section it is shown how to improve the quality of our bounds by taking into account some additional information. Define

$$\begin{aligned} \tau^* &= \max\{\langle \mathrm{diag}(Q), v\rangle :\ v \geq 0,\ v \in \mathcal{F}\}, \\ \tau_* &= \min\{\langle \mathrm{diag}(Q), v\rangle :\ v \geq 0,\ v \in \mathcal{F}\}. \end{aligned} \tag{13.2.10}$$

Note that these values are computable in polynomial time. In view of Lemma 13.2.1 we have

$$\phi_* \leq \tau_* \leq \tau^* \leq \phi^*. \tag{13.2.11}$$

Hence,

$$\beta^* \equiv \tfrac{\psi^* - \tau^*}{\psi^* - \psi_*} \in [0,1], \quad \beta_* \equiv \tfrac{\tau_* - \psi_*}{\psi^* - \psi_*} \in [0,1].$$

Using these values we can express τ^* and τ_* as follows:

$$\begin{aligned} \tau^* &= \psi^* - \beta^*(\psi^* - \psi_*) &= \psi(1-\beta^*), \\ \tau_* &= \psi_* + \beta_*(\psi^* - \psi_*) &= \psi(\beta_*). \end{aligned}$$

Denote $\omega(\beta) = \beta \arcsin(\beta) + \sqrt{1-\beta^2} \equiv 1 + \int\limits_0^\beta \arcsin(\tau) d\tau$, $\beta \in [0,1]$. This function is increasing and convex with $\omega(0) = 1$ and $\omega(1) = \frac{\pi}{2}$. In what follows we denote $\bar{\beta}$ the unique root of the following equation:

$$\tfrac{2}{\pi}\omega(\beta) = 1 - \beta, \quad \beta \in [0,1].$$

It can be shown that $\frac{23}{70} < \bar{\beta} < \frac{24}{73}$.

Theorem 13.2.2 *1. Denote*

$$\begin{aligned} \alpha^* &= \max\{\tfrac{2}{\pi}\omega(\beta_*), 1 - \beta^*\}, \\ \alpha_* &= \min\{1 - \tfrac{2}{\pi}\omega(\beta^*), \beta_*\}. \end{aligned}$$

The optimal values of the problems (13.2.1) satisfy the following relations:

$$\psi^* \geq \ \phi^* \ \geq \psi(\alpha^*), \tag{13.2.12}$$

$$\psi_* \leq \ \phi_* \ \leq \psi(\alpha_*). \tag{13.2.13}$$

2. The value $\psi(\alpha^*)$ is an implementable approximation of ϕ^* with relative accuracy

$$\mu = \tfrac{1-\alpha^*}{1-\alpha_*} \leq \tfrac{\bar{\beta}}{1-\bar{\beta}} < \tfrac{24}{49}.$$

3. Denote $\bar{\alpha} = \frac{\alpha^*(2-\alpha_*)-\alpha_*}{1+\alpha^*-2\alpha_*}$. The value $\psi(\bar{\alpha})$ is a μ-approximation of ϕ^* with

$$\mu = \tfrac{1-\alpha^*}{1+\alpha^*-2\alpha_*} \leq \tfrac{\bar{\beta}}{2-3\bar{\beta}} < \tfrac{12}{37}.$$

In Items 2 and 3 the upper bounds are achieved for $\beta^* = \beta_* = \bar{\beta}$.

Proof.
Let $X \succeq 0$ and $d \geq 0$ be feasible for the trigonometric form of the upper relaxation (13.2.3):

$$\operatorname{diag}(X) = 1_n, \quad A[d]^2 = b, \quad [d]^2 \in K.$$

Consider the matrices $X_\gamma = \gamma X + (1-\gamma)I_n$, $\gamma \in [0,1]$. Then

$$\arcsin[X_\gamma] = \arcsin[\gamma X] + (\tfrac{\pi}{2} - \arcsin(\gamma))I_n.$$

Therefore

$$\begin{aligned} \phi^* &\geq \tfrac{2}{\pi}\langle Q, \operatorname{Diag}(d)\arcsin[X_\gamma]\operatorname{Diag}(d)\rangle \\ &= \tfrac{2}{\pi}\langle Q, \operatorname{Diag}(d)\arcsin[\gamma X]\operatorname{Diag}(d)\rangle + \left(1 - \tfrac{2}{\pi}\arcsin(\gamma)\right)\langle \operatorname{diag}(Q), [d]^2\rangle. \end{aligned} \tag{13.2.14}$$

Let us choose now arbitrary $u \in K$ and $y \in R^m$ which satisfy the constraints of the dual form (13.2.5) of the lower relaxation ψ_*:

$$(u,y) \in \mathcal{F}_d = \{(u,y) \in K^* \times R^m : Q \succeq \operatorname{Diag}(u) + \operatorname{Diag}(A^T y)\}.$$

Then, in view of Corollary 3.2 [575] we have:

$$\begin{aligned} &\langle Q, \operatorname{Diag}(d)\arcsin[\gamma X]\operatorname{Diag}(d)\rangle \\ &\quad = \langle Q - \operatorname{Diag}(u) - \operatorname{Diag}(A^T y), \operatorname{Diag}(d)\arcsin[\gamma X]\operatorname{Diag}(d)\rangle \\ &\qquad + \langle \operatorname{Diag}(u + A^T y), \operatorname{Diag}(d)\arcsin[\gamma X]\operatorname{Diag}(d)\rangle \\ &\quad \geq \gamma\langle Q - \operatorname{Diag}(u) - \operatorname{Diag}(A^T y), \operatorname{Diag}(d)X\operatorname{Diag}(d)\rangle \\ &\qquad + \arcsin(\gamma)\langle u + A^T y, [d]^2\rangle \\ &\quad = \gamma\langle Q, \operatorname{Diag}(d)X\operatorname{Diag}(d)\rangle + (\arcsin(\gamma) - \gamma)\langle u + A^T y, [d]^2\rangle. \end{aligned}$$

Note that $\arcsin(\gamma) \geq \gamma$ for $\gamma \in [0,1]$. At the same time $u \in K^*$ and $[d]^2 \in K$. Therefore $\langle u, [d]^2\rangle \geq 0$. Finally, $\langle A^T y, [d]^2\rangle = \langle A[d]^2, y\rangle = \langle b, y\rangle$. Thus,

$$\begin{aligned} \langle Q, \operatorname{Diag}(d)\arcsin[\gamma X]\operatorname{Diag}(d)\rangle \geq{}& \gamma\langle Q, \operatorname{Diag}(d)X\operatorname{Diag}(d)\rangle \\ &+ (\arcsin(\gamma) - \gamma)\langle b, y\rangle. \end{aligned}$$

Substituting this inequality in (13.2.14) we get the following:

$$\begin{aligned} \phi^* \geq{}& \tfrac{2}{\pi}\left(\gamma\langle Q, \operatorname{Diag}(d)X\operatorname{Diag}(d)\rangle + (\arcsin(\gamma) - \gamma)\langle b, y\rangle\right) \\ &+ \left(1 - \tfrac{2}{\pi}\arcsin(\gamma)\right)\langle \operatorname{diag}(Q), [d]^2\rangle \\ \geq{}& \tfrac{2}{\pi}\left(\gamma\langle Q, \operatorname{Diag}(d)X\operatorname{Diag}(d)\rangle + (\arcsin(\gamma) - \gamma)\langle b, y\rangle\right) \\ &+ \left(1 - \tfrac{2}{\pi}\arcsin(\gamma)\right)\tau_*. \end{aligned}$$

Using the same reasoning as in Theorem 13.2.1, we conclude that

$$\begin{aligned}\phi^* &\geq \tfrac{2}{\pi}\gamma\psi^* + \tfrac{2}{\pi}(\arcsin(\gamma) - \gamma)\psi_* + \left(1 - \tfrac{2}{\pi}\arcsin(\gamma)\right)\tau_* \\ &= \tfrac{2}{\pi}\arcsin(\gamma)\psi\left(\tfrac{\gamma}{\arcsin(\gamma)}\right) + \left(1 - \tfrac{2}{\pi}\arcsin(\gamma)\right)\psi(\beta_*) \\ &= \psi\left(\tfrac{2\gamma}{\pi} + \left(1 - \tfrac{2}{\pi}\arcsin(\gamma)\right)\beta_*\right).\end{aligned}$$

The right-hand side of the above inequality is maximal for $\gamma^* = \sqrt{1-\beta_*^2}$. Then

$$\tfrac{2\gamma^*}{\pi} + \left(1 - \tfrac{2}{\pi}\arcsin(\gamma^*)\right)\beta_* = \tfrac{2}{\pi}\left(\sqrt{1-\beta_*^2} + \beta_*\arccos(\gamma^*)\right) = \tfrac{2}{\pi}\omega(\beta_*).$$

Thus, $\phi^* \geq \psi(\frac{2}{\pi}\omega(\beta_*))$. Combining this inequality with (13.2.11), we get the lower bound in (13.2.12). The relations (13.2.13) for ψ_* can be obtained in a similar way.

Let us prove now Item 2 of the theorem. In view of (13.2.12) and (13.2.13) we have

$$0 \leq \frac{\phi^* - \psi(\alpha^*)}{\phi^* - \phi_*} \leq \frac{\psi^* - \psi(\alpha^*)}{\psi^* - \phi_*} \leq \frac{\psi^* - \psi(\alpha^*)}{\psi^* - \psi(\alpha_*)} = \frac{\psi(1) - \psi(\alpha^*)}{\psi(1) - \psi(\alpha_*)} = \frac{1-\alpha^*}{1-\alpha_*}. \tag{13.2.15}$$

Note that

$$\begin{aligned}1 - \alpha^* &= 1 - \max\{\tfrac{2}{\pi}\omega(\beta_*), 1 - \beta^*\} &&= \min\{1 - \tfrac{2}{\pi}\omega(\beta_*), \beta^*\}, \\ 1 - \alpha_* &= 1 - \min\{1 - \tfrac{2}{\pi}\omega(\beta^*), \beta_*\} &&= \max\{\tfrac{2}{\pi}\omega(\beta^*), 1 - \beta_*\}.\end{aligned}$$

Thus, we need to find an upper bound for the ratio

$$\rho(\beta_1, \beta_2) = \frac{\min\{1 - \frac{2}{\pi}\omega(\beta_2), \beta_1\}}{\max\{\frac{2}{\pi}\omega(\beta_1), 1 - \beta_2\}}, \quad 0 \leq \beta_1, \beta_2 \leq 1.$$

Lemma 13.2.3

$$\max\{\rho(\beta_1, \beta_2) :\ 0 \leq \beta_1, \beta_2 \leq 1\} = \frac{\bar{\beta}}{1 - \bar{\beta}}.$$

Proof.
We need to prove that

$$(1 - \bar{\beta})\min\{1 - \tfrac{2}{\pi}\omega(\beta_2), \beta_1\} \leq \bar{\beta}\max\{\tfrac{2}{\pi}\omega(\beta_1), 1 - \beta_2\}, \quad 0 \leq \beta_1, \beta_2 \leq 1.$$

This is equivalent to the statement that the convex function

$$g(\beta_1, \beta_2) = \bar{\beta}\max\{\tfrac{2}{\pi}\omega(\beta_1), 1 - \beta_2\} + (1 - \bar{\beta})\max\{\tfrac{2}{\pi}\omega(\beta_2) - 1, -\beta_1\}$$

is non-negative for $0 \leq \beta_1, \beta_2 \leq 1$.

Note that $g(\bar{\beta},\bar{\beta}) = 0$ in view of the definition of $\bar{\beta}$. The subdifferential of the function $g(\cdot)$ at this point is as follows:

$$\partial g(\bar{\beta},\bar{\beta}) = \bar{\beta}\mathrm{Conv}\,\{(\tfrac{2}{\pi}\omega'(\bar{\beta}),0);\ (0,-1)\} + (1-\bar{\beta})\mathrm{Conv}\,\{(0,\tfrac{2}{\pi}\omega'(\bar{\beta}));(-1,0)\}.$$

Thus, this set contains the following points:

$$(\tfrac{2}{\pi}\bar{\beta}\omega'(\bar{\beta}) - 1 + \bar{\beta}, 0), \quad (0, \tfrac{2}{\pi}(1-\bar{\beta})\omega'(\bar{\beta}) - \bar{\beta}), \quad (\tfrac{2}{\pi}\bar{\beta}\omega'(\bar{\beta}), \tfrac{2}{\pi}(1-\bar{\beta})\omega'(\bar{\beta})).$$

Note that

$$\tfrac{2}{\pi}\omega'(\bar{\beta}) < \omega'(\bar{\beta}) < \tfrac{\bar{\beta}}{1-\bar{\beta}} < \tfrac{1-\bar{\beta}}{\bar{\beta}}.$$

Therefore the first coordinate of the first point and the second coordinate of the second point are negative. Since both coordinates of the third point are positive, we conclude that $0 \in \mathrm{int}\,\partial g(\bar{\beta},\bar{\beta})$. ■

Applying Lemma 13.2.3 to (13.2.15) we prove the statement of Item 2.

In order to prove Item 3 note that in view of inequalities (13.2.12) and (13.2.13), for any $\alpha \in [0,1]$ we have

$$\begin{aligned} \tfrac{|\psi(\alpha)-\phi^*|}{\phi^*-\phi_*} &\leq \tfrac{|\psi(\alpha)-\phi^*|}{\phi^*-\psi(\alpha_*)} = \max\left\{\tfrac{\psi(\alpha)-\phi^*}{\phi^*-\psi(\alpha_*)}, \tfrac{\phi^*-\psi(\alpha)}{\phi^*-\psi(\alpha_*)}\right\} \\ &\leq \max\left\{\tfrac{\psi(\alpha)-\psi(\alpha^*)}{\psi(\alpha^*)-\psi(\alpha_*)}, \tfrac{\psi^*-\psi(\alpha)}{\psi^*-\psi(\alpha_*)}\right\} = \max\left\{\tfrac{\alpha-\alpha^*}{\alpha^*-\alpha_*}, \tfrac{1-\alpha}{1-\alpha_*}\right\} \equiv r(\alpha). \end{aligned}$$

The minimum $\bar{\alpha}$ of the function $r(\alpha)$ is a solution of the following equation:

$$(\alpha-\alpha^*)(1-\alpha_*) = (1-\alpha)(\alpha^*-\alpha_*).$$

That is $\bar{\alpha} = \frac{\alpha^*(2-\alpha_*)-\alpha_*}{1+\alpha^*-2\alpha_*}$. Using Lemma 13.2.3 we can estimate the optimal value $r(\bar{\alpha})$ as follows:

$$r(\bar{\alpha}) = \tfrac{1}{1-\alpha_*}\left(1 - \tfrac{\alpha^*(2-\alpha_*)-\alpha_*}{1+\alpha^*-2\alpha_*}\right) = \tfrac{1-\alpha^*}{1+\alpha^*-2\alpha_*} = \tfrac{\rho(\beta^*,\beta_*)}{2-\rho(\beta^*,\beta_*)} \leq \tfrac{\bar{\beta}}{2-3\bar{\beta}}.$$

■

13.2.3 General constraints on squared variables

Let us consider now the quadratic optimization problems in the following form:

$$\begin{aligned} \text{find } \phi^* = &\quad \max\{\langle Qx,x\rangle :\ [x]^2 \in \mathcal{F}\}, \\ \text{find } \phi_* = &\quad \min\{\langle Qx,x\rangle :\ [x]^2 \in \mathcal{F}\}. \end{aligned} \tag{13.2.16}$$

where $\mathcal{F}$ is a closed convex set, which satisfies Assumption 13.2.1. Let us show that all results of Sections 13.2.1, 13.2.2 can be easily applied to the problem (13.2.16). Denote by $\xi(u)$ the support function of the set $\mathcal{F}$:

$$\xi(u) = \max\{\langle u,v\rangle :\ v \in \mathcal{F}\}.$$

Theorem 13.2.3 *The statements of Theorems 13.2.1, 13.2.2 are valid for the problem (13.2.16) with the relaxation values ψ^*, ψ_* and τ^*, τ_* defined as follows:*

$$\begin{aligned} \psi^* &= \min_u \{\xi(u) : \ \mathrm{Diag}\,(u) \succeq Q\}, \\ \psi_* &= \max_u \{-\xi(u) : \ Q + \mathrm{Diag}\,(u) \succeq 0)\}, \qquad (13.2.17) \\ \tau^* &= \xi(\mathrm{diag}(Q)), \quad \tau_* = -\xi(-\mathrm{diag}(Q)). \end{aligned}$$

Proof.

In order to prove the theorem we need to rewrite the problem (13.2.16) in a conic form. Note that in view of Assumption 13.2.1 the set $\mathcal{F}$ can be represented in the following form:

$$\mathcal{F} = \{v \in S : \ Bv = d\},$$

where S is a bounded convex set with non-empty interior, B is a non-degenerate $(m \times n)$-matrix and $d \in R^m$. Without loss of generality we can assume that

$$\{v \in \mathrm{int}\, S : \ Bv = d\} \neq \emptyset.$$

We allow also $B = 0$; in this case $d = 0$.

Let us consider a conic hull of the set S:

$$K = \{(v, \tau) : \ \tau > 0, \ \tfrac{1}{\tau} v \in S\} \bigcup \{0\}.$$

In view of our assumptions K is a closed convex cone with non-empty interior. The cone dual to K can be represented as follows (see, for example, [345]):

$$K^* = \{\hat{u} = (u, \mu) : \ \mu \geq \xi_S(-u)\},$$

where $\xi_S(\cdot)$ is a support function of the set S.

