The many faces of degeneracy
in
conic optimization

last modified on May 9, 2017

Dmitriy Drusvyatskiy
Department of Mathematics
University of Washington
ddrusv@uw.edu

Henry Wolkowicz
Faculty of Mathematics
University of Waterloo
hwolkowicz@uwaterloo.ca

Contents

List of Figures

List of Tables

Abstract

Slater's condition – existence of a "strictly feasible solution" – is a common assumption in conic optimization. Without strict feasibility, first-order optimality conditions may be meaningless, the dual problem may yield little information about the primal, and small changes in the data may render the problem infeasible. Hence, failure of strict feasibility can negatively impact off-the-shelf numerical methods, such as primal-dual interior point methods, in particular. New optimization modelling techniques and convex relaxations for hard nonconvex problems have shown that the loss of strict feasibility is a more pronounced phenomenon than has previously been realized. In this text, we describe various reasons for the loss of strict feasibility, whether due to poor modelling choices or (more interestingly) rich underlying structure, and discuss ways to cope with it and, in many pronounced cases, how to use it as an advantage. In large part, we emphasize the facial reduction preprocessing technique due to its mathematical elegance, geometric transparency, and computational potential.

1

What this paper is about

Conic optimization has proven to be an elegant and powerful modeling tool with surprisingly many applications. The classical *linear programming* problem revolutionized operations research and is still the most widely used optimization model. This is due to the elegant theory and the ability to solve in practice both small and large scale problems efficiently and accurately by the well known simplex method of Dantzig [35] and by more recent interior-point methods, e.g., [98, 148]. The size (number of variables) of linear programs that could be solved before the interior-point revolution was on the order of tens of thousands, whereas it immediately increased to millions for many applications. A large part of modern success is due to *preprocessing*, which aims to identify (primal and dual slack) variables that are identically zero on the feasible set. The article [96] is a good reference.

The story does not end with linear programming. Dantzig himself recounts in [36]: "the world is nonlinear". Nonlinear models can significantly improve on linear programs if they can be solved efficiently. Conic optimization has shown its worth in its elegant theory, efficient algorithms, and many applications e.g., [9, 20, 146]. Preprocessing to rectify possible loss of "strict-feasibility" in the primal or the dual

problems is appealing for general conic optimization as well. In contrast to linear programming, however, the area of preprocessing for conic optimization is in its infancy; see e.g., [29, 30, 107, 109, 138] and Section 1.1, below. In contrast to linear programming, numerical error makes preprocessing difficult in full generality. This being said, surprisingly, there are many specific applications of conic optimization, where the rich underlying structure makes preprocessing possible, leading to greatly simplified models and strengthened algorithms. Indeed, exploiting structure is essential for making preprocessing viable. In this article, we present the background and the elementary theory of such regularization techniques in the framework of *facial reduction (FR)*. We focus on notable case studies, where such techniques have proven to be useful.

1.1 Related work

To put this text in perspective, it is instructive to consider nonlinear programming. Nontrivial statements in constrained nonlinear optimization always rely on some regularity of the constraints. To illustrate, consider a minimization problem over a set of the form $\{x : f(x) = 0\}$ for some smooth f. How general are such constraints? A celebrated result of Whitney [143] shows that *any* closed set in a Euclidean space can written as a zero-set of some C^∞-smooth function f. Thus, in this generality, there is little difference between minimizing over arbitrary closed sets and sets of the form $\{x : f(x) = 0\}$, for smooth f. Since little can be said about optimizing over arbitrary closed sets, one must make an assumption on the equality constraint. The simplest one, eliminating Whitney's construction, is that the gradient of f is nonzero on the feasible region – the earliest form of a constraint qualification. There have been numerous papers, developing weakened versions of regularity (and optimality conditions) in nonlinear programming; some good examples are [22, 25, 62].

The Slater constraint qualification, we discuss in this text, is in a similar spirit, but in the context of (convex) conic optimization. Some good early references on the geometry of the Slater condition, and weak-

ened variants, are [19, 57, 93, 94, 144]. The concept of facial reduction for general convex programs was introduced in [23, 24], while an early application to a semi-definite type best-approximation problem was given in [145]. Recently, there has been a significant renewed interest in facial reduction, in large part due to the success in applications for graph related problems, such as Euclidean distance matrix completion and molecular conformation [6, 46, 75, 76] and in polynomial optimization [74, 110, 111, 140, 141]. In particular, a more modern explanation of the facial reduction procedure can be found in [88, 104, 107, 136, 142].

We note in passing that numerous papers show that strict feasibility holds "generically" with respect to unstructured perturbations. In contrast, optimization problems appearing in applications are often highly structured and such genericity results are of little practical use.

1.2   Outline of the paper

The paper is divided into two parts. In Part I, we present the necessary theoretical grounding in conic optimization, including basic optimality and duality theory, connections of Slater's condition to the *distance to infeasibility* and sensitivity theory, the facial reduction procedure, and the singularity degree. In Part II, we concentrate on illustrative examples and applications, including matrix completion problems (semidefinite, low-rank, and Euclidean distance), relaxations of hard combinatorial problems (quadratic assignment and max-cut), and sum of squares relaxations of polynomial optimization problems.

1.3   Reflections on Jonathan Borwein and FR

These are some reflections on Jonathan Borwein and his role in the development of the facial reduction technique, by Henry Wolkowicz. Jonathon Borwein passed away unexpectedly on Aug. 2, 2016. Jon was an extraordinary mathematician who made significant contributions in an amazing number of very diverse areas. Many details and personal memories by myself and many others including family, friends, and colleagues, are presented at the memorial website `jonborwein.org`. This was a terrible loss to his family and all his friends and colleagues,

including myself. The facial reduction process we use in this monograph originates in the work of Jon and the second author (myself). This work took place from July of 1978 to July of 1979 when I went to Halifax to work with Jon at Dalhousie University in a lectureship position. The optimality conditions for the general abstract convex program using the facially reduced problem is presented in the two papers [22, 23]. The facial reduction process is then derived in [24].

Part I

Theory

2

Convex geometry

This section collects preliminaries of linear algebra and convex geometry that will be routinely used in the rest of the manuscript. The main focus is on convex duality, facial structure of convex cones, and the primal-dual conic optimization pair. The two running examples of linear and semi-definite programming illustrate the concepts. We have tried to include proofs of important theorems, when they are both elementary and enlightening. We have omitted arguments that are longer or that are less transparent, so as not to distract the reader from the narrative.

2.1 Notation

Throughout, we will fix a Euclidean space $\mathbf{E}$ with an inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\|x\| = \sqrt{\langle x, x \rangle}$. When referencing another Euclidean space (with its own inner product), we will use the letter $\mathbf{F}$. An open ball of radius $r > 0$ around a point $x \in \mathbf{E}$ will be denoted by $B_r(x)$. The two most important examples of Euclidean spaces for us will be the space of n-vectors $\mathbf{R}^n$ with the dot product $\langle x, y \rangle = \sum_i x_i y_i$ and the space of $n \times n$ symmetric matrices $\mathbf{S}^n$ with the trace inner product

$\langle X, Y \rangle = \text{trace}(XY)$. Throughout, we let e_i be the i'th coordinate vector of $\mathbf{R}^n$. Note that the trace inner product can be equivalently written as $\text{trace}(XY) = \sum_{i,j} X_{ij} Y_{ij}$. Thus the trace inner product is itself the dot product between the two matrices stretched into vectors. A key property of the trace is invariance under permutations of the arguments: $\text{tr}(AB) = \text{tr}(BA)$ for any two matrices $A \in \mathbf{R}^{m \times n}$ and $B \in \mathbf{R}^{n \times m}$.

For any linear mapping $A \colon \mathbf{E} \to \mathbf{F}$, between Euclidean spaces $\mathbf{E}$ and $\mathbf{F}$, the *adjoint mapping* $A^* \colon \mathbf{F} \to \mathbf{E}$ is the unique mapping satisfying

$$\langle Ax, y \rangle = \langle x, A^* y \rangle \quad \text{for all } x \in \mathbf{E} \text{ and } y \in \mathbf{F}.$$

Notice that the angle brackets on the left refer to the inner product in $\mathbf{F}$, while those on the right refer to the inner product in $\mathbf{E}$.

Let us look at two important examples of adjoint maps.

Example 2.1.1 (Adjoints of mappings between $\mathbf{R}^n$ and $\mathbf{R}^m$). Consider a matrix $A \in \mathbf{R}^{m \times n}$ as a linear map from $\mathbf{R}^n \to \mathbf{R}^m$. Then the adjoint A^* is simply the transpose A^T. To make the parallel with the next example, it is useful to make this description more explicit. Suppose that the linear operator $A \colon \mathbf{R}^n \to \mathbf{R}^m$ is defined by

$$Ax = (\langle a_1, x \rangle, \ldots, \langle a_m, x \rangle), \tag{2.1}$$

where $a_1, \ldots, a_m$ are some vectors in $\mathbf{R}^n$. When thinking of A as a matrix, the vectors a_i would be its rows, and the description (2.1) corresponds to a "row-space" view of matrix-vector multiplication Ax. The adjoint $A^* \colon \mathbf{R}^m \to \mathbf{R}^n$ is simply the map

$$A^* y = \sum_i y_i a_i. \tag{2.2}$$

Again, when thinking of A as a matrix with a_i as its rows, the description (2.2) corresponds to the "column-space" view of matrix-vector multiplication $A^T y$.

Example 2.1.2 (Adjoints of mappings between $\mathbf{S}^n$ and $\mathbf{R}^m$). Consider a set of symmetric matrices $A_1, \ldots, A_m$ in $\mathbf{S}^n$, and define the linear map $A \colon \mathbf{S}^n \to \mathbf{R}^m$ by

$$A(X) = (\langle A_1, X \rangle, \ldots, \langle A_m, X \rangle).$$

We note that any linear map $A\colon \mathbf{S}^n \to \mathbf{R}^m$ can be written in this way for some matrices $A_i \in \mathbf{S}^n$. Notice the parallel to (2.1). The adjoint $A^*\colon \mathbf{R}^m \to \mathbf{S}^n$ is given by

$$A^*y = \sum_i y_i A_i.$$

Notice the parallel to (2.2). To verify that this indeed is the adjoint, simply observe the equation

$$\langle X, \sum_i y_i A_i \rangle = \sum_i y_i \langle A_i, X \rangle = \langle A(X), y \rangle,$$

for any $X \in \mathbf{S}^n$ and $y \in \mathbf{R}^n$.

The interior, boundary, and closure of any set $C \subset \mathbf{E}$ will be denoted by $\operatorname{int} C$, $\operatorname{bd} C$, and $\operatorname{cl} C$, respectively. A set C is *convex* if it contains the line segment joining any two points in C:

$$x, y \in C, \ \alpha \in [0,1] \qquad \Longrightarrow \qquad \alpha x + (1-\alpha)y \in C.$$

The minimal affine space containing a convex set C is called the *affine hull* of C, and is denoted by $\operatorname{aff} C$. We define the *relative interior* of C, written $\operatorname{ri} C$, to be the interior of C relative to $\operatorname{aff} C$. It is straightforward to show that a for a nonempty convex set C, the relative interior $\operatorname{ri} C$ is never empty.

A subset $\mathcal{K}$ of $\mathbf{E}$ is a *convex cone* if $\mathcal{K}$ is convex and is positively homogeneous, meaning $\lambda K \subseteq K$ for all $\lambda \geq 0$. Equivalently, $\mathcal{K}$ is a convex cone if, and only if, for any two points x and y in $\mathcal{K}$ and any nonnegative constants $\alpha, \beta \geq 0$, the sum $\alpha x + \beta y$ lies in $\mathcal{K}$. We say that a convex cone $\mathcal{K}$ is *proper* if $\mathcal{K}$ is closed, has nonempty interior, and contains no lines. The symbol $\mathcal{K}^\perp$ refers to the orthogonal complement of $\operatorname{aff} \mathcal{K}$. Let us look at two most important examples of proper cones for this article.

Example 2.1.3 (The nonnegative orthant $\mathbf{R}^n_+$). The nonnegative orthant

$$\mathbf{R}^n_+ := \{x \in \mathbf{R}^n : x_i \geq 0 \text{ for all } i = 1, \ldots, n\}$$

is a proper convex cone in $\mathbf{R}^n$. The interior of $\mathbf{R}^n_+$ is the set

$$\mathbf{R}^n_{++} := \{x \in \mathbf{R}^n : x_i > 0 \text{ for all } i = 1, \ldots, n\}$$

Example 2.1.4 (The positive semi-definite cone $\mathbf{S}_+^n$)**.** Consider the set of positive semi-definite matrices

$$\mathbf{S}_+^n := \{X \in \mathbf{S}^n : v^T X v \geq 0 \text{ for all } v \in \mathbf{R}^n\}.$$

It is immediate from the definition that $\mathbf{S}_+^n$ is a convex cone containing no lines. Let us quickly verify that $\mathbf{S}_+^n$ is proper. To see this, observe

$$v^T X v = \mathrm{tr}(v^T X v) = \mathrm{tr}(X v v^T) = \langle X, v v^T \rangle.$$

Thus $\mathbf{S}_+^n$ is closed because it is the intersection of the halfspaces $\{X \in \mathbf{S}^n : \langle X, v v^T \rangle \geq 0\}$ for all $v \in \mathbf{R}^n$, and arbitrary intersections of closed sets are closed. The interior of $\mathbf{S}_+^n$ is the set of positive definite matrices

$$\mathbf{S}_{++}^n := \{X \in \mathbf{S}^n : v^T X v > 0 \text{ for all } 0 \neq v \in \mathbf{R}^n\}.$$

Let us quickly verify this description. Showing that $\mathbf{S}_{++}^n$ is open is straightforward; we leave the details to the reader. Conversely, consider a matrix $X \in \mathbf{S}_+^n \setminus \mathbf{S}_{++}^n$ and let v be a nonzero vector satisfying $v^T X v = 0$. Then the matrix $X - t v v^T$ lies outside of $\mathbf{S}_+^n$ for every $t > 0$, and therefore X must lie on the boundary of $\mathbf{S}_+^n$. To summarize, we have shown that $\mathbf{S}_+^n$ is a proper convex cone.

Given a convex cone $\mathcal{K}$ in $\mathbf{E}$, we introduce two binary relations $\succeq_{\mathcal{K}}$ and $\succ_{\mathcal{K}}$ on $\mathbf{E}$:

$$x \succeq_{\mathcal{K}} y \quad \Longleftrightarrow \quad x - y \in \mathcal{K},$$
$$x \succ_{\mathcal{K}} y \quad \Longleftrightarrow \quad x - y \in \mathrm{int}\,\mathcal{K}.$$

Assuming that $\mathcal{K}$ is proper makes the relation $\succeq_{\mathcal{K}}$ into a *partial order*, meaning that for any three points $x, y, z \in \mathbf{E}$, the three conditions hold:

1. (reflexivity) $x \succeq_{\mathcal{K}} x$

2. (antisymmetry) $x \succeq_{\mathcal{K}} y$ and $y \succeq_K x \quad \Longrightarrow \quad x = y$

3. (transitivity) $x \succeq_{\mathcal{K}} y$ and $y \succeq_K z \quad \Longrightarrow \quad x \succeq_{\mathcal{K}} z.$

As is standard in the literature, we denote the partial order $\succeq_{\mathbf{R}_+^n}$ on $\mathbf{R}^n$ by $\geq$ and the partial order $\succeq_{\mathbf{S}_+^n}$ on $\mathbf{S}^n$ by $\succeq$. In particular, the

relation $x \geq y$ means $x_i \geq y_i$ for each coordinate i, while the relation $X \succeq Y$ means that the matrix $X - Y$ is positive semi-definite.

Central to conic geometry is duality. The *dual cone* of $\mathcal{K}$ is the set

$$\mathcal{K}^* := \{y \in \mathbf{E} : \langle x, y \rangle \geq 0 \text{ for all } x \in \mathcal{K}\}.$$

The following lemma will be used extensively.

Lemma 2.1.1 (Self-duality). Both $\mathbf{R}^n_+$ and $\mathbf{S}^n_+$ are self-dual, meaning $(\mathbf{R}^n_+)^* = \mathbf{R}^n_+$ and $(\mathbf{S}^n_+)^* = \mathbf{S}^n_+$.

Proof. The equality $(\mathbf{R}^n_+)^* = \mathbf{R}^n_+$ is elementary and we leave the proof to the reader. To see that $\mathbf{S}^n_+$ is self-dual, recall that a matrix $X \in \mathbf{S}^n$ is positive semi-definite if, and only if, all of its eigenvalues are nonnegative. Fix two matrices $X, Y \succeq 0$ and let $X = \sum_i \lambda_i v_i v_i^T$ be the eigenvalue decomposition of X. Then we deduce

$$\langle X, Y \rangle = \operatorname{tr} XY = \sum_i \lambda_i \operatorname{tr}(v_i v_i^T Y) = \sum_i \lambda_i (v_i^T Y v_i) \geq 0.$$

Therefore the inclusion $\mathbf{S}^n_+ \subseteq (\mathbf{S}^n_+)^*$ holds. Conversely, for any $X \in (\mathbf{S}^n_+)^*$ and any $v \in \mathbf{R}^n$ the inequality, $0 \leq \langle X, vv^T \rangle = v^T X v$, holds. The reverse inclusion $\mathbf{S}^n_+ \supseteq (\mathbf{S}^n_+)^*$ follows, and the proof is complete. $\qquad \square$

Finally, we end this section with the following two useful results of convex geometry.

