SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The duality and efficiency in semidefinite programming

av

Marie Miller

2014 - No 10

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

The duality and efficiency in semidefinite programming

Marie Miller

Självständigt arbete i matematik 15 högskolepoäng, Grundnivå

Handledare: Yishao Zhou

2014

The duality and efficiency in semidefinite programming

Marie Miller

Abstract

The purpose of this thesis is to explore the aspect of duality and efficiency in semidefinite programming. In particular, we discuss bad behaved systems in relation to the duality gap. In that sense, the impact of efficiency seems to be dependent of if there exists duality gap. There are several approaches to close up it, and we present two regularization algorithms. The first algorithm is based on abstract convex programming while the second one on semidefinite programming. Then we show how duality gap can be closed by means of facial reduction in semidefinite programming. The analysis part will end by some semidefinite programming problems.

Keywords. *Semidefinite programming, duality, efficiency.*

Acknowledgements. *I would like to express my gratitude to my supervisor Yishao Zhao at the Mathematical department, Stockholm university for her feedback, and constructive critics during the process.*

Contents

1 Introduction

Semidefinite programming is a well explored research area. The model developed around 1990 has grown fast, intensively both from the research interest and the practice perspectives. It serves many purposes and is one of the most prominent areas in mathematical programming branches, in coding theory, and finance etc [11, 19, 24].

Semidefinite programming could be classified as an extension of linear programming and is a subclass of conic programming. The extension of linear programming has made it possible in recent years to develope more efficient algorithms [11, 14, 19, 23, 24, 28].

There are some well known differences between linear programming and semidefinite programming. In the linear programming the primal optimal value always concides with the dual optimal value which does not hold necessarily in the semidefinite programming. Pataki in [20] discussed the aspects on duality in semidefinite programming in relation to bad behaved versus well behavied systems [20].

Moreover, Lustig, Marsten, and Shanno in [16], Helmberg et al in [14] have studied the interior point methods in relation to the efficiency. However in some semidefinite programming problems there exists positive duality gap and hence optimal value is not attained [20].

1.1 Problem statement

Semidefinite programming handles a finite set of inequality constraints, and variables. The model has high potential, and delivers efficiency [10, 20]. Despite of it there are some limitations primarily related to the semidefinite programming properties which cannot always be extended, interpreted, and explained in the same manner as linear programming. Hence the duality and efficiency are relevant to analyze since the equivalence of these programmings design cannot be met under the same assumptions. In addition, the structure of the semidefinite programming problems are also significant to target the duality gap [20].

The result of duality displays if there exists duality gap in semidefinite programming. To reduce the size of duality gap is related to some of the aspects; the models assumptions, structure, dualization, and regularization methods. Borwein, Wolkowicz in [8] proposed in 1981 an approach to reduce the duality gap. This regularization method is based on an abstract convex programming and the conclusion holds for subfaces. Ramana, Tunçel, and Wolkowicz in [23] validated the regularization method even for semidefinite programming. Recently, Malick et al in [17] proposed a new regularization method and the results show increased robustness. The motivation to use this method in comparison with alternative regulatization methods is based on the high level of accurancy, and speed [17].

Differences between the above described regularization methods are the following. The first method is comming from an abstract convex programming approach while the other one from a semidefinite programming. Further, the methods differ in the initiation position. Borwein and Wolkowicz [8] regularize the primal perspective in comparision with Malick et al [17] where the primal, dual perspectives are combined to construct the general algorithm. Another important issue with the regularization method is the fact that it is constructed particularly for ill-posed problem and not for general problems [12].

1.2 Research questions and aim

- How does duality gap affect the efficiency of algorithms?
- Is it possible to close up the duality gap and retain efficiency?
- What methods are suitable and why?

The main purpose of this thesis is to explore the aspects duality and efficiency in semidefinite programming.

1.3 Notations and outline

We have used the following notations. The set of symmetric $n \times n$ matrices is denoted by S^n. Similary, S^n_+ is the set of positive semidefinite $n \times n$ matrices, and S^n_{++}, the set of positive definite $n \times n$ matrices.

This thesis is structured into four chapters. The first one gives a general introduction to semidefinite programming and presents the research questions, problem statement, and aim. Next chapter is divided into two parts where the first one covers related work and the second part presents the relevant theoretical framework. Furthermore this chapter contains a section about the objective efficiency.

In the third chapter two regularization methods are introduced, and both methods are explicitly reviewed separately. In chapter four, analysis is focused on the achived results, and will also consider some notions from the theoretical perspective. The last section ends with a summery of the most important results in relation to the research questions, and propose further research about duality gap in relation to semidefinite programming.

2 Literature review and the theoretical framework

2.1 Related work

Boyd and Vanderberghe in [28] give a general review to semidefinite programming and explain the theory of primal-dual interior point method. Alizadeh in [3] used the interior point method to show that local convergence of an optimal solution holds in polynomial time. Redle in [24] considers the aspects of a duality theory in semidefinite programming and argues that duality turns out to be a key factor. Pataki [20] discusses a similiar reason on duality and points out duality generates as a certificate to obtain optimality.

There are several advantages with semidefinite programming [9, 10, 28]. First, it has many applications in diverse areas which give the theoretical framework a broader perspective, and in turn could lead to a higher level of efficiency. For instance, the interior point method in semidefinite programming. Secondly, it is possible to target many convex optimization problems by reformulating them as semidefinite programming problems. The third argument goes back to the era of semidefinite programming and beyond this powerful idea.

Kharchiyan in [10] applied the ellipsoid method in 1979 in combination with linear programming. Karmarkar in [10] further developed the idea in 1984 with an improving algoritm, and thereafter Nesterov and Nemirovski [10] have built on the method and provided important contributions to the existing and the most common used interior point methods within semidefinite programs. Many recent articles have been inspired by this interior point method and developed various types of interior point methods [3, 10, 14, 31]. For instance, Alizadeh in [3] used the interior point method in semidefinite programming in combination with combinatorial optimization.

Klerk in [11] describes the complex structure in semidefinite programming in constrast to the linear programming. Ramana in [22] explicitly highlights that the extension does not always work for general semidefinite programming and derive an exact duality theory. In addition Zhang, Chen, and Zhang in [32] have regarded the duality theory to ensure zero duality gap. From the above context we select to study the structure, duality, and efficiency, respectively, in relation to the duality gap. The impact of efficiency in semidefinite programming seems to be dependent of if there exist duality gap.

2.2 Linear programming

In order to illustrate why SDP is an extension of linear programming we present LP in standard form and its dual problem.

2.2.1 The standard and canonical form

A linear programming is the minimization problem of a linear function subject to linear constraints, it is expressed in standard or canonical form [7]. We shall first consider the standard primal linear programming problem:

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

where $c, x \in \mathbf{R}^n, A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$, and the inequality constraint is interpreted componentwise. To derive the Lagrangian dual function introduce multipliers $\lambda \in \mathbf{R}^m, \mu \in \mathbf{R}^n, \mu \geq 0$, and we obtain the Lagrange relaxed problem:

$$\begin{aligned} \theta(\lambda, \mu) &= \min\{c^t x + \lambda^t (b - Ax) - \mu^t x\} \\ &= \min\{(c^t - \lambda^t A - \mu^t)x + \lambda^t b\}, \end{aligned}$$

and the minimum value is:

$$\theta(\lambda, \mu) = \begin{cases} \lambda^t b, & \text{if} \quad c^t - \lambda^t A - \mu^t \geq 0 \\ -\infty, & \text{if} \quad (c^t - \lambda^t A - \mu^t)_i < 0 \qquad \text{for} \quad \text{some} \quad i. \end{cases}$$

The associated Lagrangian dual is:

$$\begin{aligned} \max \quad & \lambda^t b \\ \text{s.t.} \quad & c - A^t \lambda - \mu \geq 0 \\ & \mu \geq 0, \end{aligned}$$

or, equivalently, by [9]

$$\begin{aligned} \max \quad & \lambda^t b \\ \text{s.t.} \quad & A^t \lambda \leq c. \end{aligned}$$

Another approach to get the dual problem is to just lift the equality constraint into the objective function, i.e introduce the multiplier $\lambda \in \mathbf{R}^m$, we get a Lagrange relaxed problem:

$$\begin{aligned} \min \quad & (c^t - \lambda^t A)x + \lambda^t b \\ \text{s.t} \quad & x \geq 0. \end{aligned}$$

Then

$$\theta(\lambda) = \begin{cases} \lambda^t b, & \text{if} \quad c^t - \lambda^t A \geq 0 \\ -\infty, & \text{if} \quad (c^t - \lambda^t A)_j < 0 \qquad \text{for} \quad \text{some} \quad j. \end{cases}$$

The associated Lagrangian dual is:

$$\begin{aligned} \max \quad & \lambda^t b \\ \text{s.t.} \quad & c^t - \lambda^t A \geq 0. \end{aligned}$$

Thus for a linear programming problem there is a unique dual problem. This is not true in general for nonlinear programming problems. We demostrate this by an example.

Example 2.2.1. (NLP dual) Consider following NLP problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} \frac{a_i}{x_i}, \quad a_i > 0 \\ \text{s.t.} \quad & \sum_{i=1}^{n} b_i x_i = b_0, \quad b_0 > 0 \\ & l_i \leq x_i \leq u_i, \quad u_i > l_i > 0, \quad i = 1, \ldots, n. \end{aligned}$$

To obtain a Lagrange dual problem we can either lift the constraint $\sum_{i=1}^{n} b_i x_i = b_0$ or all the constraints to the objective function.

Alternative 1. Introduce λ to minimize

$$\begin{aligned} \min \quad & l(x, \lambda) = \sum_{i=1}^{n} (\frac{a_i}{x_i} + \lambda b_i x_i) - \lambda b_0 \\ \text{s.t.} \quad & l_i \leq x_i \leq u_i. \end{aligned}$$

Separate the problem and minimize for each x_i. Let $f_i(x_i) = \frac{a_i}{x_i} + \lambda b_i x_i$, $i = 1, \ldots, n$. For fixed i we have $f_i^{'}(x_i) = -\frac{a_i}{x_i^2} + \lambda b_i$, $f_i^{''}(x_i) = \frac{2a_i}{x^3} > 0$ so $f_i(x_i)$ is convex. So, the solution of $f_i^{'}(x_i) = 0$ is a minimum. Solving this equation yields $x_i^2 = \frac{a_i}{\lambda b_i}$.

Now we consider the constraints $l_i \leq x_i \leq u_i$. We have the following cases:

(1) $\lambda b_i \leq 0$, this gives optimum $\hat{x} = u_i$ because we minimize $\frac{a_i}{x_i} + \lambda b_i x_i$.

(2) $\lambda b_i > 0$ and $l_i \leq \sqrt{\frac{a_i}{\lambda b_i}} \leq u_i$, then $\hat{x_i} = \sqrt{\frac{a_i}{\lambda b_i}}$.

(3) $\lambda b_i > 0$ and $\sqrt{\frac{a_i}{\lambda b_i}} \leq l_i$, then $\hat{x_i} = l_i$.

(4) $\lambda b_i > 0$ and $\sqrt{\frac{a_i}{\lambda b_i}} \geq u_i$, then $\hat{x_i} = u_i$.

Substituting $\hat{x_1}, \ldots, \hat{x_n}$ determined in accordance above discussion in the objective function we have the dual function

$$\theta(\lambda) = \sum_{i=1}^{n} (\frac{a_i}{\hat{x_i}} + \lambda b_i \hat{x_i}) - \lambda b_0.$$

So the dual problem is

$$\max \quad \theta(\lambda)$$

which is an unconstrained problem.

Alternative 2. Introduce λ and $\underline{\mu_i} \geq 0, \bar{\mu_i} \geq 0, i = 1, \ldots, n$, and denote $\underline{\mu} = \begin{bmatrix} \underline{\mu_1} \\ \vdots \\ \underline{\mu_n} \end{bmatrix}, \bar{\mu} = \begin{bmatrix} \bar{\mu_1} \\ \vdots \\ \bar{\mu_n} \end{bmatrix}$.We minimize

$$l(\lambda, \underline{\mu}, \bar{\mu}, x) = \sum_{i=1}^{n} (\frac{a_i}{x_i} + (\lambda b_i - \underline{\mu_i} + \bar{\mu_i}) x_i) - \lambda b_0 + \sum_{i=1}^{n} \underline{\mu_i} l_i - \sum_{i=1}^{n} \bar{\mu_i} u_i.$$

Minimizing for each x_i, using the same argument as in Alternative 1, we have $x_i^2 = \frac{a_i}{\lambda b_i - \underline{\mu_i} + \bar{\mu_i}}$.

(1) If $\lambda b_i - \underline{\mu_i} + \bar{\mu_i} < 0$ we have $\hat{x_i} = \infty$.

(2) If $\lambda b_i - \underline{\mu_i} + \bar{\mu_i} \geq 0$ then $\hat{x_i} = \sqrt{\frac{a_i}{\lambda b_i - \underline{\mu_i} + \bar{\mu_i}}}$.

The minimum is achived with minimal value:

$$\Theta(\lambda, \underline{\mu}, \bar{\mu}) = \begin{cases} -\infty & \text{if} \quad \lambda b_i - \underline{\mu_i} + \bar{\mu_i} \leq 0 \\ 2\sum_{i=1}^{n} \sqrt{a_i(\lambda b_i - \underline{\mu_i} + \bar{\mu_i})} - \lambda b_0 + \sum_{i=1}^{n} (\underline{\mu_i} l_i - \bar{\mu_i} \mu_i) & \text{if} \quad \text{otherwise.} \end{cases}$$

So, the dual problem is

$$\begin{aligned} \max \quad & \Theta(\lambda, \underline{\mu}, \bar{\mu}) = 2\sum_{i=1}^{n} \sqrt{a_i(\lambda b_i - \underline{\mu_i} + \bar{\mu_i})} - \lambda b_0 + \sum_{i=1}^{n} (\underline{\mu_i} l_i - \bar{\mu_i} \mu_i) \\ \text{s.t.} \quad & \lambda b_i - \underline{\mu_i} + \bar{\mu_i} \leq 0 \\ & \underline{\mu_i} \geq 0, \bar{\mu_i} \geq 0. \end{aligned}$$

Obviously these two dual problems are different. Different dual problems will result in different efficient algorithms.

Furthermore, the linear primal standard and canonical form are equivalent. Consider the following pairs of forms below [9].

Standard form of primal and dual LP:

$$\begin{cases} \min & c^t x \\ \text{s.t.} & Ax = b, \\ & x \geq 0 \end{cases} \qquad \begin{cases} \max & \lambda^t b \\ \text{s.t.} & A^t \lambda \leq c. \\ & \end{cases}$$

Canonical form of primal and dual LP:

$$\begin{cases} \min & c^t x \\ \text{s.t.} & Ax \geq b, \\ & x \geq 0 \end{cases} \qquad \begin{cases} \max & \lambda^t b \\ \text{s.t.} & A^t \lambda \leq c, \\ & \lambda \geq 0. \end{cases}$$

We see here the canonical pair is symmetric.

Remark. *For the standard dual problem there is no sign restrictions on* λ*.*

We shall now show that these forms are equivalent by introducing the slackvariable $s \geq 0$, $s \in \mathbf{R}^m$.

$$Ax \geq b \Leftrightarrow Ax - s = b \Leftrightarrow (A| - I) \begin{bmatrix} x \\ s \end{bmatrix} = b.$$

Let $\tilde{A} := (A| - I)$, $\tilde{x} = \begin{bmatrix} x \\ s \end{bmatrix}$, $\tilde{c} = \begin{bmatrix} c \\ 0 \end{bmatrix}$ such that:

$$\begin{aligned} \min \quad & \tilde{c}^t \tilde{x} \\ \text{s.t.} \quad & \tilde{A}\tilde{x} = b \\ & \tilde{x} \geq 0. \end{aligned}$$

$$\begin{aligned} \max \quad & \lambda^t b \\ \text{s.t.} \quad & \tilde{A}^t \lambda \leq \tilde{c}, \end{aligned}$$

where the inequality constraint is:

$$\begin{aligned} &\tilde{A}^t \lambda = (A| - I)^t \lambda = \begin{bmatrix} A^t \\ -I \end{bmatrix} \lambda = \begin{bmatrix} A^t \lambda \\ -\lambda \end{bmatrix} \leq \tilde{c} = \begin{bmatrix} c \\ 0 \end{bmatrix} \\ &\Leftrightarrow A^t \lambda \leq c, \quad \underbrace{-\lambda \leq 0}_{\lambda \geq 0}, \end{aligned}$$

and the claim follows.

2.2.2 Duality properties

This section is based on the literature of Bazaraa, Sherali, and Shetty [5], Boyd, Vandenberghe [9].

Theorem 2.2.1. (Weak duality) For any feasible solution x to the primal problem and any feasible solution λ to the dual problem we have $c^t x \geq b^t \lambda$.

Proof. For any pairs of feasible solutions x, λ in the primal and its associated dual problem, we have:

$$c^t x \geq (A\lambda^t)x = \lambda^t(Ax) \geq b^t \lambda.$$

Thus, $c^t x \geq b^t \lambda$. □

Theorem 2.2.2. (Strong duality) Assume that x and λ are feasible solutions of the primal and dual problem respectively. Then they have both optimal solutions with the same objective value, i.e $c^t x = b^t \lambda$.

The following Table 1 shows the linear primal and dual perspective in case of impossible solutions. In addition, the columns and rows are associated with the infeasible, finite, infinite solutions.