Now we can rewrite the problem (13.2.16) in the following form:

$$\begin{aligned} \text{find } \phi^* = &\ \max\{\langle \hat{Q}\hat{x}, \hat{x}\rangle : \ [\hat{x}]^2 \in \hat{\mathcal{F}}\}, \\ \text{find } \phi_* = &\ \min\{\langle \hat{Q}\hat{x}, \hat{x}\rangle : \ [\hat{x}]^2 \in \hat{\mathcal{F}}\}, \end{aligned} \qquad (13.2.18)$$

where $\hat{x} \in R^{n+1}$, $\hat{Q} = \begin{pmatrix} Q & 0_n \\ 0_n^T & 0 \end{pmatrix}$, $\hat{\mathcal{F}} = \{z = (v, \tau) \in K : \ \hat{A}z = b\}$ and

$$\hat{A} = \begin{pmatrix} B & -d \\ 0_n^T & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0_m \\ 1 \end{pmatrix}.$$

Note that the problems in (13.2.18) satisfy Assumptions 13.2.1, 13.2.2. Therefore for their relaxation values ψ^* and ψ_* all statements of Theorems

13.2.1, 13.2.2 are valid. Let us find the expressions for ψ^*, ψ_*, τ^* and τ_* in terms of the initial objects of the problem (13.2.16). It is clear that

$$\begin{aligned}
\tau^* &= \max\{\langle \mathrm{diag}(\hat{Q}), z\rangle :\ z \in \hat{\mathcal{F}}\}\\
&= \max\{\langle \mathrm{diag}(Q), v\rangle :\ Bv = \tau d,\ \tau = 1,\ v/\tau \in S\} = \xi(\mathrm{diag}(Q)),\\
\tau_* &= \min\{\langle \mathrm{diag}(\hat{Q}), z\rangle :\ z \in \hat{\mathcal{F}}\}\\
&= \min\{\langle \mathrm{diag}(Q), v\rangle :\ Bv = \tau d,\ \tau = 1,\ v/\tau \in S\} = -\xi(-\mathrm{diag}(Q)).
\end{aligned}$$

Further, in view of Lemma 13.2.2 the upper relaxation value ψ^* can be represented as follows:

$$\begin{aligned}
\psi^* &= \min_{\hat{y}\in R^{m+1}, \hat{u}\in R^{n+1}} \{\langle b, \hat{y}\rangle :\ \hat{Q} + \mathrm{Diag}\,(\hat{u}) \preceq \mathrm{Diag}\,(\hat{A}^T\hat{y}),\ \hat{u} \in K^*\}\\
&= \min_{(y,\gamma)\in R^{m+1}, (u,\mu)\in R^{n+1}} \{\gamma :\ Q + \mathrm{Diag}\,(u) \preceq \mathrm{Diag}\,(B^T y),\ \mu\\
&\qquad \le \gamma - \langle d, y\rangle,\ \mu \ge \xi_S(-u)\}\\
&= \min_{u,y}\{\xi_S(-u) + \langle d, y\rangle :\ \mathrm{Diag}\,(B^T y - u) \succeq Q\}\\
&= \min_{u,y}\{\xi_S(u - B^T y) + \langle d, y\rangle :\ \mathrm{Diag}\,(u) \succeq Q\}.
\end{aligned}$$

Note that in the last expression y does not enter the constraints. Therefore we can replace the objective function of this problem by its minimum in y. That is

$$\begin{aligned}
\min_y\{\xi_S(u - B^T y) + \langle d, y\rangle\} &= \min_y \max_{v\in S}\{\langle u - B^T y, v\rangle + \langle d, y\rangle\}\\
&= \max_{v\in S}\min_y\{\langle u, v\rangle + \langle d - Bv, y\rangle\}\\
&= \max_{v\in S}\{\langle u, v\rangle :\ Bv = d\} = \xi(u).
\end{aligned}$$

Thus, we get the representation (13.2.17) for ψ^*. The representation of ψ_* can be obtained in a similar way:

$$\begin{aligned}
\psi^* &= \max_{\hat{y}\in R^{m+1},\hat{u}\in R^{n+1}} \{\langle b, \hat{y}\rangle :\ \hat{Q} \succeq \text{Diag}\,(\hat{u}) + \text{Diag}\,(\hat{A}^T\hat{y}),\ \hat{u} \in K^*\}\\
&= \max_{(y,\gamma)\in R^{m+1},(u,\mu)\in R^{n+1}} \{\gamma :\ Q \succeq \text{Diag}\,(u) + \text{Diag}\,(B^T y),\\
&\qquad 0 \geq \mu + \gamma - \langle d, y\rangle,\ \mu \geq \xi_S(-u)\}\\
&= \max_{u,y}\{-\xi_S(-u) + \langle d, y\rangle :\ Q \succeq \text{Diag}\,(u + B^T y)\}\\
&= \max_{u,y}\{-\xi_S(B^T y + u) + \langle d, y\rangle :\ Q + \text{Diag}\,(u) \succeq 0\}\\
&= \max_{u}\{-\min_{y}\{\xi_S(B^T y + u) - \langle d, y\rangle\} :\ Q + \text{Diag}\,(u) \succeq 0\}\\
&= \max_{u}\{-\xi(u) :\ Q + \text{Diag}\,(u) \succeq 0\}.
\end{aligned}$$

■

Let us present an example of application of Theorems 13.2.3, 13.2.2. Consider the following problem:

$$\begin{aligned}
&\text{find } \phi^* = \quad \max\{\langle Qx_1, x_2\rangle :\ [(x_1, x_2)]^2 \in \mathcal{F}\},\\
&\text{find } \phi_* = \quad \min\{\langle Qx_1, x_2\rangle :\ [(x_1, x_2)]^2 \in \mathcal{F}\},
\end{aligned} \qquad (13.2.19)$$

where Q is a $(k \times n)$-matrix, $x_1 \in R^k$, $x_2 \in R^n$ and $\mathcal{F}$ is a closed convex set, which satisfies Assumption 13.2.1. Since the quadratic objective function in this problem is bilinear, we conclude that $\phi_* = -\phi^*$ and $\tau^* = \tau_* = 0$.

The conic relaxation for this problem is defined as follows:

$$\begin{aligned}
\psi^* &= \min_{u=(u_1,u_2)} \left\{\xi(u) :\ \begin{pmatrix} \text{Diag}\,(u_1) & -Q^T \\ -Q & \text{Diag}\,(u_2) \end{pmatrix} \succeq 0\right\},\\
\psi_* &= \max_{u=(u_1,u_2)} \left\{-\xi(u) :\ \begin{pmatrix} \text{Diag}\,(u_1) & Q^T \\ Q & \text{Diag}\,(u_2) \end{pmatrix} \succeq 0\right\}.
\end{aligned}$$

It is clear that $\psi_* = -\psi^*$. At the same time, $\beta^* = \beta_* = \frac{1}{2}$. Therefore,

$$\alpha^* = \max\{\tfrac{2}{\pi}\omega(\beta_*), 1 - \beta^*\} = \tfrac{2}{\pi}\omega(\tfrac{1}{2}).$$

Therefore, in view of Theorem 13.2.2 we have:

$$\psi^* \geq \phi^* \geq \psi(\alpha^*) = (2\alpha^* - 1)\psi^*.$$

Note that $\alpha^* = \frac{2}{\pi}(\frac{1}{2}\arcsin\frac{1}{2} + \frac{\sqrt{3}}{2}) = \frac{\sqrt{3}}{\pi} + \frac{1}{6}$. Thus, we have proved the following theorem.

Theorem 13.2.4 *In the problem (13.2.19) the optimal and relaxation values are related as follows:*

$$\psi^* \geq \phi^* \geq \gamma\psi^*$$

with $\gamma = \frac{2\sqrt{3}}{\pi} - \frac{2}{3} > 0.43$.

13.2.4 Why the linear constraints are difficult?

In the previous sections we have got a constant relative accuracy estimates for a quadratic maximization problem with convex constraints on *squared* variables. Such type of constraints are rather specific. Therefore it is natural to try to extend the results onto the problems with convex constraints on the variables of the quadratic form. However, it appears that this is not trivial. In this section we show that even a single linear constraint can make a quadratic problem completely intractable by the presented technique.

Consider the following optimization problem:

$$\begin{aligned} \phi^* = \quad & \max \quad \langle Qx, x\rangle, \\ \text{s.t.} \quad & x \in \{-1,1\}^n, \\ & \langle c, x\rangle = \beta, \end{aligned} \tag{13.2.20}$$

where Q is an $(n \times n)$-matrix, $c \in R^n$ and $\beta > 0$. Define ϕ_* as a minimal value of the objective function in (13.2.20). A natural relaxation for this problem is as follows:

$$\psi^* = \max\{\langle Q, X\rangle : \langle Xc, c\rangle = \beta^2, \ \text{diag}(X) = 1_n, \ X \succeq 0\}. \tag{13.2.21}$$

Let us show that this relaxation can be arbitrary bad in terms of relative accuracy.

Denote by v_i, $i = 1, \ldots, 2^n$ the nodes of the boolean unit box $\{-1,1\}^n$. Let us assume that there exists only one node v_*, which satisfies the linear constraint of the problem (13.2.20). Moreover, let us assume that there are two other nodes, v_+ and v_- such that

$$0 < \langle c, v_-\rangle < \beta < \langle c, v_+\rangle. \tag{13.2.22}$$

Note that in view of our assumption we have $\phi^* = \phi_*$ independently on our choice of the matrix Q.

Let us define a convex polytope $\mathcal{P}_n$ of positive semidefinite $(n \times n)$-matrices:

$$\mathcal{P}_n = \text{Conv}\,\{V_i = v_i v_i^T, \ i = 1, \ldots, 2^n\}.$$

Lemma 13.2.4 *Any V_i is an extreme point of $\mathcal{P}_n$. Any pair of nodes V_i, V_j is connected by an exposed edge.*

Proof.

Since V_i is a rank-one matrix, the first statement is evident. In order to prove the second statement note that the edge $[V_i, V_j]$ is not exposed if and only if there exist some coefficients $\lambda_k > 0$, $k \in \mathcal{I}$, $i, j \notin \mathcal{I}$ such that

$$\alpha V_i + (1-\alpha) V_j = \sum_{k \in \mathcal{I}} \lambda_k V_k, \quad \sum_{k \in \mathcal{I}} \lambda_k = 1,$$

for some $\alpha \in (0,1)$. Since all nodes of $\mathcal{P}_n$ are positive semidefinite rank-one matrices, we conclude that

$$v_k \in \{v : \, v = \alpha v_i + \beta v_j, \, (\alpha, \beta) \in R^2\}, \quad \forall k \in \mathcal{I}.$$

A simple calculation shows that it is possible only for $v_k = \pm v_i$ or $v_k = \pm v_j$. ■

Note that in view of our assumption (13.2.22) there exists a matrix $\widetilde{V} \in \mathcal{P}_n$ such that

$$\widetilde{V} = \alpha v_- v_-^T + (1-\alpha) v_+ v_+^T, \; \alpha \in (0,1), \quad \langle \widetilde{V} c, c \rangle = \beta^2.$$

Let us choose now $Q = \widetilde{V} - v_* v_*^T$. Note that the feasible set of the relaxation problem (13.2.21) contains $\mathcal{P}_n$. Therefore

$$\psi^* \geq \langle Q, \widetilde{V} \rangle > \langle Q, v_* v_*^T \rangle = \phi^*.$$

The lower relaxation value ψ_* never exceed $\phi_* = \phi^*$. Therefore, for our example the value $\psi^* - \psi_*$ is strictly positive. This means that the relative accuracy of the value ψ^* is infinitely bad.

Note that the main source of our troubles in the above example is that the linear constraint $\langle Xc, c \rangle = \beta^2$ intersects an edge of the matrix polytope $\mathcal{P}_n$. That can happen with any value of β except $\beta = 0$. Thus, we still can hope that for the problems with homogeneous linear constraints the conic relaxation can work. In the next sections we will see some problems, for which it is true.

13.2.5 Maximization with a smooth constraint

In the previous section we have established some constant bounds on relative accuracy of the conic relaxations (13.2.4) for a quadratic maximization problem with convex constraints for the squared variables. At the same time, in Section 13.2.4 we have seen that some linear constraints on the initial variables can make the problem intractable in terms of relative accuracy. In this section we present another approach for deriving the conic relaxations. This approach is based on the standard second order optimality conditions and it allows to treat the quadratic maximization problems over l_p-boxes, $p \geq 2$, with homogeneous linear equality constraints (see Section 13.2.6). However, the quality of relaxation in this framework becomes dependent on p.

Let $f(y)$, $y \in R^m$, be a homogeneous function of degree p:

$$f(\tau y) = \tau^p f(y), \quad y \in R^m, \; \tau \geq 0. \tag{13.2.23}$$

We assume that $f(y)$ is non-negative and twice continuously differentiable at any non-zero point of R^m (notation $f \in H_p$). Recall, that for homogeneous functions we have the following simple relations.

Lemma 13.2.5 *If $f(y)$ is homogeneous of degree p then for any $y \in R^m$ and $\tau \geq 0$ we have*

$$\begin{aligned} f'(\tau y) &= \tau^{p-1} f'(y), & \textit{(13.2.24)} \\ f''(y)y &= (p-1)f'(y), & \textit{(13.2.25)} \\ \langle f'(y), y\rangle &= pf(y), & \textit{(13.2.26)} \\ \langle f''(y)y, y\rangle &= p(p-1)f(y). & \textit{(13.2.27)} \end{aligned}$$

Proof.
Indeed, if we differentiate (13.2.23) in y we get (13.2.24). If we differentiate (13.2.24) in τ and take $\tau = 1$ we get (13.2.25). In order to get (13.2.26) we differentiate (13.2.23) in τ and take $\tau = 1$. Finally, (13.2.25) and (13.2.26) give (13.2.27). ■

Let Q be a symmetric $(m \times m)$-matrix. Consider the following maximization problem:

$$\text{find } \phi^*(Q) = \max\{\langle Qy, y\rangle : \ f(y) \leq 1\}. \qquad (13.2.28)$$

If $Q \preceq 0$ then (13.2.28) is a concave maximization problem and $\phi^*(Q) = 0$. In the other cases we need some necessary conditions to characterize the local solutions of the problem (13.2.28).

Lemma 13.2.6 *Let $f \in H_p$ with $p > 0$. Then for any local maximum y_* of the problem (13.2.28) with $\langle Qy_*, y_*\rangle > 0$ we have $f(y_*) = 1$. Moreover, there exists a value $\lambda = \lambda(y_*) > 0$ such that*

$$\begin{aligned} \langle Qy_*, y_*\rangle &= p\lambda, & \textit{(13.2.29)} \\ Qy_* &= \lambda f'(y_*), & \textit{(13.2.30)} \\ Q &\preceq \lambda\left[f''(y_*) - \frac{p-2}{p} f'(y_*)f'(y_*)^T\right]. & \textit{(13.2.31)} \end{aligned}$$

Proof.
Since $\langle Qy_*, y_*\rangle > 0$ and $f(y)$ is a homogeneous function of positive degree, we necessarily have $f(y_*) = 1$. Let us write down a Lagrangean for this problem:

$$\mathcal{L}(y, \lambda) = \tfrac{1}{2}\langle Qy, y\rangle - \lambda[f(y) - 1].$$

Then, the second order necessary conditions for the problem (13.2.28) can be written as follows:

$$\begin{aligned} \mathcal{L}'_y(y_*, \lambda) &= 0, & (13.2.32) \\ \langle \mathcal{L}''_{yy}(y_*, \lambda)h, h\rangle &\preceq 0, \quad \forall h : \ \langle f'(y_*), h\rangle = 0, & (13.2.33) \end{aligned}$$

with some $\lambda \in R$. Equation (13.2.32) is exactly (13.2.30). Multiplying (13.2.30) by y_* and using (13.2.26) we get

$$\langle Qy_*, y_* \rangle = \lambda \langle f'(y_*), y_* \rangle = \lambda p f(y_*) \lambda p > 0,$$

and that is (13.2.29). Finally, since $\langle f'(y_*), y_* \rangle = p > 0$, any $h \in R^m$ such that $\langle f'(y_*), h \rangle = 0$ can be represented in the form

$$h = \left(I - \tfrac{1}{p} y_* f'(y_*)^T \right) u, \quad u \in R^n.$$

Therefore the condition (13.2.33) can be rewritten as

$$(I - \tfrac{1}{p} f'(y_*) y_*^T) \mathcal{L}''_{yy}(y_*, \lambda)(I - \tfrac{1}{p} y_* f'(y_*)^T) \preceq 0. \tag{13.2.34}$$

Note that $\mathcal{L}''_{yy}(y_*, \lambda) = Q - \lambda f''(y_*)$ and

$$\begin{aligned} &(I - \tfrac{1}{p} f'(y_*) y_*^T) Q (I - \tfrac{1}{p} y_* f'(y_*)^T) \\ &= Q - \tfrac{1}{p} f'(y_*) y_*^T Q - \tfrac{1}{p} Q y_* f'(y_*)^T + \tfrac{1}{p^2} \langle Qy_*, y_* \rangle f'(y_*) f'(y_*)^T \\ &= Q - \tfrac{\lambda}{p} f'(y_*) f'(y_*)^T \end{aligned}$$

in view of (13.2.30) and (13.2.29). Similarly, since $f(y_*) = 1$ we have

$$\begin{aligned} (I- \ &\tfrac{1}{p} f'(y_*) y_*^T) f''(y_*) (I - \tfrac{1}{p} y_* f'(y_*)^T) \\ &= f''(y_*) - \tfrac{1}{p} f'(y_*) y_*^T f''(y_*) - \tfrac{1}{p} f''(y_*) y_* f'(y_*)^T \\ &+ \tfrac{1}{p^2} \langle f''(y_*) y_*, y_* \rangle f'(y_*) f'(y_*)^T \\ \\ &= f''(y_*) - 2\tfrac{p-1}{p} f'(y_*) f'(y_*)^T + \tfrac{p(p-1)}{p^2} f'(y_*) f'(y_*)^T \\ &= f''(y_*) - \tfrac{p-1}{p} f'(y_*) f'(y_*)^T, \end{aligned}$$

in view of (13.2.25) and (13.2.27). Substituting these expressions in (13.2.34) we get

$$\begin{aligned} Q &\preceq \lambda f''(y_*) + \tfrac{\lambda}{p} f'(y_*) f'(y_*)^T - \lambda \tfrac{p-1}{p} f'(y_*) f'(y_*)^T \\ &= \lambda \left[f''(y_*) - \tfrac{p-2}{p} f'(y_*) f'(y_*)^T \right]. \end{aligned}$$

$\blacksquare$

We will use Lemma 13.2.6 in order to estimate the quality of relaxations for some non-convex maximization problems. Let $A = (a_1, \ldots, a_n) \in R^{m \times n}$ be a non-degenerate $(m \times n)$-matrix. Consider the following function:

$$f_A(y) = \sum_{i=1}^{n} \mid \langle a_i, y \rangle \mid^p,$$

where $p \geq 2$. The problem we are going to address now is as follows:

$$\text{find } \phi^*(Q,A) = \max\{\langle Qy, y\rangle : \ f_A(y) \leq 1\}. \qquad (13.2.35)$$

For this problem we can introduce the following relaxation:

$$\psi_p^*(Q,A) = \min_u\{\| u \|_q\!: \ A\mathrm{Diag}\,(u)A^T \succeq Q\}, \qquad (13.2.36)$$

where $q = (\frac{p}{2})^* = \frac{p}{p-2}$ (compare with (13.2.17)). Now we can prove the main result of this section.