Lemma 2.1.2 (Dual cone of a sum). For any two closed convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$, equalities hold:

$$(\mathcal{K}_1 + \mathcal{K}_2)^* = \mathcal{K}_1^* \cap \mathcal{K}_2^* \qquad \text{and} \qquad (\mathcal{K}_1 \cap \mathcal{K}_2)^* = \operatorname{cl}(\mathcal{K}_1^* + \mathcal{K}_2^*).$$

Lemma 2.1.3 (Double dual). A set $\mathcal{K} \subset \mathbf{E}$ is a closed convex cone if, and only if, equality $\mathcal{K} = (\mathcal{K}^*)^*$ holds.

In particular, if $\mathcal{K}$ is a proper convex cone, then so is its dual $\mathcal{K}^*$, as the reader can verify.

2.2 Facial geometry

Central to this paper is the decomposition of a cone into faces.

Definition 2.2.1 (Faces). Let $\mathcal{K}$ be a convex cone. A convex cone $F \subseteq \mathcal{K}$ is called a *face of* $\mathcal{K}$, denoted $F \trianglelefteq \mathcal{K}$, if the implication holds:

$$x, y \in \mathcal{K}, \ x + y \in F \quad \Longrightarrow \quad x, y \in F.$$

Let $\mathcal{K}$ be a closed convex cone. Vacuously, the empty set and $\mathcal{K}$ itself are faces. A face $F \trianglelefteq \mathcal{K}$ is *proper* if it is neither empty nor all of $\mathcal{K}$. One can readily verify from the definition that the intersection of an arbitrary collection of faces of $\mathcal{K}$ is itself a face of $\mathcal{K}$. A fundamental result of convex geometry shows that relative interiors of all faces of $\mathcal{K}$ form a partition of $\mathcal{K}$: every point of $\mathcal{K}$ lies in the relative interior of some face and relative interiors of any two distinct faces are disjoint. In particular, any proper face of $\mathcal{K}$ is disjoint from $\operatorname{ri} \mathcal{K}$.

Definition 2.2.2 (Minimal face). The *minimal face* of a convex cone $\mathcal{K}$ containing a set $S \subseteq \mathcal{K}$ is the intersection of all faces of $\mathcal{K}$ containing S, and is denoted by $\operatorname{face}(S, \mathcal{K})$.

A convenient alternate characterization of minimal faces is as follows. If $S \subseteq \mathcal{K}$ is a convex set, then $\operatorname{face}(S, \mathcal{K})$ is the smallest face of $\mathcal{K}$ intersecting the relative interior of S. In particular, equality $\operatorname{face}(S, \mathcal{K}) = \operatorname{face}(x, \mathcal{K})$ holds for any point $x \in \operatorname{ri} S$.

There is a special class of faces that admit "dual" descriptions. Namely, for any vector $v \in \mathcal{K}^*$ one can readily verify that the set $F = v^\perp \cap \mathcal{K}$ is a face of $\mathcal{K}$.

Definition 2.2.3 (Exposed faces). Any set of the form $F = v^\perp \cap \mathcal{K}$, for some vector $v \in \mathcal{K}^*$, is called an *exposed face* of $\mathcal{K}$. The vector v is then called an *exposing vector* of F.

The classical hyperplane separation theorem shows that any point x in the relative boundary of $\mathcal{K}$ lies in some proper exposed face. Not all faces are exposed, however, as the following example shows.

Example 2.2.1 (Nonexposed faces). Consider the set $Q = \{(x, y) \in \mathbf{R}^2 : y \geq \max(0, x)^2\}$, and let $\mathcal{K}$ be the closed convex cone generated by $Q \times \{1\}$. Then the ray $\{(0, 0)\} \times \mathbf{R}_+$ is a face of $\mathcal{K}$ but it is not exposed.

Figure 2.1: The set Q.

The following is a very useful property of exposed faces we will use.

Proposition 2.2.1 (Exposing the intersection of exposed faces)**.**
For any closed convex cone $\mathcal{K}$ and vectors $v_1, v_2 \in \mathcal{K}^*$, equality holds:

$$(v_1^{\perp} \cap \mathcal{K}) \cap (v_2^{\perp} \cap \mathcal{K}) = (v_1 + v_2)^{\perp} \cap \mathcal{K}\,.$$

Proof. The inclusion $\subseteq$ is trivial. To see the converse, note for any $x \in (v_1 + v_2)^{\perp} \cap \mathcal{K}$ we have $\langle v_1, x \rangle \geq 0$, $\langle v_2, x \rangle \geq 0$, while $\langle v_1 + v_2, x \rangle = 0$. We deduce $x \in v_1^{\perp} \cap v_2^{\perp}$ as claimed. $\square$

In other words, if the faces $F_1 \trianglelefteq \mathcal{K}$ and $F_2 \trianglelefteq \mathcal{K}$ are exposed by v_1 and v_2, respectively, then the intersection $F_1 \cap F_2$ is a face exposed by the sum $v_1 + v_2$.

A convex cone is called *facially exposed* if all of its faces are exposed. The distinction between faces and exposed faces may appear mild at first sight; however, we will see that it is exactly this distinction that can cause difficulties for preprocessing techniques for general conic problems.

Definition 2.2.4 (Conjugate face)**.**
With any face F of a convex cone $\mathcal{K}$, we associate a face of the dual cone $\mathcal{K}^*$, called the *conjugate face,* $F^{\triangle} := \mathcal{K}^* \cap F^{\perp}$.

Equivalently, $F^{\triangle}$ is the face of $\mathcal{K}^*$ exposed by any point $x \in \operatorname{ri} F$, that is $F^{\triangle} = \mathcal{K}^* \cap x^{\perp}$. Thus, in particular, conjugate faces are always exposed. Not surprisingly, one can readily verify that equality $(F^{\triangle})^{\triangle} = F$ holds if, and only if, the face $F \trianglelefteq \mathcal{K}$ is exposed.

We illustrate the concepts with our two running examples, $\mathbf{R}_+^n$ and $\mathbf{S}_+^n$, keeping in mind the parallels between the two.

Example 2.2.2 (Faces of $\mathbf{R}_+^n$). For any index set $I \subseteq \{1, \ldots, n\}$, the set

$$F_I = \{x \in \mathbf{R}_+^n : x_i = 0 \text{ for all } i \in I\}$$

is a face of $\mathbf{R}_+^n$, and all faces of $\mathbf{R}_+^n$ are of this form. In particular, observe that all faces of $\mathbf{R}_+^n$ are linearly isomorphic to $\mathbf{R}_+^k$ for some positive integer k. In this sense, $\mathbf{R}_+^n$ is "*self-replicating*". The relative interior of F_I consists of all points in F_I with $x_i > 0$ for indices $i \notin I$. The face F_I is exposed by the vector $v \in \mathbf{R}_+^n$ with $v_i = 1$ for all $i \in I$ and $v_i = 0$ for all $i \notin I$. In particular, $\mathbf{R}_+^n$ is a facially exposed convex cone. The face conjugate to F_I is F_{I^c}.

Example 2.2.3 (Faces of $\mathbf{S}_+^n$). There are a number of different ways to think about (and represent) faces of the PSD cone $\mathbf{S}_+^n$. In particular, one can show that faces $\mathbf{S}_+^n$ are in correspondence with linear subspaces of $\mathbf{R}^n$. More precisely, for any r-dimensional linear subspace $\mathcal{R}$ of $\mathbf{R}^n$, the set

$$F_{\mathcal{R}} := \{X \in \mathbf{S}_+^n : \operatorname{range} X \subseteq \mathcal{R}\} \tag{2.3}$$

is a face of $\mathbf{S}_+^n$. Conversely, any face of $\mathbf{S}_+^n$ can be written in the form (2.3), where $\mathcal{R}$ is the range space of any matrix X lying in the relative interior of the face. The relative interior of $F_{\mathcal{R}}$ consists of all matrices $X \in \mathbf{S}_+^n$ whose range space coincides with $\mathcal{R}$. Moreover, for any matrix $V \in \mathbf{R}^{n \times r}$ satisfying $\operatorname{range} V = \mathcal{R}$, we have the equivalent description

$$F_{\mathcal{R}} = V \mathbf{S}_+^r V^T.$$

In particular, $F_{\mathcal{R}}$ is linearly isomorphic to the r-dimensional positive semi-definite cone $\mathbf{S}_+^r$. The face conjugate to $F_{\mathcal{R}}$ is $F_{\mathcal{R}^\perp}$ and can be equivalently written as

$$F_{\mathcal{R}^\perp} = U \mathbf{S}_+^{n-r} U^T,$$

for any matrix $U \in \mathbf{R}^{n \times (n-r)}$ satisfying $\operatorname{range} U = \mathcal{R}^\perp$. Notice that then the matrix UU^T lies in the relative interior of $F^\triangle$ and therefore UU^T exposes the face $F_{\mathcal{R}}$. In particular, $\mathbf{S}_+^n$ is *facially exposed* and also *self-replicating*.

2.3 Conic optimization problems

Modern conic optimization draws fundamentally from "duality": every conic optimization problem gives rise to a related conic optimization problem, called its dual. Consider the *primal-dual pair*:

$$
\textbf{(P)} \quad \begin{array}{ll} \inf & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \succeq_{\mathcal{K}} 0 \end{array}
\qquad\qquad
\textbf{(D)} \quad \begin{array}{ll} \sup & \langle b, y \rangle \\ \text{s.t.} & A^* y \preceq_{\mathcal{K}^*} c \end{array}
\qquad (2.4)
$$

Here, $\mathcal{K}$ is a closed convex cone in $\mathbf{E}$ and the mapping $A \colon \mathbf{E} \to \mathbf{F}$ is linear. Eliminating the trivial case that the system $Ax = b$ has no solution, we will always assume that b lies in range A, and that $\mathcal{K}$ has nonempty interior. Two examples of conic optimization are of main importance for us: linear programming (LP) corresponds to $\mathcal{K} = \mathbf{R}^n_+$, $\mathbf{F} = \mathbf{R}^m$ and semi-definite programming (SDP) corresponds to $\mathcal{K} = \mathbf{S}^n_+$, $\mathbf{F} = \mathbf{R}^m$. The adjoint A^* in both cases was computed in Examples 2.1.3 and 2.1.4. We will also use the following notation for the *primal* and *dual feasible regions*:

$$
\mathcal{F}_p := \{ x \succeq_{\mathcal{K}} 0 : Ax = b \} \qquad \text{and} \qquad \mathcal{F}_d := \{ y : A^* y \preceq_{\mathcal{K}^*} c \}.
$$

It is important to note that the dual can be put in a primal form by introducing *slack variables* $s \in \mathbf{E}$ leading to the equivalent formulation

$$
\begin{array}{ll} \sup & \langle b, y \rangle \\ \text{s.t.} & A^* y + s = c \\ & y \in \mathbf{F}, s \in \mathcal{K}^*. \end{array}
\qquad (2.5)
$$

To a reader unfamiliar with conic optimization, it may be unclear how the dual arises naturally from the primal. Let us see how it can be done. The dual problem **(D)** can be discovered through "Lagrangian duality" in convex optimization. Define the *Lagrangian function*

$$
L(x, y) := \langle c, x \rangle + \langle y, b - Ax \rangle,
$$

and observe the equality

$$
\max_y \; L(x, y) = \begin{cases} \langle c, x \rangle & Ax = b \\ +\infty & \text{otherwise} \end{cases}.
$$

Thus the primal problem **(P)** is equivalent to

$$\min_{x \succeq_{\mathcal{K}} 0} \left(\max_y L(x, y) \right).$$

Formally exchanging min/max, yields exactly the dual problem **(D)**

$$\max_y \left(\min_{x \succeq_{\mathcal{K}} 0} L(x, y) \right) = \max_y \left(\langle b, y \rangle + \min_{x \succeq_{\mathcal{K}} 0} \langle x, c - A^* y \rangle \right)$$

$$= \max \{ \langle b, y \rangle : A^* y \preceq_{\mathcal{K}^*} c \}.$$

The primal-dual pair always satisfies the *weak duality inequality*: for any primal feasible x and any dual feasible y, we have

$$0 \leq \langle c - A^* y, x \rangle = \langle c, x \rangle - \langle b, y \rangle. \tag{2.6}$$

Thus for any feasible point y of the dual, its objective value $\langle b, y \rangle$ lower-bounds the optimal value of the primal. The weak duality inequality (2.6) leads to the following sufficient conditions for optimality.

Proposition 2.3.1 (Complementary slackness).
Suppose that (x, y) are a primal-dual feasible pair for **(P)**, **(D)** and suppose that *complementary slackness* holds:

$$0 = \langle c - A^* y, x \rangle.$$

Then x is a minimizer of **(P)** and y is a maximizer of **(D)**.

The sufficient conditions for optimality of Proposition 2.3.1 are often summarized as the primal-dual system:

$$
\begin{array}{rcll}
A^* y + s - c & = & 0 & \text{(dual feasibility)} \\
A x - b & = & 0 & \text{(primal feasibility)} \\
\langle s, x \rangle & = & 0 & \text{(complementary slackness)} \\
s \succeq_{\mathcal{K}^*} 0, \, x \succeq_{\mathcal{K}} 0 & & & \text{(nonegativity)}.
\end{array} \tag{2.7}
$$

Derivations of algorithms generally require necessary optimality conditions, i.e., failure of the necessary conditions at a current approximation of the optimum leads to improvement steps. When are sufficient conditions for optimality expressed in Proposition 2.3.1 necessary? In other words, when can we be sure that optimality of a primal

solution x can be certified by the existence of some dual feasible point y, such that the pair satisfies the complementary slackness condition? Conditions guaranteeing existence of Lagrange multipliers are called *constraint qualifications*. The most important condition of this type is called *strict feasibility*, or *Slater's condition*, and is the main topic of this article.

Definition 2.3.1 (Strict feasibility/Slater condition). We say that $(\mathbf{P})$ is *strictly feasible* if there exists a point $x \succ_{\mathcal{K}} 0$ satisfying $Ax = b$. The dual $(\mathbf{D})$ is *strictly feasible* if there exists a point y satisfying $A^*y \prec_{\mathcal{K}^*} c$.

The following result is the cornerstone of conic optimization.

Theorem 2.3.1 (Strong duality). If the primal objective value is finite and the problem $(\mathbf{P})$ is strictly feasible, then the primal and dual optimal values are equal, and the dual $(\mathbf{D})$ admits an optimal solution. In addition, for any x that is optimal for $(\mathbf{P})$, there exists a vector y such that (x, y) satisfies complementary slackness.

Similarly, if the dual objective value is finite and the dual $(\mathbf{D})$ satisfies strict feasibility, then the primal and dual optimal values are equal and the primal $(\mathbf{P})$ admits an optimal solution. In addition, for any y that is optimal for $(\mathbf{D})$, there exists a point x such that (x, y) satisfies complementary slackness.

Without a constraint qualification such as strict feasibility, the previous theorem is decisively false. The following examples show that without strict feasibility, the primal and dual optimal values may not even be equal, and even if they are equal, the optimal values may be unattained.

Example 2.3.1 (Infinite gap). Consider the following primal SDP in (2.4):
$$0 = v_p = \min_{x \in \mathbf{S}^2} \left\{ 2X_{12} \ : \ X_{11} = 0, \ X \succeq 0 \right\}.$$
The corresponding dual SDP is the infeasible problem
$$-\infty = v_d = \max_{y \in \mathbf{R}} \left\{ 0y : \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}.$$
Both the primal and the dual fail strict feasibility in this example.

Example 2.3.2 (Positive duality gap)**.** Consider the following primal SDP in (2.4):

$$v_p = \min_{X \in \mathbf{S}^3} \left\{ X_{22} \ : \ X_{33} = 0, \ X_{22} + 2X_{13} = 1, \ X \succeq 0 \right\}.$$

The constraint $X \succeq 0$ with $X_{33} = 0$ implies equality $X_{13} = 0$, and hence $X_{22} = 1$. Therefore, $v_p = 1$ and the matrix $e_2 e_2^T$ is optimal.

The corresponding dual SDP is

$$v_d = \max_{y \in \mathbf{R}^2} \left\{ y_2 \ : \ \begin{pmatrix} 0 & 0 & y_2 \\ 0 & y_2 & 0 \\ y_2 & 0 & y_1 \end{pmatrix} \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}. \qquad (2.8)$$

This time the SDP constraint implies $y_2 = 0$. We deduce that $(y_1, y_2) = (0, 0)$ is optimal for the dual and hence $v_d = 0 < v_p = 1$. There is a finite duality gap between the primal and dual problems. The culprit again is that both the primal and the dual fail strict feasibility.

Example 2.3.3 (Zero duality gap, but no attainment)**.** Consider the dual SDP

$$\begin{aligned} v_d \ &= \ \sup && y \\ &\quad \text{s.t.} && y \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

The only feasible point is $y = 0$. Thus the optimal value is $v_d = 0$ and is *attained.* The primal SDP is

$$\begin{aligned} v_p \ &= \ \inf && X_{11} \\ &\quad \text{s.t.} && 2X_{12} = 1 \\ &&& X \in \mathbf{S}_+^2. \end{aligned}$$

Notice $X_{11} > 0$ for all feasible X. On the other hand, the sequence $X^k = \begin{bmatrix} 1/k & 1/2 \\ 1/2 & k \end{bmatrix}$ is feasible and satisfies $X_{11}^k \to 0$. Thus there is no duality gap, meaning $0 = v_p = v_d$, but the primal optimal value is *not attained.* The culprit is that the dual SDP is not strictly feasible.

Example 2.3.4 (Convergence to dual optimal value)**.** Numerical solutions of problems inevitably suffer from some perturbations of the data, that is a perturbed problem is in fact solved. Moreover, often

it is tempting to explicitly perturb a constraint in the problem, so that strict feasibility holds. This example shows that this latter strategy results in the dual of the problem being solved, as opposed to the problem under consideration.