Table 1: The LP primal and dual solutions

	$\mathrm{D}_\emptyset$	D_f	D_∞
$\mathrm{P}_\emptyset$		impossible	
P_f	impossible		impossible
P_∞		impossible	impossible

The table is an immediate consequence of the strong duality except the $D_\emptyset$ and $P_\emptyset$ which is possible, seen by the following example.

Example 2.2.2. (LP duality) An example on the case where the dual and the primal problems are not feasible.

The primal problem:

$$\begin{aligned} \min \quad & -x_2 \\ \text{s.t.} \quad & x_1 - x_2 \geq 1 \\ & -x_1 + x_2 \geq 0 \\ & x_1, x_2 \geq 0, \end{aligned}$$

has no solution and neither does its dual

$$\begin{aligned} \max \quad & u_1 \\ \text{s.t.} \quad & u_1 - u_2 \leq 0 \\ & -u_1 + u_2 \leq -1 \\ & u_1, u_2 \geq 0. \end{aligned}$$

2.3 Convex programming

2.3.1 Convex sets, functions, and duality

This section presents the convex programming and it is based on the literature of Bazaraa, Sherali, and Shetty [5]. We begin by considering the constrained nonlinear problem with the equality and inequality constraints:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \\ & x \in X, \end{aligned}$$

where $f(x), g_i(x), i = 1, \dots, m, h_j(x), j = 1, \dots, l$ are functions defined on X, a subset of $\mathbf{R^n}$, and $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is a vector with n components [5]. The following definitions survey some basic notions and specify some significant properties under the assumption that $S \subseteq \mathbf{R^n}$ is not empty.

Definition 2.3.1. (Convex set) The set S is convex if the line segment between $x_1, x_2 \in S$, belongs to S, that is $\lambda x_1 + (1 - \lambda)x_2 \in S$, for all $\lambda \in [0, 1]$, and all $x_1, x_2 \in S$.

Geometrically, a straight line that passes through two distinct points inside the set S. If a part of the line segment does not belong to the set then it is not convex.

Definition 2.3.2. (Convex hull) The convex hull, denoted conv(S) is the collection of all convex combinations of S. That is, $conv(S) = \{\mathbf{x} = \sum_{i=1}^{m} \lambda_i x_i : x_i \in S, \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0 \text{ for } i = 1, \dots, m\}$, where m is a positive integer.

Definition 2.3.3. (Neighborhoods) Given $\mathbf{x}$, and an $\epsilon > 0$, the ball $N_\epsilon(\mathbf{x}) = \{\mathbf{y} : ||\mathbf{y} - \mathbf{x}|| < \epsilon\}$ is called an ϵ-neighborhood of $\mathbf{x}$.

Definition 2.3.4. (Closure) The closure of S, denoted $cl(S)$ is defined by $cl(S) = \{\mathbf{x} \in S : S \cap N_\epsilon(\mathbf{x}) \neq \emptyset \text{ for every } \epsilon > 0\}$.

Definition 2.3.5. (Affine combination) A vector $\mathbf{y}$ in $\mathbf{R}^n$ is a linear combination of $\mathbf{x_1}, \dots, \mathbf{x_k}$ in $\mathbf{R}^n$ if $\mathbf{y} = \sum_{j=1}^{k} \lambda_j x_j$ for $\lambda_1, \dots, \lambda_k$. If, in addition, $\lambda_1, \dots, \lambda_k$ satisfy $\sum_{j=1}^{k} \lambda_j = 1$, then $\mathbf{y}$ is an affine combination of $\mathbf{x_1}, \dots, \mathbf{x_k}$.

Definition 2.3.6. (Affine hull) The affine hull of S, is the collection of all affine combinations of points in S.

Definition 2.3.7. (Relative interior) The relative interior of S, denoted $ri(S)$, $ri(S) = \{\mathbf{x} \in S : N_\epsilon(\mathbf{x}) \cap \text{aff}(S) \subset S \text{ for some } \epsilon > 0\}$, where aff$(S)$ is the affine hull of S.

The following definition describes convexity for a univariate function. In parallel with convex sets is a convex function characterized as chords between two distinct points lie above its graph.

Definition 2.3.8. (Convex function) The function f defined on S is convex if $f(\lambda x_1 + (1 - \lambda x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$ for all $x_1, x_2 \in S$, and $\lambda \in [0, 1]$, where S is convex.

Furthermore, a convex function in relation with optimality determines if optimum exists and the optimal value is attained. In addition, the optimal dual value assesses to be an underestimate for the optimal primal value, it is *consistent* [6].

Before considering properties of duality, we state the primal and its Lagrangian dual:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, ..., m \\ & h_j(x) = 0, j = 1, ..., l \\ & x \in X, \end{aligned}$$

and we derive the Lagrangian dual function:

$$\theta(\lambda, \mu) = \min\{f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x) : \mathbf{x} \in X\}.$$

where λ_i, μ_j are classified as *the lagrangian multipliers*, $\lambda \geq 0, i = 1, \ldots, m$. Hence, the Lagrangian dual is then formulated:

$$\begin{aligned} \max \quad & \theta(\lambda, \mu) \\ \text{s.t.} \quad & \lambda \geq 0. \end{aligned}$$

Another important issue with duality is that maximum does not always exists, and then it is more convenient to depict maximum as supremum. In the similar way, minimum corresponds to infimum. If the primal optimal value exists, and concides with its dual then is sufficient to only examine the properties of duality [5].

Theorem 2.3.1. (Carathédory theorem) Let S be an arbitrary set in $\mathbf{R^n}$. If $\mathbf{x} \in conv(S), \mathbf{x} \in conv(\mathbf{x_1}, \dots, \mathbf{x_{n+1}})$, where $\mathbf{x}_i \in S$ for $i = 1, \dots, n+1$. Then, $\mathbf{x}$ can be represented

$$\begin{aligned}
&\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i \\
&\sum_{i=1}^{n+1} \lambda_i = 1 \\
&\lambda_i \geq 0 \quad \text{for} \quad i = 1, \dots, n+1 \\
&\mathbf{x}_i \in S \quad \text{for} \quad i = 1, \dots, n+1.
\end{aligned}$$

Example 2.3.1. ([5], Ex. 6.13) Formulate explicit the Lagrangian dual function of the following problem for which:
$X = \{(x_1, x_2, x_3, x_4) : x_1 + x_2 \leq 12, x_2 \leq 4, x_3 + x_4 \leq 6, x_1, x_2, x_3, x_4 \geq 0.$

$$\begin{aligned}
\max \quad & 3x_1 + 6x_2 + 2x_3 + 4x_4 \\
\text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 \leq 12 \\
& -x_1 + x_2 + 2x_4 \leq 4 \\
& x_1 + x_2 \leq 12 \\
& x_2 \leq 4 \\
& x_3 + x_4 \leq 6 \\
& x_1, x_2, x_3, x_4 \geq 0.
\end{aligned}$$

First, rewrite the objective function to minimize:

$$\begin{aligned}
\min \quad & -3x_1 - 6x_2 - 2x_3 - 4x_4 \\
\text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 \leq 12 \\
& -x_1 + x_2 + 2x_4 \leq 4 \\
& x_1 + x_2 \leq 12 \\
& x_2 \leq 4 \\
& x_3 + x_4 \leq 6 \\
& x_1, x_2, x_3, x_4 \geq 0.
\end{aligned}$$

Compute the Lagrangian dual function:

$$\begin{aligned}
\Theta(\lambda_1, \lambda_2) &= \min\{\mathrm{f}(\mathbf{x}) + \lambda_1 \mathrm{g}_1(\mathbf{x}) + \lambda_2 \mathrm{g}_2(\mathbf{x}) : \mathrm{x} \in \mathrm{X}\} \\
&= \min\{-3\mathrm{x}_1 - 6\mathrm{x}_2 - 2\mathrm{x}_3 - 4\mathrm{x}_4 + \lambda_1(\mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 + \mathrm{x}_4 - 12) \\
&+ \lambda_2(-x_1 + x_2 + 2x_4 - 4) : x \in X\}.
\end{aligned}$$

Divide the Lagrangian dual function into two functions:

$$\Theta_1(\lambda_1, \lambda_2) = \min\{\mathrm{x}_1(-3 + (\lambda_1 - \lambda_2) + \mathrm{x}_2(-6 + (\lambda_1 + \lambda_2)) : \mathrm{x}_1 + \mathrm{x}_2 \leq 12, \mathrm{x}_2 \leq 4, \mathrm{x}_1, \mathrm{x}_2 \geq 0\}$$
$$\Theta_2(\lambda_1, \lambda_2) = \min\{\mathrm{x}_3(-2 + \lambda_1) + \mathrm{x}_4(-4 + (\lambda_1 + 2\lambda_2) : \mathrm{x}_3 + \mathrm{x}_4 \leq 6, \mathrm{x}_3, \mathrm{x}_4 \geq 0\} - 12\lambda_1 - 4\lambda_2$$

and use the Carathédory theorem [5] where

$$\Theta_1(\lambda_1,\lambda_2)=\begin{cases} 0, & (x_1,x_2)=(0,0), \quad \text{if} \quad \lambda_1-\lambda_2\geq 3, \lambda_1+\lambda_2\geq 6\\ 4\lambda_1+4\lambda_2-24, & (x_1,x_2)=(0,4), \quad \text{if} \quad \lambda_1-\lambda_2\geq 3, \lambda_1+\lambda_2\leq 6\\ 12\lambda_1-4\lambda_2-48, & (x_1,x_2)=(8,4), \quad \text{if} \quad \lambda_1-\lambda_2\leq 3, \lambda_1+\lambda_2\leq 6\\ 12\lambda_1+12\lambda_2-36, & (x_1,x_2)=(12,0), \quad \text{if} \quad \lambda_1-\lambda_2\leq 3, \lambda_1+\lambda_2\geq 6, \end{cases}$$

$$\Theta_2(\lambda_1,\lambda_2)=\begin{cases} -12\lambda_1-4\lambda_2, & (x_3,x_4)=(0,0), \quad \text{if} \quad \lambda_1\geq 2, \lambda_1+2\lambda_2\geq 4\\ -6\lambda_1+8\lambda_2-24, & (x_3,x_4)=(0,6), \quad \text{if} \quad \lambda_1\geq 2, \lambda_1+2\lambda_2\leq 4\\ -6\lambda_1-4\lambda_2-12, & (x_3,x_4)=(6,0), \quad \text{if} \quad \lambda_1\leq 2, \lambda_1+2\lambda_2\geq 4. \end{cases}$$

where $\lambda_1, \lambda_2 \geq 0$.

The initialization step consists of rewriting the objective function in standard form, and in the main step we computed the Langrangian dual funcion, we used theorem of Carathédory. Finally, divide the Lagrangian function into two functions, and simplify to get the desired dual function.

Theorem 2.3.2. (Karush-Kuhn-Tucker Necessary conditions) Consider the primal problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \leq 0$ for $i = 1, \dots, m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$ the active index set. Suppose f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$ are continous at $\bar{\mathbf{x}}$. Furthermore, suppose that $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ are linearly independent. Then, the following KKT conditions holds true

$$\begin{aligned} \nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) &= \mathbf{0},\\ u_i g_i(\bar{\mathbf{x}}) = 0, &\quad \text{for} \quad i=1,\dots,m\\ u_i \geq 0, &\quad \text{for} \quad i=1,\dots,m, \end{aligned}$$

where $u_i g_i(\bar{\mathbf{x}}) = 0$ is the complementary slackness condition.

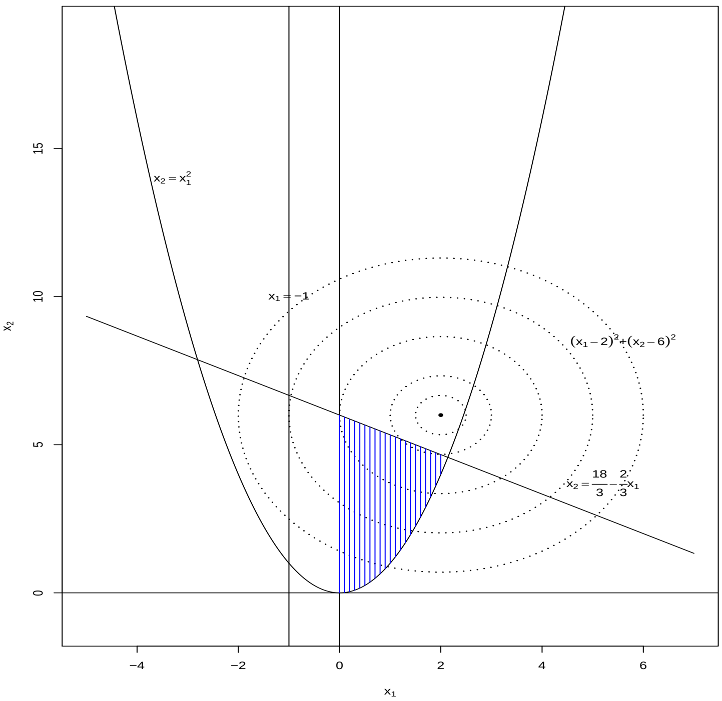
Remark. *The condition $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ is one of the constrained qualification. There are othere conditions to ensure the KKT conditions to be necessary. One commonly used is the Slaterś condition. See Definition* 2.3.16*. It turns out to be the commonly used natural condition in study of SDP.*

Example 2.3.2. ([5], Ex. 6.11) Find the optimal point, verify the KKT-conditions.

$$\begin{aligned} \min \quad & (x_1-2)^2+(x_2-6)^2\\ \text{s.t.} \quad & x_1^2-x_2\leq 0\\ & -x_1\leq 1\\ & 2x_1+3x_2\leq 18\\ & x_1,x_2\geq 0. \end{aligned}$$

For simplicity we solve the problem by geometrically. The point $(2, 6)$ is optimum without the constraints. So we enlarge the circle center at $(2, 6)$ until it tangents to the tendency of the feasible region (Figur 1).

That is, we find shortest distance from the point $(2, 6)$ to the line $2x_1 + 3x_2 = 18$, which can be parametrized by $(x_1, x_2) = (t, 6 - \frac{2}{3}t)$.



Figur 1. Graph of the objective function, and inequality constraints in R.

The shortest line segment between $(2, 6)$ and a point on the line should be orthogonal to the line, i.e. the direction $(1, -\frac{2}{3})$. So the inner product of $(1, -\frac{2}{3})$ and $(t - 2, 6 - \frac{2}{3}t - 6) = (t - 2, -\frac{2}{3}t)$ is zero, yielding $t = \frac{18}{3}$. So the minimal value is achieved at $\bar{x} = (\frac{18}{3}, \frac{66}{13})$. This shows that only one constraint is active. Let

$$\begin{aligned} g_1(x) &= x_1^2 - x_2, \\ g_2(x) &= -x_1 - 1, \\ g_3(x) &= 2x_1 + 3x_2 - 18, \\ g_4(x) &= -x_1, \\ g_5(x) &= -x_2. \end{aligned}$$

Then we have $u_3 \neq 0$ the other $u_i's$ are zero by the complementary slackness. We continue verifying the KKT-conditions, and calculate the partial derivatives:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2(x_1-2) \\ 2(x_2-6) \end{bmatrix}, \nabla g_1(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix}, \nabla g_3(\mathbf{x}) = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$
$$\nabla g_4(\mathbf{x}) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \nabla g_5(\mathbf{x}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

The first KKT-condition:

$$\nabla f(\bar{\mathbf{x}}) + u_1 \nabla g_1(\bar{\mathbf{x}}) + u_3 \nabla g_3(\bar{\mathbf{x}}) + u_4 \nabla g_4(\bar{\mathbf{x}}) + u_5 \nabla g_5(\bar{\mathbf{x}}) = 0 :$$

$$\begin{bmatrix} -\frac{8}{13} \\ -\frac{8}{13} \end{bmatrix} + u_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow u_3 = \frac{8}{13} > 0.$$

and the third KKT-condition: $u_i \geq 0$, for $i = 1,2,3,4,5$ is satisfied at $\bar{\mathbf{x}}$.

The next theorems concern some properties of duality.

Theorem 2.3.3. (Weak duality) Let $\mathbf{x}$ be a feasible solution to primal and similary let (λ, μ) be a feasible solution to the dual. Then $f(\mathbf{x}) \geq \Theta(\lambda, \mu)$.

Proof. According to the definition of the dual:

$$\begin{aligned}\Theta(\lambda,\mu) &= \min\{\mathbf{f}(\mathbf{y}) + \lambda^{\mathbf{t}}\mathbf{g}(\mathbf{y}) + \mu^{\mathbf{t}}\mathbf{h}(\mathbf{y}) : \mathbf{y} \in \mathbf{X}\} \\ &\leq \mathbf{f}(\mathbf{x}) + \lambda^{\mathbf{t}}\mathbf{g}(\mathbf{x}) + \mu^{\mathbf{t}}\mathbf{h}(\mathbf{x}) \\ &\leq \mathbf{f}(\mathbf{x}),\end{aligned}$$

and the claim follows since $\lambda \geq \mathbf{0}$, by the primal $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. □

Remark. *The dual optimal value is the lower bound of the primal. This has significance in computation.*

Theorem 2.3.4. (Strong duality) Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be convex, and let $\mathbf{h} : \mathbf{R}^n \to \mathbf{R}^l$ be affine. Suppose that the following constraint qualificaton holds true. There exists $\bar{\mathbf{x}} \in X$ s.t. $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$, and $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{0} \in \operatorname{int}\{\mathbf{h}(\mathbf{x}) : \mathbf{x} \in \mathrm{X}\}$. Then

$$\min\{\mathbf{f}(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{x} \in \mathbf{X}\} = \max\{\Theta(\lambda,\mu) : \lambda \geq \mathbf{0}\}.$$

Proof. We omit the proof. (see, for instance [5]). □

In general, strong duality holds whenever the primal optimal value is equal to its dual value. There exists a duality gap if the primal optimal value exceed dual value. These optimality criteria work also for other programmings design and not specifically developed for the convex programming.