Theorem 13.2.5 *Let the feasible set of the problem (13.2.35) be bounded. Then*

$$\frac{1}{p-1}\psi_p^*(Q,A) \leq \phi^*(Q,A) \leq \psi_p^*(Q,A). \qquad (13.2.37)$$

Moreover, any local maximum y_* of the problem (13.2.35) with positive value of the objective function satisfies inequality $\langle Qy_*, y_*\rangle \geq \frac{1}{p-1}\psi_p^*(Q,A)$.

Proof.
Indeed, let u be feasible for the problem (13.2.36). Then for any $y \in R^m$ with $f_A(y) \leq 1$ we have

$$\langle Qy, y\rangle \leq \langle A\mathrm{Diag}\,(u)A^T y, y\rangle = \langle u, [A^T y]^2\rangle.$$

At the same time,

$$\| \, [A^T y]^2 \, \|_{p/2}^{p/2} = \sum_{i=1}^n | \langle a_i, y\rangle |^p = f_A(y) \leq 1.$$

Therefore, for any feasible y we have

$$\langle Qy, y\rangle \leq \langle u, [A^T y]^2\rangle \leq \| u \|_q \cdot \| \, [A^T y]^2 \, \|_{p/2} \leq \| u \|_q \, .$$

Hence, $\phi^*(Q,A) \leq \psi_p^*(Q,A)$.

On the other hand, let y_* be a local maximum of (13.2.35) with $\langle Qy_*, y_*\rangle > 0$. Then, in view of Lemma 13.2.6 (13.2.31) for $\lambda = \lambda(y_*)$ we have:

$$Q \preceq \lambda f''(y_*) = p(p-1)\lambda \sum_{i=1}^n | \langle a_i, y_*\rangle |^{p-2} \, a_i a_i^T$$

(we have used the condition $p \geq 2$). Thus, the vector $u \in R^n$ with the components

$$u^{(i)} = p(p-1)\lambda \mid \langle a_i, y_*\rangle \mid^{p-2}, \ i = 1, \ldots, n,$$

is feasible for the problem (13.2.36). Note that

$$\begin{aligned} \| u \|_q \ &= p(p-1)\lambda \left[\sum_{i=1}^n | \langle a_i, y_*\rangle |^{(p-2)q} \right]^{1/q} \\ &= p(p-1)\lambda \left[\sum_{i=1}^n | \langle a_i, y_*\rangle |^p \right]^{1/q} \\ &= p(p-1)\lambda [f_A(y_*)]^{1/q} = p(p-1)\lambda. \end{aligned}$$

Hence, in view of (13.2.29) we have

$$\langle Qy_*, y_* \rangle = p\lambda = \frac{\| u \|_q}{p-1} \geq \frac{1}{p-1}\psi_p^*(Q, A).$$

Note that the above proof shows that under assumptions of the theorem the function $\psi_p^*(Q, A)$ is well defined.

Finally, if there is no local maximum of the problem (13.2.35) with $\langle Qy_*, y_* \rangle > 0$, then $Q \preceq 0$ and in this case we have $\psi_p^*(Q, A) = \phi^*(Q, A) = 0$. ■

Let us estimate the relative accuracy of the relaxation (13.2.36). First, we need the following trivial result.

Lemma 13.2.7 *Let for some non-negative values ϕ, ψ and γ we have the following relations:*

$$\gamma\psi \leq \phi \leq \psi.$$

Then, for $\beta = \frac{2\gamma}{1+\gamma}$ we have: $\mid \beta\psi - \phi \mid \leq (1 - \beta)\phi$.

Define $\phi_*(Q, A) = \min\{\langle Qy, y \rangle : f_A(y) \leq 1\}$.

Theorem 13.2.6 *Let $\psi^* = \frac{2}{p}\psi_p^*(Q, A)$. Then*

$$\mid \phi^*(Q, A) - \psi^* \mid \leq (1 - \tfrac{2}{p})(\phi^*(Q, A) - \phi_*(Q, A)). \qquad \textit{(13.2.38)}$$

Proof.
Note that $\phi_*(Q, A) \leq 0$. Therefore it is sufficient to prove

$$\mid \phi^*(Q, A) - \psi^* \mid \leq (1 - \tfrac{2}{p})\phi^*(Q, A).$$

Let us choose $\gamma = \frac{1}{p-1}$ and $\beta = \frac{2\gamma}{1+\gamma} = \frac{2}{p}$. Then the above inequality follows from Theorem 13.2.5 and Lemma 13.2.7. ■

Let us compare now the relaxation (13.2.36) with the conic relaxation (13.2.17). Of course, we have to choose a problem which can be treated by both approaches. Consider the problem

$$\max\{\langle Qx, x \rangle : \| x \|_p \leq 1\}, \quad p \geq 2.$$

This problem can be presented in the form (13.2.35) with $A = I_n$. On the other hand, it can be written in the form (13.2.16) with

$$\mathcal{F} = \{v : \| v \|_{p/2} \leq 1\}.$$

In this case $\xi(u) = \| u \|_q$ and we can see that (13.2.36) coincides with (13.2.17).

13.2.6 Some applications

Let us show that the results of the previous section can be extended onto the problems with linear equality constraints. Consider the following quadratic maximization problem:

$$\begin{aligned} \text{find } \phi_p^* = \quad & \max_{x \in R^n} \quad \langle Cx, x\rangle, \\ \text{s.t.} \quad & \| x \|_p \leq 1, \\ & Bx = 0, \end{aligned} \tag{13.2.39}$$

where C is an arbitrary $(n \times n)$-matrix, $p \geq 2$ and B is a non-degenerate $((n-m) \times n)$-matrix with $n > m$. Let the rows of some $(m \times n)$-matrix A span the null space of the matrix B:

$$BA^T y = 0, \quad \forall y \in R^m.$$

Then we can change variables $x = A^T y$ and obtain a problem, which is equivalent to (13.2.39):

$$\phi_p^* = \max_{y \in R^m} \{ \langle ACA^T y, y\rangle : \; f_A(y) \leq 1 \} = \phi^*(ACA^T, A).$$

Thus, in view of Theorem 13.2.5 and Lemma 13.2.7 we get the following result.

Theorem 13.2.7 *For any $p \geq 2$ we have*

$$\frac{1}{p-1}\psi_p^*(ACA^T, A) \leq \phi_p^* \leq \psi_p^*(ACA^T, A).$$

The value $\psi^* = \frac{2}{p}\psi_p^*(ACA^T, A)$ approximates the solution of the problem (13.2.39) with $(1 - \frac{2}{p})$ relative accuracy.

Now, let us consider the case when the objective function of the problem (13.2.39) has a non-zero linear term:

$$\begin{aligned} \text{find } \hat{\phi}_p^* = \quad & \max_{x \in R^n} \quad \langle Cx, x\rangle + 2\langle c, x\rangle, \\ \text{s.t.} \quad & \| x \|_p \leq 1, \\ & Bx = 0. \end{aligned} \tag{13.2.40}$$

This problem can be homogenized in a standard way:

$$\begin{aligned} & \max_{(x,\tau) \in R^{n+1}} \quad \langle Cx, x\rangle + 2\tau\langle c, x\rangle, \\ \text{s.t.} \quad & \| x \|_p \leq 1, \; | \tau | \leq 1, \\ & Bx = 0. \end{aligned} \tag{13.2.41}$$

Clearly, the optimal value of this problem is $\hat{\phi}_p^*$. However, this problem has two separate constraints for x and τ. Therefore, in order to apply the results of Section 13.2.5 we need to replace them by a single functional inequality. Consider the following problem:

$$\begin{array}{rl} \text{find } \bar{\phi}_p^* = & \max\limits_{(x,\tau)\in R^{n+1}} \quad \langle Cx, x\rangle + 2\tau\langle c, x\rangle, \\ \text{s.t.} & \| \, (x,\tau) \, \|_p \leq 1, \\ & Bx = 0, \end{array} \qquad (13.2.42)$$

Denote by $\bar{\psi}_p^*$ the value of the conic relaxation for the last problem.

Theorem 13.2.8 *Let $p \geq 2$. For $\psi_a^* = 2^{2/p}\bar{\psi}_p^*$ we have:*

$$\frac{1}{2^{2/p}(p-1)}\psi_a^* \leq \hat{\phi}_p^* \leq \psi_a^*.$$

The value $\psi_r^* = \frac{2}{p+2^{-2/p}-1}\bar{\psi}_p^*$ has at least $(1-\frac{1}{2p})$ relative accuracy.

Proof.
Note that the problems (13.2.41) and (13.2.42) have the same objective function and the same system of linear equations. Denote by $\mathcal{F}_0$ the feasible set of the problem (13.2.42) and by $\mathcal{F}_1$ the feasible set of the problem (13.2.41). Clearly, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset 2^{1/p}\mathcal{F}_0$. Therefore

$$\bar{\phi}_p^* \leq \hat{\phi}_p^* \leq 2^{2/p}\bar{\phi}_p^*.$$

On the other hand, in view of Theorem 13.2.7, we have:

$$\frac{1}{p-1}\bar{\psi}_p^* \leq \bar{\phi}_p^* \leq \bar{\psi}_p^*.$$

Hence, for $\psi_a^* = 2^{2/p}\bar{\psi}_p^*$ we obtain:

$$\psi_a^* = 2^{2/p}\bar{\psi}_p^* \geq 2^{2/p}\bar{\phi}_p^* \geq \hat{\phi}_p^* \geq \bar{\phi}_p^* \geq \tfrac{1}{p-1}\bar{\psi}_p^* = \tfrac{1}{2^{2/p}(p-1)}\psi_a^*.$$

In order to get the statement on the relative accuracy, we take $\psi = \psi_a^*$, $\phi = \hat{\phi}_p^*$, $\gamma = \frac{1}{2^{2/p}(p-1)}$ and apply Lemma 13.2.7. Then the values β and ψ_r^* can be obtained as follow:

$$\begin{array}{rl} \beta & = \frac{2\gamma}{1+\gamma} = \frac{2}{1+2^{2/p}(p-1)} \geq \frac{1}{2p}, \\ \psi_r^* & = \beta\psi_a^* = \frac{2}{p+2^{-2/p}-1}\bar{\psi}_p^*. \end{array}$$

■

We see that the quality of conic relaxation decreases as p increase. Therefore, we cannot directly apply the results of Section 13.2.5 to a problem with box constraints. However, at the same time, when p increase the shape of l_p balls becomes very close to the shape of the n-dimensional unit box. Therefore, we can use the values $\psi_p^*(ACA^T, A)$ with p large enough in order to get some bounds for ϕ_∞^*.

Theorem 13.2.9 *Let* $p = 2\ln n$, $\psi_a^* = e\psi_p^*(ACA^T, A)$ *and* $\gamma = \frac{1}{e(2\ln n - 1)}$. *Then*

$$\gamma\psi_a^* \leq \psi_\infty^* \leq \psi_a^*.$$

The value $\psi_r^* = \frac{2\gamma}{1+\gamma}\psi_a^*$ *has at least* $(1 - \frac{1}{e\ln n})$ *relative accuracy.*

Proof.
It is well known that for any two values $p \geq 2$ we have:

$$\frac{1}{n^{1/p}} \parallel x \parallel_p \leq \parallel x \parallel_\infty \leq \parallel x \parallel_p, \quad x \in R^n.$$

Therefore

$$\{x \in R^n : \parallel x \parallel_p \leq 1\} \subset \{x \in R^n : \parallel x \parallel_\infty \leq 1\} \subset \{s \in R^n : \parallel x \parallel_p \leq n^{1/p}\}.$$

Since the objective function of the problem (13.2.39) is homogeneous of degree two, this implies that $\phi_p^* \leq \phi_\infty^* \leq n^{2/p}\phi_p^*$. Thus, using Theorem 13.2.7 we obtain the following:

$$\begin{aligned} \psi_\infty^* &= e\psi_p^*(ACA^T, A) = n^{2/p}\psi_p^*(ACA^T, A) \geq n^{2/p}\phi_p^* \\ &\geq \phi_\infty^* \geq \phi_p^* \geq \tfrac{1}{p-1}\psi_p^*(ACA^T, A) = \tfrac{1}{e(2\ln n - 1)}\psi^*. \end{aligned}$$

In order to get the statement on relative accuracy we apply Lemma 13.2.7 with

$$\beta = \frac{2\gamma}{1+\gamma} = \frac{2}{1 + e(2\ln n - 1)} > \frac{1}{e\ln n}.$$

■

13.2.7 Discussion

In the previous sections we have presented some estimates for the quality of the conic relaxation for different non-convex quadratic maximization problems. The constant bounds of Sections 13.2.1, 13.2.2 can be applied to a quite large class of non-convex problems and we can expect that they can be used in many practical applications. The bounds we get in Section 13.2.5 are not so good. Indeed, they can be applied only to a rather special feasible set, that is an intersection of an l_p-ball, $p \geq 2$, with a linear subspace. Moreover, the quality of these bounds decrease as p increase.

Nevertheless, the results of Section 13.2.5 suggest some interesting conclusions. Firstly, the relative accuracy we get from the relaxation (13.2.36) is $(1 - \frac{2}{p})$. Thus, the accuracy goes to zero as p approaches two. For p small enough the results of Theorem 13.2.5 become even better than the bounds of Section 13.2.1. An important advantage of the estimates (13.2.37) is that we get the separate bounds for the minimal and the maximal value of the problem. The lower estimate for the maximal value remains positive even if the minimal value of the problems is a large negative value.

Secondly, Theorem 13.2.5 tells us that the value of the objective function of the problem (13.2.35) at *any* local solution is not worse than the lower bound we get from the conic relaxation. In fact, this statement is a kind of surprise. Indeed, if we measure a hardness of a problem as a largest ratio of the values of the objective function at the global and a local maximum, it appears that the problem (13.2.35) is not so difficult, at least for p small enough. Usually the general methods of nonlinear optimization are quite efficient in finding a local solution. Since the computational cost of such schemes is much less than that of the schemes of semidefinite programming, we can conclude that for practical applications the traditional schemes look quite attractive.[1]

Finally, in Section 13.2.6 we have shown that the results of Theorem 13.2.5 provides us with some bounds for very difficult problems. Indeed, during last years there were obtained many negative results related to the possibilities to find an approximate solution of an NP-hard problem under hypothesis that $P \neq NP$. The results relevant to the topic of our section can be found in [79]:

> Consider a quadratic optimization problem in the following form:
>
> $$\max\{\langle Cx, x\rangle : \ Bx \leq b, \ 0 \leq x \leq 1_n\}. \tag{13.2.43}$$
>
> Denote by $\widetilde{P}$ the class of languages recognizable in quasi-polynomial time.
>
> **Theorem 1.2.** *Assume $NP \not\subseteq \widetilde{P}$. Then there is a constant $\delta > 0$ such that the problem (13.2.43) has no polynomial time, $(1 - 2^{-\log^{\delta} n})$-approximation algorithm.*
>
> **Theorem 1.3.** *Assume $P \neq NP$. Then there is a constant $\mu \in (0, \frac{1}{3})$ such that a μ-approximation of the problem (13.2.43) cannot be found in polynomial time.*

In these statements the μ-approximation is understood in a weak sense. We need to compute an estimate for the value of the objective function only.

Note that using Theorems 13.2.8 and 13.2.9, we can approximate in polynomial time the optimal value of the problem

$$\max\{\langle Cx, x\rangle : \ Bx = \tfrac{1}{2}1_n, \ 0 \leq x \leq 1_n\}. \tag{13.2.44}$$

with $(1 - O(\frac{1}{\ln n}))$ relative accuracy. This result is better than the limiting bound of Theorem 1.2 [79]. At the same time, the optimization problem, which

[1] Of course, in non-convex case we cannot prove any global efficiency estimates. Moreover, in general we cannot guarantee a convergence to a point, which satisfies the necessary second order optimality conditions. This negative result is valid even for the second order methods.

is used in the proof of Theorems 1.2, 1.3 [79], has, in fact, only linear equalities constraint:

$$\max\{\langle Cx, x\rangle : \ Bx = b, \ 0 \leq x \leq 1_n\}. \tag{13.2.45}$$

Thus, the difference in the formulations (13.2.44) and (13.2.45) looks very minor. Indeed, any system of linear equations $Bx = b$ can be rewritten in the following form:

$$\bar{B}x = \tfrac{1}{2}1_n, \quad \langle a, x\rangle = 1,$$

with some matrix $\bar{B}$ and a vector $a \in R^n$. Hence, the feasible set of the problem (13.2.45) differs from the feasible set of the problem (13.2.44) just by a single linear equation, which does not pass through the center of the box. However, it appears that this linear equation makes the problem (13.2.45) completely different.