We consider the primal-dual SDP pair in Example 2.3.2. In particular, suppose first that we want to solve the dual problem. We canonically perturb the right-hand side of the dual in (2.8)

$$
v_d(\epsilon) := \sup_{y \in \mathbf{R}^2} \left\{ y_2 \ : \ \begin{pmatrix} 0 & 0 & y_2 \\ 0 & y_2 & 0 \\ y_2 & 0 & y_1 \end{pmatrix} \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \epsilon P \right\},
$$

for some matrix $P \succ 0$ and real $\epsilon > 0$. Strict feasibility now holds and we *hope* that the optimal values of the *perturbed problems* converge to that of the original one $v_d(0)$. We can rewrite feasibility for the perturbed problem as

$$
\begin{pmatrix} 0 & 0 & -y_2 \\ 0 & 1 - y_2 & 0 \\ -y_2 & 0 & -y_1 \end{pmatrix} + \epsilon P \succeq 0. \tag{2.9}
$$

A triple (y_1, y_2, y_3) is strictly feasible if, and only if, the leading principal minors M_{11}, M_{12}, M_{123} of the left-hand side matrix in (2.9) are all positive. We have $M_{11} = \epsilon P_{11} > 0$. The second leading principal minor as a function of y_2 is

$$
M_{12}(y_2) = \epsilon \left(P_{11}(1 - y_2) + \epsilon(P_{11}P_{22} - P_{12}^2) \right).
$$

In particular, rearranging we have $M_{12}(y_2) > 0$ whenever

$$
y_2 < \frac{P_{11} + \epsilon(P_{11}P_{22} - P_{12}^2)}{P_{11}}. \tag{2.10}
$$

The last minor M_{123} is positive for sufficiently negative y_1 by the Schur complement. Consequently the perturbed problems satisfy

$$
v_d(\epsilon) = 1 + \epsilon \frac{P_{11}P_{22} - P_{12}^2}{P_{11}}
$$

and therefore

$$
0 = v_d < \lim_{\epsilon \to 0} v_d(\epsilon) = 1 = v_p.
$$

That is the primal optimal value is obtained in the limit rather than the dual optimal value that is sought.

Let us look at an analogous perturbation to the primal problem. Let A be the linear operator $A(X) = (X_{33}, X_{22} + 2X_{13})$ and set $b = (0, 1)$. Consider the perturbed problems

$$v_p(\epsilon) = \min_{X \in \mathbf{S}^3} \{X_{22} : AX = b + \epsilon AP\}$$

for some fixed real $\epsilon > 0$ and a matrix $P \succ 0$. Each such problem is strictly feasible, since the positive definite matrix $\widehat{X} + \epsilon P$ is feasible for any matrix $\widehat{X}$ that is feasible for the original primal problem.

In long form, the perturbed primal problems are

$$v_p(\epsilon) = \min_{X \succeq 0} \{X_{22} \ : \ X_{33} = \epsilon P_{33}, \ X_{22} + 2X_{13} = 1 + \epsilon(P_{22} + 2P_{13})\}.$$

Consider the matrix

$$X = \begin{bmatrix} X_{11} & 0 & 1/2 + \epsilon(P_{22} + 2P_{13})/2 \\ 0 & 0 & 0 \\ 1/2 + \epsilon(P_{22} + 2P_{13})/2 & 0 & \epsilon P_{33} \end{bmatrix}.$$

This matrix satisfies the linear system $AX = b$ by construction and is positive semi-definite for all sufficiently large X_{11}. We deduce $v_p(\epsilon) = 0 = v_d$ for all $\epsilon > 0$. Again, as ϵ tends to zero we obtain the dual optimal value rather than the sought after primal optimal value.

2.4   Commentary

We follow here well-established notation in convex optimization, as illustrated for example in the monographs of Barvinok [17], Ben-Tal-Nemirovski [20], Borwein-Lewis [21], and Rockafellar [123]. The handbook of SDP   [146] and online lecture notes [85] are other excellent sources in the context of semi-definite programming. These include discussion on the facial structure. The relevant results stated in the text can all be found for instance in Rockafellar [123]. The example 2.3.2 is a modification of the example in [114]. In addition, note that the three examples 2.3.1, 2.3.2, 2.3.3 have matrices with the special *perdiagonal* structure. The universality of such special structure in "ill-posed" SDPs has recently been investigated in great length by Pataki [105].

3

Virtues of strict feasibility

We have already seen in Theorem 2.3.1 that strict feasibility is essential to guarantee dual attainment and therefore for making the primal-dual optimality conditions (2.7) meaningful. In this section, we continue discussing the impact of strict feasibility on numerical stability. We begin with the theorems of the alternative, akin to the Farkas' Lemma in linear programming, which quantify the extent to which strict feasibility holds. We then show how such systems appear naturally in stability measures of the underlying problem.

3.1 Theorem of the alternative

The definition we have given of strict feasibility (Slater) is *qualitative* in nature, that is it involves no measurements. A convenient way to measure the extent to which strict feasibility holds (i.e. its strength) arises from dual characterizations of the property. We will see that strict feasibility corresponds to inconsistency of a certain *auxiliary system*. Measures of how close the auxiliary system is to being consistent yield estimates of "stability" of the problem.

The aforementioned dual characterizations stem from the basic hy-

perplane separation theorem for convex sets.

Theorem 3.1.1 (Hyperplane separation theorem)**.** Let Q_1 and Q_2 be two disjoint nonempty convex sets. Then there exists a nonzero vector v and a real number c satisfying

$$\langle v, x \rangle \leq c \leq \langle v, y \rangle \qquad \text{for all } x \in Q_1 \text{ and } y \in Q_2.$$

When one of the sets is a cone, the separation theorem takes the following "homogeneous" form.

Theorem 3.1.2 (Homogeneous separation)**.** Consider a nonempty closed convex set Q and a closed convex cone $\mathcal{K}$ with nonempty interior. Then exactly one of the following alternatives holds.

1. The set Q intersects the interior of $\mathcal{K}$.

2. There exists a vector $0 \neq v \in \mathcal{K}^*$ satisfying $\langle v, x \rangle \leq 0$ for all $x \in Q$.

Moreover, for any vector v satisfying the alternative 2, the region $Q \cap \mathcal{K}$ is contained in the proper face $v^\perp \cap \mathcal{K}$.

Proof. Suppose that Q does not intersect the interior of $\mathcal{K}$. Then the convex cone generated by Q, denoted by $\operatorname{cone} Q$, does not intersect $\operatorname{int} \mathcal{K}$ either. The hyperplane separation theorem (Theorem 3.1.1) shows that there is a nonzero vector v and a real number c satisfying

$$\langle v, x \rangle \leq c \leq \langle v, y \rangle \quad \text{for all } x \in \operatorname{cone} Q \text{ and } y \in \operatorname{int} \mathcal{K}.$$

Setting $x = y = 0$, we deduce $c = 0$. Hence v lies in $\mathcal{K}^*$ and 2 holds.

Conversely, suppose that 2 holds. Then the inequalities $0 \leq \langle v, x \rangle \leq 0$ hold for all $x \in Q \cap \mathcal{K}$. Thus we deduce that the intersection $Q \cap \mathcal{K}$ lies in the proper face $v^\perp \cap \mathcal{K}$. Hence the alternative 1 can not hold. $\square$

Let us now specialize the previous theorem to the primal problem **(P)**, by letting Q be the affine space $\{x : Ax = b\}$. Indeed, this is the main result of this subsection and it will be used extensively in what follows.

Theorem 3.1.3 (Theorem of the alternative for the primal)**.** Suppose that $\mathcal{K}$ is a closed convex cone with nonempty interior. Then exactly one of the following alternatives holds.

1. The primal **(P)** is strictly feasible.

2. The auxiliary system is consistent:

$$0 \neq A^*y \succeq_{\mathcal{K}^*} 0 \qquad \text{and} \qquad \langle b, y \rangle \leq 0. \tag{3.1}$$

Suppose that the primal **(P)** is feasible. Then the auxiliary system (3.1) is equivalent to the system

$$0 \neq A^*y \succeq_{\mathcal{K}^*} 0 \qquad \text{and} \qquad \langle b, y \rangle = 0. \tag{3.2}$$

Moreover, then any vector v satisfying either of the equivalent systems, (3.1) and (3.2), yields a proper face $(A^*y)^\perp \cap \mathcal{K}$ containing the primal feasible region $\mathcal{F}_p$.

Proof. Set $Q := \{x : Ax = b\}$. Clearly strict feasibility of **(P)** is equivalent to alternative 1 of Theorem 3.1.2. Thus it suffices to show that the auxiliary system (3.1) is equivalent to the alternative 2 of Theorem 3.1.2. To this end, note that for any vector y satisfying (3.1), the vector $v := A^*y$ satisfies the alternative 2 of Theorem 3.1.2. Conversely, consider a vector $0 \neq v \in \mathcal{K}^*$ satisfying $\langle v, x \rangle \leq 0$ for all $x \in Q$. Fix a point $\hat{x} \in Q$ and observe the equality $Q = \hat{x} + \text{null}(A)$. An easy argument then shows that v is orthogonal to $\text{null}(A)$, and therefore can be written as $v = A^*y$ for some vector y. We deduce $\langle b, y \rangle = \langle A\hat{x}, y \rangle = \langle \hat{x}, v \rangle \leq 0$, and therefore y satisfies (3.1).

Next, assume that **(P)** is feasible. Suppose y satisfies (3.1). Then for any feasible point x of **(P)**, we deduce $0 \leq \langle x, A^*y \rangle = \langle b, y \rangle \leq 0$. Thus y satisfies the system (3.2). It follows that the two systems (3.1) and (3.2) are equivalent and that the proper face $(A^*y)^\perp \cap \mathcal{K}$ contains the primal feasible region $\mathcal{F}_p$, as claimed. $\square$

Suppose that $\mathcal{F}_p$ is nonempty. Then if strict feasibility fails, there always exists a "witness" (or "short certificate") y satisfying the auxiliary system (3.2). Indeed, given such a vector y, one immediately deduces, as in the proof, that $\mathcal{F}_p$ is contained in the proper face $(A^*y)^\perp \cap \mathcal{K}$ of $\mathcal{K}$. Such certificates will in a later section be used constructively to regularize the conic problem through the FR procedure.

The analogue of Theorem 3.1.3 for the dual **(D)** quickly follows.

Theorem 3.1.4 (Theorem of the alternative for the dual)**.** Suppose that $\mathcal{K}^*$ has nonempty interior. Then exactly one of the following alternatives holds.

1. The dual **(D)** is strictly feasible.

2. The auxiliary system is consistent:

$$0 \neq x \succeq_{\mathcal{K}} 0, \quad Ax = 0, \quad \text{and} \quad \langle c, x \rangle \leq 0. \tag{3.3}$$

Suppose that the dual **(D)** is feasible. Then the auxiliary system (3.3) is equivalent to the system

$$0 \neq x \succeq_{\mathcal{K}} 0, \quad Ax = 0, \quad \text{and} \quad \langle c, x \rangle = 0. \tag{3.4}$$

Moreover, for any vector x satisfying either of the equivalent systems, (3.3) and (3.4), yields a proper face $x^{\perp} \cap \mathcal{K}^*$ containing the feasible slacks $\{c - A^* y : y \in \mathcal{F}_d\}$.

Proof. Apply Theorem 3.1.3 to the equivalent formulation (2.5) of the dual **(D)**. □

3.2   Stability of the solution

In this section, we explain the impact of strict feasibility on stability of the conic optimization problem through quantities naturally arising from the auxiliary system (3.1). For simplicity we focus on the primal problem **(P)**, though an entirely analogous development is possible for the dual, for example by introducing slack variables.

We begin with a basic question: at what rate does the optimal value of the primal problem **(P)** change relative to small perturbations of the right-hand-side of the linear equality constraints? To formalize this question, define the *value function*

$$
\begin{aligned}
v(\Delta) := \quad &\inf \quad \langle c, x \rangle \\
&\text{s.t.} \quad Ax = b + \Delta \\
&\qquad\quad x \succeq_{\mathcal{K}} 0.
\end{aligned}
$$

The value function $v \colon \mathbf{F} \to [-\infty, +\infty]$ thus defined is convex, meaning that its epigraph

$$\operatorname{epi} v := \{(x, r) \in \mathbf{F} \times \mathbf{R} : v(x) \leq r\}$$

is a convex set. Seeking to understand stability of the primal **(P)** under perturbation of the right-hand-side b, it is natural to examine the variational behavior of the value function. There is an immediate obstruction, however. If A is not surjective, then there are arbitrarily small perturbations of b making v take on infinite values. As a result, in conjunction with strict feasibility, we will often make the mild assumption that A is surjective. These two properties taken together are refereed to as the *Mangasarian-Fromovitz Constraint Qualification (MFCQ)*.

Definition 3.2.1 (Mangasarian-Fromovitz CQ)**.** We say that the *Mangasarian-Fromovitz Constraint Qualification (MFCQ)* holds for **(P)** if A is surjective and **(P)** is strictly feasible.

The following result describes directional derivatives of the value function.

Theorem 3.2.1 (Directional derivative of the value function)**.** Suppose the primal problem **(P)** is feasible and its optimal value is finite. Let $\text{Sol}(\mathbf{D})$ be the set of optimal solutions of the dual **(D)**. Then $\text{Sol}(\mathbf{D})$ is nonempty and bounded if, and only if, MFCQ holds. Moreover, under MFCQ, the directional derivative $v'(0; w)$ of the value function at $\Delta = 0$ in direction w admits the representation

$$v'(0; w) = \sup\{\langle w, y \rangle : y \in \text{Sol}(\mathbf{D})\}.$$

In particular, in the notation of the above theorem, the local Lipschitz constant of v at the origin,

$$\limsup_{\Delta_1, \Delta_2 \to 0} \frac{|v(\Delta_1) - v(\Delta_2)|}{\|\Delta_1 - \Delta_2\|},$$

coincides with the norm of the maximal-norm dual optimal solution, and is finite if, and only if, MFCQ holds. Is there then an upper-bound on the latter that we can easily write down? Clearly, such a quantity must measure the strength of MFCQ, and is therefore intimately related to the auxiliary system (3.1). To this end, let us define the *condition number*

$$\text{cond}(\mathbf{P}) := \min_{y : \|y\| = 1} \max\{\text{dist}_{\mathcal{K}^*}(A^* y), \langle b, y \rangle\}.$$

This number is a quantitative measure of MFCQ, and will appear in latter sections as well. Some thought shows that it is in essence measuring how close the auxiliary system (3.1) is to being consistent.

Lemma 3.2.2 (Condition number and MFCQ). The condition number $\text{cond}(\mathbf{P})$ is nonzero if, and only if, MFCQ holds.

Proof. Suppose $\text{cond}(\mathbf{P})$ is nonzero. Then clearly A is surjective, since otherwise we could find a unit vector y with $A^*y = 0$ and $\langle b, y \rangle \leq 0$. Moreover, the auxiliary system (3.1) is clearly inconsistent, and therefore $(\mathbf{P})$ is strictly feasible.

   Conversely, suppose MFCQ holds. Assume for the sake of contradiction $\text{cond}(\mathbf{P}) = 0$. Then there exists a unit vector y satisfying $0 \neq A^*y \in \mathcal{K}^*$ and $\langle b, y \rangle \leq 0$. Thus (3.1) is consistent, a contradiction. $\square$

Theorem 3.2.3 (Boundedness of the dual solution set). Suppose the problem $(\mathbf{P})$ is feasible with a finite optimal value **val**. If the condition number $\text{cond}(\mathbf{P})$ is nonzero, then the inequality

$$\|y\| \leq \frac{\max\{\|c\|, -\mathbf{val}\}}{\text{cond}(\mathbf{P})}$$

holds for all dual optimal solutions y.

Proof. Consider an optimal solution y of the dual $(\mathbf{D})$. The inclusion $c - A^*y \in \mathcal{K}^*$ implies $\text{dist}_{\mathcal{K}^*}(-A^*y/\|y\|) \leq \|c\|/\|y\|$. Moreover, we have $\langle b, \frac{-y}{\|y\|} \rangle = \frac{-\mathbf{val}}{\|y\|}$. We deduce $\text{cond}(\mathbf{P}) \leq \frac{\max\{\|c\|, -\mathbf{val}\}}{\|y\|}$ and the result follows. $\square$

   Thus the Lipschitz constant of the value function depends on the extent to which MFCQ holds through the condition number. What about stability of the solution set itself? The following theorem, whose proof we omit, answers this question.

Theorem 3.2.4 (Stability of the solution set). Suppose $(\mathbf{P})$ satisfies MFCQ. Let $\mathcal{F}_p(g)$ be the solution set of the perturbed system

$$Ax = b + g, \quad x \succeq_{\mathcal{K}} 0.$$

Fix a putative solution $\bar{x} \in \mathcal{F}_p$. Then there exist constants $c > 0$ and $\epsilon > 0$ so that the inequality

$$\text{dist}(x; \mathcal{F}_p(g)) \leq c \cdot \|(Ax - b) - g\|$$

holds for any $x \in \mathcal{K} \cap B_\epsilon(\bar{x})$ and $g \in B_\epsilon(0)$. The infimal value of c over all choices of $\epsilon > 0$ so that the above inequalities hold is exactly

$$\sup_{y:\|y\|=1} \frac{1}{\text{dist}(-A^*y; \text{face}(\bar{x}, \mathcal{K})^\triangle)}. \tag{3.5}$$

In particular, under MFCQ, we can be sure that for any point $\bar{x} \in \mathcal{F}_p$, there exist c and $\epsilon > 0$ satisfying

$$\text{dist}(\bar{x}; \mathcal{F}_p(g)) \leq c \cdot \|g\| \qquad \text{for all } g \in B_\epsilon(0).$$

In other words, the distance $\text{dist}(\bar{x}; \mathcal{F}_p(g))$, which measures the how far $\mathcal{F}_p(g)$ has moved relative to $\bar{x}$, is bounded by a multiple of the perturbation parameter $\|g\|$. The proportionality constant c is fully governed by the strength of MFCQ, as measured by the quantity (3.5).