2.3.2 Convex cones

This subsection proceed with convex cones. In particular, we explore convex cones, check the validity, and utilize the result to semidefinite cones. Here we have assumed that $K \subseteq V$, an inner product space.

Definition 2.3.9. (A cone,[9]) The set K is a cone if every $x \in K$ and $\lambda \in [0,1]$ imply $\lambda x \in K$.

Definition 2.3.10. (A convex cone, [9]) The set K is called a convex conve if it is closed and convex, i.e for any $x_1, x_2 \in K$ and $\lambda_1, \lambda_2 \geq 0$ we have $\lambda_1 x_1 + \lambda_2 x_2 \in K$.

Definition 2.3.11. (Alternative Definition of a convex cone, [2]) The cone K is convex if it is closed under addition $x_1, x_2 \in K \Rightarrow x_1 + x_2 \in K$.

Clearly, these two definition are equivalent. The following definitions is taken from e.g. Ahron, Nemiroviski [2].

Definition 2.3.12. (A pointed cone) The convex cone is pointed if $x_1 \in K, -x_1 \in K$ imply $x_1 = 0$.

Definition 2.3.13. (A proper cone) The cone is proper if the following conditions holds; convex, closed, pointed, and has a nonempty interior.

Example 2.3.3. (A proper cone, [2]). The nonnegative orthant $K = \{x \in \mathbf{R}^{\mathbf{n}} : x_i \geq 0, i = 1, \ldots, n\}$ is a proper cone. S^n_+ is also a proper cone.

A proper cone K induces a generalized inequality (or partial ordering) as follows [9]:

$$
\begin{aligned}
x_1 \preceq_K x_2 \Leftrightarrow x_2 - x_1 \in K \\
x_1 \prec_K x_2 \Leftrightarrow x_2 - x_1 \in \mathbf{int} K,
\end{aligned}
$$

where $\mathbf{int}K$ is the interior of K. According to Boyd, Vandenberghe [9] the generalized inequality $\preceq_K$ satisfies following properties:

- reflexive: $x_1 \preceq_K x_1$.
- antisymmetric: if $x_1 \preceq_K x_2$ and $x_2 \preceq_K x_1$ then $x_1 = x_2$.
- transitive: if $x_1 \preceq_K x_2$ and $x_2 \preceq_K x_3$ then $x_1 \preceq_K x_3$.
- preserved under addition: if $x_1 \preceq_K x_2, u_1 \preceq_K u_2$, then $x_1 + u_1 \preceq_K x_2 + u_2$.
- If $x_1 \preceq_K x_2$ then $\lambda x_1 \preceq \lambda x_2$ for all $\lambda > 0$.

Example 2.3.4. (The generalized inequality,[9]) For $K = R^n_+$, $x \preceq_K y$ means $x_i \preceq y_i, i = 1, \ldots, n$; for $S^n_+, X \preceq_K Y$ means $Y - X$ is positive semidefinite.

Definition 2.3.14. (Dual cone, [9]) Let K be a cone. The set $K^* = \mathbf{y} : \mathbf{x^t y} \succeq 0, \forall x \in K\}$ is called the dual cone of K.

Definition 2.3.15. *If $K^* = K$ then K is said to be self dual.*

Example 2.3.5. (Cones and their dual cones, [9]) The aim of this example is to show that proving a set is to show that matrix formulation is sometimes very effective in proving properties of cones. Therefore we are going to give two alternative ways to prove some properties of the following special cone.

(1) $\mathbf{R}^n_+$ is self-dual.

(2) Icecream cone is self-dual.

(3) $(\mathbf{S^n_+})^* = \mathbf{S^n_+}$.

Example 2.3.6. (Cones and their dual cones)

$$K = \{(x_1, x_2, x_3) \in \mathbf{R}^3 : x_1 \geq 0, x_2 \geq 0, x_1 x_2 \geq x_3^2\}.$$

Alternatively, $K = S^3_+$ can be defined as the following set

$$K = \{x_1, x_2, x_3) \in \mathbf{R}^3 : \begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix} \succeq_{S^3_+} 0\}.$$

Proposition 2.3.1. *K is a closed convex cone.*

Proof. We apply the alternative definition to show the closedness. We need to show that the complement is open. If we have a symmetric matrix $M = \begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix}$ that is not positive semidefinite, there exists $\tilde{x} \in \mathbf{R}^2$ such that $\tilde{x}^t M \tilde{x} < 0$ and this inequality still sholds for all matrices $M^{'}$ in a sufficiently small neighborhood of M.

Now, we show the convex cone properties:

(i) $\forall x \in K, \forall \lambda \geq 0$ (real) we have $\lambda x \in K$, (i.e K is a cone).

(ii) $\forall x, x^{'} \in K$ we have $x + x^{'} \in K$, (i.e K is convex, since if $x, x^{'} \in K$ and $\lambda \in [0, 1]$, then $(1-\lambda)x, \lambda x^{'} \in K$ by (i) and then (ii) shows that $(1-\lambda)x + \lambda x^{'} \in K$ as required by convexity.)

Now, if $x^t M x \geq 0$ and $x^t M^{'} x \geq 0$ then also $x^t(\lambda M)x = \lambda x^t M x \geq 0$ for $\lambda \geq 0$ and $x^t(M + M^{'})x = x^t M x + x^t M^{'} x \geq 0$. □

Remark. *We can prove the proposition using the original definition, but the proof is not as simple as given above. For example, to show (ii) (the sum property) we compute*

$$\begin{aligned}(x_1 + x_1^{'})(x_2 + x_2^{'}) &= x_1x_2 + x_1x_2^{'} + x_1^{'}x_2 + x_1^{'}x_2^{'} \\ &\geq x_3^2 + 2\frac{x_1x_2^{'} + x_1^{'}x_2}{2} + x_3^{'2} \\ &\geq x_3^2 + x_3^{'2} + 2\sqrt{x_1x_2^{'}x_1^{'}x_2} \\ &\geq x_3^2 + x_3^{'2} + 2\sqrt{x_3^2x_3^{'2}} \\ &= x_3^2 + x_3^{'2} + 2|x_3||x_3^{'}| \\ &\geq x_3^2 + x_3^{'2} + 2x_3x_3^{'} = (x_3 + x_3^{'})^2,\end{aligned}$$

where in the second inequality above, we used the Arithmetic-Geometric Mean Inequality (AGM).

Proposition 2.3.2. *The dual cone of K is*

$$K^* = \{(x_1, x_2, x_3) \in \mathbf{R}^3 : x_1 \geq 0, x_2 \geq 0, x_1x_2 \geq \frac{x_3^2}{4}\} \subseteq \mathbf{R}^3.$$

Proof. We show first the inclusion $\supseteq$. Again we use AMG inequality. Let us fix $\tilde{y} = (\tilde{x_1}, \tilde{x_2}, \tilde{x_3})$ such that $\tilde{x_1} \geq 0, \tilde{x_2} \geq 0, \tilde{x_1}\tilde{x_2} \geq \frac{\tilde{x_3}^2}{4}$. Then for $x = (x_1, x_2, x_3) \in K$ chosen arbitrarily, we get

$$\begin{aligned}\tilde{y}^t x &= \tilde{x_1}x_1 + \tilde{x_2}x_2 + \tilde{x_3}x_3 \\ &= 2\frac{\tilde{x_1}x_1 + \tilde{x_2}x_2}{2} + \tilde{x_3}x_3 \\ &\geq 2\sqrt{\tilde{x_1}x_1\tilde{x_2}x_2} + \tilde{x_3}x_3 \\ &= 2\frac{|\tilde{x_3}|}{2}|x_3| + \tilde{x_3}x_3 \geq 0.\end{aligned}$$

This means that $\tilde{y} \in K^*$.

For $\subseteq$ let us fix $\tilde{y} = (\tilde{x_1}, \tilde{x_2}, \tilde{x_3})$ such that $\tilde{x_1} < 0$ or $\tilde{x_2} < 0$ or $\tilde{x_1}\tilde{x_2} < \frac{\tilde{x_3}^2}{4}$. We need to find a proof for $\tilde{y} \in K^*$. If $\tilde{x_1} < 0$ we choose $x = (1, 0, 0) \in K$ and get the desired $\tilde{x_2}^t x < 0$. If $\tilde{x_2} < 0$, $x = (0, 1, 0)$ will do the job. In case of $\tilde{x_1}, \tilde{x_2} \geq 0$, but $\tilde{x_1}\tilde{x_2} < \frac{\tilde{x_3}^2}{4}$, let us first assume $\tilde{x_3} \geq 0$ and set $x = (\tilde{x_2}, \tilde{x_1}, -\sqrt{\tilde{x_1}\tilde{x_2}}) \in K$. Then

$$\tilde{y}^t x = 2\tilde{x_1}\tilde{x_2} - \tilde{x_3}\sqrt{\tilde{x_1}\tilde{x_2}} < 2\tilde{x_1}\tilde{x_2} - 2\tilde{x_1}\tilde{x_2} = 0.$$

For $\tilde{x_3} < 0$, we pick $x = (\tilde{x_2}, \tilde{x_1}, \sqrt{\tilde{x_1}\tilde{x_2}}) \in K$. □

Remark. *We can see that the proof will be much easier by alternative definition using matrices.*

2.3.3 Constraint qualifications

In this section we consider the constrained convex programming problem under the inequality constraint [5]:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, ..., m \\ & x \in X. \end{aligned}$$

As seen in the Karush-Kuhn-Tucker necessary theorem, there is an extra condition that makes the KKT conditions to be necessary for the local optimum, that is, the set of gradients of g_i is linear independent at the KKT point $\bar{\mathbf{x}}$ where i is the active constraint index. In this section we consider several other such conditions.

Definition 2.3.16. (Slaterś constraint qualification) We say the above nonlinear programming problem satifies the Slater condition if $g_1, \ldots, g_m$ are convex, and there is a point $\bar{x}$ in the open set X satisfying $g_i(\bar{x}) < 0, i = 1, \ldots, m$.

Bazaraa, Sherali, and Shetty in [5] describe several constraint qualifications (CQś) and their relations. The top level starts with the strongest condition the slaterś CQ and the linear independence CQ.

Definition 2.3.17. (Linear independence constraint qualfication) The set X is open, each g_i for $i \notin I$ is continous at $\bar{x}$, and $\nabla g_i(\bar{x})$ for $i \in I$ are linearly independent.

Example 2.3.7. ([5], Ex. 6.11, revisited) Check if slaterś CQ, LICQ hold for the following problem:

$$\begin{aligned} \min \quad & (x_1 - 2)^2 + (x_2 - 6)^2 \\ \text{s.t.} \quad & x_1^2 - x_2 \leq 0 \\ & 2x_1 + 3x_2 \leq 18 \\ & -x_1 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0. \end{aligned}$$

Now $X = \mathbf{R}^2$. As before let $g_1 = x_1^2 - x_2, g_2 = 2x_1 + 3x_2 - 18, g_3 = -x_1 - 1, g_4 = -x_1, g_5 = -x_2$. Clearly all $g_i's$ are convex. At the point $(1, 2)$ all functions $g_i < 0$. So the Slater conditon holds. Next we consider $g_1 = g_2 = 0$ which gives a solution at $\bar{x} = (\frac{\sqrt{55}-1}{3}, \frac{-2\sqrt{55}+56}{9})$. At this point the gradients of g_1 and g_2 are computed as follows:

$$\nabla g_1(\bar{x}) = \begin{bmatrix} \frac{2(\sqrt{55}-1)}{3} \\ -1 \end{bmatrix}$$

$$\nabla g_2(\bar{x}) = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

which are linear independent, and the LICQ is satisfied.

2.4 Abstract convex programming

2.4.1 The abstract convex programming

The abstract convex programming is characterized as an *extension* of the convex programming [26]. In this section we formulate the general abstract convex programming according to Borwein, Wolkowicz [8]:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & g(x) \preceq_S 0, \\ & x \in \Omega \end{aligned}$$

where f: is an extended convex function on $\mathbf{R^n}$, g: an extended S-convex function on $\mathbf{R^n} \to \mathbf{R^m}$, $\Omega \subset \mathbf{R^n}$ is convex, $S \subset \mathbf{R^m}$ is a convex cone. Furthermore the convex cone S is pointed, and defines a generalized inequality [8].

Definition 2.4.1. (S-convex function, [9]) An S-convex function is represented by a convex function w.r.t a proper cone S. More precisely, $f(\lambda x_1 + (1-\lambda)x_2) \preceq_S \lambda f(x_1) + (1-\lambda) f(x_2)$.

Example 2.4.1. [9] Example of an abstract convex programming problem. Boyd, Vandenberghes form of the abstract convex programming problem:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t} \quad & f_i(x) \leq 0 \quad \text{for} \quad i = 1, \dots, m, \\ & a_i^t x = b_i, \quad \text{for} \quad i = 1, \dots, p, \end{aligned}$$

where $f_0, \dots, f_m$ are convex. It seems that it is very restrictive but many problems can be reformulated in this form. For example

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t} \quad & \frac{x_1}{1+x_2^2} \leq 0, \\ & (x_1 + x_2)^2 = 0, \end{aligned}$$

we can see that $f_1 = \frac{x_1}{1+x_2^2}$ is not convex, show the inequality constraint in the abstract form is not convex by utilize the hessian:

$$\begin{aligned} \frac{\partial g_1(x_1, x_2)}{\partial x_1} &= \frac{1}{1+x_2^2} \\ \frac{\partial g_1(x_1, x_2)}{\partial x_2} &= \frac{-2x_1x_2}{(1+x_2^2)^2} \\ \Rightarrow \nabla g_1(x_1, x_2) &= \begin{bmatrix} \frac{1}{1+x_2^2} \\ \frac{-2x_1x_2}{(1+x_2^2)^2} \end{bmatrix} \end{aligned}$$

$$\Rightarrow \nabla^2 g_1(x_1, x_2) = H = \begin{bmatrix} 0 & \frac{-2x_2}{(1+x_2^2)^2} \\ \frac{-2x_2}{(1+x_2^2)^2} & \frac{-2x_1(1+x_2^2)+4x_1x_2}{(1+x_2^2)^3} \end{bmatrix}$$

The leading principle minor of the matrix H is:

$$H_1 = 0, det(H) = \begin{vmatrix} 0 & \frac{-2x_2}{(1+x_2^2)^2} \\ \frac{-2x_2}{(1+x_2^2)^2} & \frac{-2x_1(1+x_2^2)+4x_1x_2}{(1+x_2^2)^2} \end{vmatrix} = 0 - \frac{4x_2^2}{(1+x_2^2)^4} = -\frac{4x_2^2}{(1+x_2^2)^4},$$

and $det(H) = -\frac{4x_2^2}{(1+x_2^2)^4}$ implies H is negative semidefinite, g_1 is concave for all $x_1, x_2 \geq 0$, and $(x_1 + x_2)^2$ is not affine. So, it is not a convex programming problem. But it can be transformed to a convex programming by its equivalent form [9].

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t} \quad & x_1 \leq 0, \\ & x_1 + x_2 = 0. \end{aligned}$$

2.4.2 Subcones and faithfully convex function

This section concerns subcones contained in the convex cone, faithfully convex function. Subfaces is an important issue in the abstract regularization method, and assesses as the core. All definitions below are based on Borwein, Wolkowiczs work [8], Moskowitz, Paligiannis [18], and Boyd and Vandenberghe [9].

Definition 2.4.2. (A face, [23]) A subcone K of S is a face of S, and denoted $K \lhd S$, $x_1, x_2 \in S, x_1 + x_2 \in K \Rightarrow x_1, x_2 \in K$.

Definition 2.4.3. (An exposed face) A face of S is exposed if there exist ψ in S^* such that $K = \{s \in S : \langle \psi, s \rangle = 0\}$. Furthermore, the convex cone S is called facially exposed if every face of S is exposed.

Definition 2.4.4. (Faithful convex) The S-convex functions g is faithfully convex with respect to the face E if g is not affine along any line segment in E unless they are affine along the entire line extending the segment.

Definition 2.4.5. (Real analytic at $\bar{x}$, [18]) A smooth function f which is represented by the Taylor series

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\bar{x})}{k!}(x - \bar{x})^k$$

in a neighborhood of $\bar{x}$, is called real analytic at $\bar{x}$. Furthermore, if f is analytic at every point $\bar{x} \in \Omega$, we say f is real analytic on Ω.