Let us look at the concrete form of the problem (13.2.45) ([79], p.438). Denote by X and Y two $(n \times n)$-matrices. And let $\phi(X, Y)$ be a bilinear form in X and Y with all non-negative coefficients. Then the problem (13.2.45) is as follows:

$$\begin{aligned} &\max \ \phi(X, Y), \\ \text{s.t.} \quad & X1_n = 1_n, \ Y1_n = 1_n, \\ & 0 \leq X, Y \leq 1_{n \times n}. \end{aligned} \tag{13.2.46}$$

Now we can see the source of our troubles. Indeed, the technique of Section 13.2.5 can be applied only to l_p boxes with $p \geq 2$. However, if we will try to approximate the feasible set of the problem (13.2.46) with the boxes $\mathcal{B}_p = \{x : \| \ x - \frac{1}{2}1_n \ \|_p \leq \frac{1}{2}\}$, we need to choose p very large. It is necessary to take $p = O(n \ln n)$ just to have a non-empty intersection of the box $\mathcal{B}_p$ with the system of linear constraints in (13.2.46).

Thus, we conclude that the feasible set of the problem (13.2.46) is too far from the center of the box. On the other hand, it is clear that the box structure in (13.2.46) is quite artificial: the constraint X, $Y \leq 1_{n \times n}$ can be eliminated without changing the feasible set of the problem. Note that we can easily rewrite the problem (13.2.46) in a more symmetric form:

$$\begin{aligned} &\max \ \phi(X, Y), \\ \text{s.t.} \quad & \| \ Xe_i \ \|_1 \leq 1, \ i = 1, \ldots n, \\ & \| \ Ye_i \ \|_1 \leq 1, \ i = 1, \ldots n. \end{aligned} \tag{13.2.47}$$

Since the coefficients of the form $\phi(X, Y)$ are non-negative, the optimal value of the problem (13.2.47) is the same as that of (13.2.43). The polyhedral structure of the feasible set in (13.2.47) can be seen as a combination of l_∞-structure with l_1-structure. However, it appears the latter structure is exactly that one, for which no reasonable bounds for quadratic problems are known.

Thus, the above discussion highlights the following unsolved problem:

Find some bounds for the optimal value of the following quadratic problem:

$$\phi^* = \max\{\langle Qx, x\rangle : \| x \|_p \leq 1,\ x \in R^n\}, \quad 1 \leq p < 2. \tag{13.2.48}$$

For an indefinite Q a trivial bound for ϕ^* is given by its maximal eigenvalue $\lambda_{\max}(Q)$:

$$\lambda_{\max}(Q) \geq \phi^* \geq \lambda_{\max}(Q) \cdot n^{1-\frac{2}{p}}, \quad 1 \leq p \leq 2.$$

For $p = 1$ we can suggest for the problem (13.2.48) a kind of semidefinite relaxation:

$$\begin{aligned} \psi^* &= \max_{X,u}\{\langle Q, X\rangle :\ \mathrm{Diag}\,(u) \succeq X,\ \langle 1_n, u\rangle \leq 1,\ X \succeq 0\} \\ &= \min_{S,\lambda}\{\lambda :\ \lambda 1_n = \mathrm{diag}(S),\ S \succeq Q,\ S \succeq 0\}. \end{aligned} \tag{13.2.49}$$

Note that for any x, $\| x \|_1 \leq 1$, the pair $(X = xx^T, u = \mathrm{abs}[x])$ is feasible for the primal form of the relaxation (13.2.49). Therefore we can guarantee that $\psi^* \geq \phi^*$. However, the relative accuracy of such a bound is not known.

13.3 QUADRATIC CONSTRAINTS

Yinyu Ye

Consider the quadratic programming (QP) problem with diagonally quadratic equality and inequality constraints

$$(QP) \qquad \begin{array}{rll} \bar{q}(Q) := & \text{Maximize} & q(x) := x^T Q x \\ & \text{Subject to} & \sum_{j=1}^n a_{ij} x_j^2 = b_i,\ i = 1, \ldots, m, \\ & & \sum_{j=1}^n c_{ij} x_j^2 \leq d_i,\ i = 1, \ldots, p \end{array}$$

where the symmetric matrix $Q \in \mathcal{S}^n$, $A = \{a_{ij}\} \in \mathcal{M}_{m,n}$, $C = \{c_{ij}\} \in \mathcal{M}_{p,n}$, $b \in \Re^m$, and $d \in \Re^p$ are given. We assume that the QP problem is feasible and its feasible set is bounded (this can be checked by a linear program considering x_j^2 as nonnegative variables). Let $\bar{x}(Q)$ be a maximizer of the problem.

The (QP) problem has applications in combinatorial and global optimization problems, see, e.g., Gibbons et al. [273]. Note that this quadratic problem includes the max-cut problem by letting $x_j^2 = 1,\ j = 1, ..., n$, be the quadratic constraints. Also note that perturbing the diagonal of Q may change the objective function on the feasible set of the problem.

Normally, there is a linear term in the objective function:

$$\begin{array}{ll} \text{Maximize} & x^T Qx + c^T x \\ \\ \text{Subject to} & \sum_{j=1}^n a_{ij} x_j^2 = b_i,\ i = 1, \ldots, m, \\ & \sum_{j=1}^n c_{ij} x_j^2 \leq d_i,\ i = 1, \ldots, p \end{array}$$

However, the problem can be homogenized as

$$\begin{array}{ll} \text{Maximize} & x^T Qx + tc^T x \\ \\ \text{Subject to} & \sum_{j=1}^n a_{ij} x_j^2 = b_i,\ i = 1, \ldots, m,\ t^2 = 1, \\ & \sum_{j=1}^n c_{ij} x_j^2 \leq d_i,\ i = 1, \ldots, p \end{array}$$

by adding a scalar variable t. There always is an optimal solution $(\bar{x}, \bar{t})$ for this problem in which $\bar{t} = 1$ or $\bar{t} = -1$. If $\bar{t} = 1$, then $\bar{x}$ is also optimal for the non-homogeneous problem; if $\bar{t} = -1$, then $-\bar{x}$ is optimal for the non-homogeneous problem. Thus, without loss of generality, we can let $q(x) = x^T Qx$ throughout this Section 13.3.

The function $q(x)$ has a minimizer and a maximizer over the bounded feasible set

$$\mathcal{F} := \{x \in \Re^n : \sum_{j=1}^n a_{ij} x_j^2 = b_i,\ i = 1, \ldots, m,\ \sum_{j=1}^n c_{ij} x_j^2 \leq d_i,\ i = 1, \ldots, p\}.$$

Let $\underline{q} := -\bar{q}(-Q)$ and $\bar{q} := \bar{q}(Q)$ denote their minimal and maximal objective values, respectively. An ϵ-maximal solution or ϵ-maximizer, $\epsilon \in [0, 1]$, for (QP) is defined as an $x \in \mathcal{F}$ such that

$$\frac{\bar{q} - q(x)}{\bar{q} - \underline{q}} \leq \epsilon.$$

Recently, there were several significant results on approximating specific quadratic problems. Goemans and Williamson [285] (also see Frieze and Jerrum [255]) proved an approximation result for the Maxcut problem where $\epsilon \leq 1 - 0.878$ when all arc weights are nonnegative. Nesterov [572] generalized their result to approximating a boolean QP problem

$$\begin{array}{ll} \text{Maximize} & q(x) = x^T Qx \\ \\ \text{Subject to} & x_j^2 = 1,\ j = 1, \ldots, n, \end{array}$$

where $\epsilon \leq 4/7$. Ye [859] extended the 4/7 result to solving continuous nonconvex QP problems, such as,

$$\begin{array}{ll} \text{Maximize} & q(x) = x^T Qx \\ \\ \text{Subject to} & x_j^2 \leq 1, \; j = 1, \ldots, n. \end{array}$$

Note that some negative results on this problem were given by Bellare and Rogaway [79]. Other results can be found in Fu, Luo and Ye [258], Pardalos and Rosen [619], Vavasis [818], and Ye [855].

In this Section 13.3, we, based on the analyses of Ye [859] and Nesterov [575], further generalize the 4/7 result to approximating (QP) containing (diagonally) quadratic constraints. These constraints have added a few difficulties in analyzing the problem, and they frequently appear in some practical applications.

13.3.1 Positive Semi-Definite Relaxation

The approximation algorithm for (QP) is to solve a positive semi-definite programming (SDP) relaxation problem

$$(SDP) \quad \begin{array}{lll} \bar{p}(Q) := & \text{Maximize} & \langle Q, X \rangle \\ & \text{Subject to} & \langle D(a_i), X \rangle = b_i, \; i = 1, \ldots, m, \\ & & \langle D(c_i), X \rangle \leq d_i, \; i = 1, \ldots, p. \end{array} \qquad (13.3.50)$$

Here, $a_i = (a_{i1}, \ldots, a_{in}) \in \Re^n$, $c_i = (c_{i1}, \ldots, c_{in}) \in \Re^n$, and unknown $X \in \Re^{n \times n}$ is a symmetric matrix. Furthermore, $\langle \cdot, \cdot \rangle$ is the matrix inner product $\langle Q, X \rangle = \text{trace}(Q^T X)$, $D(a)$ is the diagonal matrix of vector a, and $X \succeq Z$ means that $X - Z$ is positive semi-definite. Since the original QP problem is feasible and bounded, so is the SDP relaxation.

The dual of the problem is

$$\begin{array}{lll} \bar{p}(Q) = & \text{Minimize} & d^T z + b^T y \\ \\ & \text{Subject to} & \sum_{i=1}^p z_i D(c_i) + \sum_{i=1}^m y_i D(a_i) \succeq Q, \; z \geq 0. \end{array} \qquad (13.3.51)$$

Note that the primal is feasible and bounded and the dual has an interior so that there is no duality gap between the primal and dual. Denote by $\bar{X}(Q)$ and $(\bar{y}(Q), \bar{z}(Q))$ an optimal solution pair for the primal (13.3.50) and dual (13.3.51).

The positive semi-definite relaxation was first proposed by Lovász and Shrijver [497], also see recent papers by Alizadeh [17], Fujie and Kojima [260] and Polijak, Rendl and Wolkowicz [635]. This relaxation problem pair can be solved in polynomial time, e.g., see Nesterov and Nemirovskii [583] and Alizadeh [17].

We have the following relations between (QP) and (SDP) from Ye [859].

Proposition 13.3.1 *Let $\bar{q} = \bar{q}(Q)$, $\underline{q} = -\bar{q}(-Q)$, $\bar{p} = \bar{p}(Q)$, $\underline{p} = -\bar{p}(-Q)$, and*
$(\underline{y}, \underline{z}) = (-\bar{y}(-Q), -\bar{z}(-Q))$. Then, $\underline{q}$ is the minimal objective value of x^TQx in the feasible set of (QP) and $\underline{p} = d^T\underline{z} + b^T\underline{y}$ is the minimal objective value of $\langle Q, X \rangle$ in the feasible set of (SDP). Furthermore,

$$\underline{p} = -\bar{p}(-Q) \leq \underline{q} = -\bar{q}(-Q) \leq \bar{q}(Q) = \bar{q} \leq \bar{p}(Q) = \bar{p}.$$

In what follows, we let $\bar{x} = \bar{x}(Q)$, $\bar{X} = \bar{X}(Q)$. Since $\bar{X}$ is positive semidefinite, there is a factorization matrix $\bar{V} = (\bar{v}_1, \ldots, \bar{v}_n) \in \Re^{n \times n}$, i.e., $\bar{v}_j$ is the jth column of $\bar{V}$, such that $\bar{X} = \bar{V}^T\bar{V}$. The algorithm (Goemans and Williamson [285], Nesterov [572], and Ye [859]) generates a random vector u uniformly distributed on the n-dimensional unit ball and then assigns

$$\hat{x} = \bar{D}\sigma(\bar{V}^T u), \tag{13.3.52}$$

where

$$\bar{D} = \mathrm{diag}(\|\bar{v}_1\|, \ldots, \|\bar{v}_n\|) = \mathrm{diag}(\sqrt{\bar{x}_{11}}, \ldots, \sqrt{\bar{x}_{nn}}),$$

and for any $x \in \Re^n$, $\sigma(x)$ is the vector whose components are $\mathrm{sign}(x_j)$, $j = 1, \ldots, n$, that is,

$$\mathrm{sign}(x_j) = \begin{cases} 1 & \text{if } x_j \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$

It is easily seen that $\hat{x}$ is a feasible point for (QP) and we will show later that the expected objective value, $\mathrm{E}_u q(\hat{x})$, satisfies

$$\frac{\bar{q} - \mathrm{E}_u q(\hat{x})}{\bar{q} - \underline{q}} \leq \frac{\pi}{2} - 1 \leq \frac{4}{7}.$$

13.3.2 Approximation Analysis

The following lemma is an analogue to the lemma of Nesterov [572] and Ye [859].

Lemma 13.3.1 *Let u be uniformly distributed on the n-dimensional unit ball. Then,*

$$\begin{aligned} \bar{q}(Q) = \quad & \textit{Maximize} && \mathrm{E}_u(\sigma(V^T u)^T DQD\sigma(V^T u)) \\ & \textit{Subject to} && \langle D(a_i), V^T V \rangle = b_i, \; i = 1, \ldots, m, \\ & && \langle D(c_i), V^T V \rangle \leq d_i, \; i = 1, \ldots, p, \end{aligned}$$

where

$$D = \mathit{diag}(\|v_1\|, \ldots, \|v_n\|).$$

Proof. Since, for any feasible V, $D\sigma(V^T u)$ is a feasible point for (QP), we have

$$\bar{q}(Q) \geq \mathrm{E}_u(\sigma(V^T u)^T DQD\sigma(V^T u)).$$

On the other hand, for any fixed u with $||u|| = 1$, we have

$$\mathrm{E}_u(\sigma(V^Tu)^T DQD\sigma(V^Tu)) = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}||v_i||||v_j||\mathrm{E}_u(\sigma(v_i^Tu)\sigma(v_j^Tu)). \tag{13.3.53}$$

Let us choose $v_i = \frac{\bar{x}_i}{||\bar{x}||}\bar{x}$, $i = 1, \ldots, n$. (Note that V is feasible for the problem above.) Then

$$\mathrm{E}_u(\sigma(v_i^Tu)\sigma(v_j^Tu)) = \begin{cases} 1 & \text{if } \sigma(\bar{x}_i) = \sigma(\bar{x}_j) \\ -1 & \text{otherwise.} \end{cases}$$

Thus,

$$||v_i||||v_j||\mathrm{E}_u(\sigma(v_i^Tu)\sigma(v_j^Tu)) = \bar{x}_i\bar{x}_j$$

which implies that for this particular feasible V

$$\bar{q}(Q) = q(\bar{x}) \le \mathrm{E}_u(\sigma(V^Tu)^T DQD\sigma(V^Tu)).$$

These two relations give the desired result. ■

For any function of one variable $f(t)$ and $X \in \Re^{n\times n}$, let $f[X] \in \Re^{n\times n}$ be the matrix with the components $f(x_{ij})$. Nesterov [572] has proved the next technical lemma.

Lemma 13.3.2 *Let $X \succeq 0$ and $d(X) \le 1$. Then* $\arcsin[X] \succeq X$. ■

Now we are ready to prove the following theorem.

Theorem 13.3.1

$$\begin{array}{rll} \bar{q}(Q) = & \sup & \frac{2}{\pi}\langle Q, D\arcsin[D^{-1}XD^{-1}]D\rangle \\ & \textit{Subject to} & \langle D(a_i), X\rangle = b_i,\ i = 1, \ldots, m, \\ & & \langle D(c_i), X\rangle \le d_i,\ i = 1, \ldots, p, \\ & & X \succ 0, \end{array}$$

where

$$D = \mathrm{Diag}\,(\sqrt{x_{11}}, \ldots, \sqrt{x_{nn}}).$$

Proof. For any $X = V^TV \succ 0$, we have

$$\begin{aligned} \mathrm{E}_u(\sigma(v_i^Tu)\sigma(v_j^Tu)) &= 1 - 2\mathrm{Pr}\{\sigma(v_i^Tu) \ne \sigma(v_j^Tu)\} \\ &= 1 - 2\mathrm{Pr}\{\sigma(\tfrac{v_i^Tu}{||v_i||}) \ne \sigma(\tfrac{v_j^Tu}{||v_j||})\}. \end{aligned}$$

From Lemma 1.2 of Goemans and Williamson [285], we have

$$\mathrm{Pr}\{\sigma(\frac{v_i^Tu}{||v_i||}) \ne \sigma(\frac{v_j^Tu}{||v_j||})\} = \frac{1}{\pi}\arccos(\frac{v_i^Tv_j}{||v_i||||v_j||}).$$

Using the above lemma and equality (13.3.53) and noting $\arcsin(t)+\arccos(t) = \frac{\pi}{2}$ give the desired result. ■

We have used *Supremum* and $X \succ 0$ in the problem above merely for the technical presentation of D^{-1}. The feasible set of this problem can be closed if we rewrite it in terms of variable $Y = D^{-1}XD^{-1}$

Theorem 13.3.1 leads to our main result.