3.3 Distance to infeasibility

In numerical analysis, the notion of stability is closely related to the "distance to infeasibility" – the smallest perturbation needed to make the problem infeasible. A simple example is the problem of solving an equation $Lx = b$ for an invertible matrix $L \colon \mathbf{R}^n \to \mathbf{R}^n$. Then the Eckart-Young theorem shows equality

$$\min_{G \in \mathbf{R}^{n \times n}} \{\|G\| : L + G \text{ is singular}\} = \frac{1}{\|L^{-1}\|}.$$

Here $\|G\|$ denotes the operator norm of G. The left-hand-side measures the smallest perturbation G needed to make the system $(L + G)x = b$ singular, while the right-hand-side measures the Lipschitz dependence of the solution to the linear system $Lx = b$ relative to perturbations in b, and yet the two quantities are equal. An entirely analogous situation holds in conic optimization, with MFCQ playing the role of invertibility.

Definition 3.3.1 (Distance to infeasibility)**.** The *distance to infeasibility* of **(P)** is the infimum of the quantity $\max\{\|G\|, \|g\|\}$ over linear mappings G and vectors g such that the system

$$(A + G)x = b + g, \quad x \succeq_{\mathcal{K}} 0 \qquad \text{is infeasible.}$$

This quantity does not change if instead of the loss of feasibility, we consider the loss of strict feasibility. The following fundamental result equates the condition number (measuring the strength of MFCQ) and the distance to infeasibility.

Theorem 3.3.1 (Strict feasibility and distance to infeasibility)**.** The following exact equation is always true:

$$\text{distance to infeasibility} = \text{cond}(\mathbf{P}).$$

3.4   Commentary

The classical theorem of the alternative is Farkas Lemma that appears in proofs of duality in linear programming, as well as in more general nonlinear programming, after linearizations. This and more general theorems of the alternative are given in e.g., the 1969 book by Mangasarian [92] and in the 1969 survey paper by Ben-Israel [18]. The specific theorems of the alternative that we use are similar to the one used in the FR development in [22, 23, 24].

The Mangasarian-Fromovitz CQ was introduced in [91]. This condition and its equivalence to stability with respect to perturbations in the data and compactness of the multiplier set has been the center of extensive research, e.g., [54]. The analogous conditions for general nonlinear convex constraint systems is the *Robinson regularity condition*, e.g.,[121, 122]. The notion of distance to infeasibility and relations to condition numbers was initiated by Renegar e.g., [108, 118, 119, 120]. The relation with Slater's condition is clear. Theorem 3.3.1, as stated, appears in [44], though in essence it is present in Renegar's work [118, 120].

4

Facial reduction

Theorems 3.1.3 and 3.1.4 have already set the stage for the "Facial Reduction" procedure, used for regularizing degenerate conic optimization problems by restricting the problem to smaller and smaller dimensional faces of the cone $\mathcal{K}$. In this section, we formalize this viewpoint, empathizing semi-definite programming. Before we proceed with a detailed description, it is instructive to look at the simplest example of Linear Programming. In this case, a single iteration of the facial reduction procedure corresponds to finding redundant variables (in the primal) and implicit equality constraints (in the dual).

4.1 Preprocessing in linear programming

Improvements in the solution methods for large-scale linear programming problems have been dramatic since the late 1980's. A technique that has become essential in commercial software is a *preprocessing step* for the linear program before sending it to the solver. The preprocessing has many essential features. For example, it removes redundant variables (in the primal) and implicit equality constraints (in the dual) thus potentially dramatically reducing the size of the problem while

simultaneously improving the stability of the model. These steps in linear programming are examples of the Facial Reduction procedure, which we will formalize shortly.

Example 4.1.1 (primal facial reduction). Consider the problem

$$
\begin{aligned}
\min \quad & \begin{pmatrix} 2 & 6 & -1 & -2 & 7 \end{pmatrix} x \\
\text{s.t.} \quad & \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
& x \geq 0.
\end{aligned}
$$

If we sum the two constraints we see

$$
x \geq 0 \quad \text{and} \quad 2x_1 + x_4 + x_5 = 0 \quad \implies \quad x_1 = x_4 = x_5 = 0.
$$

Thus the coordinates x_1, x_4, and x_5 are identically zero on the entire feasible set. In other words, the feasible region is contained in the proper face $\{x \geq 0 : x_1 = x_4 = x_5 = 0\}$ of the cone $\mathbf{R}^n_+$. The zero coordinates can easily be eliminated and the corresponding columns discarded, yielding the equivalent simplified problem in the smaller face:

$$
\begin{aligned}
\min \quad & \begin{pmatrix} 6 & -1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \\
\text{s.t.} \quad & \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
& x_2, x_3 \geq 0, \quad x_1 = x_4 = x_5 = 0.
\end{aligned}
$$

The second equality can now also be discarded as it is is equivalent to the first.

How can such implicit zero coordinates be discovered systematically? Not surprisingly, the auxiliary system (3.2) provides the answer:

$$
0 \neq A^T y \geq 0, \qquad b^T y = 0.
$$

Suppose y is feasible for this auxiliary system. Then for any x feasible for the problem, we deduce $0 \leq \sum_i x_i (A^T y)_i = x^T (A^T y) = b^T y = 0$. Thus all the coordinates x_i, for which the strict inequality $(A^T y)_i > 0$ holds, must be zero.

Example 4.1.2 (dual facial reduction). A similar procedure applies to the dual. Consider the problem

$$\max \quad \begin{pmatrix} 2 & 6 \end{pmatrix} y$$
$$\text{s.t.} \quad \begin{bmatrix} -1 & -1 \\ 1 & 1 \\ 1 & -1 \\ -2 & 2 \end{bmatrix} y \leq \begin{pmatrix} 1 \\ 2 \\ 1 \\ -2 \end{pmatrix}.$$

Twice the third row plus the fourth row sums to zero. We conclude that the last two constraints are implicit equality constraints. Thus after substituting $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we obtain a simple univariate problem. Again, this discovery of implicit equality constraints can be done systematically by considering the auxiliary system (3.4):

$$0 \neq x \geq 0, \quad Ax = 0, \quad \text{and} \quad c^\top x = 0.$$

Suppose we find such a vector x. Then for any feasible vector y we deduce $0 \leq \sum_i x_i (c - A^T y)_i = x^T (c - A^T y) = 0$. Thus for each positive component $x_i > 0$, the corresponding inequality $(c - A^T y)_i \geq 0$ is fulfilled with equality along the entire feasible region.

4.2 Facial reduction in conic optimization

Keeping in mind the example of Linear Programming, we now formally describe the Facial Reduction procedure. To do this, consider the primal problem **(P)** failing Slater's condition. Our goal is to find an equivalent problem that does satisfy Slater's condition. To this end, suppose that we had a description of $\text{face}(\mathcal{F}_p, \mathcal{K})$ – the minimal face of $\mathcal{K}$ containing the feasible region $\mathcal{F}_p$. Then we could replace $\mathcal{K}$ with $\mathcal{K}' := \text{face}(\mathcal{F}_p, \mathcal{K})$, **E** with $\mathbf{E}' := \text{span } \mathcal{K}'$, and A with its restriction to $\mathbf{E}'$. The resulting smaller dimensional primal problem would automatically satisfy Slater's condition, since $\mathcal{F}_p$ intersects the relative interior of $\mathcal{K}'$.

The Facial Reduction procedure is a conceptual method that at termination discovers $\text{face}(\mathcal{F}_p, \mathcal{K})$. Suppose that $\mathcal{K}$ has nonempty interior. In the first iteration, the scheme determines any vector y satisfying the

auxiliary system (3.2). If no such vector exists, Slater's condition holds and the method terminates. Else, Theorem 3.1.3 guarantees that $\mathcal{F}_p$ lies in proper face $\mathcal{K}' := (A^*y)^\perp \cap \mathcal{K}$. Treating $\mathcal{K}'$ as a subset of the ambient Euclidean space $\mathbf{E}' := \text{span}\,\mathcal{K}'$ yields the smaller dimensional reformulation of $(\mathbf{P})$:

$$\begin{aligned} \min_{x \in \mathbf{E}'} \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K}'. \end{aligned} \tag{4.1}$$

We can now repeat the process on this problem. Since the dimension of the problem decreases with each facial reduction iteration, the procedure will terminate after at most $\dim \mathbf{E}$ steps.

Definition 4.2.1 (Singularity degree)**.** The *singularity degree* of $(\mathbf{P})$, denoted $\text{sing}(\mathbf{P})$, is the minimal number of iterations that are necessary for the Facial Reduction to terminate, over all possible choices of certificates generated by the auxiliary systems in each iteration.

The singularity degree of linear programs is at most one, as we will see shortly. More generally, such as for semi-definite programming problems, the singularity degree can be much higher.

The Facial Reduction procedure applies to the dual problem $(\mathbf{D})$ by using the equivalent primal form (2.5) and using Theorem 3.1.4. We leave the details to the reader.

4.3   Facial reduction in semi-definite programming

Before discussing further properties of the Facial Reduction algorithm in conic optimization, let us illustrate the procedure in semi-definite programming. To this end, consider the primal problem $(\mathbf{P})$ with $\mathcal{K} = \mathbf{S}_+^n$. Suppose that we have available a vector y feasible for the auxiliary system (3.2). Form now an eigenvalue decomposition

$$A^*y = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U & V \end{bmatrix}^T,$$

where $\begin{bmatrix} U & V \end{bmatrix} \in \mathbf{R}^{n \times n}$ is an orthogonal matrix and $\Lambda \in \mathbf{S}_{++}^r$ is a diagonal matrix. Then as we have seen in Example 2.2.3, the matrix

A^*y exposes the face $V\mathbf{S}^{n-r}V^T$ of $\mathbf{S}^n_+$. Consequently, defining the linear map $\widetilde{A}(Z) = A(VZV^T)$, the primal problem **(P)** is equivalent to the smaller dimensional SDP

$$\begin{aligned} \min_Z \quad & \langle V^TCV, Z\rangle \\ \text{s.t.} \quad & \widehat{A}(Z) = b \\ & Z \in \mathbf{S}^{n-r}_+. \end{aligned} \tag{4.2}$$

Thus one step of facial reduction is complete. Similarly let us look at the dual problem **(D)**. Suppose that X is feasible for the auxiliary system (3.4). Let us form an eigenvalue decomposition

$$X = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U & V \end{bmatrix}^T,$$

where $\begin{bmatrix} U & V \end{bmatrix} \in \mathbf{R}^{n\times n}$ is an orthogonal matrix and $\Lambda \in \mathbf{S}^r_{++}$ is a diagonal matrix. The face exposed by X, namely $V\mathbf{S}^{n-r}V^T$, contains all the feasible slacks $\{C - A^*y : y \in \mathcal{F}_d\}$ by Theorem 3.1.4. Thus defining the linear map $L(y) := V^T(A^*y)V$, the dual **(D)** is equivalent to the problem

$$\begin{aligned} \max_y \quad & \langle y, b\rangle \\ \text{s.t.} \quad & L(y) \preceq_{\mathbf{S}^{n-r}} V^TCV. \end{aligned} \tag{4.3}$$

Thus one step of Facial Reduction is complete and the process can continue.

To drive the point home, the following simple example shows that for SDP the singularity degree can indeed be strictly larger than one.

Example 4.3.1 (Singularity degree larger than one). Consider the primal SDP feasible region

$$\mathcal{F}_p = \{X \in \mathbf{S}^3_+ : X_{11} = 1, X_{12} + X_{33} = 0, X_{22} = 0\}.$$

Notice the equality $X_{22} = 0$ forces the second row and column of X to be zero, i.e. they are redundant. Let us see how this will be discovered by Facial Reduction.

The linear map $A\colon \mathbf{S^3} \to \mathbf{R}^3$ has the form

$$AX = (\operatorname{trace}(A_1X), \operatorname{trace}(A_2X), \operatorname{trace}(A_3X))^T$$

for the matrices

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Notice that $\mathcal{F}_p$ is nonempty since it contains the rank 1 matrix $e_1 e_1^T$. The auxiliary system (3.2) then reads

$$0 \neq \begin{bmatrix} v_1 & v_2/2 & 0 \\ v_2/2 & v_3 & 0 \\ 0 & 0 & v_2 \end{bmatrix} \succeq 0 \quad \text{and} \quad v_1 = 0.$$

Looking at the second principal minor, we see $v_3 = 0$. Thus all feasible v are positive multiples of the vector $e_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^T$. One step of Facial Reduction using the exposing vector $A^* e_3 = e_2 e_2^T$ yields the equivalent reduced region

$$\left\{ \begin{bmatrix} X_{11} & X_{13} \\ X_{13} & X_{33} \end{bmatrix} \succeq 0 : X_{11} = 1, X_{33} = 0 \right\}.$$

This reduced problem clearly fails Slater's condition, and Facial Reduction can continue. Thus the singularity degree of this problem is exactly two.

The pathological Example 4.3.1 can be generalized to higher dimensional space with $n = m$, by nesting, leading to problems with singularity degree $n - 1$; the construction is explained in Tunçel [136, page 43].

4.4   What facial reduction actually does

There is a direct and enlightening connection between Facial Reduction and the geometry of the image set $A\mathcal{K}$. To elucidate this relationship, we first note the following equivalent characterization of Slater's condition.

Proposition 4.4.1 (Range space characterization of Slater)**.** The primal problem **(P)** is strictly feasible if, and only if, the vector b lies in the relative interior of $A(\mathcal{K})$.

The following is the central result of this section.

Theorem 4.4.1 (Fundamental description of the minimal face).
Assume that the primal **(P)** is feasible. Then a vector v exposes a proper face of $A(\mathcal{K})$ containing b if, and only if, y satisfies the auxiliary system (3.2). Defining for notational convenience $N := \text{face}(b, A(\mathcal{K}))$, the following are true.

(I)   We always have:

$$\text{face}(\mathcal{F}_p, \mathcal{K}) = \mathcal{K} \cap A^{-1} N.$$

(II)   For any vector $y \in \mathbf{F}$ the following equivalence holds:

$$y \text{ exposes } N \quad \Longleftrightarrow \quad A^* y \text{ exposes } \text{face}(\mathcal{F}_p, \mathcal{K}).$$

In particular, the inequality $\text{sing}(\mathbf{P}) \leq 1$ holds if, and only if, $\text{face}(b, A(\mathcal{K}))$ is an exposed face of $A(\mathcal{K})$.

Some commentary is in order. First, as noted in Proposition 4.4.1, the primal **(P)** is strictly feasible if, and only if, the right-hand-side b lies in $\text{ri}\, A(\mathcal{K})$. Thus when strict feasibility fails, the set $N = \text{face}(b, A(\mathcal{K}))$ is a proper face of the image $A(\mathcal{K})$. The theorem above yields the exact description $\text{face}(\mathcal{F}_p, \mathcal{K}) = \mathcal{K} \cap A^{-1} N$ of the object we are after. On the other hand, determining a facial description of $A(\mathcal{K})$ is a difficult proposition. Indeed, even when $\mathcal{K}$ is a simple cone, the image $A(\mathcal{K})$ can be a highly nontrivial object. For instance, the image $A(\mathbf{S}_+^n)$ may fail to be facially exposed or even closed; examples are forthcoming.

Seeking to obtain a description of N, one can instead try to find "certificates" y exposing a proper face of $A(\mathcal{K})$ containing b. Such vectors y are precisely those satisfying the auxiliary system (3.2). In particular, part II of Theorem 4.4.1 yields a direct obstruction to having low singularity degree: $\text{sing}(\mathbf{P}) \leq 1$ if, and only if, $\text{face}(b, A(\mathcal{K}))$ is an exposed face of $A(\mathcal{K})$. Thus the lack of facial exposedness of the image $A(\mathcal{K})$ can become an obstruction. For the cone $\mathcal{K} = \mathbf{R}_+^n$, the image $A(\mathcal{K})$ is polyhedral and is therefore facially exposed. On the other hand, linear images of the cone $\mathcal{K} = \mathbf{S}_+^n$ can easily fail to be facially exposed. This in essence is the reason why preprocessing for general

conic optimization is much more difficult than its linear programming counterpart (having singularity degree at most one).

The following two examples illustrate the possibly complex geometry of image sets $A(\mathbf{S}^n)$.

Example 4.4.1 (Linear image not closed). Define the linear map $A \colon \mathbf{S}^2 \to \mathbf{R}^2$ by

$$A(X) := \begin{pmatrix} 2X_{12} \\ X_{22} \end{pmatrix}$$

Then the image $A(\mathbf{S}^2_+)$ is not closed, since

$$A\left(\begin{bmatrix} k & 1 \\ 1 & \frac{1}{k} \end{bmatrix}\right) \to \begin{pmatrix} 2 \\ 0 \end{pmatrix} \notin A\left(\mathbf{S}^2_+\right).$$

More broadly, it is easy to see the equality

$$A\left(\mathbf{S}^2_+\right) = \{(0,0)\} \cup (\mathbf{R} \times \mathbf{R}_{++}).$$

Example 4.4.2 (Linear image that is not facially exposed). Consider the feasible region in Example 4.3.1. There we showed that the singularity degree is equal to two. Consequently, by Theorem 4.4.1 we know that the minimal face of $A(\mathcal{K})$ containing $b = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$ must be nonexposed.