Definition 2.4.6. (Taylors theorem in several variables, [18]) Let $f : \Omega \subseteq \mathbf{R^n} \to \mathbf{R}$ be a continious, differentiable function, that is C^2 on the open convex set Ω of $\mathbf{R^n}$ and $\bar{x}$, and x such that

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{1}{2!} \langle H_f(c)(x - \bar{x}), x - \bar{x} \rangle.$$

Example 2.4.2. (Faithfully convex function, [29]) Consider the function f defined by

$$f(x_1, x_2, x_3) = -\sqrt{(4 + (x_1 + x_2)^2)} + x_1 + x_2 + x_3^2$$

A faithfully convex function is convex, analytic. Recall that the Definition 2.3.8. Since we have a multi variable function we apply the hessian to verify that f is convex. Start by calculating the partial derivatives for f:

$$\begin{aligned}
\frac{\partial f(x_1, x_2, x_3)}{\partial x_1} &= 1 - \frac{(x_1 + x_2)}{\sqrt{4 + (x_1 + x_2)^2}} \\
\frac{\partial f(x_1, x_2, x_3)}{\partial x_2} &= 1 - \frac{(x_1 + x_2)}{\sqrt{4 + (x_1 + x_2)^2}} \\
\frac{\partial f(x_1, x_2, x_3)}{\partial x_3} &= 2x_3. \\
\Rightarrow \nabla f(x_1, x_2, x_3) &= \begin{bmatrix} 1 - \frac{(x_1+x_2)}{\sqrt{4+(x_1+x_2)^2}} \\ 1 - \frac{(x_1+x_2)}{\sqrt{4+(x_1+x_2)^2}} \\ 2x_3 \end{bmatrix}
\end{aligned}$$

$$\Rightarrow \nabla^2 f(x_1, x_2, x_3) = H = \begin{bmatrix} \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} & \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} & 0 \\ \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} & \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The leading principal minors of the matrix H are:

$$H_1 = \frac{4}{(4 + (x_1 + x_2)^2)^{\frac{3}{2}}} > 0, H_2 = \begin{vmatrix} \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} & \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} \\ \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} & \frac{4}{(4+(x_1+x_2)^2)^{\frac{3}{2}}} \end{vmatrix} = 0,$$

$det(H) = 0$ implies H is positive semidefinite, f is convex for all $x_1, x_2, x_3 \geq 0$. The analycity is clear, since f has a Taylor series.

2.4.3 The extended slater constraint

This section is built on previous sections. The main result is the extended slaterś constraint in terms of extended inequality $\prec_S$. Again we refer to Borwein, Wolkowicz [8].

Theorem 2.4.1. (The extended slaterś constraint) Suppose that g is continuous and weakly faithfully S-convex on Ω, Ω is the intersection of a polyhedrar set and a closed linear manifold, and P satisfies the generalized slaters conditions: there exists $\bar{x} \in \Omega$ with $g(\bar{x}) \prec_S 0$. Then the standard Lagrange multiplier theorem holds, that is,

(a) assume that μ is the finite optimal value of

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & g(x) \preceq_S 0, \\ & x \in \Omega. \end{aligned}$$

Then $f(x) + \lambda g(x) \geq \mu$ for all $x \in \Omega$ for some $\lambda \in S^*$.

(b) If μ is attained by $f(a)$, $a \in \Omega$, then $\lambda g(a) = 0$.

Proof. We omit the proof. (See, [8]). □

2.5 Semidefinite programming

2.5.1 Positive semidefinite matrices

This section describes semidefinite matrices, semidefinite programming in relation to the primal and duality. Following definitions, theorems are according to the literature of Boyd, Vandenberghe [9], Aharon, Nemiroviski [2].

Definition 2.5.1. (Positive semidefinite matrices, [2]) A positive semidefinite matrix (PSD) is denoted $A \succeq 0$ with following properties:
(i) A is symmetric,
(ii) $x^t Ax \geq 0$ for any $x \in \mathbf{R}^n$.

This definition is equivalent to all eigenvalues of A denoted $\lambda(A)$ are nonnegative, i.e $\lambda(A) \geq 0$. Similary, the matrix A is positive definite if $x^t Ax > 0$, all eigenvalues $\lambda(A) > 0$.

Example 2.5.1. [9] *The cone of positive semidefinite $n \times n$ matrices, is a convex cone.*

Proof. According to the Definition 2.3.10 if $\lambda \in [0, 1], A, B \in S^n_+$ then $\lambda A + (1 - \lambda)B \in S^n_+$. Insert the convex expression in Definition 2.5.1 and hence $x^t Ax = x^t(\lambda A + (1 - \lambda)B)x = \lambda x^t Ax + (1 - \lambda)x^t Bx \geq 0$. □

Definition 2.5.2. (Inner product) The inner product of matrices S^n is defined as $A \bullet B$:

$$A \bullet B = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij} = \mathbf{tr}(A^t B)$$

This definition can be justifed to satisfy the axioms of inner product.

2.5.2 Dual problems, equivalence of SDP problems

In literature, there are often two standard forms of SDP. We state them as two definitions following Vandenberghe and Boyd [28].

Remark. *Sometimes, especially we compute, we also use the notation $\langle A, B \rangle$ for the inner product of S^n for simplicity. And we use these notations interchangeably. Also we use the same notation for inner product $\langle a, b \rangle$ for $a, b \in \mathbf{R}^n$.*

Definition 2.5.3. (Conic standard form) A primal SDP in the conic standard form is (SDP_c) :

$$\begin{aligned} \min \quad & \mathbf{tr}(CX) \\ \text{s.t.} \quad & \mathbf{tr}(A_iX) = b_i, i = 1, ..., m \\ & X \succeq_{S_+^n} 0 \Leftrightarrow X \in S_+^n \end{aligned}$$

where $C, A_i \in S^n$, and S_+^n is the cone of symmetric positive semidefinite matrices. Clearly, this is an obvious parallel to the standard form for LP, where the only difference is that the cone $\mathbf{R}_+^n$ is substituted by the cone S_n^+. To see this we use the notation:

$$\mathbf{vect}(A) := (a_{11}, a_{21}, \ldots, a_{m1}, a_{12}, a_{22}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{nn})^t$$

for any matrix of $m \times n$, that is we stack the columns of A on the top of each other. Then $\mathbf{tr}(CX) = \sum_{i,j=1}^n c_{ij}x_{ij} = c^tx$ and $\mathbf{tr}(A_iX) = a_i^tx, i = 1, \ldots, m$ where $x = \mathbf{vect}(X)$, $a_i = \mathbf{vect}(A_i)$ and $c = \mathbf{vect}(C)$. Thus (SDP_c) can be rewritten as

$$\begin{aligned} \min \quad & c^tx \\ \text{s.t.} \quad & a_i^tx = b_i, i = 1, ..., m \\ & X \succeq_{S_+^n} 0. \end{aligned}$$

Remark. *Malick et al in [17] call it linear semidefinite programming. Apperantly, this can be considered as a generalization of LP problem in standard form. The only difference is we use the variable in the cone S_+^n.*

Definition 2.5.4. (SDP in inequality standard form) A primal SDP is in inequality standard form is $(SDP_{\preceq})$

$$\begin{aligned} \min \quad & c^tx \\ \text{s.t.} \quad & x_1B_1 + \cdots + x_kB_k \preceq B \end{aligned}$$

where $B_1, \ldots B_k, B \in S^n$, and we call $B(x) := x_1B_1 + \cdots + x_kB_k \preceq B$ linear matrix inequality (LMI). Note that it is here the same semidefinite program comes from ([28]).

Apply Lagrange duality theory, we can derive the associate dual problem to (SDP_c) and $(SDP_{\preceq})$, respectively. For (SDP_c) we introduce $y \in \mathbf{R}^m$ and construct the Lagrange dual function:

$$\begin{aligned}
\Theta(y) &= \min_{X\in S_n^+}\{\mathbf{tr}(CX)+\sum_{i=1}^{m}(b_i^t-\mathbf{tr}(A_iX))y_i\} \\
&= \min_{X\in S_n^+}\{\langle C,X\rangle-\sum_{i=1}^{m}\langle A_i,X\rangle y_i\}+b^ty \\
&= \min_{X\in S_n^+}\{\langle C,X\rangle-\langle\sum_{i=1}^{m}A_iy_i,X\rangle\}+b^ty \\
&= \min_{X\in S_n^+}\{\langle C-\sum_{i=1}^{m}A_iy_i,X\rangle\}+b^ty \\
&= \begin{cases} b^ty & \text{if } \langle C-\sum_{i=1}^{m}A_iy_i,X\rangle\geq 0 \\ -\infty & \text{otherwise} \end{cases} .
\end{aligned}$$

Therefore we obtain the dual problem

$$\begin{aligned}
\max \quad & \Theta(y) \\
\text{s.t.} \quad & \langle C-\sum_{i=1}^{m}A_iy_i,X\rangle\geq 0
\end{aligned}$$

Explicitly, in the matrix form:

$$\begin{aligned}
(DSDP_c) \quad \max \quad & b^ty \\
\text{s.t.} \quad & C-\sum_{i=1}^{m}y_iA_i\succeq_{S_n^+}0.
\end{aligned}$$

The last constraint is true due to the fact that $(S_+^n)^*=S_+^n$, and the characterizations of positive semidefinite matrices: A is positive semidefinite if and only if $\langle A,B\rangle\geq 0$ for all $B\in S_n^+$. Now we have a pair of primal-dual problem for $(DSDP_c)$ in Definition 2.5.3.

Similarly, we can derive the dual problem of $(SDP_{\preceq})$ as follows:

$$B(x)\preceq B\Leftrightarrow\langle B(x)-B,X\rangle\leq 0,\forall X\in S_+^m.$$

Then the dual objective function is, for $X \succeq 0$:

$$
\begin{aligned}
f(Y) &:= \min_x \{c^t x + \langle B(x) - B, X\rangle\} \\
&= \min_x \{c^t x + \sum_{i=1}^{k} \langle x_i B_i, X\rangle - \langle B, X\rangle\} \\
&= \min_x \{c^t x + \sum_{i=1}^{k} x_i \langle B_i, X\rangle\} - \langle B, X\rangle \\
&= \min_x \sum_{i=1}^{k} (c_i + \langle B_i, X\rangle) x_i - \langle B, X\rangle \\
&= \begin{cases} -\langle B, X\rangle & \text{if all } c_i + \langle B_i, X\rangle \geq 0 \\ -\infty & \text{otherwise} \end{cases} .
\end{aligned}
$$

So, the dual problem of $(SDP_{\preceq})$ is:

$$
\begin{aligned}
\max \quad & -\langle B, X\rangle = -\mathbf{tr}(BX) \\
\text{s.t.} \quad & \mathbf{tr}(B_i X) + c_i \geq 0, i = 1, \ldots k. \\
& X \succeq_{S^n_+} 0.
\end{aligned}
$$

To get equality constraints, we introduce the slackvariables $s_1, \ldots, s_n \geq 0$:

$$
\begin{aligned}
\max \quad & -\mathbf{tr}(BX) \\
\text{s.t.} \quad & \mathbf{tr}(B_i X) + c_i - s_i = 0, i = 1, \ldots k. \\
& X \succeq_{S^n_+} 0, s_i \geq 0, i = 1, \ldots, k.
\end{aligned}
$$

To put this in the form of (SDP_c), we construct a new matrix variable $\tilde{X} = \begin{bmatrix} X & 0 \\ 0 & \operatorname{diag}(s_1, \ldots, s_k) \end{bmatrix} \succeq_{S^{n+k}_+} 0$ and accordingly $\tilde{B} = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$ and $\tilde{B}_i = \begin{bmatrix} B_i & 0 \\ 0 & 0 \end{bmatrix}$, where $\operatorname{diag}(s_1, \ldots, s_k)$ is the diagonal matrix with $s_1, \ldots, s_k$ as diagonal elements. Hence the dual of $(SDP_{\preceq})$ is:

$$
\begin{aligned}
(DSDP_{\preceq}) \quad \max \quad & -\mathbf{tr}(\tilde{B}\tilde{X}) \\
\text{s.t.} \quad & \mathbf{tr}(\tilde{B}_i \tilde{X}) + c_i = 0 \\
& \tilde{X} \succeq_{S^{n+k}_+} 0.
\end{aligned}
$$

Remark. *We notice that there is an obvious relation between $(SDP_{\preceq})$ and $(DSDP_c)$, respectively, between (SDP_c) and $(DSDP_{\preceq})$.*

So, we raise the question: Is the dual problem $(DSDP_{\preceq})$ still a SDP problem? In fact we can show that the two forms $(SDP_{\preceq})$ and (SDP_c) are equivalent. In other words, we can convert one form to the other.

Proposition 2.5.1. $(SDP_{\preceq}) \Leftrightarrow (SDP_c)$.

Proof. For the direction $(SDP_c) \Rightarrow (SDP_{\preceq})$, we assume for simplicitly that the matrices $A_1, \ldots, A_k$ are linearly independent. Then we can expess the affine set

$$\{Z : Z \in S^n, \mathbf{tr}(A_i Z) = b_i, i = 1, \ldots, k\}$$

in the form

$$\{G(y) = G_0 + y_1 G_1 + \cdots + y_p G_p : y \in \mathbf{R}^p\}$$

where $p = \frac{n(n+1)}{2} - k$, and G_i are appropriate matrices, that is, the solution set can be parametrized by p parameters $y_1, \ldots, y_p$ (due to linearly indepence of $A_1, \ldots, A_k$). Applying this fact to (SDP_c) we have $X = G(y) \geq 0$ and $\mathbf{tr}(C^t G((y))$ should be minimized which is

$$\begin{aligned}\mathbf{tr}(C^t(G_0 + y_1 G_1 + \cdots + y_p G_p)) =& \mathbf{tr}(C^t G_0 + (C^t G_1) y_1 + \cdots + (C^t G_p) y_p) \\ =& \mathbf{tr}(C^t G_0) + \mathbf{tr}(C^t G_1) y_1 + \cdots + \mathbf{tr}(C^t G_p) y_p.\end{aligned}$$

Since $\mathbf{tr}(C^t G_0)$ is a constant, the (SDP_c) is equivalent to:

$$\begin{aligned} \min \quad & c_1 y_1 + \cdots + c_p y_p \\ \text{s.t.} \quad & G(y) \succeq 0, \end{aligned}$$

where $c_i = \mathbf{tr}(C^t G_i), i = 1, \ldots, p$. This is in the form of $(SDP_{\preceq})$.To show the other direction, let $Z := B - \sum_{i=1}^k x_i B_i$. This will be our new variable matrix. Thus, we have the variables Z and k scalars $x_1, \ldots, x_k$. Denote $Z = (z_{ij}), B = (b_{ij}), B_l = (b^l_{ij}), l = 1, \ldots, k, i, j = 1, \ldots, n$. Then

$$Z = B - \sum_{l=1}^k x_l B_l \Leftrightarrow z_{ij} = b_{ij} - \sum_{l=1}^k x_l b^l_{ij}, i, j = 1, \ldots, n.$$

Clearly, they are linear constraints. Hence $(SDP_{\preceq})$ becomes

$$\begin{aligned} \max \quad & c_k^t x \\ \text{s.t.} \quad & \sum_{l=1}^k b^l_{ij} x_l + z_{ij} = b_{ij}, i, j = 1, \ldots n. \\ & Z \succeq 0. \end{aligned}$$

Since $x_i' s$ free variables, we use the usual trick to convert them to positive ones by introducing $x_i^+ \geq 0, x_i^- \geq 0$ such that $x_i = x_i^+ - x_i^-$, i.e $x = x^+ - x^-$. Then the above program is turned to:

$$\begin{aligned} \max \quad & c^t(x^+ - x^-) \\ \text{s.t.} \quad & \sum_{l=1}^k b^l_{ij}(x_l^+ - x_l^-) + z_{ij} = b_{ij}, i, j = 1, \ldots n \\ & Z \succeq 0, x_l^+ \geq 0, x_l^- \geq 0. \end{aligned}$$

Finally, we define:

$$C = \begin{bmatrix} O_n & 0 & 0 \\ 0 & \mathrm{diag}(c) & 0 \\ 0 & 0 & -\mathrm{diag}(c) \end{bmatrix}$$

$$A_{ij} = \begin{bmatrix} E_{ij} & 0 & 0 \\ 0 & \mathrm{diag}(a_{ij}^1, \ldots, a_{ij}^k) & 0 \\ 0 & 0 & -\mathrm{diag}(a_{ij}^1, \ldots, a_{ij}^k) \end{bmatrix}, i, j = 1, \ldots, n$$

where O_n is the $n \times n$-zero matrix and E_{ij} has all entries 0 expect 1 at (i, j) and diag(c) is the diagonal matrix with elements c at diagonal. Thus $(SDP_{\preceq})$ is converted to

$$\begin{aligned} \text{mim} \quad & \mathbf{tr}(C^t X) \\ \text{s.t.} \quad & \mathbf{tr}(A_{ij}^t X) = b_{ij}, i, j = 1, \ldots n. \\ & X \succeq 0, \end{aligned}$$

with the variable matrix

$$X = \begin{bmatrix} Z & 0 & 0 \\ 0 & \mathrm{diag}(x^+) & 0 \\ 0 & 0 & \mathrm{diag}(x^-) \end{bmatrix}.$$

Note that C, A_{ij} and X are of size $(n+2k) \times (n+2k)$. Thus $X \in S_+^{n+2k}$. □

Rendel in [24] discusses the duality for the semidefinite programming, and apply it to approximate integer problems. Another strength is based on transformations into semidefinite programming. The following lemma is very powerful in both the process of transformations and in the constructions of primal-dual algorithms.

Lemma 2.5.1. (Schur complement, [2]) Let X be decomposed as follows

$$X = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix} \succeq 0 \Leftrightarrow S = C - B^t A^{-1} B \succeq 0,$$

where the matrices $A, C \succeq 0$, symmetric,and det(A) $\neq 0 \Leftrightarrow A - BC^t B \succeq 0$.