Theorem 13.3.2 *We have*

1.

$$\bar{q} - \underline{p} \ge \frac{2}{\pi}(\bar{p} - \underline{p}).$$

2.

$$\bar{p} - \underline{q} \ge \frac{2}{\pi}(\bar{p} - \underline{p}).$$

3.

$$\bar{p} - \underline{p} \ge \bar{q} - \underline{q} \ge \frac{4-\pi}{\pi}(\bar{p} - \underline{p}).$$

Proof. Recall $\underline{z} = -\bar{z}(-Q) \le 0$, $\underline{p} = -\bar{p}(-Q) = d^T\underline{z} + b^T\underline{y}$, and

$$Q - \sum_{i=1}^{p} \underline{z}_i D(c_i) - \sum_{i=1}^{m} \underline{y}_i D(a_i) \succeq 0.$$

Thus, for any $X \succ 0$ feasible for (SDP), and $D = \operatorname{diag}(\sqrt{x_{11}}, \ldots, \sqrt{x_{nn}})$, we have from Theorem 13.3.1

$$\begin{aligned} \tfrac{\pi}{2}\bar{q} &= \tfrac{\pi}{2}\bar{q}(Q) \\ &\ge \langle Q, D\arcsin[D^{-1}XD^{-1}]D\rangle \\ &= \Bigg\langle Q - \sum_{i=1}^{p}\underline{z}_i D(c_i) - \sum_{i=1}^{m}\underline{y}_i D(a_i) + \sum_{i=1}^{p}\underline{z}_i D(c_i) \\ &\qquad + \sum_{i=1}^{m}\underline{y}_i D(a_i), D\arcsin[D^{-1}XD^{-1}]D \Bigg\rangle \end{aligned}$$

$$
\begin{aligned}
&= \left\langle Q - \sum_{i=1}^{p} \underline{z}_i D(c_i) - \sum_{i=1}^{m} \underline{y}_i D(a_i),\, D \arcsin[D^{-1} X D^{-1}] D \right\rangle \\
&+ \left\langle \sum_{i=1}^{p} \underline{z}_i D(c_i) + \sum_{i=1}^{m} \underline{y}_i D(a_i),\, D \arcsin[D^{-1} X D^{-1}] D \right\rangle \\
&\geq \left\langle Q - \sum_{i=1}^{p} \underline{z}_i D(c_i) - \sum_{i=1}^{m} \underline{y}_i D(a_i),\, D D^{-1} X D^{-1} D \right\rangle \\
&+ \left\langle \sum_{i=1}^{p} \underline{z}_i D(c_i) + \sum_{i=1}^{m} \underline{y}_i D(a_i),\, D \arcsin[D^{-1} X D^{-1}] D \right\rangle \\
&\Bigg(\text{since } Q - \sum_{i=1}^{p} \underline{z}_i D(c_i) - \sum_{i=1}^{m} \underline{y}_i D(a_i) \succeq 0 \\
&\quad \text{and } \arcsin[D^{-1} X D^{-1}] \succeq D^{-1} X D^{-1} \Big) \\
&= \left\langle Q - \sum_{i=1}^{p} \underline{z}_i D(c_i) - \sum_{i=1}^{m} \underline{y}_i D(a_i),\, X \right\rangle \\
&+ \left\langle \sum_{i=1}^{p} \underline{z}_i D(c_i) + \sum_{i=1}^{m} \underline{y}_i D(a_i),\, D \arcsin[D^{-1} X D^{-1}] D \right\rangle \\
&= \langle Q, X \rangle - \left\langle \sum_{i=1}^{p} \underline{z}_i D(c_i) + \sum_{i=1}^{m} \underline{y}_i D(a_i),\, X \right\rangle \\
&+ \left\langle \sum_{i=1}^{p} \underline{z}_i D(c_i) + \sum_{i=1}^{m} \underline{y}_i D(a_i),\, D \arcsin[D^{-1} X D^{-1}] D \right\rangle \\
&= \langle Q, X \rangle - \sum_{i=1}^{p} \underline{z}_i \langle D(c_i), X \rangle - \sum_{i=1}^{m} \underline{y}_i \langle D(a_i), X \rangle \\
&+ \sum_{i=1}^{p} \underline{z}_i \langle D(c_i), D \arcsin[D^{-1} X D^{-1}] D \rangle \\
&\qquad + \sum_{i=1}^{m} \underline{y}_i \langle D(a_i), D \arcsin[D^{-1} X D^{-1}] D \rangle \\
&= \langle Q, X \rangle - \sum_{i=1}^{p} \underline{z}_i \langle D(c_i), X \rangle - \underline{y}^T b + \sum_{i=1}^{p} \underline{z}_i \left(\frac{\pi}{2} \langle D(c_i), X \rangle\right) + \underline{y}^T \left(\frac{\pi}{2} b\right) \\
&= \langle Q, X \rangle + \left(\tfrac{\pi}{2} - 1\right) \sum_{i=1}^{p} \underline{z}_i \langle D(c_i), X \rangle + \left(\frac{\pi}{2} - 1\right) \underline{y}^T b \\
&\geq \langle Q, X \rangle + \left(\tfrac{\pi}{2} - 1\right) (\underline{z}^T d + \underline{y}^T b) \\
&(\text{since } \langle D(c_i), X \rangle \leq d_i \;\; i = 1, \ldots, p, \text{ and } \underline{z} \leq 0) \\
&= \langle Q, X \rangle + \left(\tfrac{\pi}{2} - 1\right) \underline{p}.
\end{aligned}
$$

Let X converge to $\bar{X}$, then $\langle Q, X \rangle \to \bar{p}$ and we have the desired first inequality.

Replacing Q with $-Q$ proves the second inequality in the theorem.

Adding the first two inequalities gives the third statement in the theorem. ■

The result indicates that the positive semi-definite relaxation value $\bar{p} - \underline{p}$ is a constant approximation of $\bar{q} - \underline{q}$.

Similarly, the following corollary can be devised.

Corollary 13.3.1 *Let $X = V^T V \succ 0$, $\langle D(a_i), X\rangle \le d_i$ $(i = 1, \ldots, p)$, $\langle D(a_i), X\rangle = b_i$ $(i = 1, \ldots, m)$, $D = diag(\sqrt{x_{11}}, \ldots, \sqrt{x_{nn}})$, and $\hat{x} = D\sigma(V^T u)$ where u with $\|u\| = 1$ is a random vector uniformly distributed on the unit ball. Moreover, let $X \to \bar{X}$. Then,*

$$\lim_{X \to \bar{X}} \mathrm{E}_u(q(\hat{x})) = \lim_{X \to \bar{X}} \frac{2}{\pi}\langle Q, D \arcsin[D^{-1} X D^{-1}] D\rangle \ge \frac{2}{\pi}\bar{p} + (1 - \frac{2}{\pi})\underline{p}.$$

Finally, we have

Theorem 13.3.3 *Let $\hat{x}$ be generated above from $X = \bar{X}$. Then*

$$\frac{\bar{q} - \mathrm{E}_u q(\hat{x})}{\bar{q} - \underline{q}} \le \frac{\pi}{2} - 1.$$

Proof. The proof is similar to that in Nesterov [572] and Ye [859]. We include it here for completeness. Since

$$\bar{p} \ge \bar{q} \ge \frac{2}{\pi}\bar{p} + (1 - \frac{2}{\pi})\underline{p} \ge (1 - \frac{2}{\pi})\bar{p} + \frac{2}{\pi}\underline{p} \ge \underline{q} \ge \underline{p}$$

we have

$$\begin{aligned}
\frac{\bar{q} - \mathrm{E}_u q(\hat{x})}{\bar{q} - \underline{q}} &\le \frac{\bar{q} - \frac{2}{\pi}\bar{p} - (1 - \frac{2}{\pi})\underline{p}}{\bar{q} - \underline{q}} \\
&\le \frac{\bar{q} - \frac{2}{\pi}\bar{p} - (1 - \frac{2}{\pi})\underline{p}}{\bar{q} - (1 - \frac{2}{\pi})\bar{p} - \frac{2}{\pi}\underline{p}} \\
&\le \frac{\bar{p} - \frac{2}{\pi}\bar{p} - (1 - \frac{2}{\pi})\underline{p}}{\bar{p} - (1 - \frac{2}{\pi})\bar{p} - \frac{2}{\pi}\underline{p}} \\
&= \frac{(1 - \frac{2}{\pi})(\bar{p} - \underline{p})}{\frac{2}{\pi}(\bar{p} - \underline{p})} \\
&= \frac{(1 - \frac{2}{\pi})}{\frac{2}{\pi}} = \frac{\pi}{2} - 1.
\end{aligned}$$

■

13.3.3 Results for Other Quadratic Problems

Consider now another nonconvex QP problem:

$$\begin{array}{ll}\text{Maximize} & x^T Q x + c^T x \\ \\ \text{Subject to} & x^T A_i x + c_i^T x \leq b_i, \;\; i = 1, \ldots, m,\end{array}$$

where given symmetric matrices $A_i \in \Re^{n \times n}$. We summarize approximation results for solving this problem.

- If $m = 1$, $A_1 = I$, the identity matrix, and $c_1 = 0$, then the problem is polynomially solvable. That is, there is an algorithm to generate an ϵ-solution for any $\epsilon > 0$, and its running time is polynomial in n and $\log(1/\epsilon)$, see an early proof by Vavasis [818] and Ye [855] and a later by Rendl and Wolkowicz [661]. (Ye [856] further reduced the complexity time dependency on ϵ to $\log\log(1/\epsilon)$.)

- If all A_i are mutually commutative (they can be simultaneously diagonalized) and all $c_i = 0$, then the problem can be transformed into a problem with only diagonally quadratic constraints, and thus can be approximated for $\epsilon = 4/7$ according to our early analysis, also see Ye [859] and Nesterov [575].

- If all A_i are positive semidefinite, then the problem can be approximated for $\epsilon = 1 - \frac{constant}{m^2}$ by Fu et al. [258]; and in addition, if all $c_i = 0$, then it can be approximated for $\epsilon = 1 - \frac{constant}{\log(mn)}$ by Nemirovskii et al. [568].

13.4 RELAXATIONS OF Q^{2}P

Henry Wolkowicz

In this part of the chapter we look at several different instances of Q^{2}P. In particular, we start with several different tractable relaxations for the max-cut problem and show that, surprisingly, they are all equal to the Lagrangian (**and** SDP) relaxation.

We then illustrate a recipe for constructing relaxations for QQPs by finding a strengthened SDP bound for the max-cut problem.

Other instances discussed are the quadratic assignment and graph partitioning problems.

We then consider trust region type problems and discuss when strong duality holds. This includes problems where orthogonal constraints arise, e.g. orthogonal relaxations of the quadratic assignment and graph partitioning problems. In particular, this part of the chapter emphasizes the theme about the strength of the Lagrangian relaxation.

13.4.1 Relaxations for the Max-cut Problem

The success of the SDP relaxation (equivalently Lagrangian relaxation) over the last few years is exemplified by the success on the *Max-Cut Problem.* Let $G = (V, E)$ be an undirected graph with edge set $V = \{v_i\}_{i=1}^n$ and weights w_{ij} on the edges $(v_i, v_j) \in E$. We want to find the index set $\mathcal{I} \subset \{1, 2, \ldots n\}$, to maximize the weight of the edges with one end point with index in $\mathcal{I}$ and the other in the complement. This is equivalent to

$$(MC) \qquad \max\ \tfrac{1}{2}\textstyle\sum_{i<j} w_{ij}(1 - x_i x_j), \quad x \in \mathcal{F},$$

where $\mathcal{F} := \{\pm 1\}^n$, and $x_i = 1$ if $i \in \mathcal{I}$ and -1 otherwise. The objective function is a (homogeneous) quadratic form, x^TQx.

Several *Different* Relaxations. We now look at several different tractable relaxations of MCQ, (13.4.54). These have different motivations. For example, one bound relaxes the constraints to the unit ball of radius $\sqrt{n}$, while another relaxes the constraints to the convex hull, i.e. to the unit cube. Following [638, 635], we observe that several quadratic type bounds considered in the literature are actually equal. The key to the simple proofs is the strong duality result for the trust region subproblem, see [747]. A similar phenomenon occurs for linearizations of (P), such as in *roof duality,* see e.g. [324], where many bounds obtained from various linearizations have been shown to be equal and, in fact, they have been shown to be equal to the Lagrangian dual of a linearized problem, see [4]. (The quality of the SDP bounds is the main topic in the first two parts of this chapter; see above.)

We allow a more general objective function, i.e. we consider the ±1 constrained quadratic program

$$(MCQ) \qquad \mu^* := \max_{x \in \mathcal{F}} q_0(x) \quad (:= x^TQx - 2c^Tx). \tag{13.4.54}$$

The bounds are derived using the fact that we can perturb the objective function q_0 and exploit the fact that $x_i^2 = 1$ on the feasible set $\mathcal{F}$. Note that

$$\begin{array}{rcl} q_u(x) & := & x^T(Q + \mathrm{Diag}\,(u))x - 2c^Tx - u^Te \\ & = & q_0(x), \quad \forall x \in \mathcal{F}. \end{array} \tag{13.4.55}$$

For each u we get a trivial upper bound obtained from ignoring the constraints and allowing the diagonal perturbations, i.e. we have

$$\mu^* \leq f_0(u) := \max_x q_u(x). \tag{13.4.56}$$

But, the function f_0 can take on the value $+\infty$. Let

$$S := \left\{u : u^T e = 0, Q + \text{Diag}\,(u) \preceq 0\right\}.$$

We then get the following trivial bound.

$$\mu^* \leq B_0 := \min_u f_0(u) \qquad \left(= \min_{u^T e=0} f_0(u), \text{ if } S \neq \emptyset\right). \tag{13.4.57}$$

Note that if the set S is not empty, then we can minimize over the unconstrained parameter u or add the restriction to $u^T e = 0$. This can be seen from the optimality conditions for min-max problems. This comment is true for the following bounds as well. (Details can be found in [638].)

In addition we can restrict the parameters and avoid infinite values for the inner maximization problem by adding the hidden semidefinite constraint, i.e. we use the fact that a quadratic function is unbounded above if the Hessian is not negative semidefinite. (Note that a quadratic function is bounded above if and only if the Hessian is negative semidefinite and the stationarity equation is consistent.) The following is a tractable bound since we minimize a convex function over a convex set.

$$\mu^* \leq B_0 = \min_{Q+\text{Diag}\,(u)\preceq 0} f_0(u). \tag{13.4.58}$$

Next we relax the feasible set to the sphere of radius $\sqrt{n}$. We get

$$\mu^* \leq f_1(u) := \max_{||x||^2=n} q_u(x). \tag{13.4.59}$$

And our next bound is

$$\mu^* \leq B_1 := \min_u f_1(u). \tag{13.4.60}$$

The inner maximization problem is the trust region subproblem and is tractable, see e.g. Section 13.4.3 below. Thus we have our second tractable bound.

We can replace the spherical constraint with the box constraint.

$$\mu^* \leq f_2(u) := \max_{|x_i|\leq 1} q_u(x). \tag{13.4.61}$$

After adding the semidefinite constraint to make the bound tractable, i.e. to make the calculation of f_2 tractable, we get our next bounds.

$$\mu^* \leq \min_u f_2(u) \tag{13.4.62}$$

and

$$\mu^* \leq B_2 := \min_{Q+\text{Diag}\,(u)\preceq 0} f_2(u). \tag{13.4.63}$$

Given Q and c, define the $(n+1)\times(n+1)$-matrix Q^c by adding a $0-th$ row and column, so that

$$\begin{array}{ll} q^c_{00} = 0 & \\ q^c_{0i} = q^c_{i0} = -c_i & \text{for } i > 0 \\ q^c_{ij} = q_{ij} & \text{for } i,j > 0, \end{array}$$

i.e.

$$Q^c := \begin{bmatrix} 0 & -c^T \\ -c & Q \end{bmatrix}. \qquad (13.4.64)$$

In order to have analogous functions $q^c_u(y)$ and $f_i(u)$ as in the previous cases, let us introduce

$$q^c_u(y) := y^T(Q^c + \operatorname{diag}(u))y - u^T e. \qquad (13.4.65)$$

Note that q^c_u reduces to q_u if the first component y_0 is ± 1. The equivalent relaxed problem is

$$\mu^* \leq f^c_1(u) := \max_{||y||^2=n+1} q^c_u(y) = (n+1)\lambda_{\max}(Q^c + \operatorname{diag}(u)) - u^T e, \quad (13.4.66)$$

where $\lambda_{\max}$ denotes the maximum eigenvalue. Now another bound is

$$\mu^* \leq B^c_1 := \min_u f^c_1(u). \qquad (13.4.67)$$

Similarly, we get equivalent bounds B^c_0 and homogenized bounds for the other models.

The above argument shows that we can homogenize the problem by moving into a higher dimension. Therefore, we can consider the special case that $c = 0$. We now look at the SDP bound, see also Section 13.2 above for the performance guarantees. The relaxation comes from the fact that the trace is commutative, i.e.

$$x^T Qx = \operatorname{Trace} x^T Qx = \operatorname{Trace} Qxx^T$$

and, for $x \in \mathcal{F}$, $y_{ij} = x_i x_j$ defines a symmetric, rank one, positive semidefinite matrix Y with diagonal elements 1. Therefore, we can *lift* the problem into the higher dimensional space of symmetric matrices and relax the rank one constraint. This yields the following relaxation and our bound 3.

$$\begin{array}{rcl} B_3 := & \max & \operatorname{Trace} QY \\ & \text{subject to} & \operatorname{diag}(Y) = e \\ & & Y \succeq 0. \end{array} \qquad (13.4.68)$$

This SDP is a convex programming problem and is tractable.

Now we replace the ± 1 constraints with $x_i^2 = 1, \forall i$. This does not change the feasible set of the original problem. In [638, 635] it is shown that all the above relaxations and bounds for MC come from the Lagrangian dual of (P_E), the following equivalent problem to MCQ. Thus we enforce our theme about the strength of the Lagrangian relaxation. The strong duality result for the trust region subproblem is the key to the proofs.

$$\begin{array}{lll} (P_E) & \max & q_0(x) = x^TQx - 2c^Tx \\ & \text{subject to} & x_i^2 = 1, \quad i = 1, \cdots, n. \end{array} \tag{13.4.69}$$

Note that the Lagrangian dual of P_E yields precisely our trivial first bound B_0 in (13.4.57).