Let us verify this directly. To this end, we can without loss of generality treat A as mapping into $\mathbf{S}^2$ via

$$A(X) = \begin{bmatrix} X_{11} & X_{12} + X_{33} \\ X_{12} + X_{33} & X_{22} \end{bmatrix}$$
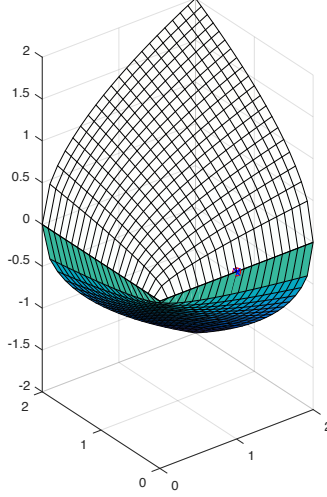
and identify b with $e_1 e_1^T$. Then the image $A(\mathbf{S}^3_+)$ is simply the sum,

$$A(\mathbf{S}^3_+) = \mathbf{S}^2_+ + \mathrm{cone} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

See Figure 4.1.

Consider the set

$$G := \mathrm{cone} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Figure 4.1: Nonexposed face of the image set

We claim that G is a face of $A(\mathbf{S}_+^3)$ and is therefore the minimal face containing $e_1 e_1^T$. Indeed, suppose we may write

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} + X_{33} \\ X_{12} + X_{33} & X_{22} \end{bmatrix} + \begin{bmatrix} X'_{11} & X'_{12} + X'_{33} \\ X'_{12} + X'_{33} & X'_{22} \end{bmatrix},$$

for some matrices $X, X' \in \mathbf{S}_+^3$. Comparing the $2,2$-entries, we deduce $X_{22} = X'_{22} = 0$ and consequently $X_{12} = X'_{12} = 0$. Comparing the 1,2-entries yields $X_{33} = X'_{33} = 0$. Thus both summands lie in G; therefore G is a face of the image $A(\mathbf{S}_+^3)$. Next, using Lemma 2.1.2, observe

$$A(\mathbf{S}_+^3)^* = \left(\mathbf{S}_+^2 + \mathrm{cone} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right)^* = \{ Y \in \mathbf{S}_+^2 : Y_{12} \geq 0 \}.$$

Consequently, any matrix exposing G must lie in the set

$$A(\mathbf{S}^3)^* \cap G^\perp = \mathrm{cone}\, (e_2 e_2^T).$$

On the other hand, the set

$$(e_2 e_2^T)^\perp \cap A(\mathbf{S}_+^3) = \mathrm{cone} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \mathrm{cone} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is clearly strictly larger than G. Hence G is not an exposed face.

4.5   Singularity degree and the *Hölder error bound* in SDP

For semi-definite programming, the singularity degree plays an especially important role, controlling the *Hölderian stability of the feasible region.* Consider two sets Q_1 and Q_2 in $\mathbf{E}$. A convenient way to understand the regularity of the intersection $Q_1 \cap Q_2$ is to determine the extent to which the computable residuals, $\mathrm{dist}(x, Q_1)$ and $\mathrm{dist}(x, Q_2)$, bound the *error* $\mathrm{dist}(x, Q_1 \cap Q_2)$. Relationships of this sort are commonly called *error bounds* of the intersection and play an important role for convergence and stability of algorithms. Of particular importance are Hölderian error bounds – those asserting the inequalities

$$\mathrm{dist}(x, Q_1 \cap Q_2) \leq \mathcal{O}\left(\mathrm{dist}^q(x, Q_1) + \mathrm{dist}^q(x, Q_2)\right)$$

on compact sets, for some powers $q \geq 0$. For semi-definite programming, the singularity degree precisely dictates the Hölder exponent q.

Theorem 4.5.1 (Hölderian error bounds from the singularity degree). Consider a feasible primal SDP problem **(P)** and define the affine space

$$\mathcal{V} := \{X : A(X) = b\}.$$

Set $d := \mathrm{sing}(\mathbf{P})$. Then for any compact set U, there is a real $c > 0$ so that

$$\mathrm{dist}_{\mathcal{F}_p}(X) \leq c\left(\mathrm{dist}^{2^{-d}}(X, \mathbf{S}_+^n) + \mathrm{dist}^{2^{-d}}(X, \mathcal{V})\right), \quad \text{for all } x \in U.$$

What is remarkable about this result is that neither the dimension of the matrices n, the number of inequalities m, nor the rank of the matrices in the region $\mathcal{F}_p$ determines the error bound. Instead, it is only the single quantity, the singularity degree, that drives this regularity concept.

Example 4.5.1 (Worst-case example). Consider the SDP feasible region

$$\mathcal{F}_p := \{X \in \mathbf{S}_+^n : X_{22} = 0 \text{ and } X_{k+1,k+1} = X_{1,k} \text{ for } k = 2, 3, \ldots, n-1\}.$$

For any feasible X, the constraint $X_{22} = 0$ forces $0 = X_{12} = X_{33}$. By an inductive argument, then we deduce $X_{1k} = 0$ and $X_{k,k} = 0$ for all $k = 2, \ldots, n$. Thus the feasible region $\mathcal{F}_p$ coincides with the ray $\mathrm{cone}\,(e_1 e_1^T)$.

Given $\epsilon > 0$, define the matrix

$$
X(\epsilon) = \begin{bmatrix}
n & \epsilon^{2^{-1}} & \epsilon^{2^{-2}} & \dots & \epsilon^{2^{-(n-1)}} \\
\epsilon^{2^{-1}} & \epsilon & 0 & \dots & 0 \\
\epsilon^{2^{-2}} & 0 & \epsilon^{2^{-1}} & & \\
\vdots & \vdots & & \ddots & \\
\epsilon^{2^{n-1}} & 0 & & & \epsilon^{2^{-(n-2)}}
\end{bmatrix}
$$

Observe that $X(\epsilon)$ violates the linear constraints only in the requirement $X_{22} = 0$. Consequently, the distance of $X(\epsilon)$ to the linear space $\mathcal{V} := \{X \in \mathbf{S}^n : A(X) = b\}$ is on the order of ϵ. On the other hand, the distance of $X(\epsilon)$ to the solution set is at least on the order of $\epsilon^{2^{-(n-1)}}$. This example shows that the Hölder exponent in this case is at least $2^{-(n-1)}$. Combined with Theorem 4.5.1 and the fact that the feasible region contains rank one matrices, we deduce $\text{sing}(\mathbf{P}) = n - 1$ and the Hölder exponent guaranteed by the theorem is sharp.

4.6 Towards computation

The Facial Reduction procedure is conceptual. To implement it, since the error compounds along the iterations, one must be able to either solve the auxiliary systems (3.2) (resp. (3.4)) to machine precision in each iteration or certify that the systems are inconsistent. On the other hand, in general, there is no reason to believe that solving a single auxiliary system is any easier than solving the original problem **(P)**.

One computational approach for facial reduction in SDP, explored by Permenter-Parrilo [110], is to *relax* the auxiliary problems to ones that are solvable. Instead of considering (3.2), one can choose a convex cone $\widehat{\mathcal{K}} \subseteq \mathbf{S}^n_+$ so that consistency of the system can be checked:

$$
0 \neq A^* y \in \widehat{\mathcal{K}} \qquad \text{and} \qquad \langle b, y \rangle = 0.
$$

If a vector y satisfying the system is found, then one can perform one step of facial reduction. If not, the scheme quits, possibly without having successfully deduced that Slater's condition holds. Simple examples of $\widehat{\mathcal{K}}$ are the sets $\text{Diag}(\mathbf{R}^n_+)$ and the cone dual to $\{X \in \mathbf{S}^n : \text{every } 2 \times 2 \text{ minor of } X \text{ is PSD}\}$, where PSD denote positive semi-definite. The above feasibility problem is then an instance of

linear programming in the first case and of second-order cone programming in the second. More details are provided in [110]. Readers may be skeptical of this strategy since this technique will work only for special types of degeneracy. For example, the first relaxation $\mathrm{Diag}(\mathbf{R}_+^n)$ can only detect that some diagonal elements of X are identically zero on the feasible region $\mathcal{F}_p$. On the other hand, it does appear that degeneracy typically arising in applications *is* highly structured, and promising numerical results has been reported in [110].

There exist other influential techniques for regularizing conic optimization problems that are different from the facial reduction procedure. Two notable examples are Ramana's extended dual [114] and the homogeneous self-dual embedding e.g., [39, 109]. The latter, in particular, is used by MOSEK [8] and SeDuMi [128]. A dual approach, called the *conic expansion approach* is discussed at length in [142], see also [88, 90, 104, 127]

We do not discuss these techniques here. Instead, we focus on the most promising class of conic optimization problems – those having singularity degree at most one. In the rest of the manuscript, we provide a series of influential examples, where the structure of the problem enables one to obtain feasible points of the auxiliary systems without having to invoke any solvers. Numerical illustrations show that the resulting reduced subproblems are often much smaller and more stable than the original.

4.7   Commentary

Preprocessing is essential in making LP algorithms efficient. A main ingredient is identifying primal and/or dual slack variables that are identically zero on the feasible set. This is equivalent to facial reduction that reduces the problem to faces of the nonnegative orthant, e.g., [67] The facial reduction procedure for general conic optimization started in [22, 23, 24]. The procedure provides a primal-dual pair of conic optimization problems that are *proper* in the sense that the dual of the dual yields the primal. Example 4.3.1 and extensions can be found in Tunçel [136, page 43]. The notation of singularity degree and its con-

nection to error bounds (Theorem 4.5.1) was discovered by Sturm [129, Sect. 4]; Example 4.5.1 appears in this paper as well. Example 4.4.2 is motivated by Example 1 in [106]. Theorem 4.4.1 appeared in [47]. As mentioned previously, there are many approaches to "regularization" of conic optimization problem, aside from facial reduction, including the self-dual embedding and the approximation approaches in [109, 110, 112]. An alternate view of obtaining a dual without a constraint qualification was given in [114, 115], though a relationship to facial reduction was later explained in [116].

Part II

Applications and illustrations

In this chapter, we discuss a number of diverse and important computational problems. This includes various matrix completion problems and discrete optimization problems such as the quadratic assignment problem, graph partitioning, and the strengthened relaxation of the maximum cut problem. In the final section, we also discuss sum of squares relaxations for polynomial optimization problems. In each case, we use the structure of the problem to determine a face of the positive semi-definite cone containing the entire feasible region. One exception is the matrix completion problem, where we instead determine a face containing the *optimal face*, as opposed to the entire feasible region. Numerical illustrations illustrate the efficacy of the approach.

5

Matrix completions

We begin with various matrix completion problems. Broadly speaking, the goal is to complete a partially specified matrix, while taking into account a priori known structural properties such as a given rank or sparsity. There is a great variety of references on matrix completion problems; see for example [68, 81, 84, 95].

5.1 Positive semi-definite matrix completion

We begin with a classical problem of completing a PSD matrix from partially observed entries. To model this problem, consider an undirected graph $G = (V, E)$ with a vertex set $V = \{1, \ldots, n\}$ and *graph* edge set $E \subseteq \{ij : 1 \leq i \leq j \leq n\}$. The symbols ij and ji always refer to the same edge. Notice we allow self-loops $ii \in E$. For simplicity, we in fact assume that E contains ii for each node $i \in V$. Elements ω of $\mathbf{R}^E$ are called *partial matrices*, as they specify entries of a partially observed symmetric $n \times n$ matrix. Given a partial matrix $\omega \in \mathbf{R}^E$, the *PSD completion problem* asks to determine, if possible, a matrix X in the set

$$\mathcal{F}_p := \{X \in \mathbf{S}_+^n : X_{ij} = \omega_{ij} \quad \text{for all indices } ij \in E\}.$$

That is, we seek to complete the partial matrix ω to an $n \times n$ positive semi-definite matrix. When do such PSD completions exist, that is, when is $\mathcal{F}_p$ nonempty? Clearly, a necessary condition is that ω is a *partial PSD matrix*, meaning that its restriction to any specified principal submatrix is PSD. This condition, however, is not always sufficient.

A graph $G = (V, E)$ is called *PSD completable* if every partial PSD matrix $\omega \in \mathbf{R}^E$ is completable to a PSD matrix. It turns out that PSD completable graphs are precisely those that are chordal. Recall that a graph G is called *chordal* if any cycle of four or more nodes has a chord – an edge joining any two nodes that are not adjacent in the cycle.

Theorem 5.1.1 (PSD completable and chordal graphs). The graph G is PSD completable if, and only if, G is chordal.[1]

Chordal graphs, or equivalently those that are PSD completable, play a fundamental role for the PSD completion problem. For example, on such graphs, the completion problem admits an efficient combinatorial algorithm [61, 83, 124].

Next, we turn to Slater's condition. Consider the completion problem

$$\begin{bmatrix} 1 & 1 & ? & ? \\ 1 & 1 & 1 & ? \\ ? & 1 & 1 & -1 \\ ? & ? & -1 & 2 \end{bmatrix}. \tag{5.1}$$

The question marks ? denote the unknown entries. The underlying graph on four nodes is a path and is therefore chordal. The known entries make up a partial PSD matrix since the three specified principal minors are PSD. Thus by Theorem 5.1.1, the completion problem is solvable. Does Slater's condition hold? The answer is no. The first leading principal minor is singular, and therefore any PSD completion must be singular.

By the same logic, any singular specified principal minor of a partial matrix $\omega \in \mathbf{R}^E$ certifies that strict feasibility fails. Much more is true, however. We now show how any singular specified principal minor of a

---

[1] If all the self-loops are not included in E, one needs to add that the two subgraphs with self-loops and without are disconnected from each other, see e.g., [47].

partial matrix $\omega \in \mathbf{R}^E$ yields a face of $\mathbf{S}^n_+$ containing the entire feasible region, allowing one to reduce the dimension of the problem.

To see how this can be done, let us introduce some notation. Define the *coordinate projection* map $\mathcal{P}_E \colon \mathbf{S}^n \to \mathbf{R}^E$ by setting

$$\mathcal{P}_E(X) = (X_{ij})_{ij \in E}.$$

In this notation, we can write

$$\mathcal{F}_p := \{X \in \mathbf{S}^n_+ : \mathcal{P}_E(X) = \omega\}.$$

We will now see how the geometry of the image set $\mathcal{P}_E(\mathbf{S}^n_+)$, along with Theorem 4.4.1, helps us discover a face of $\mathbf{S}^n_+$ containing the feasible region. We note in passing that the image $\mathcal{P}_E(\mathbf{S}^n_+)$ is always closed.[2]

Proposition 5.1.1 (Closure of the image). The image $\mathcal{P}_E(\mathbf{S}^n_+)$ is closed.

The reader can check that the adjoint $\mathcal{P}_E^* \colon \mathbf{R}^E \to \mathbf{S}^n$ simply pads partial matrices in $\mathbf{R}^E$ with zeros:

$$(\mathcal{P}_E^*(y))_{ij} = \begin{cases} y_{ij}, & \text{if } ij \in E \text{ or } ji \in E \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

For any subset of vertices $\alpha \subseteq V$, we let $E[\alpha] := \{ij \in E : i, j \in \alpha\}$ be the edge set induced by G on α and we set $\omega[\alpha]$ to be the restriction of ω to $E[\alpha]$. Define the *relaxed region*

$$\mathcal{F}_p(\alpha) := \{X \in \mathbf{S}^n_+ : P_{E[\alpha]}(X) = \omega[\alpha]\}. \tag{5.3}$$

Clearly $\alpha \subseteq V$ means we have fewer constraints and

$$\mathcal{F}_p = \mathcal{F}_p(V) \subseteq \mathcal{F}_p(\alpha), \qquad \text{for all } \alpha \subseteq V.$$

A subset $\alpha \subseteq V$ is called a *clique* if for any two nodes $i, j \in \alpha$ the edge ij lies in E. Specified principal minors of ω correspond precisely to cliques in the graph G. We can moreover clearly identify $\mathbf{R}^{E[\alpha]}$ with the matrix space $\mathbf{S}^k$. Suppose now that $\omega[\alpha]$ has rank $r < k$, i.e., the principal submatrix of ω indexed by α is singular. Then the right-hand-side

---

[2]If some elements ii do not lie in E, contrary to our simplifying assumption, then the image $\mathcal{P}_E(\mathbf{S}^n_+)$ can fail to be closed. A precise characterization is given in [47].

$\omega[\alpha]$ of (5.3) lies in the boundary of the image set $\mathcal{P}_{E[\alpha]}(\mathbf{S}^n_+) \subseteq \mathbf{S}^k_+$. Let V_α be an exposing vector of face$(\omega[\alpha], \mathbf{S}^k_+)$. Then by Theorem 4.4.1, we can be sure that $W_\alpha := \mathcal{P}^*_\alpha(V_\alpha)$ exposes the minimal face of $\mathbf{S}^n_+$ containing the entire region $\mathcal{F}_p(\alpha)$. Given a collection of cliques $\alpha_1, \alpha_2, \ldots, \alpha_l$, we can perform the same procedure and deduce that the entire feasible region $\mathcal{F}_p$ lies in the face

$$(W^\perp_{\alpha_1} \cap \mathbf{S}^n_+) \cap (W^\perp_{\alpha_2} \cap \mathbf{S}^n_+) \cap \ldots \cap (W^\perp_{\alpha_l} \cap \mathbf{S}^n_+),$$

which by Proposition 2.2.1 admits the equivalent description

$$(W_{\alpha_1} + W_{\alpha_2} + \ldots + W_{\alpha_l})^\perp \cap \mathbf{S}^n_+.$$

The following example will clarify the strategy.

Example 5.1.1 (Reducing the PSD completion problem). Let $\mathcal{F}_p$ consist of all matrices $X \in \mathbf{S}^4_+$ solving the PSD completion problem (5.1). There are three nontrivial cliques in the graph, all of size 2. The minimal face of $\mathbf{S}^2_+$ containing the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

is exposed by

$$\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Moreover, the matrix $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ is definite and hence the minimal face of $\mathbf{S}^2_+$ containing this matrix is exposed by the all-zero matrix.