Proof. Apply Definition 2.5.1, and simplify with the rules of algebra:

$$0 \preceq [x, y]^t \begin{bmatrix} A & B \\ B^t & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x, y]^t \begin{bmatrix} Ax + By \\ B^t x + Cy \end{bmatrix} = x^t Ax + 2x^t B^t y + y^t Cy.$$

This is equivalent to $f(x, y) = x^t Ax + 2x^t By + y^t Cy \succeq 0$, where A is positive definite by assumption. Differentiate f w.r.t x, and solve for $x = -A^{-1}By$, and with optimal value $-y^t BA^{-1}B^t y + y^t Cy = y^t(C - BA^{-1}B^t)y$. □

Example 2.5.2. [9] *Derive an expression for the inverse Schur complement.*

$$\begin{bmatrix} A & B \\ B^t & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}.$$

where A is invertible, i.e det$(A) \neq 0$. Consider the following two systems of equations:

$$\begin{aligned} Ax + By &= u \\ B^t x + Cy &= v. \end{aligned}$$

Solve the first equation for x:

$$\begin{aligned} Ax &= u - By \\ x &= A^{-1}(u - By), \end{aligned}$$

and substitute in the second equation:

$$\begin{aligned} B^t x + Cy = B^t(A^{-1}(u - By)) + Cy &= B^t A^{-1} u - A^{-1} B^t B y + Cy \\ &= B^t A^{-1} u + y(C - B^t A^{-1} B) \\ &= B^t A^{-1} u + yS \\ &= v, \end{aligned}$$

where S is the Schur complement of A in x. Solve this equation for y:

$$\begin{aligned} v &= B^t A^{-1} u + yS \\ yS &= v - B^t A^{-1} u \\ y &= S^{-1}(v - B^t A^{-1} u), \end{aligned}$$

and insert the above expression in x:

$$\begin{aligned} x &= A^{-1}(u - By) \\ &= A^{-1}(u - B(S^{-1}(v - B^t A^{-1} u) \\ &= u(A^{-1} + A^{-1} B S^{-1} B^t A^{-1}) - A^{-1} B S^{-1} v. \end{aligned}$$

$$\begin{bmatrix} A & B \\ B^t & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1} B S^{-1} B^t A^{-1} & -A^{-1} B S^{-1} \\ -S^{-1} B^t A^{-1} & S^{-1} \end{bmatrix}.$$

Example 2.5.3. ([9], Ex. 4.40) Transform the second order cone programming to a semidefinite programming problem.

$$\begin{aligned} \min \quad & f^t x \\ \text{s.t.} \quad & ||A_i x + b_i|| \leq c_i^t x + d_i, i = 1, \ldots, m. \end{aligned}$$

Introduce the variable t, and let $||A_i + b_i|| \preceq t^2 I$.

$$\begin{array}{ll} \min & f^t x \\ \text{s.t.} & \begin{bmatrix} (c_i^t x + d_i)I & A_i x + b_i \\ (A_i x + b_i)^t & (c_i^t + d_i)I \end{bmatrix} \succeq 0. \end{array}$$

Example 2.5.4. [15] *Transform the matrix fractional programming to a semidefinite programming problem.*

$$\begin{array}{ll} \min & (Ax+b)^t (A_0 + x_1 A_1 + ... + x_n A_n)^{-1} (Ax+b) \\ \text{s.t.} & A_0 + x_1 A_1 + ... + x_n A_n \succ 0 \\ & x \geq 0, \end{array}$$

where $A(x) = A_0 + x_1 A_1 + ... + x_n A_n$, and the inequality constraint defines a strict matrix inequality, i.e $(A \succ B)$. Apply Definition 2.5.3, and Lemma 2.5.1:

$$\begin{array}{ll} \min & t \\ \text{s.t.} & \begin{bmatrix} A(x) & Ax+b \\ (Ax+b)^t & t \end{bmatrix} \succeq 0. \end{array}$$

Remark. *In the case when $A \succeq B$ then the matrix fractional problem could instead be restated more efficiently as a second order cone programming* [15].

The first appearances of semidefinite programming is the Shannon capacity problem of a graph G, and it is related to efficiency [10, 7].

Definition 2.5.5. (Strong product of graphs, [10]) The strong product of $G_1 \boxplus G_2$ of two graphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ has vertex set

$$V_1 \times V_2 = \{(u_1, u_2) : u_1 \in V_1, u_2 \in V_2\}$$

with $(u_1, u_2) \neq (v_1, v_2)$ adjacent if and only if $u_i = v_i$ or $u_i v_i \in E_i$ for $i = 1, 2$.

Definition 2.5.6. (Shannon capacity of a graph, [10]) The Shannon capacity of a graph $G = (V, E)$ is defined as

$$\vartheta(G) := \mathrm{lim_{r \to \infty}} \alpha(\mathrm{G^r})^{\frac{1}{\mathrm{r}}},$$

where $\alpha(G^r)$ is the maximum number of words of length r in G.

Theorem 2.5.1. (Lovász theta function of G, [10]) Let two graphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ be given. Then

$$\vartheta(G_1 * G_2) = \vartheta(G_1)\vartheta(G_2).$$

Proposition 2.5.2. (The $\vartheta(G)$ is the optimum of the following semidefinite programming problem, [7])

$$\begin{array}{llll} \min & \vartheta(G) = t & & \\ \text{s.t.} & Y \preceq tI & & \\ & Y_{ii} = 1, & i \in V, & \\ & Y_{ij} = 1, & (i,j) \notin E & \end{array} \qquad \begin{array}{llll} \max & \vartheta(G) = \mathbf{tr}(JX) & & \\ \text{s.t.} & \mathbf{tr}(X) = 1 & & \\ & X_{ij} = 0, & (i,j) \in E, & \\ & X \succeq 0, & & \end{array}$$

where the matrix J has all entries equal to 1.

2.5.3 Duality of SDP

Proposition 2.5.3. (Weak duality, [27]) If X is feasible in the primal standard form and (y, S) in the dual then

$$\mathbf{tr}(CX) - b^t y = \mathbf{tr}(XS) \geq 0.$$

Proof. Apply the Definitions 2.5.1, 2.5.2, and 2.5.3:

$$\begin{aligned} \mathbf{tr}(CX) - b^t y &= \mathbf{tr}((\sum_{i=1}^{m} y_i A_i + S)X) - b^t y \\ &= \sum_{i=1}^{m} \mathbf{tr}((A_i X) y_i) + \mathbf{tr}(SX) - b^t y \\ &= \mathbf{tr}(SX) \\ &= \mathbf{tr}(XS) \geq 0 \end{aligned}$$

Since $X \in S^n_+$ and $S \in S^n_+$. $\square$

Theorem 2.5.2. (Slaterś regularity condition, [10]) If there exists feasible point in the primal and dual problem s.t. $X \succ 0$, $S \succ 0$ then $p^* = d^*$.

Zhang, Chen and Zhang in [32] emphasizes nonattainment of optimal value occurs in semidefinite programming, motivated to utilize slaterś constraint qualification. Ramana in [22] have also regarded the duality in semidefinite programming, proposed to examine an exact duality theory, and investigate its consequences. The differences among these duality approaches are embedded in the variation to verify the absence of duality gap.

2.5.4 The duality gap from a geometric point of view

In constrast to LP, it is no longer true that optimality implies that the optimal dual objective value is equal to the optimal value of the primal as shown in Table 1.

So, the optimal value of dual (if feasible) is a lower bound of the optimal value of the primal, called weak duality. The gap between a dual feasible solution (y, S) and a primal feasible solution X is

$$\mathbf{tr}(CX) - \sum_{i=1}^{m} b_i y_i = \mathbf{tr}((\sum_{i=1}^{m} y_i A_i + S)X) - \sum_{i=1}^{m} y_i \mathbf{tr}(A_i X) = \mathbf{tr}(SX) \geq 0$$

If $\mathbf{tr}(SX) = 0$, then this primal-dual pair is an optimal solution. Unlike LP, it is no longer true that optimality implies $\mathbf{tr}(SX) = 0$. Consider the following example:

Example 2.5.5. (A finite duality gap by a straight algebraic derivation of the dual problem, [13])

$$\begin{aligned} \min \quad & x_{12} \\ \text{s.t.} \quad & \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0. \end{aligned}$$

In matrix form

$$C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Dualization yields

$$\begin{aligned} \max \quad & y_1 \\ \text{s.t.} \quad & Z = C - y_1 A_1 - y_2 A_2 - y_3 A_3 - y_4 A_4 \succeq 0. \end{aligned}$$

$$\begin{aligned} \Leftrightarrow \max \quad & y_1 \\ \text{s.t.} \quad & Z = \begin{bmatrix} -y_2 & \frac{1+y_2}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0. \end{aligned}$$

If the primal matrix is PSD, then $x_{12} = 0$, because $x_{11} = 0$. In the same manner, $z_{22} = 0$ implies that $y_1 = -1$ in the dual problem. The duality gap is 1.

Next, we show the gap is closed if a smart transformation is made to an equivalent problem. If $X \succeq 0$ then $X = PWP^t \succeq 0$, for $W \succeq 0$, where we choose for example $P = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$. The choice of P will be discussed later. So replacement of $X \succeq 0$ by $\tilde{X} = PWP^t \succeq 0$ with $W \succeq 0$ does not change the primal problem, since

$$\tilde{X} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & w_{11} & w_{12} \\ 0 & w_{21} & w_{22} \end{bmatrix},$$

and $\mathbf{tr}(C\tilde{X}) = 0, \mathbf{tr}(A_1\tilde{X}) = w_{22}, \mathbf{tr}(A_2\tilde{X}) = 0, \mathbf{tr}(A_3\tilde{X}) = 0, \mathbf{tr}(A_4\tilde{X}) = 2w_{12} = 0$. This yields the optimal solution $\tilde{X}^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & w_{11} & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Now dualizing min $\mathbf{tr}(C\tilde{X})$ s.t $\mathbf{tr}(A_1\tilde{X}) = 1, \mathbf{tr}(A_2\tilde{X}) = 0, \mathbf{tr}(A_3\tilde{X}) = 0, \mathbf{tr}(A_4\tilde{X}) = 0, \tilde{X} \succeq 0$. We have

$$\begin{aligned} \max \quad & y_1 \\ \text{s.t.} \quad & Z = \begin{bmatrix} -y_2 & \frac{1+y_2}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0. \end{aligned}$$

So, Z must be psd with respect to the subspace spanned by the columns of P. So

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -y_2 & \frac{1+y_1}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -y_4 \\ -y_4 & -y_1 \end{bmatrix} \succeq 0$$

Showing that $y_1 = 0$. Now the duality gap is 0. This shows that a pure algebraic derivation of the dual problem, as done in the beginning is not sufficient, we have to consider the geometry of the feasible set. This motivates the study of *faces* of semidefinite cones. To make the later section on first regulaizaton method meaningful for SDP we are going to give a more detail discussion on faces of semidefinite cones. As demostrated above, it has heavy geometric arguments. We carry out most properties by matrix theory.

2.5.5 Characterization of faces of the semidefinite cone

We start by the cone of the nonnegative orthant in $\mathbf{R}^n$. Then we discuss the cone of positive semidefinite matrices.

Definition 2.5.7. (A nice cone and face, [21]) A closed convex cone K is called nice, if the set $K^* + F^{\perp}$ is closed for all $F \lhd K$.

Example 2.5.6. (The nonnegative orthant $\mathbf{R}^n_+$) Let $x \in \mathbf{R}^n_+$. Then the minimal face of $\mathbf{R}^n_+$ contaning $\{x\}$ is the set

$$F(x, \mathbf{R}^n_+) := \{y \in \mathbf{R}^n_+ : y_i = 0, \forall i \quad such \quad that \quad x_i = 0\}$$

This face can be transformed to the form $F(\begin{bmatrix} e \\ 0 \end{bmatrix}, \mathbf{R}^n_+)$, by permutation of components, for an e appropriate size.

Example 2.5.7. (Semidefinite cone S^n_+). For $P \in S^n_+$, the minimal face of S^n_+ consisting P is the set

$$F(P, S^n_+) := \{X \in S^n_+ : \mathcal{R}(X) \subseteq \mathcal{R}(P)\}$$

where $\mathcal{R}(\cdot)$ is the range space (or image) of the matrix.

The proof of this characterization of minimal faces is much involved in e.g.[4]. But it is useful. This characterization can be reformulated as follows:

Theorem 2.5.3. *F is a face of S^n_+ if and only if*

$$F = \{O_{m\times n}\} \quad or \quad F = \{X : X = PWP^t, W \in S^k_+\}$$

for some $k \in \{1, 2, \ldots, n\}$, $P \in \mathbf{R}^{m\times k}$ with rank k.

Remark. *This theorem explains why we choose such a matrix $P = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ in the motivating Example 2.5.5.*

Since any matrix $A \in S^n_+$ can be transformed by a full rank assuming k matrix V such that V^tAV in the form $\begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix}$ for $x \in S^k_+$, the face can be in the form

$$F = \{\begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix} x \in S^k_+\}.$$

Here we give an elementary proof.

Proof. Observe that F is a convex cone in S^n_+. Next observation is we can choose P with orthonormal columns, so $P^tP = F_k$. Now, suppose $X, Y \in S^n_+$ such that $X + Y \in F$. Then $X + Y = PWP^t$ for some $W \in S^k_+$. Then $P^tXP + P^tYP = W$ and $P^tXP \succeq 0$ and $P^tYP \succeq 0$ since $X, Y \in S^n_+$. These inequalities imply that

$$P^tXP = B_X \quad \text{and} \quad P^tYP = B_Y$$

for some $B_X, B_Y \in S^k_+$ therefore $X, Y \in F$ so $F \lhd S^n_+$. $\square$

Definition 2.5.8. (Exposed cone) A face $F \lhd K$ is exposed if there is a $\phi \in K^*$ such that

$$F = \{x \in K : \langle \phi, x \rangle = 0\} = k \cap \{\phi\}^\perp$$

A cone is facially exposed if every face $F \lhd K$ is exposed. We need following result to describe exposed face containing the face F.

Proposition 2.5.4. *Let K and F be arbitrary convex cones. If $F \subseteq K$, then $F^c := F^\perp \cap K^* \lhd K^*$. We call F^c the conjugate face.*

Proof. Clearly $F^c \subseteq K^*$ by definition. Since K^* and $F^\perp$ are convex cones, the intersection of them is a convex cone. Thus F^c is a convex cone. To show F^c is a face of K^*, pick up $x, y \in K^*$ such that $x + y \in F^c$ (since K^* is convex), i.e $\langle x + y, z \rangle = 0$ for $z \in F$. Then $z \in F \subset K$ implies $\langle x, z \rangle \geq 0, \langle y, z \rangle \geq 0$ and $0 = \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$. Hence $\langle x, z \rangle = \langle y, z \rangle = 0, \forall z \in F$ so $x, y \in F^\perp$. Now, $x, y \in K^*$, we have $x, y \in F^c$. Thus $F^c \lhd K^*$. $\square$

Proposition 2.5.5. *If $F \lhd K, \phi \in F^c$, then*

$$F \lhd K \cap \{\phi\}^\perp \lhd K.$$

Remark. *This result shows that each point in F^c defines an exposed face containing F.*

Proof. Since F is a face, $F \subseteq K$, $\phi \in F^c$ implies $\phi \in F^\perp$. So, $F \subseteq K \cap \{\phi\}^\perp$ which is clearly a convex cone. Let $x, y \in K \cap \{\phi\}^\perp$ such that $x + y \in F$. Since $x, y \in K$ and $F \lhd K$, we have $x, y \in F$. Showing that $F \lhd K \cap \{\phi\}^\perp$.

Next, we show that $K \cap \{\phi\}^\perp$ is a face of K. Note that $K \cap \{\phi\}^\perp \subset K$. Let $x, y \in K$ such that $x + y \in K \cap \{\phi\}^\perp$. Then $\phi \in K^*$ implies that $\langle \phi, x \rangle \geq 0, \langle \phi, y \rangle \geq 0$ and $0 = \langle \phi, x+y \rangle = \langle \phi, x \rangle + \langle \phi, y \rangle$ so $\langle \phi, x \rangle = \langle \phi, y \rangle = 0$ proving that $x, y \in K \cap \{\phi\}^\perp$, hence $K \cap \{\phi\}^\perp \lhd K$. $\square$

Now, we can rephrase Theorem 7.1 in [8]. To make it comparable with Farkas lemma , we set up more compact notation. We define the linear map $\mathcal{A} : S^n \to \mathbf{R}^m$ by

$$\mathcal{A}X = \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix}$$

and the adjoint operator to $\mathcal{A}$, denoted by $\mathcal{A}^* : \mathbf{R}^m \to S^n$ defined by $\mathcal{A}^* y = \sum_{i=1}^m y_i A_i$.

Theorem 2.5.4. *Exactly one of the following systems has solution:*

I. $0 \neq X \succeq 0, \mathcal{A}X = 0, \langle C, X\rangle \leq 0.$

II. $\mathcal{A}^* y \prec C.$

Furthermore, if there exists $X \in S^n_+$ such that $\mathcal{A}X = 0$ and $\langle C, X\rangle < 0$, then the system $\mathcal{A}^* X \preceq C$ is also infeasible.