Theorem 13.4.1 *All the bounds for MCQ discussed above are equal to the optimal value of the Lagrangian dual of the equivalent program P_E.*

A Strengthened Bound for MC. From the results above, it would appear that we might have the strongest possible tractable bound. However, adding redundant constraints can strengthen bounds. The following bound is motivated by the strong duality results presented in Section 13.4.3 below and is presented in [41]. The SDP bound (13.4.68) for MCQ arises from a lifting procedure, i.e. identifying

$$0 \preceq X = xx^T \text{ and } x^TQx = \operatorname{Trace} X.$$

Discarding the rank one condition on X results in the tractable SDP bound. It is not clear what constraints one can add to P_E in order to strengthen the Lagrangian relaxation, i.e. linear combinations of the constraints will not help since they are already included in the Lagrangian. But, in the space of matrices, it is also true that

$$X^2 = xx^Txx^T = nX.$$

Therefore we can use the following equivalent quadratic matrix model for MCQ.

$$\begin{array}{rll} \mu^* := & \max & \operatorname{Trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X^2 - nX = 0, \end{array}$$

where X is a symmetric matrix. This problem is equivalent to P_E since $X^2 = nX$ and $\operatorname{Trace} X = n$ implies X is rank one. Therefore we are including the rank one information from the original problem. However, this problem is a nonconvex problem and cannot be solved in general. Note that if $X^2 = nX$, then $\operatorname{Trace} QX = (1/n)\operatorname{Trace} QX^2$, and $\operatorname{diag}(X^2) = ne$. As a result, the above quadratic model is equivalent to the model:

$$\begin{array}{rll} \mu^* = & \max & \frac{1}{n}\operatorname{Trace} QX^2 \\ & \text{s.t.} & x_i^Tx_i = n, \quad i = 1, \ldots, n \\ & & X^2 - nx_0X = 0 \\ & & x_0^2 = 1, \end{array} \tag{13.4.70}$$

where x_i^T, $i = 1, \ldots, n$ denotes the ith row of X, and x_0 is a scalar. Having a quadratic objective is an advantage only if it results in a larger class of available

Lagrange multipliers. Therefore, the sign of the eigenvalues of Q will determine whether this objective or $\frac{1}{n}\text{Trace}\,QXx_0$ is better. (Note that if $x_0 = -1$ then changing x_0 to 1 and replacing X with $-X$ leaves the objective and constraints in (13.4.70) unchanged.) We will obtain an upper bound $\mu_2 \geq \mu^*$ by applying a Lagrangian procedure to all of the constraints in (13.4.70). Using multipliers u_i for the constraints $x_i^T x_i = n$, $i = 1, \ldots, n$, u_0 for the constraint $x_0^2 = 1$, and a symmetric matrix S for the matrix equality $X^2 - nX = 0$, we obtain a Lagrangian problem

$$\begin{array}{rl}\mu_2 := \min_{u_0,u,S} & u_0 + nu^T e + \max_{x_0,X} \frac{1}{n}\text{Trace}\,QX^2 - \text{Trace}\,UX^2 \\ & \qquad + \text{Trace}\,SX^2 - nx_0\text{Trace}\,SX - u_0x_0^2,\end{array}$$

where $U = \text{Diag}\,(u)$. Letting $\bar{x}^T = (x_0, \text{vec}\,(X)^T)$, this problem can be written in Kronecker product form as

$$\mu_2 = \min_{u_0,u,S} \; u_0 + ne^T u + \max_{\bar{x}} \; \bar{x}^T \bar{Q} \bar{x},$$

where

$$\bar{Q} = \begin{pmatrix} -u_0 & -\frac{n}{2}\text{vec}\,(S)^T \\ -\frac{n}{2}\text{vec}\,(S) & I \otimes \left(\frac{1}{n}Q - U + S\right) \end{pmatrix}.$$

Applying the hidden semidefinite constraint $\bar{Q} \preceq 0$, we obtain an equivalent problem

$$\begin{array}{rll}\mu_2 = & \min & u_0 + ne^T u \\ & \text{s.t.} & \begin{pmatrix} u_0 & \frac{n}{2}\text{vec}\,(S)^T \\ \frac{n}{2}\text{vec}\,(S) & I \otimes \left(-\frac{1}{n}Q + U - S\right) \end{pmatrix} \succeq 0 \\ & & S = S^T. \end{array} \qquad (13.4.71)$$

Note that if we take $S = 0$ in (13.4.71), then $u_0 = 0$ is clearly optimal, and the problem reduces to

$$\begin{array}{rcl} \min & e^T u & \\ & -Q + U & \succeq \; 0, \end{array}$$

which is exactly the dual of the usual SDP relaxation for MC. It follows that we have obtained an upper bound μ_2 which is a strengthening of the usual SDP bound, i.e. $\mu_2 \leq B_0$.

Alternative Strengthened Relaxation. This presents an alternative strengthened SDP relaxation for the max-cut problem, i.e. this continues from the above Section 13.4.1 but tries to fully exploit the rank-one condition in the Lagrangian.

We use the notation: For $S \in \mathcal{S}^n$, the vector $s = \text{svec}\,(S) \in \Re^{t(n)}$, is formed (columnwise) from S while ignoring the strictly lower triangular part of S. Its inverse is the operator $S = \text{sMat}\,(s)$. The adjoint of svec is the operator hMat (v) which forms a symmetric matrix where the off-diagonal terms are multiplied by a half, i.e. this satisfies

$$\text{svec}\,(S)^T v = \text{Trace}\, S\, \text{hMat}\,(v), \qquad \forall S \in \mathcal{S}^n, v \in \Re^{t(n)},$$

where $t(n) = n(n+1)/2$. The adjoint of sMat is the operator dsvec (S) which works like svec except that the off diagonal elements are multiplied by 2, i.e. this satisfies

$$\text{dsvec}\,(S)^T v = \text{Trace}\, S\, \text{sMat}\,(v), \qquad \forall S \in \mathcal{S}^n, v \in \Re^{t(n)}.$$

For notational convenience, we define the vectors sdiag $(s) :=$ diag(sMat (s)) and vsMat $(s) =$ vec (sMat (s)); the adjoint of vsMat is then given by

$$\text{vsMat}^*(s) = \text{dsvec}\,\left(\left(\text{Mat}\,(v) + \text{Mat}\,(v)^T\right)/2\right).$$

As above we can start with the following equivalent program

$$\text{MC}_\text{O} \qquad \begin{array}{rcc} \mu^* = & \max & \frac{1}{2}\text{Trace}\, QX \\ & \text{s.t.}¡ & \text{diag}(X) = e \\ & & X \circ X = e \\ & & X^2 - nX = 0. \end{array} \qquad (13.4.72)$$

There are many redundant constraints. However, it is uncertain which of these become redundant in the SDP relaxation. The recipe is to throw in redundant constraints; then take the Lagrangian dual twice and delete redundant constraints at the end. At the end one has an SDP with linear constraints and one can often remove the redundancy using the structure of the problem. This illustrates the strength of the Lagrangian relaxation approach. This is done in [34]. (See also [635] and more recently [473].) The result after deleting redundant constraints is the simplified SDP relaxation (see [34]):

$$\text{MCPSDP2} \qquad \begin{array}{rcl} \nu_2^* = & \max & \text{Trace}\, H_c Y \\ & \text{s.t.} & \text{diag}(Y) = e \\ & & Y_{0,t(i)} = 1, \quad \forall i = 1, \ldots, n \\ & & \sum\limits_{k=1}^{i} Y_{t(i-1)+k,t(j-1)+k} + \sum\limits_{k=i+1}^{j} Y_{t(k-1)+i,t(j-1)+i} \\ & & \quad + \sum_{k=j+1}^{n} Y_{t(i-1)+i,t(k-1)+j} - nY_{0,t(j-1)+i} = 0 \\ & & \qquad \forall 1 \le i < j \le n \\ & & Y \succeq 0, Y \in S^{t(n)+1}. \end{array}$$

(13.4.73)

This problem has $2t(n)-1$ constraints. In fact, there is still some redundancy as it can be shown that Slater's constraint qualification fails for this problem. This can be further exploited by projecting the problem onto the space determined by the *minimal face* of the problem, see [34].

13.4.2 General Q^2P

We now move on to applying the Lagrangian relaxation to *general* quadratic constrained quadratic problems, denoted Q^{2}P ; and, we apply it to several specific instances: the quadratic assignment, graph partitioning, max-clique problems. The general Q^{2}P problem is also studied in e.g. [260, 442] and [652, 451, 449, 510].

Quadratic bounds using a Lagrangian relaxation have been extensively studied and applied in the literature, for example in [444] and, more recently, in [445]. The latter calls the Lagrangian relaxation the "best convex bound". Discussions on Lagrangian relaxation for nonconvex programs also appear in [245]. More references are given throughout this chapter.

Remark 13.4.1 *Any equality constraints are written as two inequality constraints; any linear equality constraints, $Ax = b$, is transformed to a quadratic constraint via $||Ax - b||^2 = 0$. The reason for these transformations for linear equality constraints is discussed in [635], i.e. the Lagrangian dual essentially ignores linear constraints as can be seen from: $-\infty = \max_\lambda \min_x -x^2 + \lambda x$, which is the dual of the problem* $\min\{-x^2 : x = 0\}$.

We now recall the Q^{2}P in x.

$$(\mathrm{Q}^2\mathrm{P}_x) \qquad \begin{array}{rll} q^* := & \min & q_0(x) := x^TQ_0x + 2g_0^Tx + \alpha_0 \\ & \text{subject to} & q_k(x) := x^TQ_kx + 2g_k^Tx + \alpha_k \le 0 \\ & & k \in \mathcal{I} := \{1, \ldots, m\} \\ & & x \in \Re^n, \end{array} \qquad (13.4.74)$$

where the matrices Q_k are symmetric. The *feasible set* is

$$F_x := \{x \in \Re^n : q_k(x) \le 0, \forall k \in \mathcal{I}\}.$$

(Note that though the feasible set F_x may be empty, the feasible set of the relaxation may not be.) The objective function and the constraints are not convex, necessarily. Therefore the feasible set can be a very "nasty" set. This problem is a very hard problem to solve in general, see e.g. [614].

Let

$$P_k := \begin{bmatrix} \alpha_k & g_k^T \\ g_k & Q_k \end{bmatrix} \qquad (13.4.75)$$

and, by abuse of notation, define

$$q_k(y) := y^TP_ky, \;\; k = 0, 1, \ldots, m.$$

Then an equivalent homogenized formulation to (Q^2P_x) is

$$(Q^2P_y) \qquad \begin{array}{rll} q^* = & \min & q_0(y) \\ & \text{subject to} & q_k(y) \le 0, k \in \mathcal{I} \\ & & y_0^2 = 1 \\ & & y = \begin{pmatrix} y_0 \\ x \end{pmatrix} \in \Re^{n+1}. \end{array}$$

It is clear that the optimal values of the two equivalent formulations are equal. In fact, if $y_0 = -1$ is optimal, then we can replace y by $-y$. This is because the objective function and all but the last constraint are homogeneous.

We will refer to both equivalent formulations of Q^2P in the sequel. The correct reference will be clear from the context.

Remark 13.4.2 *Note that we could replace the constraint $y_0^2 = 1$ by $y_0 = 1$. (The constraint $y_0 = 1$ is used in [260].) In the latter case, the feasible sets of the two formulations coincide exactly, while in the former case they can differ by a sign, i.e. $x \in F_x$ implies that both $\begin{pmatrix} -1 \\ -x \end{pmatrix}$ and $\begin{pmatrix} 1 \\ x \end{pmatrix}$ are in F_y, i.e. are feasible for the homogenized problem Q^2P_y.*

The Lagrangian Relaxation of a General Q^{2}P. The Lagrangian relaxation of the homogenized problem Q^{2}P$_y$ provides a simple technique for obtaining the SDP relaxation. In addition, an application of the strong duality result for the trust region subproblem shows that both the SDP and Lagrangian relaxation are equal. The Lagrangian of Q^{2}P$_y$ is

$$L(y,\mu,\lambda) := y^TP_0y - \mu(y_0^2-1) + \sum_{k\in\mathcal{I}} \lambda_k y^TP_ky.$$

The Lagrangian relaxation of Q^{2}P$_y$ is

$$(DQ^2P_y) \qquad d^* := \max_{\substack{\mu \\ \lambda\geq 0}} \min_y y^TP_0y - \mu(y_0^2-1) + \sum_{k\in\mathcal{I}} \lambda_k y^TP_ky.$$

Note that

$$\begin{aligned} d^* &= \max_{\lambda\geq 0}\max_{\mu}\min_y y^TP_0y - \mu(y_0^2-1) + \sum_{k\in\mathcal{I}} \lambda_k y^TP_ky \\ &= \max_{\lambda\geq 0}\min_{y_0^2=1} y^TP_0y + \sum_{k\in\mathcal{I}} \lambda_k y^TP_ky, \end{aligned}$$

from strong duality of the trust region subproblem, see [747]. Therefore, we get equivalence of the dual values for the problems in x and in y. (This is similar to the approaches in [849, 733].)

$$(DQ^2P_x) \qquad d^* = \max_{\lambda\geq 0}\min_x q_0(x) + \sum_{k\in\mathcal{I}} \lambda_k q_k(x).$$

We immediately conclude that *weak duality* holds

$$d^* \leq q^* = \min_y \max_{\substack{\mu \\ \lambda\geq 0}} y^TP_0y - \mu(y_0^2-1) + \sum_{k\in\mathcal{I}} \lambda_k y^TP_ky.$$

Therefore, if the optimal μ^*, λ^* can be found, we have found a single quadratic function whose minimal value approximates the original minimal value q^*, i.e.

$$q^* \geq d^* = \min_y y^TP_0y - \mu^*(y_0^2-1) + \sum_{k\in\mathcal{I}} \lambda_k^* y^TP_ky. \qquad (13.4.76)$$

Moreover, in the dual program, the Lagrangian is a quadratic function of y. Therefore, the outer maximization problem has the nonnegativity and an additional hidden semidefinite constraint

$$P_0 - \mu E_{00} + \sum_{k \in \mathcal{I}} \lambda_k P_k \succeq 0, \quad \lambda \geq 0, \tag{13.4.77}$$

where E_{00} is the zero matrix with 1 in the top left corner and $M \succeq 0$ denotes the Löwner partial order, i.e. that the symmetric matrix M is positive semidefinite. The minimum of the minimization subproblem, in this case, is attained by $y = 0$. Therefore the Lagrangian dual is equivalent to the SDP problem

$$(DSDP) \qquad \begin{array}{rll} d^* := & \max & \mu \\ & \text{subject to} & \mu E_{00} - \sum_{k \in \mathcal{I}} \lambda_k P_k \preceq P_0 \\ & & \lambda \geq 0. \end{array}$$

Valid Inequalities. Using the above approach we see that more constraints $q_k(y)$ means that we have a stronger dual. This can be phrased as adding redundant constraints to get new valid inequalities to strengthen the relaxation. We will see how this occurs when we look at orthogonally constrained problems below. Another approach is also specified in detail in Kojima and Tuncel [442, 440].

For problems that also have linear equality constraints, one can use the notion of copositivity to strengthen the SDP relaxation. However, this does not result in a tractable relaxation in general, see [649].

Specific Instances of SDP Relaxation. We now study four specific instances and show how to apply the recipe for relaxations. In each case we derive a min-max eigenvalue problem from the Lagrangian dual of an appropriately chosen quadratic constrained program. The dual of this dual problem provides a semidefinite relaxation for the original problem. Adding redundant constraints at the start helps in reducing the duality gap. These redundant constraints are automatically deleted at the end, i.e. in the SDP relaxation, by ensuring full row rank and Slater's condition. We do this for: the quadratic assignment problem; graph partitioning; max-clique problem; and the stable set problem.

Quadratic Assignment Problem

Typical relaxations for QAP, see the definition in Section 13.1, try to exploit the trace formulation and use perturbations on A, B separately. Current approaches have two serious drawbacks. They completely discard the nonnegativity constraints and then they derive a bound from the sum of two bounds obtained by treating the quadratic and linear parts of the objective function separately, see e.g. [610]. However, the Lagrangian relaxations and homogenization for the special case $S = \Re^n$ shows that we should consider more general perturbations and, in particular, we should consider perturbations that arise from Lagrangian quadratic relaxations. This approach does not have the two drawbacks mentioned above.

We now use the fact that the set of permutation matrices is equal to the intersection of the orthogonal matrices with the 0,1 matrices. We get the following equivalent program to QAP.

$$\begin{array}{llc} & \mu^* := \quad \max & q(X) = \operatorname{Trace}(AXB-2C)X^T \\ (QAP_E) & \quad\text{subject to} & XX^T = I \\ & & X_{ij}^2 - X_{ij} = 0, \quad \forall i,j. \end{array} \qquad (13.4.78)$$

We could also consider the square of the norm of the residual of the (redundant) linear constraints

$$Xe = e, \; X^T e = e.$$

Other relaxations and bounds can be obtained by adding redundant constraints such as

$$\operatorname{Trace} XX^T = n, \quad X^T X = I,$$

or

$$0 \leq X_{ij} \leq 1, \; \forall i,j.$$

We now devote our attention to homogenization since that results in a min-max eigenvalue problem and an equivalent semidefinite programming problem. We have seen that we can homogenize by increasing the dimension of the problem by 1. We first add the 0,1 constraints to the objective function using Lagrange multipliers W_{ij}.

$$\min_{W} \max_{XX^T=I} \operatorname{Trace}(AXB-2C)X^T + \sum_{ij} W_{ij}(X_{ij}^2 - X_{ij}). \qquad (13.4.79)$$

We now homogenize the objective function by multiplying by a constrained scalar x.

$$\min_{W} \max_{XX^T=I, x^2=1} \operatorname{Trace}\left[AXBX^T + W(X \circ X)^T - x(2C+W)X^T\right]. \quad (13.4.80)$$

We can now use Lagrange multipliers to get a parametrized min-max eigenvalue problem in dimension n^2+1. We get the following bound. The parameters are: the symmetric $n \times n$ matrix $\Lambda = \Lambda^T$, the general $n \times n$ matrix W and the scalar α.

$$\begin{array}{rcl} B_{QAP} & := & \min_{\Lambda,W,\alpha} \max_{X} \operatorname{Trace}\,[\\ & & AXBX^T + \Lambda XX^T + W^T(X \circ X) + \alpha x^2 \\ & & -x(2C+W)X^T \;] - \alpha - \operatorname{Trace}\Lambda. \end{array} \qquad (13.4.81)$$

We have grouped the quadratic, original linear, and constant terms together. The hidden semidefinite constraint now yields a semidefinite programming problem.

$$\begin{array}{lc} \min & -\operatorname{Trace}\Lambda - \alpha \\ \text{subject to} & L_Q + \operatorname{Arrow}(\alpha, \operatorname{vec}(W)) + \mathrm{B}^0\mathrm{Diag}(\Lambda) \preceq 0, \end{array} \qquad (13.4.82)$$

where we define the matrix

$$L_Q := \begin{bmatrix} 0 & -\text{vec}\,(C)^T \\ -\text{vec}\,(C) & B \otimes A \end{bmatrix}, \tag{13.4.83}$$

and the linear operators

$$\text{Arrow}\,(\alpha, \text{vec}\,(W)) := \begin{bmatrix} \alpha & -\frac{1}{2}\text{vec}\,(W)^T \\ -\frac{1}{2}\text{vec}\,(W) & \text{Diag}\,(\text{vec}\,(W)) \end{bmatrix}, \tag{13.4.84}$$

$$\text{B}^0\text{Diag}\,(\Lambda) := \begin{bmatrix} 0 & 0 \\ 0 & I \otimes \Lambda \end{bmatrix}. \tag{13.4.85}$$

We can now introduce the $(n^2+1) \times (n^2+1)$ dual variable matrix $Y \succeq 0$ and derive the dual program to this min-max eigenvalue problem, i.e.

$$\max_{Y \succeq 0} \min_{\Lambda, W, \alpha} -\text{Trace}\,\Lambda - \alpha + \text{Trace}\,Y(L_Q + \text{Arrow}\,(\alpha, \text{vec}\,(W)) + \text{B}^0\text{Diag}\,(\Lambda)).$$

The inner minimization problem is unconstrained and linear in the variables. Therefore, after reorganizing the variables, we can differentiate to get the dual problem to this dual problem, or the semidefinite relaxation to the original QAP. (Recall that $Y_{i,j:k}$ refers to the i-th row and columns j to k of the matrix Y; and $\text{b}^0\text{diag}\,(Y)$ is the block diagonal sum of Y which ignores the first row.) The derivatives with respect to α and W yields the first constraint and the derivative with respect to Λ yields the second constraint in the following program. Equivalently, the constraints are the adjoints of the linear operators Arrow and B^0Diag.

$$\begin{array}{cc} \max & \text{Trace}\,L_Q Y \\ \text{subject to} & \text{diag}(Y) = (1, Y_{0,1:n^2})^T \\ & \text{b}^0\text{diag}\,(Y) = I \\ & Y \succeq 0. \end{array} \tag{13.4.86}$$

Another primal-dual pair can be obtained using a trust region subproblem as the inner maximization problem, rather than homogenizing to an eigenvalue problem. This is done by adding the redundant trust region constraint $\text{Trace}\,XX^T = n$. Also, as mentioned above, we can add the redundant constraint

$$||Xe - e||^2 + ||X^Te - e||^2 = 0.$$

This type of constraint is discussed below for the graph partitioning problem. A primal-dual interior point method based on the these types of dual pairs of programs, such as (13.4.86),(13.4.82), are being tested and studied in [870].