The intersection of exposed faces is exposed by the sum of their exposing vectors. We deduce that $\mathcal{F}_p$ is contained in the face of $\mathbf{S}^4_+$ exposed by the sum

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

After finding the nullspace of this matrix, we deduce

$$\mathcal{F}_p \subseteq \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S}_+^2 \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}^T .$$

The following lemma is another nice consequence of the procedure described in the above example.

Lemma 5.1.2 (Completion of banded all ones matrices)**.** The matrix of all ones is the unique positive semi-definite matrix satisfying $X_{ij} = 1$ for all indices with $|i - j| \leq 1$.

Proof. Consider the edge set $E = \{ij : |i - j| \leq 1\}$ and let $\omega \in \mathbf{R}^E$ be a partial matrix of all ones. Observe ω has n specified 2×2-principal submatrices, each having rank 1. By the same logic as in Example 5.1.1, it follows that the feasible region $\mathcal{F}_p$ is zero-dimensional, as claimed. $\square$

The strategy outlined above suggests an algorithm for finding the minimal face based on exploiting cliques in the graph. This strategy is well-founded at least for chordal graphs.

Theorem 5.1.3 (Finding the minimal face on chordal graphs)**.** Suppose that G is chordal and consider a partial PSD matrix $\omega \in \mathbf{R}^E$. Then the equality

$$\text{face}(\mathcal{F}_p, \mathbf{S}_+^n) = \bigcap_{\alpha \in \Theta} \text{face}(\mathcal{F}_p(\alpha), \mathbf{S}_+^n) \qquad \text{holds,}$$

where Θ denotes the set of all maximal cliques in G.

On the other hand, it is important to realize that when the graph is not chordal, the minimal face $\text{face}(\mathcal{F}_p, \mathbf{S}_+^n)$ is not always guaranteed to be found from cliques alone. The following example shows a PSD completion problem that fails Slater's condition but where all the faces arising from cliques are trivial.

Example 5.1.2 (Slater condition & nonchordal graphs, [47])**.** Let $G = (V, E)$ be the graph with $V = \{1, 2, 3, 4\}$ and $E = \{12, 23, 34, 14\} \cup$

$\{11, 22, 33, 44\}$. Define the corresponding PSD completion problems $C(\epsilon)$, parametrized by $\epsilon \geq 0$:

$$
C(\epsilon): \qquad
\begin{bmatrix}
1+\epsilon & 1 & ? & -1 \\
1 & 1+\epsilon & 1 & ? \\
? & 1 & 1+\epsilon & 1 \\
-1 & ? & 1 & 1+\epsilon
\end{bmatrix}.
$$

Let $\omega(\epsilon) \in \mathbf{R}^E$ denote the corresponding partial matrices. From Lemma 5.1.2, the PSD completion problem $C(0)$ is infeasible, that is $\omega(0)$ lies outside of $\mathcal{P}_E(\mathbf{S}_+^4)$. On the other hand, for all sufficiently large ϵ, the partial matrices $\omega(\epsilon)$ lie in $\operatorname{int} \mathcal{P}(\mathbf{S}_+^4)$ by diagonal dominance. Since $\mathcal{P}_E(\mathbf{S}_+^4)$ is closed by Proposition 5.1.1, we deduce that there exists $\hat{\epsilon} > 0$, so that $\omega(\hat{\epsilon})$ lies on the boundary of $\mathcal{P}_E(\mathbf{S}_+^4)$, that is Slater's condition <u>fails</u> for the completion problem $C(\hat{\epsilon})$. In fact, it can be shown that the smallest such ϵ is $\hat{\epsilon} = \sqrt{2} - 1$ with completion values of 0 in the ? positions in $C(\hat{\epsilon})$. On the other hand, all specified principal matrices of $\omega(\epsilon)$ for $\epsilon > 0$ are clearly positive definite, and therefore all the corresponding faces are trivial. Thus we have found a partial matrix $\omega(\hat{\epsilon})$ that has a singular completion but the minimal face cannot be found from an intersection using the cliques of the graph.

Given the importance of singularity degree, the following question arises naturally. Which graphs $G = (V, E)$ have the property that the cone $\mathcal{P}_E(\mathbf{S}_+^n)$ is facially exposed? Equivalently, on which graphs $G = (V, E)$ does every feasible PSD completion problem have singularity degree at most one? Let us make the following definition. The *singularity degree of a graph* $G = (V, E)$ is the maximal singularity degree among all completion problems with PSD completable partial matrices $\omega \in \mathbf{R}^E$.

Chordal graphs have singularity degree one [47], and surprisingly these are the only graphs with this property [132].

Corollary 5.1.4 (Singularity degree of chordal completions).
The graph G has singularity degree one if, and only if, G is chordal.
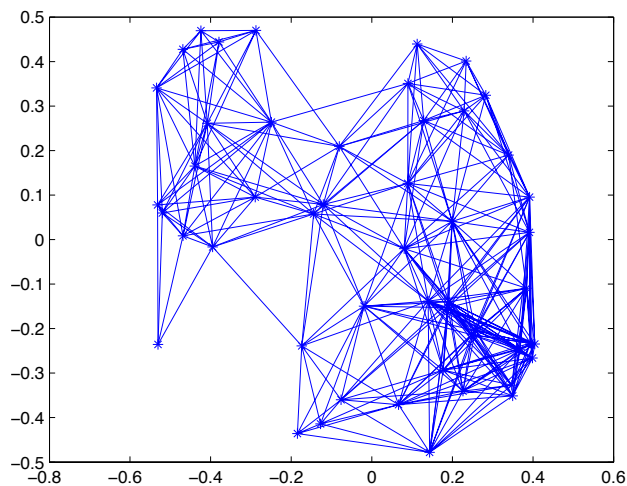
5.2   Euclidean distance matrix completion, EDMC

In this section, we discuss a problem that is closely related to the PSD completion problem of the previous section, namely the Euclidean distance matrix completion, EDMC, problem. As we will see, the EDMC problem inherently fails Slater's condition, and facial reduction once again becomes applicable by analyzing certain cliques in a graph.

Setting the stage, fix an undirected graph $G = (V, E)$ on a vertex set $V = \{1, 2, \ldots, n\}$ with an edge set $E \subseteq \{ij : 1 \leq i < j \leq n\}$. Given a partial matrix $d \in \mathbf{R}^E$, the *Euclidean distance matrix completion problem* asks to determine if possible an integer k and a collection of points $p_1, \ldots, p_n \in \mathbf{R}^k$ satisfying

$$d_{ij} = \|p_i - p_j\|^2 \quad \text{for all } ij \in E.$$

See figure 5.1 for an illustration.

Figure 5.1: Instance of EDMC

We now see how this problem can be modeled as an SDP. To this end, let us introduce the following notation. A matrix $D \in \mathbf{S}^n$ is called a *Euclidean distance matrix, EDM*, if there exists an integer k and points $p_1, \ldots, p_n \in \mathbf{R}^k$ satisfying

$$D_{ij} = \|p_i - p_j\|^2 \quad \text{for all } i, j = 1, \ldots, n.$$

Such points $p_1, \ldots, p_n$ are said to *realize D in* $\mathbf{R}^k$. The smallest integer k such that there exist points in $\mathbf{R}^k$ realizing D is called the *embedding dimension of D*, and is denoted by embdim D. We let $\mathcal{E}^n$ denote the set of all $n \times n$ EDM matrices. In this language the EDM completion problem reads: given a partial matrix $d \in \mathbf{R}^E$ determine a matrix in the set

$$\{D \in \mathcal{E}^n : D_{ij} = d_{ij} \quad \text{for all } ij \in E\}.$$

Thus the EDM completion problem is a conic feasibility problem. Since we are interested in facial reduction, the facial structure of $\mathcal{E}^n$ is central.

Notice that $\mathcal{E}^n$ has empty interior since it is contained in the space of *hollow matrices*

$$\mathbf{S}_H^n := \{D \in \mathbf{S}^n : \operatorname{diag}(D) = 0\}.$$

A fundamental fact often used in the literature is that $\mathcal{E}^n$ is linearly isomorphic to $\mathbf{S}^{n-1}$. More precisely, consider the mapping

$$\mathcal{K} \colon \mathbf{S}^n \to \mathbf{S}^n$$

defined by

$$\mathcal{K}(X)_{ij} := X_{ii} + X_{jj} - 2X_{ij}. \tag{5.4}$$

Clearly $\mathcal{K}$ maps into the space of hollow matrices $\mathbf{S}_H^n$. One can quickly verify that the adjoint is given by

$$\mathcal{K}^*(D) = 2(\operatorname{Diag}(De) - D).$$

Moreover, the range of the adjoint $\mathcal{K}^*$ is the space of *centered matrices*

$$\mathbf{S}_c^n := \{X \in \mathbf{S}^n : Xe = 0\}.$$

The following result is fundamental.

Theorem 5.2.1 (Parametrization of the EDM cone)**.** The map $\mathcal{K} \colon \mathbf{S}_c^n \to \mathbf{S}_H^n$ is a linear isomorphism carrying $\mathbf{S}_c^n \cap \mathbf{S}_+^n$ onto $\mathcal{E}^n$. The inverse $\mathcal{K}^\dagger \colon \mathbf{S}_H^n \to \mathbf{S}_c^n$ is the map $\mathcal{K}^\dagger(D) = -\frac{1}{2}JDJ$, where $J := I - \frac{1}{n}ee^T$ is the orthogonal projection onto $e^\perp$.[3]

---

[3]In fact, we can consider the map as $\mathcal{K} \colon \mathbf{S}^n \to \mathbf{S}^n$. Then we still have $\mathcal{K}(\mathbf{S}_+^n) = \mathcal{E}^n$ and the *Moore-Penrose pseudoinverse* $\mathcal{K}^\dagger(D) = -\frac{1}{2}J \operatorname{offDiag}(D) J$, where offDiag zeros out the diagonal.

In particular, the cone $\mathcal{E}^n$ is linearly isomorphic to the cone $\mathbf{S}_c^n \cap \mathbf{S}_+^n$. On the other hand, observe for any matrix $X \in \mathbf{S}_+^n$ the equivalence

$$Xe = 0 \iff e^T X e = 0 \iff \operatorname{trace}(Xee^T) = 0 \iff X \in (ee^T)^\perp.$$

Thus $\mathcal{E}^n$ is linearly isomorphic to the face

$$\mathbf{S}_c^n \cap \mathbf{S}_+^n = (ee^T)^\perp \cap \mathbf{S}_+^n$$

as claimed. More explicitly, forming an $n \times n$ orthogonal matrix $\left[\frac{1}{\sqrt{n}}e \quad U\right]$ yields the equality $\mathbf{S}_c^n \cap \mathbf{S}_+^n = U\mathbf{S}_+^{n-1}U$.

Thus the EDM completion problem amounts to finding a matrix X in the set

$$\mathcal{F}_p = \{X \in \mathbf{S}_c^n \cap \mathbf{S}_+^n : \mathcal{P}_E \circ \mathcal{K}(X) = d\}. \tag{5.5}$$

To see how to recover the realizing points p_i of $d \in \mathbf{R}^E$, consider a matrix $X \in \mathcal{F}_p$ and form a factorization $X = PP^T$ for some matrix $P \in \mathbf{R}^{n \times k}$ with $k = \operatorname{rank} X$. Let $p_1, \ldots, p_n \in \mathbf{R}^k$ be the rows of P. Then X lying in $\mathbf{S}_c^n$ implies $\sum_{i=1}^n p_i = 0$, that is, the points p_i are centered around the origin, while the constraint $\mathcal{P}_E \circ \mathcal{K}(X) = d$ implies

$$d_{ij} = X_{ii} + X_{jj} - 2X_{ij} = \|p_i\|^2 + \|p_j\|^2 - 2p_i^T p_j = \|p_i - p_j\|^2,$$

for all $ij \in E$. Hence the points $p_1, \ldots, p_n$ solve the EDM completion problem.

Let us turn now to understanding (strict) feasibility of $\mathcal{F}_p$. A vector $d \in \mathbf{R}^E$ is called a *partial EDM* if the restriction of d to every specified principal submatrix is an EDM. The graph G is *EDM completable* if every partial EDM $d \in \mathbf{R}^E$ is *EDM completable.* The following result, proved in [15], is a direct analogue of Theorem 5.1.1 for the PSD completion problem.

Theorem 5.2.2 (EDM completability & chordal graphs)**.** The graph G is EDM completable if, and only if, G is chordal.

We also mention in passing the following observation from [47].

Theorem 5.2.3 (Closedness of the projected EDM cone)**.** The projected image $\mathcal{P}(\mathcal{E}^n)$ is always closed. $\qquad\square$

Given a clique α in G, we let $\mathcal{E}^\alpha$ denote the set of $|\alpha| \times |\alpha|$ Euclidean distance matrices indexed by α. In what follows, given a partial matrix $d \in \mathbf{R}^E$, the restriction $d[\alpha]$ can then be thought of either as a vector in $\mathbf{R}^{E(\alpha)}$ or as a hollow matrix in $\mathbf{S}^\alpha$. We also use the symbol $\mathcal{K}_\alpha \colon \mathbf{S}^\alpha \to \mathbf{S}^\alpha$ to indicate the mapping $\mathcal{K}$ acting on $\mathbf{S}^\alpha$. The following recipe provides a simple way to discover faces of the PSD cone containing the feasible region from specified cliques in the graph.

Theorem 5.2.4 (Clique facial reduction for EDM completions). Let α be any k-clique in the graph G. Let $d \in \mathbf{R}^E$ be a partial Euclidean distance matrix and define the relaxation

$$\mathcal{F}_p(\alpha) \quad := \quad \{X \in \mathbf{S}_+^n \cap \mathbf{S}_c^n : P_{E[\alpha]} \circ \mathcal{K}(X) = d[\alpha]\}.$$

Then for any matrix V_α exposing face $\big(\mathcal{K}_\alpha^\dagger(d[\alpha]), \mathbf{S}_+^\alpha \cap \mathbf{S}_c^\alpha\big)$, the matrix

$$\mathcal{P}_{E[\alpha]}^* V_\alpha \quad \text{exposes} \quad \text{face}(\mathcal{F}_p(\alpha), \mathbf{S}_+^n \cap \mathbf{S}_c^n).$$

In other words, the recipe is as follows. Given a clique α in G, consider the matrix $\mathcal{K}_\alpha^\dagger(d[\alpha]) \in \mathbf{S}_+^\alpha \cap \mathbf{S}_c^\alpha$. Let $V_\alpha \in \mathbf{S}_+^\alpha \cap \mathbf{S}_c^\alpha$ be an exposing vector of face $\big(\mathcal{K}_\alpha^\dagger(d[\alpha]), \mathbf{S}_+^\alpha \cap \mathbf{S}_c^n\big)$. Then $\mathcal{P}_{E[\alpha]}^* V_\alpha$ is an extension of V_α to $\mathbf{S}^n$ obtained by padding V_α with zeroes. The above theorem guarantees that the entire feasible region of (5.5) is contained in the face of $\mathbf{S}_c^n \cap \mathbf{S}_+^n$ exposed by $\mathcal{P}_{E[\alpha]}^* V_\alpha$.

In direct analogy with Theorem 5.1.3 for PSD completions, the minimal face is sure to be discovered in this way for chordal graphs.

Theorem 5.2.5 (Clique facial reduction for EDM is sufficient). Suppose that G is chordal, and consider a partial Euclidean distance matrix $d \in \mathbf{R}^E$ and the region

$$F := \{X \in \mathbf{S}_c^n \cap \mathbf{S}_+^n : [\mathcal{K}(X)]_{ij} = d_{ij} \text{ for all } ij \in E\}.$$

Let Θ denote the set of all maximal cliques in G, and for each $\alpha \in \Theta$ define

$$F_\alpha := \{X \in \mathbf{S}_c^n \cap \mathbf{S}_+^n : [\mathcal{K}(X)]_{ij} = d_{ij} \text{ for all } ij \in E[\alpha]\}.$$

Then the equality

$$\text{face}(F, \mathbf{S}_c^n \cap \mathbf{S}_+^n) = \bigcap_{\alpha \in \Theta} \text{face}(F_\alpha, \mathbf{S}_c^n \cap \mathbf{S}_+^n) \qquad \text{holds.}$$

Corollary 5.2.6 (Singularity degree of chordal completions)**.** If the graph $G = (V, E)$ is chordal, then the EDM completion problem has singularity degree at most one, when feasible.

Finally, in analogy with Corollary 5.2.7, the following is true. Define the *singularity degree of a graph* $G = (V, E)$ to be the maximal singularity degree among all EDM completion problems with EDM completable partial matrices $d \in \mathbf{R}^E$.

Corollary 5.2.7 (Singularity degree of chordal completions)**.**
The graph G has singularity degree one if, and only if, G is chordal.

5.2.1 EDM and SNL with exact data

The material above explains in part the surprising success of the algorithm in [76] for the *sensor network localization problem, SNL*. The SNL problem differs from the EDM completion problem only in that some of the points or *sensors p_i* that define the problem are in fact *anchors* and their positions are known. The algorithm proceeds by iteratively finding faces of the PSD cone from cliques and intersecting them two at a time, thereby decreasing the dimension of the problem in each step. In practice, this procedure often terminates with a unique solution of the problem. We should mention that the anchors are a *red herring.* Indeed, they should only be treated differently than the other sensors after *all* the sensors have been localized. In the post-processing step, a so-called Procrustes problem is solved to bring the putative anchors as close as possible to their original (known) positions and thus rotating the sensor positions appropriately. Another important point in applications is that the distances for sensors that are close enough to each other are often known. This suggests that there are often many local cliques in the graph. This means that the resulting SDP relaxation is highly degenerate but this degeneracy can be exploited as we have seen above.