Proof. Suppose there exists $0 \neq X \succeq 0$ such that $\mathcal{A}X = 0$ and $\langle C, X\rangle \leq 0$. Assume contradiction there is $y \in \mathbf{R}^m$ s.t. $\mathcal{A}^* y \prec C$. Then

$$0 < \langle C - \mathcal{A}^* y, X\rangle = \langle C, X\rangle - \langle \mathcal{A}^* y, X\rangle = \langle C, X\rangle \leq 0.$$

This is a contradiction so there is no such $y \in \mathbf{R}^m$ such that $\mathcal{A}^* y \prec C$. Suppose that $C - \mathcal{A}^* y \in S^n_{++}$ for all $y \in \mathbf{R}^m$. First we note that $C \notin S^n_{++} + \mathcal{A}^*(\mathbf{R}^m)$. Since $S^n_{++} + \mathcal{A}^*(\mathbf{R}^m)$ is a convex cone, by hyperplane separation there exists $X \neq 0$ and $\beta \in \mathbf{R}$ such that

$$\langle C, X\rangle \leq \beta \leq \langle S + \mathcal{A}^* y, X\rangle, \forall S \in S^n_{++}, y \in \mathbf{R}^m$$

Taking $y = 0$ we have $X \succeq 0$. To show this, we assume contradiction that $X \notin S^n_+$, that is, there is a $v \neq 0$ such that $v^t X v < 0$, then take $S = tvv^t + I \succ 0$ for $t > 0$, we have

$$\begin{aligned} \beta &\leq \langle tvv^t + I, X\rangle \\ &= \langle tvv^t, X\rangle + \langle I, X\rangle \\ &= tv^t xv + \langle I, X\rangle \end{aligned}$$

by the property of the trace (i.e. $\mathbf{tr}(AB) = \mathbf{tr}(BA)$). Letting $t \to \infty$ yields $tV^t XV + \langle I, X\rangle \to -\infty$ so $\beta \leq -\infty$, contradicting $\beta \geq \langle C, X\rangle$ a fixed number so X must be PSD.

Taking $S = \frac{1}{t}I, y = -t\mathcal{A}X$ for $t > 0$, we would have $\mathcal{A}X = 0$. Otherwise

$$\beta \leq \langle \frac{1}{t}I + \mathcal{A}^*(-t\mathcal{A}X), X\rangle = \frac{1}{t}\langle I, X\rangle - t(\mathcal{A}X, X) \to -\infty \quad \text{as} \quad t \to \infty,$$

a contradiction. If we choose $S = \frac{1}{t}I$ for $t > 0$ and $y = 0$, we have $\langle C, X\rangle \leq 0$, because

$$\langle C, X\rangle \leq \beta \leq \langle \frac{1}{t}I, X\rangle \to 0 \quad \text{as} \quad t \to \infty$$

Therefore, there is $X \neq 0, X \succeq 0$ such that $\mathcal{A}X = 0$ and $\langle C, X\rangle \leq 0$.

Finally, assume there is an $X \in S^n_+, \mathcal{A}X = 0, \langle C, X\rangle < 0$. To show the last statement we assume again contradiction that there is $\tilde{y}$ such that $\mathcal{A}^*\tilde{y} \preceq C$. Set $\tilde{S} := C - \mathcal{A}^*\tilde{y}$. Then $\tilde{S} \succeq 0$. Consequently

$$0 > \langle C, X\rangle = \langle \tilde{S} + \mathcal{A}^*\tilde{y}, X\rangle = \langle \tilde{S}, X\rangle + \langle \tilde{y}, AX\rangle = \langle \tilde{S}, X\rangle$$

which is impossible. $\square$

Since the conjugate face can be viewed as a collection of exposed faces (Proposition 2.5.4) it is our intention to give such a description of the conjugate face of the minimal face of a feasible set of SDP problem. The theorem above provides such a possibility.

Proposition 2.5.6. *Let $F_D := face(\mathcal{F}_D)$ where*

$$F_D := \{S \in S^n_+ : S = C - \mathcal{A}^*y, \quad for \quad some \quad y \in \mathbf{R}^m\}.$$

If $\mathcal{F}_D \neq \emptyset$, then face $(\{X \in S^n_+ : \mathcal{A}X = 0, \langle C, X\rangle = 0\}) \lhd F^c_D$. Here $face(\mathcal{F}_D)$ and the similar stands for the minimal face of $\mathcal{F}_D$.

Proof. Assume $X \in S^n_+$ such that $\mathcal{A}X = 0$ and $\langle C, X\rangle = 0$. If $S = C - \mathcal{A}^*y \in \mathcal{F}_D$, then compute $\langle S, X\rangle$

$$\begin{aligned}\langle S, X\rangle &= \langle C - \mathcal{A}^*y, X\rangle = \langle C, X\rangle - \langle \mathcal{A}^*y, X\rangle \\ &= \langle C, X\rangle - \langle y, \mathcal{A}^*X\rangle = \langle C, X\rangle = 0.\end{aligned}$$

Therefore $\langle S, X\rangle = 0$, $\forall S \in \mathcal{F}_D$, implying $X \in \mathcal{F}_D^\perp$. Now $\mathcal{F}_D \neq \emptyset$. So there is some $\bar{S} \in ri(\mathcal{F}_D)$. We have for $\bar{S}$ the decomposition $\bar{S} = U\Lambda U^t$ with $U \in \mathbf{R}^{m\times k}$ whose columns are orthonormal and Λ being diagonal in S^k_{++}. Then $F_D = US^k_+U^t$. Further, $0 = \bar{S}X = U\Lambda U^tX$ implies $\Lambda U^tX = 0 \Rightarrow U^tX = 0 \Rightarrow X \in F_D^\perp \Rightarrow X \in F^c_D$. Hence

$$\{X \in S^n_+ : \mathcal{A}X = 0, \langle C, X\rangle = 0\} \subseteq F^c_D$$

so

$$face(\{X \in S^n_+ : \mathcal{A}X = 0, \langle C, X\rangle = 0\}) \lhd F^c_D.$$

$\square$

An immediate consequence is the following result:

Corollary 2.5.1. *Let $F_D := face(\mathcal{F}_D)$. If $\mathcal{F}_D \neq \emptyset$ and $X \in S^n_+$ satisfying $\mathcal{A}X = 0$ and $\langle C, X\rangle = 0$, then*

$$F_D \lhd S^n_+ \cap \{X\}^\perp = US^k_+U^t,$$

where $U \in \mathbf{R}^{n\times k}$ is of full column rank and $\mathcal{R}(U) = \mathcal{N}(X)$.

Proof. By Proposition 2.5.6, $X \in F^c_D$. By Proposition 2.5.5, $F_D \lhd S^n_+ \cap \{X\}^\perp$. It remains to prove

$$S^n_+ \cap \{X\}^\perp = US^k_+U^t$$

with the required properties for U. Let $F := US^k_+U^t$. To process, we prove the following things:

(a) Let $U \in \mathbf{R}^{n\times k}$. Then

$$Y \in US^k_+U^t \Leftrightarrow Y \succeq 0 \quad \text{and} \quad \mathcal{R}(Y) \subseteq \mathcal{R}(U).$$

(b) $F = \{X \in S^n_+ : \mathcal{R}(X) \subseteq \mathcal{R}(U)\} = \{X \in S^n_+ : \mathcal{R}(V) \subseteq \mathcal{N}(X)\}$ where V is such that the matrix (U, V) is orthogonal.

(c) $ri(F) = US^k_+U^t = \{X \in S^n_+ : \mathcal{R}(X) = \mathcal{R}(U)\}; ri(F^c) = VS^{n-k}_{++}V^t = \{Y \in S^n_+\}$.

Proof of (a): Assume $Y \in US^k_+U^t$. Then $Y \succeq 0$. For any $y \in \mathcal{R}(Y)$ i.e. there exists an $x \in \mathbf{R}^n$ such that

$$y = Yx = USU^tx \in \mathcal{R}(U) \quad \text{for} \quad S \in S^k_+.$$

So $\mathcal{R}(Y) \subseteq \mathcal{R}(U)$. Now, assume $Y \succeq 0$ and $\mathcal{R}(Y) \subseteq \mathcal{R}(U)$. Since $Y \succeq 0$ then there exists V such that $Y = VV^t$ with $V \succeq 0$ and obviously $\mathcal{R}(V) = \mathcal{R}(Y)$ so there exists $\Phi \in \mathbf{R}^{k\times n}$ such that $V = U\Phi$ so that $Y = UZU^t$ with $Z := \Phi\Phi^t \in S^k_+$ from which we conclude that $Y \in US^k_+U^t$.

Proof of (b): The first equality follows from (a). The second follows from the fact that $\mathcal{R}(X) \subset \mathcal{R}(U) \Leftrightarrow \mathcal{R}(U)^\perp \subset \mathcal{R}(X)^\perp \Leftrightarrow \mathcal{R}(V) \subset \mathcal{N}(X)$.

Proof of (c): Assume $X = U\Sigma U^t \in ri(F)$. Let $Y = UU^t$. Then $Y \in F$. So there is $\mu > 1$ such that $(1-\mu)Y + \mu X \in F$ [25]. Hence there is $Z \in S^k_+$ such that

$$UZU^t = (1-\mu)Y + \mu X = U((1-\mu)I + \mu\Sigma)U^t$$

Since $U^tU = I$, we have $Z = (1-\mu)I + \mu\Sigma \Leftrightarrow \Sigma = \frac{1}{\mu}Z + \frac{\mu-1}{\mu}I \in S^n_{++}$. Therefore, $X \in US^k_{++}U^t$. Since U has full column rank, by (a) we have $\mathcal{R}(X) = \mathcal{R}(U)$.

Conversely, assume $X \in S_+^n$ such that $\mathcal{R}(X) = \mathcal{R}(U)$. We know that U is of full column rank by (a), $X = U\Sigma U^t$ for some $\Sigma \in S_{++}^k$. To prove $X \in ri(F)$, we can prove that for all $Y \in F$, there is some $\mu > 1$ such that $(1-\mu)Y + \mu X \in F$. Let $Y = U\Phi U^t \in F$. Obviously $S_{++}^n = int(S_+^n)$. If $\Sigma \in S_{++}^k, \Phi \in S_+^k$, there is some $\mu > 1$ such that $(1-\mu)\Phi + \mu\Sigma \in S_+^k$. Hence

$$(1-\mu)Y + \mu X = U((1-\mu)\Phi + \mu\Sigma)U^t \in US_+^k U^t = F$$

implying $X \in ri(F)$. The other part is proved in similar manner. Now assume $X = V\Phi V^t \in ri(F^c)$. If $Y = U\Sigma U^t \in F$ then $YX = 0$ since $U^tV = 0$. Then $Y \in S_+^n \cap \{X\}^\perp$. On the other direction, $Y \in S_+^n \cap \{X\}^\perp$. Then $V\Phi V^t Y = XY = 0 \Rightarrow V^tY = 0$ due to the fact $V^tV = I$ and $I \succ 0 \Rightarrow \mathcal{R}(Y) \subseteq \mathcal{N}(V^t) = \mathcal{R}(U)$. By (b) we have $Y \in F$ so $F = S_+^n \cap \{X\}^\perp$

$\square$

2.6 Efficiency

It has almost taken ten years to achive the efficiency with interior point methods [10]. Several studies have discussed the interior point method and efficiency under the composed name *“state of the art”* [10, 16]. Thus a series of improvments of several algorithms have reinforced aspects such as control, and stability.

Lustig, Marsten, and Shanno in [16] highlight the implementation process to develop a successful code. An efficient algorithm or code is indeed a technical priori, and to match the input and output optimal.

Vandenberghe, Boyd [28] have also noted the objective efficiency in relation to interior point methods, and divide efficiency into three levels. The first level *”practical efficiency“* corresponds to a competitive factor. This level represents an effective approach for small, medium, and large scale problems in finite steps. The second level is called *”theoretical efficiency“*, and it responds to semidefinite programming, interior point methods based on a worse case analysis. The last level considers each solution iteration step individually, and is realized as *”ability to exploit problem structure“* [28].

To accomplish zero duality gap is related to regularization methods [8, 17]. For instance, Malick et al [17] use the quadratic term to handle the aspect stability, and Borwein, Wolkowicz [8] apply instead succesive problem reduction. Another argued approach for these reasons is for instance the extended Lagrange slater dual [32, 23]. A third approach consider consequently an exact duality theory [22]. These three approaches demonstrate the theoretical and methodological perspectives to handle the duality gap efficiently.

3 Regularization methods

This section presents two regularization methods. The first method is based on abstract convex analysis, and the second utilizes the semidefinite programming. These methods have in common the primal regularization technique.

3.1 Abstract convex regularization

Borwein and Wolkowicz in [8] present an algorithm to regularize the primal problem based on abstract convex analysis. The main idea with the algorithm is to transform the primal problem into a new primal problem on an exposed subface contained in the minimal cone. Hence, the new primal problem validated the extended slaterś constraint qualification. Lustig, Marsten, and Shannon in [16] highlight problem size reduction as an important factor for large scale problems. Wolkowicz in [30] validated another approach regarding problem size, to regularize by adding or substitution of a finite number of linear constraints.

The algorithm in [8] consists of determining two cones, the minimal cone of the feasible set S^f, and the cone of direction of constancy of g $D_g^{=}(S^f)$. The algorithm holds in the case g is weakly faithfully convex. Although g is not weakly faithfully convex, the algorithm works as well by adding an additional condition. In the case of faithfully S-convex function, the algorithm is modified. Furthermore, in the case of the not weakly faithfully convex and the faithfully S-convex adjustments required affect the algorithm speed [8].

Borwein and Wolkowicz in [8] have also clarified the limitations of the regularization method. The first limitation is concerned with the new primal problem, the extended slaters constraint qualification which holds valid for subfaces. Secondly, if the optimal point does not satisfy the Kuhn-Tucker condition, then it could affect the stability [8].

Before we present the algorithm, regard some notations. The annihilator to the cone K is marked by $K^{\perp} = K^{+} \cap (-K^{+})$, where K^{+} is the nonnegative dual cone. The generalized inverse of the matrix A is expressed as $A^{\top}$. In this section we also use $\phi^{\perp}$ instead of $\{\phi\}^{\perp}$. We consider explicit the real case of analytic functions in several variables.

3.1.1 Algorithm I

The following algorithm describes how to determine the minimal cone, cones of directions of constancy. Assume g is weakly faithfully convex. The iterations are repeated until all $m_i = 0$ [8].

Initialization step Let $\bar{x}$ be an optimal point in the feasible set. Set

$$\begin{aligned} \Omega_0 &= \Omega - \bar{x}, \\ m_0 &= \dim \mathbf{R}^m, \\ Q_0 &= I_{m_0 \times m_0}, \\ S_0 &= S, \\ n_0 &= \dim \mathbf{R}^n, \\ P_0 &= I_{n_0 \times n_0}, \\ i &= 0, \end{aligned}$$

and preceed to main step.

Main step i-th step ($0 \leq i \leq t$). If $m_i > 0$ consider the system

$$\begin{gathered} \Omega_i^+ \cap [\partial \phi_i Q_i g(\bar{x})] P_i \neq \emptyset \\ \phi_i Q_i g(\bar{x}) = 0, \quad 0 \neq \phi_i = S_i^+. \end{gathered}$$

(a) If the system is consistent, use algorithm A to find n_i-by-n_{i+1} matrix A_{i+1} satisfied by:

$$\mathcal{R}(A_{i+1}) = D^{=}_{(\phi_i Q_i g) P_i}$$

Then set:

$$\begin{aligned} m_{i+1} &= m_i - 1 = \dim \phi_i^{\perp} \\ B_{i+1} &: \phi_i \xrightarrow{onto} \mathbf{R}^{m_i+1} \\ \mathcal{N}(B_{i+1}) &= \operatorname{span}\{\phi_i\}, \text{with} \quad B_{i+1} = [1], \quad \text{if} \quad m_{i+1} = 0 \\ P_{i+1} &= P_i A_{i+1} \\ Q_{i+1} &= B_{i+1} Q_i \\ E_i &= S_i \cap \phi_i^{\perp} \\ S_{i+1} &= B_{i+1} E_i \\ \Omega_{i+1} &= A_{i+1}^{+} \{\Omega_i \cap \mathcal{R}(A_{i+1})\} \end{aligned}$$

and iterate the main step $i+1$.

(b) If the system is inconsistent or $m_i = 0$ stop.

$$\begin{aligned} S^f &= B_1^+ B_2^+ \dots B_i^+ S_i \\ D_g^{=}(S^f) &= \mathcal{R}(P_i). \end{aligned}$$

The regularized primal and dual problem becomes:

$$\begin{aligned} \inf \quad & f(x) \\ \text{s.t} \quad & g(x) \preceq_{S_t} 0 \\ & x \in \Omega_t \end{aligned}$$

where,

$$S_t = S^f, \phi_i \in S_i^t, S_{i+1} = (S_i \cap \phi_i^{\perp}), \Omega_t = (x + \cap_{i=1}^t D_{\phi_i g}^{=}) \cap \Omega.$$

The associated regularized dual is given by:

$$\begin{aligned} \sup \quad & L^H(\lambda) \\ \text{s.t} \quad & \lambda \in (S^f)^+, \end{aligned}$$

and the restricted Lagrangian is $L^H(\lambda) = \inf\{f(x) + \lambda g(x) : x \in \hat{x} + D_{\phi g}^{=}\}$.