Graph Partitioning

Let $G = (V, E)$ be an undirected graph as in the description for (MC). The graph partitioning problem is the problem of partitioning the node set V into k disjoint subsets of specified sizes so as to minimize the total weight of the edges connecting nodes in distinct subsets of the partition. Let $A = (a_{ij})$ be the weighted adjacency matrix of G, i.e.

$$a_{ij} = \begin{cases} w_{ij} & ij \in E \\ 0 & \text{otherwise.} \end{cases}$$

The graph partitioning problem can be described by the following (0,1)-quadratic program see e.g. [660].

$$(\mathbf{GP}) \quad \begin{array}{cc} w(E_{uncut}) = \max & \frac{1}{2}\mathrm{Trace}\, X^t A X \\ \text{subject to} & X e_k = e_n \\ & X^T e_n = m \\ & X_{ij} \in \{0,1\}, \ \forall ij, \end{array}$$

where e_k is the vector of ones of appropriate size and m is the vector of ordered set sizes

$$m_1 \geq \ldots \geq m_k \geq 1 \quad \text{and} \quad k < n.$$

The columns of the 0,1 $n \times k$ matrices X are the indicator vectors for the sets. We can replace the 0,1 constraints by quadratic and also change the linear constraints to quadratic by squaring. We get the following equivalent program.

$$\begin{array}{cc} w(E_{uncut}) = \max & \frac{1}{2}\mathrm{Trace}\, X^t A X \\ \text{subject to} & ||Xe_k - e_n||^2 + ||X^T e_n - m||^2 = 0 \\ & X_{ij}^2 - X_{ij} = 0, \ \forall ij. \end{array}$$

The Lagrangian relaxation yields the following bound.

$$\begin{array}{rcl} B_{GP} & := & \min_{\alpha, W} \max_{X} \mathrm{Trace}\, [\\ & & \frac{1}{2} X^T A X + \alpha(e_k e_k^T X^T X + X^T e_n e_n^T X) + W^T(X \circ X) \\ & & -2\alpha(e_k e_n^T X + m e_n^T X) - W^T X \,] \\ & & +\alpha(n + \sum_i m_i^2). \end{array} \tag{13.4.87}$$

We can now homogenize the problem by adding a variable x.

$$\begin{array}{rcl} B_{GP} & := & \min_{\alpha, W} \max_{\substack{X \\ x^2=1}} \mathrm{Trace}\, [\\ & & \frac{1}{2} X^T A X + \alpha(e_k e_k^T X^T X + X^T e_n e_n^T X) + W^T(X \circ X) \\ & & +x(-2\alpha(e_k e_n^T X + m e_n^T X) - W^T X) \,] \\ & & +\alpha(n + \sum_i m_i^2). \end{array}$$

We now lift the variable x into the Lagrangian to get a min-max eigenvalue problem.

$$\begin{array}{rcl} B_{GP} & := & \min_{\alpha, W, \delta} \max_{X, x} \mathrm{Trace}\, [\\ & & \frac{1}{2} X^T A X + \alpha(e_k e_k^T X^T X + X^T e_n e_n^T X) + W^T(X \circ X) + \delta x^2 \\ & & +x(-2\alpha(e_k e_n^T X + m e_n^T X) - W^T X) \,] \\ & & +\alpha(n + \sum_i m_i^2) - \delta. \end{array}$$

The above has a hidden semidefinite constraint.

$$\begin{array}{cc} \min & \alpha(n+\sum_i m_i^2)-\delta \\ \text{subject to} & L_A+\operatorname{Arrow}(\delta,\operatorname{vec}(W))+\alpha L_\alpha \preceq 0, \end{array} \tag{13.4.88}$$

where we define the matrices

$$L_A := \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2}I\otimes A \end{bmatrix}, \tag{13.4.89}$$

$$v = \operatorname{vec} e_n m^T,$$

$$L_\alpha := \begin{bmatrix} 0 & -(e+v)^T \\ -(e+v) & (e_k e_k^T I\otimes I + I\otimes e_n e_n^T) \end{bmatrix}, \tag{13.4.90}$$

and the linear operator

$$\operatorname{Arrow}(\delta,\operatorname{vec}(W)) := \begin{bmatrix} \delta & -\frac{1}{2}(\operatorname{vec}(W))^T \\ -\frac{1}{2}(\operatorname{vec}(W)) & \operatorname{Diag}(\operatorname{vec}(W)) \end{bmatrix}. \tag{13.4.91}$$

The dual program yields the semidefinite relaxation of (GP).

$$\begin{array}{cc} \max & \operatorname{Trace} L_A Y \\ \text{subject to} & \operatorname{diag}(Y)=(1,Y_{0,1:n})^T \\ & \operatorname{Trace} Y L_\alpha = 0 \\ & Y\succeq 0. \end{array} \tag{13.4.92}$$

Max-Clique and Stable Set

Consider again the undirected graph $G=(E,V)$ defined above. The *max-clique* problem consists in finding the largest connected subgraph. We let $\omega(G)$ denote the size of the largest clique in G. A *stable set* is a subset of vertices of V such that no two vertices are adjacent. We denote the size of the largest stable set in $\bar{G}$, the complement of G, by $\alpha(\bar{G})$. Clearly

$$\alpha(\bar{G})=\omega(G).$$

Bounds for these problems and relationships to the theta function, or Lovász number of the graph, are described in the expository paper e.g. [425]; see also [701].

In this section we show that the Lovasz bound on $\omega(G)$ can be alternatively obtained from two distinct 01-programs (13.4.93) and (13.4.96) by Lagrangian relaxations. Let A be the incidence matrix of the graph, i.e. $A=(a_{ij})$ with $a_{ij}=1$ if $ij\in E$ and 0 otherwise. If x is the indicator vector for the largest clique in G of size k, A then $x^T(I+A)x/x^Tx=k^2/k=k$. A quadratic formulation of the max-clique problem is the following (0,1)-quadratic program.

$$\begin{array}{ccl} \omega(G)= & \max & \frac{x^T(I+A)x}{x^Tx} \\ & \text{subject to} & x_ix_j=0,\ \text{if } ij\notin E,\ i\neq j \\ & & x_i\in\{0,1\},\ \forall i. \end{array} \tag{13.4.93}$$

Therefore, a quadratic relaxation of the max-clique problem is the following quadratic constrained program.

$$\begin{array}{rcl} \omega(G) \leq \omega_1^* := & \max & x^T(I+A)x \\ & \text{subject to} & x_i x_j = 0, \text{ if } ij \notin E, \ i \neq j \\ & & x^T x = 1. \end{array} \tag{13.4.94}$$

The Lagrangian relaxation for this problem is the perturbed min-max eigenvalue problem and the equivalent semidefinite program

$$\begin{array}{rcll} \omega_1^* & \leq & \min\limits_{W_{ij}=0, \text{ if } ij \in E, \text{ or } i=j} & \lambda_{\max}(I+A+W) - \alpha x^T x + \alpha \\ & = & \min\limits_{w,\alpha} \max\limits_{x} x^T(I+A)x + \sum\limits_{ij \notin E, \ i \neq j} w_{ij} x_i x_j - \alpha x^T x + \alpha & \\ & = & \min\limits_{\substack{I+A+W \preceq \alpha I \\ W_{ij}=0, \text{ if } ij \in E, \text{ or } i=j}} & \alpha \end{array}$$

i.e. minimize the max eigenvalue over perturbations in the off-diagonal elements corresponding to disjoint nodes. This bound is equal to the Lovasz theta function on the complementary graph.

$$\vartheta(\bar{G}) = \min_{A \in \mathcal{A}} \lambda_{\max}(A), \tag{13.4.95}$$

where $\mathcal{A} = \{A : A \text{ symmetric } \ n \times n \text{ matrix with } \ A_{ij} = 1, \text{ if } ij \in E, \text{ or } i = j\}$.

By considering the (optimal) indicator vector for the largest clique, we see that a (0,1)-quadratic program that describes the max-clique problem exactly is the following one. Note that if node i is not in the largest clique, then necessarily, $x_i x_j = 0$ for some j with node j in the clique, i.e. necessarily $x_i = 0$ in the indicator vector.

$$\begin{array}{rcl} \omega(G) = & \max & x^T x \\ & \text{subject to} & x_i x_j = 0, \text{ if } ij \notin E, \ i \neq j \\ & & x_i^2 - x_i = 0, \ \forall i. \end{array} \tag{13.4.96}$$

The Lagrangian relaxation yields the bound

$$B_{\text{clique}} := \min_{W,\lambda} \max_{x} x^T x + \sum_{ij \notin E, \ i \neq j} w_{ij} x_i x_j + \sum_i \lambda_i (x_i^2 - x_i).$$

We let W be an $n \times n$ matrix with zeros in positions where $ij \in E$. We can homogenize by adding the constraint $y^2 = 1$ and then lifting it into the Lagrangian.

$$\min_{\alpha,W,\lambda} \max_{x,y} x^T x + \sum_{ij \notin E} w_{ij} x_i x_j + \sum_i \lambda_i x_i^2 + \alpha y^2 - y \sum_i \lambda_i x_i - \alpha.$$

We now exploit the hidden semidefinite constraint to get the semidefinite program.

$$\begin{array}{rcc} B_{\text{clique}} = & \min\limits_{W,\lambda,\alpha} & -\alpha \\ & \text{subject to} & L_A + L_W(W) + \text{Arrow}\,(\alpha, \lambda) \preceq 0 \\ & & W_{ij} = 0, \ \forall ij \in E, \ \text{ or } i = j, \end{array} \tag{13.4.97}$$

where the matrix

$$L_A := \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}, \tag{13.4.98}$$

and the linear operators

$$L_W(W) := \begin{bmatrix} 0 & 0 \\ 0 & W \end{bmatrix}, \tag{13.4.99}$$

$$\operatorname{Arrow}(\alpha, \lambda) := \begin{bmatrix} \alpha & -\frac{1}{2}\lambda^T \\ -\frac{1}{2}\lambda & \operatorname{Diag}(\lambda) \end{bmatrix}. \tag{13.4.100}$$

The dual of the above min-max eigenvalue problem yields the semidefinite relaxation for the max-clique problem with $Y \in \mathcal{S}_{n+1}$.

$$\begin{array}{cc} \max & \operatorname{Trace} L_A Y \\ \text{subject to} & \operatorname{diag}(Y) = (1, Y_{0,1:n})^T \\ & Y_{ij} = 0, \ \forall ij \notin E \\ & Y \succeq 0. \end{array} \tag{13.4.101}$$

The equivalence of the bounds (13.4.95) and (13.4.101) was shown in lemma 2.17 of [497].

Consider the program (13.4.93) with an additional redundant constraint

$$x_i x_j \geq 0 \text{ for } ij \in E \tag{13.4.102}$$

That is

$$\begin{array}{lcl} \omega(G) = & \max & \frac{x^T(I+A)x}{x^T x} \\ & \text{subject to} & x_i x_j = 0, \ \text{if } ij \notin E, \ i \neq j \\ & & x_i x_j \geq 0, \ \text{if } ij \in E, \\ & & x_i \in \{0, 1\}, \ \forall i. \end{array} \tag{13.4.103}$$

A quadratic relaxation of the max-clique problem is the following quadratic constrained program.

$$\begin{array}{lcl} \omega(G) \leq \omega_1^* := & \max & x^T(I+A)x \\ & \text{subject to} & x_i x_j = 0, \ \text{if } ij \notin E, \ i \neq j \\ & & x_i x_j \geq 0, \ \text{if } ij \in E, \\ & & x^T x = 1. \end{array} \tag{13.4.104}$$

The Lagrangian relaxation for this problem is equal to the Schrijver's improvement [701] of the theta function on the complementary graph.

$$\vartheta'(\bar{G}) = \min_{A \in \mathcal{A}'} \lambda_{\max}(A),$$

where $\mathcal{A}' = \{A : A$ symmetric $n \times n$ matrix with $A_{ij} \geq 1$, if $ij \in E$, or $i = j\}$. Haemmers [321] constructed graphs where $\vartheta'(\bar{G})$ is strictly smaller than $\vartheta(\bar{G})$.

Analogously, it is possible to modify the program (13.4.96) by adding the constraint (13.4.102).

13.4.3 Strong Duality

In the case of strong duality (zero duality gap and dual attainment), our bounds are exact. As expected, this holds (generically) in the convex case. Surprisingly, there are several cases on nonconvex quadratic programs where this holds as well. In this Section 13.4.3 we amplify on our theme that illustrates the strength of the Lagrangian relaxation, i.e. that a tractable bound implies a Lagrangian relaxation is at work.

Recall the general quadratically constrained quadratic program (13.4.74). For simplicity we have replaced each equality constraint by two inequality constraints. We will use equality constraints when absolutely required. We let $\mathcal{F}$ denote the feasible set.

We define the Lagrangian

$$L(x,\lambda) := q_0(x) + \sum_{k=1}^{m} \lambda_k q_k(x),$$

and the dual functional

$$\phi(\lambda) := \min_x L(x,\lambda).$$

The Lagrangian is linear in λ and so the dual function is a minimum of linear functions, i.e. it is a concave function of λ. Thus the maximum of this concave function is a tractable problem if the dual functional can be evaluated efficiently. For each $\lambda \geq 0$, we have the lower bound

$$\begin{array}{rcl} \mu^* & = & \min\limits_{x\in\mathcal{F}} q_0(x) \\ & \geq & \min\limits_{x\in\mathcal{F}} L(x,\lambda) \\ & \geq & \min\limits_{x} L(x,\lambda) \\ & \geq & \nu^* := \max\limits_{\lambda\geq 0} \phi(\lambda). \end{array}$$

Thus we have defined our dual problem

$$\mu^* \geq \nu^* = \max_{\lambda\geq 0} \phi(\lambda),$$

which provides a lower bound for our primal problem. If, in addition, we have found the feasible $\bar{x} \in \mathcal{F}$ with attainment in the Lagrangian $\bar{x} \in \operatorname{argmin}_x L(x,\bar{\lambda})$ and with complementary slackness $\sum_k \bar{\lambda}_k q_k(\bar{x}) = 0$, then

$$\begin{array}{rcl} \mu^* & \geq & \nu^* = L(\bar{x},\bar{\lambda}) \\ & = & q_0(\bar{x}) \\ & \geq & \mu^*, \end{array}$$

i.e. we have found an optimum $\bar{x}$ and have a zero duality gap when these sufficiency conditions (feasibility, attainment, complementary slackness) hold. Note that since we are dealing with an unconstrained minimum of a quadratic

Lagrangian, we obtain the interesting statement: *necessary conditions for the sufficiency conditions to hold*, i.e. we need stationarity of the Lagrangian and positive semidefiniteness of the Hessian of the Lagrangian. Thus, when these two conditions are incompatible we lose strong duality; we can even expect a duality gap.

We now present several Q^2P problems where the Lagrangian relaxation is important and well known. In all these cases, the Lagrangian dual provides an important theoretical tool for algorithmic development, even where the duality gap may be nonzero. We continue to emphasize our theme that illustrates that the Lagrangian relaxation is best.

Convex Quadratic Programs. We start with the easy case; consider the convex quadratic program

$$\begin{array}{lrll} \text{CQP} & \mu^* := & \min & q_0(x) \\ & & \text{s.t.} & q_k(x) \leq 0, \quad k = 1, \ldots m, \end{array}$$

where all $q_i(x)$ are convex quadratic functions. We now see that Lagrangian duality can always solve this problem.