Some numerical results from the year 2010 in [76] appear in Table 5.1. These results are on random, noiseless problems using a 2.16 GHz Intel Core 2 Duo, 2 GB of RAM. The embedding dimension is $r = 2$ and the sensors are in a square region $[0, 1] \times [0, 1]$ with $m = 9$

# sensors	# anchors	radio range	RMSD	Time
20000	9	.025	$5e{-}16$	25s
40000	9	.02	$8e{-}16$	1m 23s
60000	9	.015	$5e{-}16$	3m 13s
100000	9	.01	$6e{-}16$	9m 8s

Table 5.1: Empirical results for SNL

anchors. We use the Root Mean Square Deviation to measure the quality of the solution:

$$\text{RMSD} = \left(\frac{1}{n} \sum_{i=1}^{n} \| p_i - p_i^{\text{true}} \|^2 \right)^{1/2}$$

The *huge* expected number of constraints and variables in the four problems in Table 5.1 are

$$
\begin{aligned}
M &= \begin{pmatrix} 3,078,915 & 12,315,351 & 27,709,309 & 76,969,790 \end{pmatrix} \\
N &= 10^9 \begin{pmatrix} 0.2000 & 0.8000 & 1.8000 & 5.0000 \end{pmatrix},
\end{aligned}
$$

respectively[4].

5.2.2 Extensions to noisy EDM and SNL problems

When there is noise in the distance measurements – the much more realistic setting – the approach requires an intriguing modification. Let us see what goes wrong, in the standard approach. Given a clique α, let us form $\mathcal{K}_\alpha^\dagger(d[\alpha])$ as in Theorem 5.2.4. The difficulty is that this matrix is no longer PSD. On the other hand, it is simple enough to find the nearest matrix W of $\mathbf{S}_+^\alpha \cap \mathbf{S}_c^n$ to $\mathcal{K}_\alpha^\dagger(d[\alpha])$. Let then V_α be a vector exposing face$(W, \mathbf{S}_+^\alpha \cap \mathbf{S}_c^n)$. Letting Θ be the collection of cliques under consideration, we thus obtain faces F_α exposed by $\mathcal{P}_{E[\alpha]}^* V_\alpha$ for $\alpha \in \Theta$. In the noiseless regime, the entire feasible region is contained in the

---

[4]The 2016 tarfile with MATLAB codes is available:

www.math.uwaterloo.ca/~hwolkowi//henry/reports/SNL2010codes.d/SNLSDPclique_ver05.tar

intersection $\bigcap_{\alpha \in \Theta} F_\alpha$. In the noisy regime, this intersection likely consists only of the origin for the simple reason that randomly perturbed faces typically intersect only at the origin. Here is an elementary fix that makes the algorithm robust to noise. Form the sum

$$V := \sum_{\alpha \in \Theta} \mathcal{P}^*_{E[\alpha]} V_\alpha.$$

Again in the noiseless regime, Proposition 2.2.1 implies that V exposes precisely the intersection $\bigcap_{\alpha \in \Theta} F_\alpha$. When noise is present, the matrix V will likely have only one zero eigenvalue corresponding to the vector of all ones e and the rest of the eigenvalues will be strictly positive. Suppose we know that the realization of the graph should lie in r-dimensional space. Then we can find a rank $n - r + 1$ best PSD approximation of V and use it to expose a face of the PSD cone. Under appropriate conditions, this procedure is indeed provably robust to noise and extremely effective in practice. A detailed description of such a scheme is presented in [46].

5.3   Low-rank matrix completions

In this section, we consider another example inspired by facial reduction. We will be considering matrices $Z \in \mathbf{R}^{m \times n}$; for convenience, we will index the rows of Z by $i \in \{1, \ldots, m\}$ and the columns using $j \in \{m + 1, \ldots, m + n\}$. Consider two vertex sets $V_1 = \{1, \ldots, m\}$ and $V_2 := \{m + 1, \ldots, m + n\}$ and a bipartite graph $G = (V_1 \cup V_2, E)$.

Given a partial matrix $z \in \mathbf{R}^E$, the *low-rank matrix completion problem, LRMC*, aims to find a rank r matrix Z from the partially observed elements z. A common approach (with statistical guarantees) is to instead solve the convex problem:

$$\min_{Z \in \mathbf{R}^{m \times n}} \|Z\|_* \quad \text{subject to } P_E(Z) = z, \tag{5.6}$$

where $\|Z\|_*$ is the *nuclear norm* – the sum of the singular values of Z. Throughout the section, we will make the following assumption: the solution of the convex problem (5.6) coincides with the rank r matrix Z that we seek. There are standard statistical assumptions that one makes in order to guarantee this to be the case [27, 50, 117].

It is known that this problem (5.6) can be solved efficiently using SDP. At first glance it appears that this does not fit into our framework for problems where strict feasibility fails; indeed strict feasibility holds under the appropriate reformulation below. We will see, however, that one can exploit the special structure at the *optimum* and discover a face of the PSD cone containing an optimal solution, thereby decreasing the dimension of the problem. Even though, this is not facial reduction exactly, the ideas behind facial reduction play the main role.

Let us first show that the problem (5.6) can be written equivalently as the SDP:

$$
\begin{aligned}
\min \quad & \tfrac{1}{2}\operatorname{tr}(Y) \\
\text{s.t.} \quad & Y = \begin{bmatrix} A & Z \\ Z^T & B \end{bmatrix} \\
& \mathcal{P}_E(Z) = z \\
& Y \in \mathbf{S}_+^{m+n}
\end{aligned}
\tag{5.7}
$$

To see this, we recall a classical fact that the operator norm $\|\cdot\|_2$ of the matrix is dual to the nuclear norm $\|\cdot\|_*$, that is

$$
\|Z\|_* = \sup\{\operatorname{tr}(Z^T X) : \|X\|_2 \le 1\}.
$$

Note the equivalence

$$
\|X\|_2 \le 1 \iff I - XX^T \succeq 0 \iff \begin{bmatrix} I & X \\ X^T & I \end{bmatrix} \succeq 0.
$$

Thus we may represent the nuclear norm through an SDP:

$$
\|Z\|_* = \sup_X \left\{ \operatorname{tr}(Z^T X) : \begin{bmatrix} I & X \\ X^T & I \end{bmatrix} \succeq 0 \right\}.
$$

The dual of this SDP is

$$
\begin{aligned}
\min \quad & \tfrac{1}{2}\operatorname{tr}(Y) \\
\text{s.t.} \quad & Y = \begin{bmatrix} A & Z \\ Z^T & B \end{bmatrix} \\
& Y \in \mathbf{S}_+^{m+n}
\end{aligned}
$$

Thus the problems (5.6) and (5.7) are indeed equivalent. Let us moreover make the following important observation. Suppose that Z is optimal for (5.6). Let $Z = U\Sigma V^T$ be a compact SVD of Z and set

$A := U\Sigma U^T$ and $B := V\Sigma V^T$. Then the triple (Z, A, B) is feasible for (5.7) since

$$Y = \begin{bmatrix} A & Z \\ Z^T & B \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \begin{bmatrix} U \\ V \end{bmatrix}^T \succeq 0.$$

Moreover $\operatorname{rank}(Y) = \operatorname{rank}(Z)$ and $\frac{1}{2}\operatorname{tr}(Y) = \frac{1}{2}(\operatorname{tr}(A) + \operatorname{tr}(B)) = \operatorname{tr}(\Sigma) = \|Z\|_*$. Thus Y is optimal for (5.7).

Let us see now how we can exploit the structure and target rank r of the problem to find an *exposing vector* of a face containing an optimal solution of the SDP. Fix two numbers $p, q > r$ and let α be any $p \times q$ complete bipartite subgraph of G. Let also $z[\alpha]$ be the restriction of z to α. Thus $z[\alpha]$ corresponds to a fully specified submatrix.

For almost any[5] rank r underlying matrix Z, it will be the case that $\operatorname{rank}(z[\alpha]) = r$. Without loss of generality, after row and column permutations if needed, we can assume that α encodes the bottom left corner of Z:

$$Z = \begin{bmatrix} Z_1 & Z_2 \\ z[\alpha] & Z_3 \end{bmatrix},$$

that is $\alpha = \{m - p + 1, \dots, m\} \times \{m + 1, \dots, m + q\}$. Form now the factorization $z[\alpha] = \bar{P}\bar{D}\bar{Q}^T$ obtained using the compact SVD. Both $\bar{P}, \bar{Q}$ have rank r.

Let $Z = U\Sigma V^T$ be a compact SVD of Z and define

$$Y = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \begin{bmatrix} U \\ V \end{bmatrix}^T.$$

As we saw previously, Y is optimal for the SDP (5.7). Subdividing U and V into two blocks each, we deduce

$$0 \preceq Y = \begin{bmatrix} U_1 \\ U_2 \\ V_1 \\ V_2 \end{bmatrix} \Sigma \begin{bmatrix} U_1 \\ U_2 \\ V_1 \\ V_2 \end{bmatrix}^T = \left[\begin{array}{c|c|c|c} U_1\Sigma U_1 & U_1\Sigma U_2^T & U_1\Sigma V_1^T & U_1\Sigma V_2^T \\ \hline U_2\Sigma U_1^T & U_2\Sigma U_2^T & U_2\Sigma V_1^T & U_2\Sigma V_2^T \\ \hline V_1\Sigma U_1^T & V_1\Sigma U_2^T & V_1\Sigma V_1^T & V_1\Sigma V_2^T \\ \hline V_2\Sigma U_1^T & V_2\Sigma U_2^T & V_2\Sigma V_1^T & V_2\Sigma V_2^T \end{array} \right].$$

$$\tag{5.8}$$

---

[5]This is in the sense of Lebesgue measure on the factors $P \in \mathbf{R}^{m \times r}$, $Q \in \mathbf{R}^{n \times r}$ satisfying $Z = PQ^T$.

Therefore, we conclude that $z[\alpha] = U_2 \Sigma V_1^T = \bar{P}\bar{D}\bar{Q}^T$. Taking into account that $z[\alpha]$ has rank r and the matrices $U_2, V_1, \bar{P}, \bar{Q}$ have exactly r columns we deduce

$$
\begin{aligned}
\text{range}(z[\alpha]) &= \text{range}(U_2) = \text{range}(\bar{P}), \\
\text{range}(z[\alpha]^T) &= \text{range}(V_1) = \text{range}(\bar{Q}).
\end{aligned}
\tag{5.9}
$$

We can now use the *exposing vector* form of FR formed from $\bar{P}$ and/or $\bar{Q}$. Using the calculated $\bar{P}, \bar{Q}$, let $\bar{E} \in \mathbf{R}^{p \times (p-r)}$ and $\bar{F} \in \mathbf{R}^{q \times (q-r)}$ satisfy range $\bar{E} = (\text{range } \bar{P})^\perp$ and range $\bar{F} = (\text{range } \bar{Q})^\perp$. Define then the PSD matrix

$$
\overline{W} = \left[\begin{array}{c|ccc|c} 0 & 0 & 0 & 0 \\ \hline 0 & \bar{E}\bar{E}^T & 0 & 0 \\ 0 & 0 & \bar{F}\bar{F}^T & 0 \\ \hline 0 & 0 & 0 & 0 \end{array} \right] \succeq 0.
$$

By construction $\overline{W}Y = 0$. Hence $\overline{W}$ exposes a face of the PSD cone containing the optimal Y.

Performing this procedure for many specified submatrices $z \in \mathbf{R}^E$, can yield a dramatic decrease in the dimension of the final SDP that needs to be solved. When noise is present, the strategy can be made robust in exactly the same way as for the EDM problem in Section 5.2.2.

We include one of the tables of numerics from [66] in Table 5.2, page 59. Results are for the average of five instances. We have recovered the correct rank 4 each time without calling an SDP solver at all. Note that the largest matrices recovered have $(2,500) \times (20,000) = 50,000,000$ elements.

Table 5.2: noiseless: $r = 4$; $m \times n$ size; density p.

Specifications			Time (s)	Rank	Residual ($\%Z$)
m	n	mean(p)			
700	2000	0.36	12.80	4.0	1.5217e-12
1000	5000	0.36	49.66	4.0	1.0910e-12
1400	9000	0.36	131.53	4.0	6.0304e-13
1900	14000	0.36	291.22	4.0	3.4847e-11
2500	20000	0.36	798.70	4.0	7.2256e-08

5.4   Commentary

The work using chordal graphs for PSD completions was done in [61] and extended the special case of banded structure in [48]. Surveys for matrix completion are given in e.g., [4, 32, 33, 63, 68, 69, 70, 71]. A survey specifically related to chordality is given in [139]. More details on early algorithms for PSD completion are in e.g., [4, 72].

The origin of distance geometry problems can be traced back to the work of Grassmann in 1896 [60]. More recent work appeared in e.g., [34, 45, 58, 59, 64]. Many of these papers emphasized the relationships with molecular conformation. Chordality and relations with positive definiteness are studied in [73] and more recently in [80, 82]. The book [16] has a chapter on matrix completions with the connections to EDM completions, see also [134] for the relations with faces. An excellent online reference for EDM is the book by Dattorro [37]. In addition, there are many excellent survey articles, e.g., [2, 43, 77, 87, 97]. The survey [86] contains many open problems in EDMC and references for application areas.

Early work using SDP interior point algorithms for EDM completion problems is given in [3]. Exploiting the clique structure for SNL type problems is done in e.g., [42, 75, 76]. The improved robust algorithm based on averaging approximate exposing vectors was developed in [46], while a parallel viewpoint based on rigidity theory was developed in [126]. In fact, a parallel view on facial reduction is based on rigidity theory, e.g., [1, 31, 55, 56]. The facial structure for the EDM cone is studied in e.g., [5, 133]. Applications of the technique to molecular conformation are in [7].

The LRMC problem has parallels in the compressed sensing framework that is currently of great interest. The renewed interest followed the work in [27, 50, 51, 117] that used the nuclear norm as a convex relaxation of the rank function. Exploiting the structure of the optimal face using FR is introduced recently in [66]. An alternative approach, which applies much more broadly, is described in [112].

6

Hard combinatorial problems

6.1 Quadratic assignment problem, QAP

The *quadratic assignment problem, QAP*, is arguably the hardest of the
so-called NP-hard combinatorial optimization problems. The problem
can best be described in terms of facility location. We have n given
facilities that need to be located among n specified locations. As input
data, we have information on the distances D_{ij} between pairs of loca-
tions i, j and the flow values (weights) F_{st} between pairs of facilities
s, t. The (quadratic) cost of a possible location is the flow between each
pair of facilities multiplied by the distance between their assigned loca-
tions. Surprisingly, problems of size $n \geq 30$ are still considered hard to
solve. As well, we can have a (linear) cost C_{kl} of locating facility k in
location l. The unknown variable that decides which facility goes into
which location is an $n \times n$ *permutation matrix* $X = (X_{kl}) \in \Pi$ with

$$X_{kl} = \begin{cases} 1 & \text{if facility } k \text{ is assigned to location } l \\ 0 & \text{otherwise} \end{cases}.$$

This problem has the elegant trace formulation

$$\min_{X \in \Pi} \text{tr}(FXDX^T) + \text{tr}(CX^T).$$

Notice that the objective is a quadratic function, and typically the quadratic form, $\mathrm{tr}(FXDX^T)$, is indefinite.[1]

Notice also that the feasible region consists of permutation matrices, a discrete set. There is a standard strategy for forming a semi-definite programming relaxation for such a problem. Consider the *vectorization* $x := \mathrm{vec}(X) \in \mathbf{R}^{n^2}$ and define the *lifting* to the rank one *block matrix*

$$Y = \begin{pmatrix} 1 \\ x \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}^T = \begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix} = \begin{bmatrix} 1 & x^T \\ x & \left[X_{:,i}X_{:,j}^T\right] \end{bmatrix} \in \mathbf{S}_+^{n^2+1},$$

where the matrix consists of $n^2, n \times n$ blocks $X_{:,i}X_{:,j}^T$ beginning in row and column 2. The idea is then to reformulate the objective and a relaxation of the feasible region linearly in terms of Y, and then simply insist that Y is PSD, though not necessarily rank one. In particular, the objective function can easily be rewritten as a *linear function* of Y, namely $\mathrm{tr}(LY)$, where

$$L := \begin{bmatrix} 0 & \frac{1}{2}\,\mathrm{vec}(C)^T \\ \frac{1}{2}\,\mathrm{vec}(C) & D \otimes F \end{bmatrix},$$

and we denote the *Kronecker product, $D \otimes F$*.

Next we turn to the constraints. We seek to replace the set of permutation matrices by more favorable constraints that permutation matrices satisfy. For example, observe that the permutation matrices are doubly stochastic and hence the row sums and column sums are one, yielding the following *linear assignment constraints*

$$Xe = X^T e = e. \tag{6.1}$$

There are of course many more possible constraints one can utilize; the greater their number, even if redundant, the tighter the SDP relaxation

---

[1]One can perturb the objective function by exploiting the structure of the permutation matrices and obtain positive definiteness of the quadratic form. However, this can result in deterioration of the bounds from any relaxations.

in general. Some prominent ones, including the ones above, are

$$Xe = X^T e = e \tag{6.2a}$$

$$X_{ij}^2 - X_{ij} = 0, \tag{6.2b}$$

$$XX^T = X^T X = I, \tag{6.2c}$$

$$X(:,i) \circ X(:,j) = 0, X(k,:) \circ X(l,:) = 0, \quad \text{for all } i \neq j, k \neq l, \tag{6.2d}$$

where $\circ$ denotes the *Hadamard (elementwise) product*. Note that including both equivalent orthogonality constraints $XX^T = X^T X = I$ is not redundant in the relaxations.