Example 3.1.1. [8, 30] Let $S = \mathbf{R}_+^7$, $g : \mathbf{R}^5 \to \mathbf{R}^7$. Determine S^f and $D_g^{=}$.

$$\begin{aligned} g_1(x) &= e^{x_1} + x_2^2 - 1 \leq 0 \\ g_2(x) &= x_1^2 + x_2^2 + e^{-x_3} - 1 \leq 0 \\ g_3(x) &= x_1 + x_4^2 + x_5^2 - 1 \leq 0 \\ g_4(x) &= e^{-x_2} - 1 \leq 0 \\ g_5(x) &= (x_1 - 1)^2 + x_2^2 - 1 \leq 0 \\ g_6(x) &= x_1 + e^{-x_4} - 1 \leq 0 \\ g_7(x) &= x_2 + e^{-x_5} - 1 \leq 0, \\ & x_1, x_2, x_3, x_4, x_5 \geq 0. \end{aligned}$$

We begin by consider the set of active inequality constraints g_1, g_3, g_4, g_5, and an optimal point $\bar{\mathbf{x}} = (0, 0, 1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Next, compute the active constraints partial derivatives at the optimal point:

$$\nabla g_1(\bar{\mathbf{x}}) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \nabla g_3(\bar{\mathbf{x}}) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} \quad \nabla g_4(\bar{\mathbf{x}}) = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \nabla g_5(\bar{\mathbf{x}}) = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Initialization step

According to initialization step, we define the following:

$$
\begin{aligned}
\Omega_0 &= \mathbf{R}^5,\\
m_0 &= 7,\\
Q_0 &= I_{7\times 7},\\
S_0 &= \mathbf{R}^7_+,\\
n_0 &= 5,\\
P_0 &= A_0 = I_{5\times 5},\\
i &= 0.
\end{aligned}
$$

Step 0

Solve the following system of equations:

$$
\lambda_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} + \lambda_4 \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_5 \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},
$$

where $\lambda_1 + \lambda_3 + \lambda_4 + \lambda_5 = 1, \lambda_k \geq 0.$

$$
\begin{cases}
\lambda_1 + \lambda_3 - 2\lambda_5 = 0 \\
-\lambda_4 = 0 \\
\sqrt{2}\lambda_3 = 0 \\
\sqrt{2}\lambda_3 = 0,
\end{cases}
$$

which provides $\lambda_3 = 0, \lambda_4 = 0, \lambda_5 = \frac{1}{3}, \lambda_1 = \frac{2}{3}$ for $i = 1, 3, 4, 5.$

$$
\phi_0^t = (\frac{2}{3}, 0, 0, 0, \frac{1}{3}, 0, 0)
$$

$$
\Rightarrow [\phi_0^t Q_0 g] P_0 = \phi_0^t \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ g_6 \\ g_7 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
$$

$$= \left[\tfrac{2}{3},0,0,0,\tfrac{1}{3},0,0\right] \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ g_6 \\ g_7 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{2}{3}g_1 + \frac{1}{3}g_5.$$

Next, select $P_1 = P_0A_1 = A_0A_1 = A_1$ such that

$$P_1 = A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Q_1 = B_1Q_0 = B_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\begin{aligned} E_0 &= \{s = (s_i) \in \mathbf{R}^7_+ : s_1 = s_5 = 0\} \\ S_1 &= \{s = (s_i) \in \mathbf{R}^6_+ : s_4 = 0\} \\ \Omega_1 &= \mathbf{R}^3. \end{aligned}$$

Step 1

The vector $\phi_1^t = (0,0,1,0,0,0)$ solves the main step system.

$$[\phi_1^t Q_1 g]P_1 = [g_4]P_1.$$

Then

$$A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$P_2 = P_1A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} E_1 &= \{s = (s_i) \in \mathbf{R}_+^6 : s_3 = s_4 = 0\} \\ S_2 &= \{s = (s_i) \in \mathbf{R}_+^5 : s_3 = 0\} \\ \Omega_2 &= \mathbf{R}^3. \end{aligned}$$

Step 2

The vector $\phi_2^t = (0, 0, 1, 0, 0)$ solve the main step system.

$$[\phi_2^t Q_2 g] P_3 = 0.$$

Then

$$A_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$P_3 = P_2$$

$$B_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\begin{aligned} E_2 &= S_2 \\ S_3 &= \mathbf{R}_+^4 \\ \Omega_3 &= \mathbf{R}^3. \end{aligned}$$

Step 3

$$\phi_3^t = (\lambda_i) \in S_3^\perp.$$

The main step system becomes $\lambda_i \geq 0$,

$$\begin{aligned}
&0 = \phi_3^t Q_3 g(\bar{x}) = (\lambda_1 g_2 + \lambda_2 g_3 + \lambda_3 g_6 + \lambda_4 g_7)(\bar{x}) \\
&0 \in [\partial \phi_3^t Q_3 g(\bar{x})] P_3 = [(\lambda_1 \nabla g_2 + \lambda_2 \nabla g_3 + \lambda_3 \nabla g_6 + \lambda_4 \nabla g_7)(\bar{x})] P_3,
\end{aligned}$$

and since g_2, g_6, g_7 is in the complement of the active constraints, we get:

$$\begin{aligned}
\lambda_1 = \lambda_3 = \lambda_4 = 0, \quad \lambda_2 > 0 \\
0 = [\lambda_2 \nabla g_3(\bar{x})] P_3,
\end{aligned}$$

which is *inconsistent*.

Conclusion

The minimal cone:

$$\begin{aligned}
S^f &= B_0^+ B_1^+ B_2^+ B_3^+ S_3 \\
&= B_1^t B_2^t B_3^t \mathbf{R}_+^4 \\
&= \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \{s = (s_i) \in \mathbf{R}_+^7 : s_1 = s_4 = s_5 = 0\}.
\end{aligned}$$

The cone of direction of constancy:

$$D_g^{=}(D^f) = \mathcal{R}(P_3) = \{d = (d_i) \in \mathbf{R}^5 : d_1 = d_2 = 0\}.$$

3.1.2 Facial reduction in SDP

For SDP the facial is much easier. Now we turn to facial reduction in SDP. We consider the conic standard form

$$\begin{array}{ll} \min & \langle C, X\rangle \\ \text{s.t.} & \langle A_i, X\rangle = b_i, i = 1, ..., m \\ & X \in S^n_+ \end{array}$$

where $A_i \in S^n, i = 1, \ldots, m$ are linearly independent, $b \in \mathbf{R}^m$, and $C \in S^n$. Let $U \in \mathbf{R}^{n\times k}$ have full column rank such that

$$face(\{X \in S^n_+ : \langle A_i, X\rangle = b_i, \forall i\}) = US^k_+U^t$$

Then the (SDP_c) is equivalent to the strictly feasible problem

$$\begin{array}{ll} \min & \langle C, UZU^t\rangle \\ \text{s.t.} & \langle A_i, UZU^t\rangle = b_i, i = 1, ..., m \\ & Z \in S^n_{++} \end{array}$$

$$\begin{array}{ll} \Leftrightarrow \min & \langle UCU^t, Z\rangle \\ \text{s.t.} & \langle UA_iU^t, Z\rangle = b_i, i = 1, ..., m \\ & Z \in S^n_{++} \end{array}$$

Set $\tilde{C_i} := U^tCU, \tilde{A_i} := U^tA_iU$, we have

$$\begin{array}{ll} \min & \langle \tilde{C}, Z\rangle \\ \text{s.t.} & \langle \tilde{A_i}, Z\rangle = b_i, i = 1, ..., m \\ & Z \in S^n_{++} \end{array}$$

Let $\mathcal{A} : S^n \to \mathbf{R}^m$ be defined by $\mathcal{A}X = \begin{bmatrix} \langle \tilde{A_1}, Z\rangle \\ \vdots \\ \langle \tilde{A_m}, Z\rangle \end{bmatrix}$. By Theorem 2.5.4, if (SDP_c) is feasible then it is not strictly feasible if and only if there is a $y \in \mathbf{R}^m$ such that $\mathcal{A}^*y \neq 0$ and $\mathcal{A}^*y \succ 0$ and $b^ty = 0$. Furthermore, if $y \in \mathbf{R}^m$ is such that $\mathcal{A}^*y \neq 0$ and $\mathcal{A}^*y \succeq 0$ and $b^ty = 0$ then

$$face\{X \in S^n_+ : \mathcal{A}X = b\} \lhd S^n_+ \cap \{\mathcal{A}^*y\}^\perp = US^k_+U^t \lhd S^m_+$$

where $U \in \mathbf{R}^{n\times k}$ is of full column rank and $\mathcal{R}(U) = \mathcal{N}(\mathcal{A}^*y)$. If the set of matrices $\{\tilde{A_i}, i = 1, \ldots, m\}$ is linearly dependent, we can choose a maximal subset $I \subset \{1, 2, \ldots, m\}$ such that $\{A_i\}_{i\in I}$ is linearly independent.

Assume that the number of I is l. Then $\bar{b} \in \mathbf{R}^l$. Therefore the (SPD_c) is reduced to:

$$\begin{aligned} \min \quad & \langle \bar{C}, Z \rangle \\ \text{s.t.} \quad & \bar{\mathcal{A}} Z = \bar{b} \\ & Z \in S_+^k \end{aligned}$$

where $\bar{A} : S^k \to \mathbf{R}^l$ is defined by $\bar{\mathcal{A}} Z = \begin{bmatrix} \langle \bar{A}_1, Z \rangle \\ \vdots \\ \langle \bar{A}_l, Z \rangle \end{bmatrix}$.

In conclusion, facial reduction is a method of regularization of bad SDP problem, i.e. we can close the duality gap by solving the problem on a smaller positive semidefinite cone.

3.2 Quadratic regularization

Malick et al in [17] present a regularization algorithm for standard semidefinite programming, and it fits for several classes of large scale problems. A general regularization algorithm is created by combining two separate regularize algorithms. These parts assign to regularize the primal semidefinite problem by Moreeau Yosida regularization, and its dual by the augmented Lagrangian method. Both these approaches are based on quadratic regularization, and are equivalent [17].

The idea to apply quadratic regulariztion to linear semidefinite programming is to stabilize the problems. In addition, under certain conditions the augumented dual Lagrangian function concides with the dual of the Moreau Yosida regulariztion [17].

We begin to consider the equivalent SDP problem [17]:

$$\begin{aligned} \min \quad & \langle C, X\rangle + \frac{1}{2t}||X - Y||^2 \\ \text{s.t.} \quad & \langle A_i, X\rangle = b_i, i = 1, \ldots, m, X \succeq 0, Y \in S_n. \end{aligned}$$

Simplfy the norm $||\cdot||$:

$$||X - Y||^2 = \langle X - Y, X - Y\rangle = \langle X, X\rangle - 2\langle X, Y\rangle + \langle Y, Y\rangle.$$

Insert in the Moreau Yosida regularization $F_t(Y)$, and let $t > 0$:

$$\begin{aligned} F_t(Y) &= \min\{\langle C, X\rangle + \frac{1}{2t}(\langle X, X\rangle - 2\langle X, Y\rangle + \langle Y, Y\rangle) : X \succeq 0, \mathcal{A}X = b\} \\ &= \min\{\langle C, X\rangle + \frac{\langle X, X\rangle}{2t} - \frac{\langle X, Y\rangle}{t} + \frac{\langle Y, Y\rangle}{2t} : X \succeq 0, \mathcal{A}X = b\} \end{aligned}$$

We have the Lagrangian dual function:

$$\begin{aligned} \Theta_t(\lambda, Z) &= \inf\{\langle C, X\rangle + \frac{\langle X, X\rangle}{2t} - \frac{\langle X, Y\rangle}{t} + \frac{\langle Y, Y\rangle}{2t} - \langle \lambda, \mathcal{A}X - b\rangle - \langle Z, X\rangle : X \in S^n\} \\ &= \inf\{\langle C, X\rangle + \frac{\langle X, X\rangle}{2t} - \frac{\langle X, Y\rangle}{t} + \frac{\langle Y, Y\rangle}{2t} + \langle \lambda, b\rangle - \langle \mathcal{A}^*\lambda, X\rangle - \langle Z, X\rangle : X \in S^n\}, \end{aligned}$$

where $\mathcal{A}X = \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix}$, $\lambda \in \mathbf{R}^m$, $Z \in S^n_+$. Now, we compute $\Theta_t(\lambda, Z)$.

Let

$$L_\lambda(X, Z) := \langle C, X\rangle + \frac{\langle X, X\rangle}{2t} - \frac{\langle X, Y\rangle}{t} + \frac{\langle Y, Y\rangle}{2t} + \langle \lambda, b\rangle - \langle \mathcal{A}^*\lambda, X\rangle - \langle Z, X\rangle$$

Note that the objective function in Moreau Yosida regularization $F_t(Y)$ is convex, so, the minimum exists and unique and determined by:

$$\frac{\partial L_\lambda(X,Z)}{\partial X} = C + \frac{1}{t}(X-Y) - \mathcal{A}^*\lambda - Z = 0 \quad \Rightarrow X(\lambda,Z) = t(Z+\mathcal{A}^*\lambda - C) + Y.$$

$$\begin{aligned}
\Rightarrow \Theta_t(\lambda,Z) &= \langle C, X(\lambda,Z)\rangle - \langle \mathcal{A}^*\lambda, X(\lambda,Z)\rangle - \langle Z, X(\lambda,Z)\rangle + \frac{1}{2t}||X(\lambda,Z)-Y||^2 + b^t\lambda \\
&= \langle C-\mathcal{A}^*\lambda - Z, t(Z+\mathcal{A}^*\lambda - C)+Y\rangle + \frac{1}{2t}||(t(Z+\mathcal{A}^*\lambda-C)+Y)-Y||^2 + b^t\lambda \\
&= -t\langle Z+\mathcal{A}^*\lambda - C, Z+\mathcal{A}^*\lambda - C\rangle + \langle C-\mathcal{A}^*\lambda - Z, Y\rangle + \frac{1}{2t}||t(Z+\mathcal{A}^*\lambda - C)||^2 + b^t\lambda \\
&= -t\langle Z+\mathcal{A}^*\lambda - C, Z+\mathcal{A}^*\lambda - C\rangle + \langle C-\mathcal{A}^*\lambda - Z, Y\rangle + \frac{t}{2}||Z+\mathcal{A}^*\lambda - C||^2 + b^t\lambda \\
&= -t||Z+\mathcal{A}^*\lambda - C||^2 - \langle Y, Z+\mathcal{A}^*\lambda - C\rangle + \frac{t}{2}||Z+\mathcal{A}^*\lambda - C||^2 + b^t\lambda \\
&= b^t\lambda - \langle Y, Z+\mathcal{A}^*\lambda - C\rangle - \frac{t}{2}||Z+\mathcal{A}^*\lambda - C||^2,
\end{aligned}$$

its associated dual problem is [17]:

$$\begin{aligned}
\max \quad & b^t\lambda - \langle Y, Z+\mathcal{A}\lambda - C\rangle - \frac{t}{2}||Z+\mathcal{A}^*\lambda - C)||^2 \\
\text{s.t.} \quad & \lambda \in \mathbf{R}^m, Z \succeq 0.
\end{aligned}$$

To this end, we compute $||Z+\mathcal{A}^*\lambda - C||^2$:

$$\begin{aligned}
||Z+\mathcal{A}^*\lambda - C||^2 &= \langle Z - C + \mathcal{A}^*\lambda, Z - C + \mathcal{A}^*\lambda\rangle \\
&= \langle Z-C, Z-C\rangle + \langle \mathcal{A}^*\lambda, Z-C\rangle + \langle Z-C, \mathcal{A}^*\lambda\rangle + \langle \mathcal{A}^*\lambda, \mathcal{A}^*\lambda\rangle \\
&= \langle Z-C, Z-C\rangle + 2\langle \mathcal{A}(Z-C), \lambda\rangle + \langle \mathcal{A}\mathcal{A}^*\lambda, \lambda\rangle.
\end{aligned}$$

Thus

$$\Theta_t(\lambda,Z) = b^t\lambda - \langle Y, Z-C\rangle - \langle \mathcal{A}Y, \mathcal{A}^*\lambda\rangle - \frac{t}{2}\langle Z-C, Z-C\rangle - t\langle \mathcal{A}(Z-C),\lambda\rangle - \frac{t}{2}\langle \mathcal{A}\mathcal{A}^*\lambda,\lambda\rangle$$

So,

$$\begin{aligned}
\frac{\partial \Theta_t(\lambda,Z)}{\partial \lambda} &= b - \mathcal{A}Y - t\mathcal{A}(Z-C) - t\mathcal{A}\mathcal{A}^*\lambda \\
&= b - \mathcal{A}Y - t\mathcal{A}(Z - C + \mathcal{A}^*\lambda) \\
&= b - \mathcal{A}(Y + t(Z+\mathcal{A}^*\lambda - C)).
\end{aligned}$$

Clearly $\nabla_Z\Theta_t(\lambda,Z) = -t(Z+\mathcal{A}^*\lambda - C) + Y)$. So, we have proved

Proposition 3.2.1. (Inner dual function, [17]) The dual function $\Theta_t(\lambda,Z)$ is equal to $b^t\lambda - \mathbf{tr}(Y(Z+\mathcal{A}^*\lambda - C)) - \frac{t}{2}||Z+\mathcal{A}^*\lambda - C)||^2$. Furthermore, $\Theta_t(\lambda,Z)$ is differentiable, and its gradient w.r.t λ is:

$$\nabla\Theta_t(\lambda,Z) = b - A(t(Z+\mathcal{A}^*\lambda - C) + Y,$$

it gradient with w.r.t Z is:

$$\nabla \Theta_t(\lambda, Z) = -(t(Z + \mathcal{A}^*\lambda - C) + Y).$$

If we compute the derivatives of the primal Moreau Yosida problem, w.r.t Y, then

$$\frac{\partial \Theta_t(\lambda, Z)}{\partial Y} = \frac{1}{2t}(-2X + 2Y) = \frac{1}{t}(Y - X).$$

Definition 3.2.1. (Lipschitz continuity, [1]) A function $f : A \to \mathbf{R}$ is called Lipschitz continous if there exists a bound $L > 0$ such that

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2| \quad \forall x_1, x_2 \in A.$$

Definition 3.2.2. (Uniform continuity, [1]) A function $f : A \to \mathbf{R}$ is uniformly continous on A if for every $\epsilon > 0$ there exists a $\delta > 0$ such that any $x_1, x_2 \in A$ satisfying $|x_1 - x_2| < \delta$ implies $|f(x_1) - f(x_2)| < \epsilon$.