The dual is

$$\text{DCQP} \qquad \nu^* := \max_{\lambda \geq 0} \ \min_{x} \ q_0(x) + \sum_{k=1}^{m} \lambda_k q_k(x).$$

If ν^* is attained at λ^*, x^*, then a *sufficient* condition for x^* to be optimal for CQP is primal feasibility and complementary slackness, i.e.

$$\sum_{k=1}^{m} \lambda_k^* q_k(x^*) = 0.$$

In addition, it is well known that the Karush-Kuhn-Tucker (KKT) conditions are sufficient for global optimality, and under an appropriate constraint qualification the KKT conditions are also necessary. Therefore strong duality holds if a constraint qualification is satisfied, i.e. in this case there is no duality gap and the dual is attained.

However, surprisingly, *if the primal value of CQP is bounded then it is attained and there is no duality gap*, see e.g. [776, 630, 631, 629]. (This can be considered to be an extension of the Frank-Wolfe Theorem, [510].) However, the dual may not be attained, e.g. consider the convex program

$$0 = \min\{x : x^2 \leq 0\}$$

and its (unattained) dual

$$0 = \max_{\lambda \geq 0} \min_{x} x + \lambda x^2 = \max_{\lambda > 0} \min_{x} x + \lambda x^2.$$

Algorithmic approaches based on Lagrangian duality appear in e.g. [363, 509, 583].

Nonconvex Quadratic Programs.

Rayleigh Quotient. Suppose that $A = A^T \in \mathcal{S}^n$. It is well known that the smallest eigenvalue λ_1 of A is obtained from the Rayleigh quotient, i.e.

$$\lambda_1 = \min\{x^T A x : x^T x = 1\}. \qquad (13.4.105)$$

Since A is not necessarily positive semidefinite, this is the minimization of a nonconvex function on a nonconvex set. However, the Rayleigh quotient forms the basis for many algorithms for finding the smallest eigenvalue, and these algorithms are very efficient. In fact, it is easy to see that there is no duality gap for this nonconvex problem, i.e.

$$\lambda_1 = \max_{\lambda} \min_{x} \; x^T A x - \lambda(x^T x - 1) = \max_{A-\lambda I \succeq 0} \lambda. \qquad (13.4.106)$$

To see this note that the inner minimization problem in (13.4.106) is unconstrained. This implies that the outer maximization problem has the hidden semidefinite constraint (an ongoing theme in the chapter)

$$A - \lambda I \succeq 0,$$

i.e. λ is at most the smallest eigenvalue of A. With λ set to the smallest eigenvalue, the inner minimization yields the eigenvector corresponding to λ_1. Thus, we have an example of a *nonconvex problem for which strong duality holds.* Note that the problem (13.4.105) has the special norm constraint, and a homogeneous quadratic objective.

Trust Region Subproblem. We will next see that strong duality holds for a larger class of seemingly nonconvex problems. The trust region subproblem, TRS, is the minimization of a quadratic function subject to a norm constraint. No convexity or homogeneity of the objective function is assumed. We allow for a further extension, i.e. we do not assume convexity of the constraint and allow indefinite quadratic functions for both objective and constraint. (See e.g. [155] for applications of indefinite quadratic forms.) This problem is important in nonlinear programming, e.g. [552, 551].

$$\begin{array}{lll} \text{TRS} \qquad & \mu^* := \min & q_0(x) = x^T Q_0 x - 2c_0^t x \\ & \quad \text{s.t.} & x^T x - \delta^2 \leq 0 \text{ (or } = 0). \end{array}$$

or the generalized trust region subproblem [747, 549].

$$\begin{array}{lll} \text{GTRS} \qquad & \mu^* := \min & q_0(x) = x^T Q_0 x - 2c_0^t x \\ & \quad \text{s.t.} & q_1(x) \leq 0 \text{ (or } = 0), \end{array}$$

where q_1 is another quadratic function. In addition, one can have two sided constraints $\alpha \leq q_1(x) \leq \beta$, which are used in trust region algorithms as well.

For TRS, assuming that the constraint is written "$\leq$," the Lagrangian dual is:

$$\text{DTRS} \qquad \nu^* := \max_{\lambda \geq 0} \min_{x} \; q_0(x) + \lambda(x^T x - \delta^2).$$

This is equivalent to (see [747]) the (concave) nonlinear semidefinite program

$$\begin{array}{llll} \text{DTRS} & \nu^* := \max & c_0^T (Q_0 + \lambda I)^\dagger c_0 - \lambda \delta^2 \\ & \quad\text{s.t.} & Q_0 + \lambda I \succeq 0 \\ & & \lambda \geq 0. \end{array}$$

where $\cdot^\dagger$ denotes Moore-Penrose inverse. It is shown in [747] that strong duality holds for TRS, i.e. there is a zero duality gap $\mu^* = \nu^*$, and the dual is attained. (The primal is also attained.) Thus, as in the eigenvalue case, we see that this is an example of a nonconvex program where strong duality holds. In addition, this implies that this problem can be solved efficiently; polynomial time results are presented in [854].

Proof.

We include a short proof of strong duality, for the inequality constrained case, based on the outline in [478], i.e. we fall back on the convex case after a perturbation. Note that the key to the proof is being able to pass between the inequality and equality constraints.

Without loss of generality, we can assume that TRS is nonconvex. (Otherwise, we apply the convex results discussed above.) Therefore μ^* is attained on the boundary of the feasible set and the smallest eigenvalue of Q_0, denoted γ, is negative. Then TRS is equivalent to

$$\begin{array}{rcll} \mu^* & = & \displaystyle\min_{x^T x \leq \delta^2} & x^T(Q_0 - \gamma I)x - 2c_0^t x + \gamma x^T x \\ & = & \displaystyle\min_{x^T x = \delta^2} & x^T(Q_0 - \gamma I)x - 2c_0^t x + \gamma x^t x, \quad (Q_0 \text{ is indefinite}) \\ & = & \displaystyle\min_{x^T x = \delta^2} & x^T(Q_0 - \gamma I)x - 2c_0^t x + \gamma \delta^2 \\ & = & \displaystyle\min_{x^T x \leq \delta^2} & x^T(Q_0 - \gamma I)x - 2c_0^t x + \gamma \delta^2, \quad (Q_0 - \gamma I \text{ is singular}) \\ & = & \displaystyle\max_{\lambda \geq 0} \min_{x} & x^T(Q_0 - \gamma I)x - 2c_0^t x + \lambda(x^T x - \delta^2) + \gamma\delta^2 \text{ (convex case)} \\ & = & \displaystyle\max_{\lambda \geq 0} \min_{x} & x^T Q_0 x - 2c_0^t x + (\lambda - \gamma)(x^T x - \delta^2) \\ & \leq & \displaystyle\max_{\lambda \geq \gamma} \min_{x} & x^T Q_0 x - 2c_0^t x + (\lambda - \gamma)(x^T x - \delta^2) \quad (\gamma < 0) \\ & = & \nu^* \leq \mu^*. & \end{array} \tag{13.4.107}$$

■

As mentioned above, extensions of this result to a two-sided general, possibly nonconvex, constraint are discussed in [747, 549]. An algorithm based on Lagrangian duality appears in [661] and (implicitly) in [551, 691]. These algorithms are extremely efficient for the TRS problem, i.e. they solve this problem almost as quickly as an eigenvalue problem.

The fact that we can solve the TRS efficiently even though the objective and constraint may be nonconvex is surprising. In fact, in [524] Martinez shows that the TRS can have at most one local and nonglobal optimum, and the Lagrangian at this point has one negative eigenvalue. Therefore, it is even more surprising that the Lagrangian dual (relaxation) allows one to find the global minimum without ever getting trapped near the local minimum.

In fact, for GTRS we still have a 0 duality gap, though strong duality may fail, e.g. consider the simple program $\min x$ s.t. $x^2 \leq 0$. The results in [747] provide strong duality for GTRS with a two sided constraint using the constraint qualification that $\alpha < \beta$. In [549], necessary and sufficient optimality conditions are presented for GTRS using the constraint qualification that $\min q_0(x) < \max q_0(x)$. Using these results in combination with the extension of the Frank-Wolfe result (e.g. [510]) gives us the following.

Theorem 13.4.2 *Consider GTRS: a zero duality gap always holds and, moreover, if the optimal value is finite, then it is attained.* ■

Two Trust Region Subproblem. The two trust region subproblem, TTRS, consists in minimizing a (possibly nonconvex) quadratic function subject to a norm and a least squares constraint, i.e. two convex quadratic constraints. This problem arises in solving general nonlinear programs using a sequential quadratic programming approach, and is often called the CDT problem, see [154].

In contrast to the above single TRS, the TTRS can have a nonzero duality gap, see e.g. [626, 862, 863, 864]. This is closely related to quadratic theorems of the alternative, e.g. [177]. In addition, if the constraints are not convex, then the primal may not be attained, see e.g. [510].

As mentioned above, Martinez [524] shows that the TRS can have at most one local and nonglobal optimum, and the Lagrangian at this point has one negative eigenvalue. Therefore, if we have such a case and add another ball constraint that contains the local, nonglobal, optimum in its interior and also makes this point the global optimum, we obtain a TTRS where we cannot have a zero duality gap due to the negative eigenvalue. It is uncertain what constraints could be added to close this duality gap. In fact, it is still an open problem whether TTRS is an NP-hard or a polynomial time problem.

General $\mathbf{Q^2P}$. The general, possibly nonconvex, $\mathrm{Q^2P}$ has many applications in modeling and approximation theory, see e.g. the applications to SQP methods in [451]. Examples of approximations to $\mathrm{Q^2P}$ also appear in [258].

The Lagrangian relaxation of a $\mathrm{Q^2P}$ is equivalent to the SDP relaxation, and is sometimes referred to as the Shor relaxation, see [733]. The Lagrangian relaxation can be written as an SDP if one takes into the account the hidden semidefinite constraint, i.e. a quadratic function is bounded below only if the Hessian is positive semidefinite. The SDP relaxation is then the Lagrangian dual of this semidefinite program. It can also be obtained directly by *lifting* the problem into matrix space using the fact that $x^T Qx = \text{Trace}\, x^T Qx = \text{Trace}\, Qxx^T$, and relaxing xx^T to a semidefinite matrix X.

One can relate the geometry of the original feasible set of Q^{2}P with the feasible set of the SDP relaxation. The connection is through *valid quadratic inequalities*, i.e. nonnegative (convex) combinations of the quadratic functions; see [260, 442] and our Section 13.4.2.

Orthogonally Constrained Programs with Zero Duality Gaps. We now follow the approach in [41, 37, 36] and consider the *orthonormal type constraints*

$$X^T X = I, \qquad X \in \mathcal{M}_{m,n}$$

(sometimes known as the Stiefel manifold, e.g. [203]) and the trust region type constraint

$$X^T X \preceq I, \qquad X \in \mathcal{M}_{m,n}.$$

Applications and algorithms for optimization on orthonormal sets of matrices are discussed in [203].) In this section we will show that for $m = n$, strong duality holds for a certain (nonconvex) quadratic program defined over orthonormal matrices. Because of the similarity of the orthonormality constraint to the norm constraint $x^T x = 1$, the results of this section can be viewed as a matrix generalization of the strong duality result for the Rayleigh Quotient problem (13.4.105).

Let A and B be $n \times n$ symmetric matrices, and consider the orthonormally constrained homogeneous Q^{2}P

$$\text{QQP}_\text{O} \qquad \mu^O := \begin{array}{ll} \min & \text{Trace}\, AXBX^T \\ \text{s.t.} & XX^T = I. \end{array} \tag{13.4.108}$$

This problem can be solved exactly using Lagrange multipliers, see e.g. [318], or using the classical Hoffman-Wielandt inequality, e.g. [112].

Proposition 13.4.1 *Suppose that the orthogonal diagonalizations of A, B are $A = V\Sigma V^T$ and $B = U\Lambda U^T$, respectively, where the eigenvalues in Σ are ordered nonincreasing, and the eigenvalues in Λ are ordered nondecreasing. Then the optimal value of* QQP$_\text{O}$ *is $\mu^O = \text{Trace}\, \Sigma\Lambda$, and the optimal solution is obtained using the orthogonal matrices that yield the diagonalizations, i.e. $X^* = VU^T$.* ■

The Lagrangian dual of QQP$_\text{O}$ is

$$\max_{S=S^T} \min_X \text{Trace}\, AXBX^T - \text{Trace}\, S(XX^T - I). \tag{13.4.109}$$

However, there can be a nonzero duality gap for the Lagrangian dual, see [870] for an example. The inner minimization in the dual problem (13.4.109) is an unconstrained quadratic minimization in the variables vec(X), with hidden constraint on the Hessian

$$B \otimes A - I \otimes S \succeq 0.$$

The first order stationarity conditions are equivalent to $AXB = SX$ or $AXBX^T = S$. Once can easily construct examples where the semidefinite condition and the stationarity are in conflict and result in a duality gap. In order to close the duality gap, we need a larger class of quadratic functions.

Note that in $\mathrm{QQP_O}$ the constraints $XX^T = I$ and $X^TX = I$ are equivalent. Adding the redundant constraints $X^TX = I$, we arrive at

$$\begin{array}{llll} \mathrm{QQP_{OO}} & \mu^O := & \min & \mathrm{Trace}\, AXBX^T \\ & & \text{s.t.} & XX^T = I,\ X^TX = I. \end{array}$$

Using symmetric matrices S and T to relax the constraints $XX^T = I$ and $X^TX = I$, respectively, we obtain a dual problem

$$\begin{array}{llll} \mathrm{DQQP_{OO}} & \mu^O \geq \mu^D := & \max & \mathrm{Trace}\, S + \mathrm{Trace}\, T \\ & & \text{s.t.} & (I \otimes S) + (T \otimes I) \preceq (B \otimes A) \\ & & & S = S^T,\ T = T^T. \end{array}$$

Theorem 13.4.3 *Strong duality holds for* $\mathrm{QQP_{OO}}$ *and* $\mathrm{DQQP_{OO}}$, *i.e.,* $\mu^D = \mu^O$ *and both primal and dual are attained.* ■

A further relaxation of the above orthogonal relaxation is the trust region relaxation studied in [398]

$$\begin{array}{lll} \mu^*_{QAPT} := & \min & \mathrm{Trace}\, AXBX^T \\ & \text{s.t.} & XX^T \preceq I. \end{array}$$

The constraints are convex with respect to the Löwner partial order and so it is hoped that solving this problem would be useful. Also, this problem is visually similar to the TRS discussed above. And so we would like to find a characterization of optimality.

The set

$$\{X : W = XX^T \preceq I\}$$

is studied separately in [604, 233] and is useful in eigenvalue variational principles.

We now study the matrix trust-region relaxation of QAP:

$$\begin{array}{lll} \mu^*_{SDPT} = & \min & \mathrm{Trace}\, AXBX^T \\ & \text{s.t.} & XX^T \preceq I. \end{array}$$

The following generalization of the Hoffman-Wielandt inequality holds.

Theorem 13.4.4 *For any $XX^T \leq I$, we have*

$$\sum_{i=1}^n \min\{\lambda_i\mu_{n-i+1}, 0\} \leq trAXBX^T \leq \sum_{i=1}^n \max\{\lambda_i\mu_i, 0\}$$

And, the upper bound is attained if

$$X = PDiag(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)Q^T, \qquad (13.4.110)$$

where

$$\varepsilon_i = \begin{cases} 1, & \lambda_i\mu_i > 0, \\ \alpha \in [0,1], & \lambda_i\mu_i = 0, \\ 0, & \lambda_i\mu_i < 0; \end{cases} \qquad (13.4.111)$$

The lower bound is attained if

$$X = PDiag(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)JQ^T, \qquad (13.4.112)$$

where

$$\varepsilon_i = \begin{cases} 1, & \lambda_i\mu_{n-i+1} < 0, \\ \alpha \in [0,1], & \lambda_i\mu_{n-i+1} = 0, \\ 0, & \lambda_i\mu_{n-i+1} > 0. \end{cases} \qquad (13.4.113)$$

■

For a scalar ξ, let $\xi^- := \min\{0, \xi\}$. The lower bound in the above theorem states that $\mu^*_{SDPT} = \sum_{i=1}^n [\lambda_i\mu_i]^-$. Since the Theorem provides the feasible point of attainment, i.e. an upper bound for the relaxation problem, we will prove the theorem by proving another theorem that shows that the value μ^*_{SDPT} is also attained by a Lagrangian dual program. Note that since XX^T and X^TX have the same eigenvalues, $XX^T \preceq I$ if and only if $X^TX \preceq I$. Explicitly using both sets of constraints, as in [41], we obtain

$$\begin{array}{llll} \text{QAPTR} & \mu^*_{QAPT} := & \min & \text{Trace}\, AXBX^T \\ & & \text{s.t.} & XX^T \preceq I, \quad X^TX \preceq I. \end{array}$$

Next we apply Lagrangian relaxation to QAPTR, using matrices $S \succeq 0$ and $T \succeq$ to relax the constraints $XX^T \preceq I$ and $X^TX \preceq I$, respectively. This results in the dual problem

$$\begin{array}{llll} \text{DQAPTR} & \mu^*_{QAPT} \geq \mu^D_{QAPT} := & \max & -\text{Trace}\, S - \text{Trace}\, T \\ & & \text{s.t.} & (B \otimes A) + (I \otimes S) + (T \otimes I) \succeq 0 \\ & & & S \succeq 0, \ T \succeq 0. \end{array}$$

To prove that $\mu^*_{QAPT} = \mu^D_{QAPT}$ we will use the following simple result.

Lemma 13.4.1 *Let $\lambda \in \Re^n$, $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. For $\gamma \in \Re^n$ consider the problem*

$$\min \quad z_\pi := \sum_{i=1}^n [\lambda_i \gamma_{\pi(i)}]^-,$$

where $\pi(\cdot)$ *is a permutation of* $\{1,\ldots,n\}$*, Then the permutation that minimizes* z_π *satisfies* $\gamma_{\pi(1)} \geq \gamma_{\pi(2)} \geq \cdots \gamma_{\pi(n)}$. ■

Theorem 13.4.5 *Strong duality holds for QAPTR and DQAPTR, i.e.,* $\mu^D_{QAPT} = \mu^*_{QAPT}$ *and both primal and dual are attained.* ■

The above results illustrate the theme about the strength of the Lagrangian relaxation, i.e. that tractable problems can be solved using Lagrangian duality in some form.