Example 3.2.1. ([1], Ex. 4.4.9) If $f : A \to \mathbf{R}$ is Lipschitz, then it is uniformly continous on A.

Proof. If $\epsilon > 0$ there exists $\delta > 0$ such that $|x_1 - x_2| < \delta$. Let $\delta = \frac{\epsilon}{L}$. By definition 3.2.1 $|f(x_1) - f(x_2)| < L\frac{\epsilon}{L} = \epsilon$. □

Proposition 3.2.2. (Regularization properties, [17]) The function F_t is finite everywhere, convex, and differentiable. Its gradient at $Y \in S^n$ is

$$\nabla F_t(Y) = \frac{1}{t}(Y - P_t(Y)),$$

where $P_t(Y) = t(Z + \mathcal{A}^* - C) + Y$ is the proximal point of Y with parameter t, and $\nabla F_t(Y)$ are Lipschitz continuous.

If we differentiate the dual of Moreau Yosida problem, w.r.t Y, then

$$\begin{aligned} \max \quad & b^t\lambda - \frac{t}{2}||Z + \mathcal{A}^*\lambda - C)||^2 \\ \text{s.t.} \quad & C - \mathcal{A}^*\lambda = Z, Z \succeq 0, \end{aligned}$$

which resembles the augumented dual Lagrangian [17]:

$$\begin{aligned} \max \quad & b^t\lambda - \frac{\sigma}{2}||Z + \mathcal{A}^*\lambda - C)||^2 \\ \text{s.t.} \quad & C - \mathcal{A}^*\lambda = Z, Z \succeq 0, \end{aligned}$$

where $\Theta_\sigma(Y) = b^t\lambda - \frac{\sigma}{2}||Z + \mathcal{A}^*\lambda - C)||^2$. According Malick et al [17] there is an outer connection between the primal and dual:

Proposition 3.2.3. (Outer connection, [17]) If $t = \sigma$ then $\Theta_\sigma(Y) = F_t(Y)$ for all $Y \in S^n$.

Remark. [17] *By the propositions 3.2.1, 3.2.3 there is no duality gap .*

3.2.1 Algorithm II

The following algorithm describe the primal perspective to regularize linear semidefinite programming, and it is built upon the primal Moreau Yosida regularization [17].

Initialization step Let $t > 0$, and $Y_{S_n} \preceq 0$.

Main step Repeat until $\frac{1}{t}||Y - P_t(Y)||$ is small.

1. Solve the inner problem to the primal Moreau Yosida regularization to get $X = P_t(Y)$.

2. Set $Y = X$, and update t.

4 Numerical illustrations

4.1 Diagonal matrices

This section begins by consider the aspect *structure* for symmetric matrices in semidefinite programming. There are two specific forms of symmetric matrices. The first matrice is expressed in the diagonal form, the other one is not diagonal. Thereafter, we apply the theory of *duality*, characterized a duality gap.

Case 1. Diagonal matrices, Definition 2.5.3, zero duality gap.

$$C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

We express the primal SDP problem in the standard form:

$$\begin{aligned} \min \quad & \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle \\ \text{s.t.} \quad & \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 0, \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 1, \\ & X \in S^n_+. \end{aligned}$$

A feasible solution X is then:

$$X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and minimum attains to be 1. The assoicated dual for the SDP is:

$$\begin{aligned} \max \quad & y_2 \\ \text{s.t.} \quad & y_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + y_2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ & S \succeq 0. \end{aligned}$$

This is equivalent to:

$$\begin{aligned} \max \quad & y_2 \\ \text{s.t.} \quad & \begin{bmatrix} -y_1 - y_2 & 0 & 0 \\ 0 & 1 - y_2 & 0 \\ 0 & 0 & -y_1 - y_2 \end{bmatrix} \succeq 0 \Leftrightarrow \begin{cases} -y_1 - y_2 \geq 0 \\ 1 - y_2 \geq 0 \\ -y_1 - y_2 \geq 0 \end{cases} \end{aligned}$$

and a unique feasible solution is $(y_1, y_2) = (-1, 1)$ and maximum attains to be 1. Since $p^* - d^* = 1 - 1 = 0$ we conclude there is *zero duality gap*.

If we interchange the values in the vector b, the conclusion remains same. Furthermore, the primal and dual optimal values are attained. On the other hand, if the vector b is equal to zero, then the dual optimal value is not attained. However, the primal optimal value attains its minimum. If $b^t = (0, -1)$ then there is again zero duality gap. Pataki in [20], Zhang, Chen and Zhang in [32], and Ramana in [22] describe these internal aspects within semidefinite programming.

Pataki in [20] have also noted the matrices partition into bad semidefinite programming. This is accompliched by delete the second row, the second column for the three dimensions matrices. In our case, this will return the identity matrices. The next case, do not however give the identity matrices.

Case 1.1 Case 1, with additional constraints.

$$C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, b^t = \begin{bmatrix} 0, 1, 0, 1, 0, 1 \end{bmatrix}.$$

We express the primal SDP problem in the standard form:

$$\begin{aligned}
\min \quad & \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle \\
\text{s.t.} \quad & \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 0, \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 1, \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 0, \\
& \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle = 1, \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle = 0, \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 1 \\
& X \in S_+^n.
\end{aligned}$$

A feasible solution X is then:

$$X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and minimum attains to be 1. The assoicated dual for the SDP is:

$$\begin{aligned}
\max \quad & y_2 + y_4 + y_6 \\
\text{s.t.} \quad & y_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + y_2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + y_3 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + y_4 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + y_5 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
& + y_6 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
& S \succeq 0.
\end{aligned}$$

This is equivalent to:

$$\begin{aligned}
\max \quad & y_2 + y_4 + y_6 \\
\text{s.t.} \quad & \begin{bmatrix} -y_1 - y_2 - y_5 & 0 & 0 \\ 0 & 1 - y_2 - y_4 - y_6 & 0 \\ 0 & 0 & -y_1 - y_2 - y_3 - y_6 \end{bmatrix} \succeq 0 \\
& \Leftrightarrow \begin{cases} -y_1 - y_2 - y_5 \geq 0 \\ 1 - y_2 - y_4 - y_6 \geq 0 \\ -y_1 - y_2 - y_3 - y_6 \geq 0, \end{cases}
\end{aligned}$$

and the feasible solution is $(y_1, y_2, y_3, y_4, y_5, y_6) = (0, 0, 0, 1, 0, 0)$ and maximum attains to be 1. Since $p^* - d^* = 1 - 1 = 0$ we conclude there is *zero duality gap*.

Case 1.2 Case 1.1, with an additional constraint.

$$C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$A_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_7 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} b^t = \begin{bmatrix} 0, 1, 0, 1, 0, 1, 0 \end{bmatrix}.$$

We express the primal SDP problem in the standard form:

$$\begin{aligned}
\min \quad & \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle \\
\text{s.t.} \quad & \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 0, \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 1, \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 0, \quad \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle = 1, \\
& \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle = 0, \langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \rangle = 1, \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \rangle = 0, \\
& X \in S^n_+.
\end{aligned}$$

The last inequality constraint do not satisfy the other constraints. Hence, the primal is infeasible.

We turn now to the case, non-diagonal matrices.

4.2 Non-diagonal matrices

Case 2. Non-diagonal matrices, Definition 2.5.4, zero duality gap.

$$c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, A_1 = \begin{bmatrix} 8 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 10 \end{bmatrix}, A_2 = \begin{bmatrix} 2 & 4 & 0 \\ 4 & 12 & 6 \\ 0 & 6 & 10 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

We express the primal SDP problem on the inequality form:

$$\begin{aligned}
\min \quad & x_1 + x_2 \\
\text{s.t.} \quad & x_1 \begin{bmatrix} 8 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 10 \end{bmatrix} + x_2 \begin{bmatrix} 2 & 4 & 0 \\ 4 & 12 & 6 \\ 0 & 6 & 10 \end{bmatrix} \preceq \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

The feasible solution is $(x_1, x_2) = (0, 0)$ and minimum attains to be 0. The assoicated dual for the SDP is:

$$\begin{aligned} \max \quad & \langle \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \lambda \rangle \\ \text{s.t.} \quad & \langle \begin{bmatrix} 8 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 10 \end{bmatrix}, \lambda \rangle = 1, \langle \begin{bmatrix} 2 & 4 & 0 \\ 4 & 12 & 6 \\ 0 & 6 & 10 \end{bmatrix}, \lambda \rangle = 1. \end{aligned}$$

A feasible solution λ is then:

$$\lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{10} \end{bmatrix}$$

and maximum attains to be 0. Since $p^* - d^* = 0 - 0 = 0$ we conclude it is *zero duality gap*.

4.3 Mixed matrices

Case 3. Mixed matrices, Definition 2.5.4, zero duality gap.

$$c = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, A_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_2 = \begin{bmatrix} 8 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 8 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We express the primal SDP problem on the inequality form:

$$\begin{aligned} \min \quad & x_1 + 4x_2 \\ \text{s.t.} \quad & x_1 \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 8 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 8 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The feasible solution is $(x_1, x_2) = (\frac{1}{2}, 0)$ and minimum attains to be $\frac{1}{2}$. The assoicated dual for the SDP is:

$$\begin{aligned} \max \quad & \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda \rangle \\ \text{s.t.} \quad & \langle \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda \rangle = 1, \langle \begin{bmatrix} 8 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 8 \end{bmatrix}, \lambda \rangle = 4. \end{aligned}$$

A feasible solution λ is then:

$$\lambda = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and maximum attains to be $\frac{1}{2}$. Since $p^* - d^* = \frac{1}{2} - \frac{1}{2} = 0$ we conclude it is *zero duality gap*.

Malick et al [17] have also regarded the matrices solutions. In particular, if there is small pertubutations in the semidefinite programming or if the starting position vary with different initial conditions then it could affect the systems of equations solutions. According the theory of matrices, it is well known a system of equations could have unique, infinitely many or no solutions.

Furthermore, Ramana, Tunçel, and Wolkowicz in [23] disusses the solutions in semidefinite programming. The solutions might consist of rational or irrational numbers. The previous two cases illustrate this more concretely.

Case 3.1 [20] Mixed matrices, Definition 2.5.4, duality gap.

$$c = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We express the primal SDP problem on the inequality form:

$$\begin{aligned} \min \quad & -x_2 \\ \text{s.t.} \quad & x_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The feasible solution is $(x_1, x_2) = (1, 0)$ and minimum attains to be 0. The assoicated dual for the SDP is:

$$\begin{aligned} \max \quad & -\langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda \rangle \\ \text{s.t.} \quad & \langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda \rangle = 0, \langle \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \lambda \rangle = 1. \end{aligned}$$

A feasible solution λ is then:

$$\lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and maximum attains to be -1. Since $p^* - d^* = 0 - (-1) = 1$ we conclude that *duality gap* exists. This is due to either strong duality property is not satisfied or slaterś condition fails. Nevertheless, the optimal values are attained [10].

4.4 Concluding comments

We have described two primal regularization methods to close up the duality gap. The first method is based on abstract convex programming, and the other by semidefinite programming. The algorithms are according Borwein, Wolkowitz [8], and Malick et al [17]. These methods shows various ways to close up the duality gap depending on problem size, i.e small, medium, and large scale problems.

Furthermore, we explore in the analysis part a combination of the aspects duality, and matrices structure for some semidefinite programming problems to characterize a duality gap.

In conclusion, the duality and efficiency in semidefinite programming are important aspects both from the theoretical and practical perspectives. Further reserach in the theme is suggested to gain insight among the complex phenomenas, and clarify the efficiency levels relation to regularization methods.

References

[1] Abbott S. *Understanding analysis*, New York, Springer, cop. 2001.

[2] Ahron B., Nemiroviski A. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, MPS-SIAM series on optimization, 2013.

[3] Alizadeh F., Interior point methods in semidefinite programming with applications to combinatorial optimization, *SIAM Journal on optimization* Vol. 5, Issue 1, 1995, p. 13-51.

[4] Barker G., Carlson D., Cones of diagonally dominant matrices, *Pacific Journal of Mathematics* Vol. 57, Issue 1, 1975, p. 15-32.

[5] Bazaraa M., Sherali H., Shetty C.. *Nonlinear programming theory and algorithms*, New York, John Wiley, cop., 2006.

[6] Bertsekas D., Nedić A., Ozdaglar A. *Convex analysis and optimization*, Belmont, Mass., Athena Scientific, cop., 2003.

[7] Blekherman G., Parrilo P., Thomas R. *Semidefinite optimization and convex algebraic geometry*, Philadelphia, Society for Industrial and applied mathematics, cop., 2013.

[8] Borwein J., Wolkowicz H., Regularizing the Abstract Convex Program, *Journal of Mathematical analysis and applications*, Vol.83, Issue 2, 1981, p. 495-530.

[9] Boyd S., Vandenberghe L. *Convex optimization*, Cambridge, Cambridge University Press, 2004.

[10] De Klerk E., *Aspects of Semidefinite programming. Interior point Algorithms and selected applications*, Dordecht; Boston, Kluwer Academic Publishers, 2002.

[11] De Klerk E., Exploiting special structure in semidefinite programming: a survey of theory and applications, *European Journal of Operational Research*, Vol. 201, Issue 1, 2010, p. 1-10.

[12] Floudas C., Pardalos P. *Encyclopedia of optimization*, New York, Springer, 2008.

[13] Helmberg C. Semidefinite Programming, *European Journal of operational research*, Vol. 137, Issue 3, 2002, p. 461-482.

[14] Helmberg C., Rendl F., Vanderbei R., Wolkowicz H. An interior-point method for semidefinite programming, *SIAM Journal on Optimization*, Vol. 6, Issue 2, 1996, p. 342-361.

[15] Lobo M. *Robust and convex optimization with applications in finance*, Ph.D thesis, Department of Electrical Engieering, Stanford University, Stanford, 2000.

[16] Lustig I., Marsten R., Shanno D., Interior point methods for linear programming: computational state of the art, *ORSA Journal of computing*, Vol. 6, Issue 1, 1994, p. 1-14.

[17] Malick J., Poyh J., Rendle F., Wiegele A. Regularization methods for semidefinite programming, *SIAM Journal Optimization*, Vol. 20, Issue 1, 2009, p. 336-356.

[18] Moskowitz M., Paliogiannis F., *Functions of several real variables*, Chichester, Ellis Horwood, 1991.

[19] Overton M., Wolkowicz H., Semidefinite programming-Foreword, *Mathematical programming* Vol. 77, Issue 2, 1997, p. 105-109.

[20] Pataki G. Bad Semidefinite programs: they all look the same, *arXiv* 2011, p. 1-14.

[21] Pataki G. On the connection of facially exposed and nice cones, *Journal of Matematical analysis and applications* 2013, p. 211-221.

[22] Ramana M., An exact duality theory for semidefinite programming and its complexity implications, *Mathematical programming*, Vol. 77, Issue 2, 1997, p. 129-162.

[23] Ramana M.,Tunçel L., Wolkowicz H. Strong duality for semidefinite programming, *SIAM Journal Optimization*, Vol. 7, Issue 3, 1997, p. 641-662.

[24] Rendle F. Semidefinite programming and combinatorial optimization, *Applied numerical mathematics*, Vol. 29, Issue 3, 1999, p. 255-281.

[25] Rockafellar R. *Convex analysis*, Princeton, N.J., Princeton Univ. Press, 1970.

[26] Singer I. *Abstract convex analysis* New York, Wiley, cop., 1997.

[27] Todd M. Semidefinite optimization, *Acta Numerica*, Vol. 10, 2001, p. 515-560.

[28] Vandenberghe L.,Boyd S. Semidefinite programming, *SIAM Review*, Vol. 38, Issue 1, 1996, p. 49-95.

[29] Wolkowicz H. Calculating the cone of directions of constancy, *Journal of optimization theory and applications*, Vol. 25, Issue 3, 1978, p. 451-457.

[30] Wolkowicz H. Geometry of optimality conditions and constraint qualifications: The convex case, *Mathematical programming*, Vol. 19, Issue 1, 1980, p. 32-80.

[31] Zhadan V., Orlov A. An admissible dual internal point method for a linear semidefinite programming problem, *Automation and Remote Control*, Vol. 73, Issue 2, 2012, p. 232-246.

[32] Zhang Q., Chen G., Zhang T., Duality formulations in semidefinite programming, *Journal of industrial and management optimization*, Vol. 6, Issue 4, 2010, p. 881- 893.