Let us see how to reformulate the constraints linearly in Y. We first consider the n linear row sum constraints $(Xe - e)_i = 0$ in (6.2a). To this end, observe

$$
\begin{aligned}
0 &= (Xe - e)_i \\
&= e_i^T (Xe - e) \\
&= \operatorname{tr}(X^T e_i e^T) - 1 \\
&= \begin{pmatrix} 1 \\ x \end{pmatrix}^T \begin{pmatrix} -1 \\ \operatorname{vec}(e_i e^T) \end{pmatrix}.
\end{aligned}
$$

We obtain

$$
0 = \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T \begin{pmatrix} -1 \\ \operatorname{vec}(e_i e^T) \end{pmatrix} \begin{pmatrix} -1 \\ \operatorname{vec}(e_i e^T) \end{pmatrix}.
$$

Defining now the matrix

$$
E_r = \begin{bmatrix} -1 & -1 & \ldots & -1 \\ \operatorname{vec}(e_1 e^T) & \operatorname{vec}(e_2 e^T) & \ldots & \operatorname{vec}(e_n e^T) \end{bmatrix},
$$

we obtain the equivalent linear homogeneous constraint $\operatorname{tr}(Y E_r E_r^T) = 0$. Similarly the n linear column sum constraints $(X^T e - e)_i = 0$ amount to the equality $\operatorname{tr}(Y E_c E_c^T) = 0$, where

$$
E_c = \begin{bmatrix} -1 & -1 & \ldots & -1 \\ \operatorname{vec}(e e_1^T) & \operatorname{vec}(e e_2^T) & \ldots & \operatorname{vec}(e e_n^T) \end{bmatrix}.
$$

Thus the feasible region of the SDP relaxation lies in the face of $\mathbf{S}_+^{n^2+1}$ exposed by $D_0 := E_r E_r^T + E_c E_c^T$. Henceforth, let $\widehat{V}$ be full column rank and satisfying $\operatorname{range}(\widehat{V}) = \operatorname{null}(D_0)$.

The other three constraints in (6.2) can be rephrased linearly in Y as well: (6.2b) results in the so-called *arrow constraint* (the first row (column) and the diagonal of Y are equal); the constraints (6.2c) yield the *block-diagonal* constraint (diagonal blocks sum to the identity matrix) and the *off-diagonal* contraint (the traces of the off-diagonal blocks are zero); and the Hadamard orthogonality constraints (6.2d) are called the *gangster constraints* and guarantee that the diagonal blocks are diagonal matrices and the diagonals of the off-diagonal blocks are zero. We omit further details[2] but denote the resulting constraints with the additional $Y_{00} = 1$ in the form $\mathcal{A}(Y) = b$. We note that the transformation $\mathcal{A}$ without the Hadamard orthogonality constraints is onto while $\mathcal{A}$ is not. We numerically test both settings with and without the gangster constraints and together with and without facial reduction below in this section.

Now the standard relaxation of the problem is obtained by letting Y be a positive semi-definite matrix with no constraint on its rank:

$$
\begin{aligned}
\min \quad & \operatorname{tr}(LY) \\
\text{s.t.} \quad & \mathcal{A}(Y) = b \\
& Y \in \mathbf{S}_+^{n^2+1}.
\end{aligned} \tag{6.3}
$$

All in all, the number of linear constraints is

$$
m_A = 1 + n + (n(n-1)/2) + n^2 + 2(n(n(n-1)/2)) + 1,
$$

i.e.,

$$
\mathcal{A} : \mathbf{S}_+^{n^2+1} \to \mathbf{R}^{(n^3 + \frac{n^2}{2} + \frac{n}{2} + 2)}.
$$

As discussed above, the matrix D_0 certifies that this relaxation fails strict feasibility. Indeed the entire feasible region lies in the face of $\mathbf{S}_+^{n^2+1}$ exposed by D_0. Surprisingly, after restricting to the face exposed by D_0, the constraints $\mathcal{A}(Y) = b$ simplify dramatically. The resulting equivalent formulation becomes

$$
\begin{aligned}
\min \quad & \operatorname{tr}(\widehat{V}^T L \widehat{V} R) \\
\text{s.t.} \quad & \mathcal{G}(\widehat{V} R \widehat{V}^T) = e_0 e_0^T \\
& R \in \mathbf{S}_+^{(n-1)^2+1},
\end{aligned} \tag{6.4}
$$

---

[2]See more details in [149].

where e_0 is the first unit vector, as we start indexing at 0, and

$$\mathcal{G}(Y)_{ij} = \begin{cases} 1 & \text{if } ij \in \bar{\mathcal{J}} \\ 0 & \text{otherwise} \end{cases}$$

and $\bar{\mathcal{J}}$ is an appropriately defined index set; see [149]. Roughly speaking, this index set guarantees that the diagonal blocks of $Y = \widehat{V} R \widehat{V}^T$ are diagonal matrices and the diagonal elements of the off-diagonal blocks of Y are all zero. In particular, one can show that the resulting linear constraint is surjective.

In fact, the *gangster operator* and gangster constraints guarantee that most of the Hadamard product constraints in (6.2d) hold. And the constraints corresponding to the linear constraints in (6.2a), the arrow constraint in (6.2b), the block-diagonal and off-diagonal constraints in (6.2c) and some of the gangster constraints in (6.2d) have all become redundant, thereby illuminating the strength of the facial reduction together with the gangster constraints (6.2d).

Moreover, we can rewrite the linear constraints in (6.4) as

$$\text{tr}\left(Y\left(\widehat{V}_{j:}^T \widehat{V}_{i:} + \widehat{V}_{i:}^T \widehat{V}_{j:}\right)\right) = 0, \quad \forall ij \in \bar{\mathcal{J}}.$$

We see that these *low rank*[3] constraints are linearly independent and the number of constraints has been reduced from $m_A = n^3 + \frac{n^2}{2} + \frac{n}{2} + 2$ to

$$|\bar{\mathcal{J}}| = 1 + n(n(n-1)/2) + n(n(n-1)/2 - (n-1) - 1) = n^3 - 2n^2 + 1,$$

i.e., the number of constraints is still $O(n^3)$ but has decreased by $1 + \frac{5n^2 + n}{2}$.

Finally, we should mention that the singularity degree of the SDP relaxation of the QAP is $d = 1$. The problem (6.4) has a strictly feasible point $\widehat{R}$. Moreover, one can show that the dual of (6.4) also has a strictly feasible point, see [100, 149].

Let us illustrate empirically the improvement in accuracy and cputime for the facially reduced SDP relaxation of the QAP. We use the model in (6.3) and compare it to the simplified facially reduced

---

[3]These constraints are rank two. Low rank constraints can be exploited in several of the current software packages for SDP.

model in (6.4). See Figure 6.1, page 66, and Figure 6.2, page 67. The improvement in accuracy and cputime is evident.
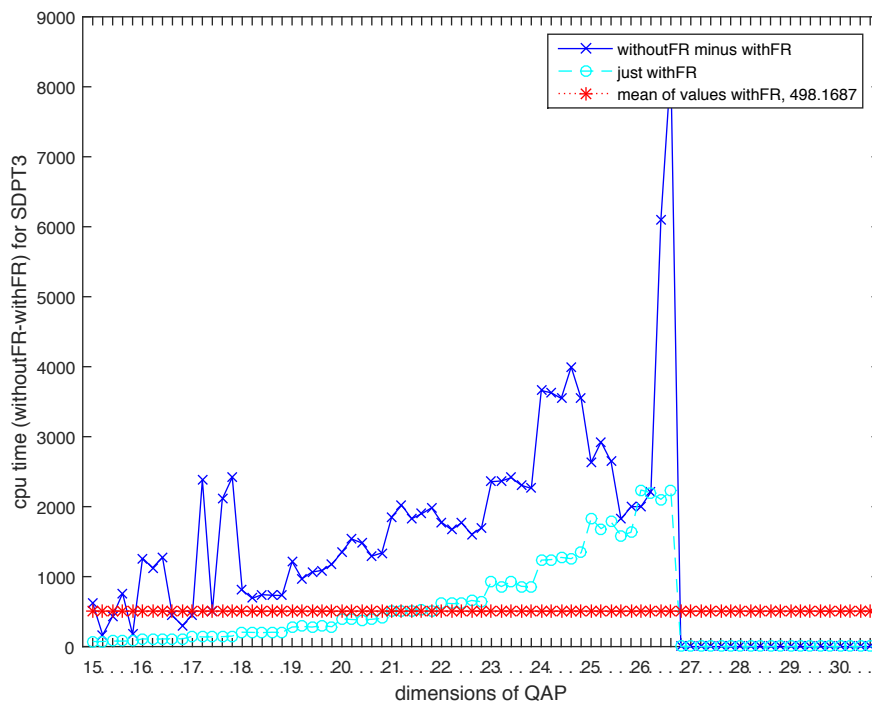

Figure 6.1: Difference in cpu seconds (without FR− with FR)

6.2   Second lift of Max-Cut

Recall that for a given weighted undirected graph $G = (V, E, W)$, the maximum cut problem is to determine a vertex set S such that the total weight of the edges between S and its complement S^c is as large as possible. Thus enumerating the vertices $V = \{1, \ldots, n\}$, we are interested in the problem

$$
(\mathbf{MC}) \qquad \begin{array}{ll} \max & \frac{1}{2} \sum_{i<j} w_{ij}(1 - x_i x_j) \\ \text{s.t.} & x_i \in \{\pm 1\}, \qquad \text{for } i = 1, \ldots, n. \end{array}
$$

Here, we have $x_i = 1$ for $i \in S$ and $x_i = -1$ for $i \notin S$. Notice the constraints $x_i \in \{\pm 1\}$ can equivalently be written with the quadratic
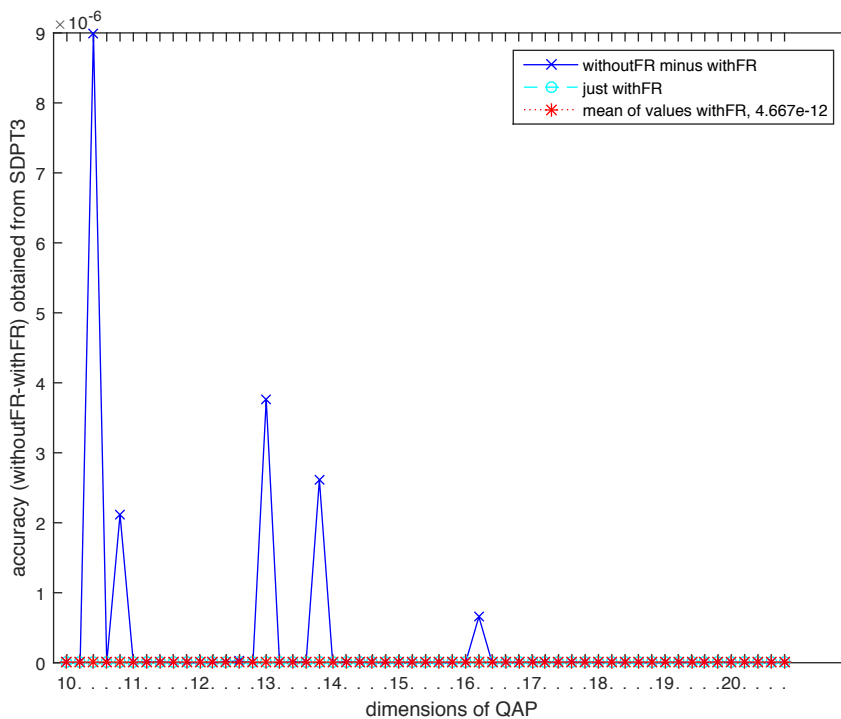
Figure 6.2: Difference in accuracy values (without FR− with FR)

constraint

$$\operatorname{diag}(xx^T) = e. \tag{6.5}$$

Relaxing xx^T to a positive semi-definite matrix X, we arrive at the celebrated SDP relaxation of Max-Cut:

$$\begin{array}{ll} \max & \frac{1}{4}\operatorname{tr}(LX) \\ \text{s.t.} & \operatorname{diag}(X) = e \\ & X \succeq 0 \end{array} \tag{6.6}$$

Here L denotes the weighted Laplacian matrix of the graph, which will not play a role in our discussion. This SDP is clearly strictly feasible.

Another idea now to improve the accuracy of the relaxation is to "extend the lifting". Namely, with the goal of tightening the approximation to the original Max-Cut problem, we can certainly *add* the

following quadratic constraints to the SDP relaxation:

$$X_{ij}X_{jk} = X_{ik}, \qquad \text{for all } i, j, k. \tag{6.7}$$

Let us see how to form a relaxation with these nonlinear constraints.

For $X \in \mathbf{S}^n$, let s2vec(X) denote the vector formed from the upper triangular part of X taken columnwise with the strict upper triangular part multiplied by $\sqrt{2}$. By abuse of notation, we let $x = $ s2vec(X) and define the matrix $Y = \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T$. We can now form a new SDP relaxation by insisting that Y is PSD (though not rank one) and rewriting the constraints linearly in Y. The nonlinear constraints (6.7) can indeed be written linearly in Y; we omit the details. On the other hand, note that the i-th constraint in the original SDP relaxation (6.6) is equivalent to

$$\begin{aligned} 0 &= \langle e_i, \text{diag}(X) - e \rangle \\ &= \langle \text{Diag}(e_i), X \rangle - 1 \\ &= x^T \, \text{s2vec}(\text{Diag}(e_i)) - 1. \end{aligned}$$

Exactly, the same way as in Section 6.1, we can define the matrix

$$E := \begin{bmatrix} -1 & -1 & \ldots & -1 \\ \text{s2vec}(\text{Diag}(e_1)) & \text{s2vec}(\text{Diag}(e_2)) & \ldots & \text{s2vec}(\text{Diag}(e_n)) \end{bmatrix},$$

which certifies that strict feasibility fails and that the entire feasible region lies in the face of $\mathbf{S}_+^{\frac{n(n+1)}{2}+1}$ exposed by EE^T. It turns out that this second lift of max-cut, in practice, provides much tighter bounds than the original SDP relaxation (6.6), and the elementary facial reduction step using EE^T serves to stabilize the problem.

6.3 General semi-definite lifts of combinatorial problems

Let us next look at a general recipe often used for obtaining SDP relaxations of NP-hard problems; elements of this technique were already used in the previous sections. Consider a nonconvex feasible region of the form

$$\mathcal{F} := \left\{ x \in \mathbf{R}^n : \mathcal{A} \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} = 0 \right\}$$

where $\mathcal{A}\colon \mathbf{S}^n \to \mathbf{R}^m$ is a linear transformation. An SDP relaxation of this region is the set

$$\widehat{\mathcal{F}} = \left\{ X \in \mathbf{S}^n_+ : \mathcal{A}(X) = 0,\ X_{11} = 1 \right\}.$$

Indeed, $\mathcal{F}$ is the image of a linear projection of the intersection of $\widehat{\mathcal{F}}$ with rank one matrices. For this reason $\widehat{\mathcal{F}}$ is often called an *SDP lift* of $\mathcal{F}$.

In applications, such as the ones in the previous sections, the affine hull of $\mathcal{F}$ may not be full dimensional. For example, the affine hull of the set of permutation matrices (used for QAP) has empty interior. To this end, suppose that the affine hull of $\mathcal{F}$ is given by $\{x : Lx = l\}$, where L is a linear transformation and l a vector. Define the matrix $\widehat{L} = \begin{bmatrix} -l & L \end{bmatrix}$. Then clearly there is no harm in including the redundant constraint

$$\left\langle \widehat{L}^T \widehat{L}, \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \right\rangle = 0 \tag{6.8}$$

in the very definition of $\mathcal{F}$. Notice then $\widehat{\mathcal{F}}$ is clearly contained in the face of $\mathbf{S}^n_+$ exposed by $\widehat{L}^T \widehat{L}$. Indeed, this is the minimal face of $\mathbf{S}^n_+$ containing $\widehat{\mathcal{F}}$. To see this, suppose that the affine span of $\mathcal{F}$ has dimension d, and consider any affinely independent vectors $x_1, \ldots, x_{d+1} \in \mathcal{F}$. Then the vectors $\begin{pmatrix} 1 \\ x_1 \end{pmatrix}, \begin{pmatrix} 1 \\ x_2 \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ x_{d+1} \end{pmatrix}$ are linearly independent, and therefore the barycenter

$$\frac{1}{d+1} \sum_{i=1}^{d+1} \begin{pmatrix} 1 & x_i^T \\ x_i & x_i x_i^T \end{pmatrix}$$

is a rank $d+1$ matrix lying in $\widehat{\mathcal{F}}$. On the other hand, it is immediate that the face of $\mathbf{S}^n_+$ exposed by $\widehat{L}^T \widehat{L}$ also has dimension $d+1$. The claimed minimality follows. It is curious to note that if the constraint (6.8) were not explicitly included in the definition $\mathcal{F}$, then the SDP lift $\widehat{\mathcal{F}}$ could nevertheless be strictly feasible, and hence unnecessarily large.

Example 6.3.1 (Strictly feasible SDP lifts). Consider the region:

$$\left\{ (x, y, z) \in \mathbf{R}^3 : x^2 = y^2 = z^2 = 1, \quad xy + yz = 0 \right\}.$$

There are only four feasible points, namely $\left\{ \pm \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \pm \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right\}$, and
they affinely span the two dimensional subspace perpendicular to the
vector $\begin{pmatrix} 1 & 0 & 1 \end{pmatrix}^T$. If this constraint is not included explicitly, then the
SDP lift is given by

$$\{X \in \mathbf{S}_+^4 : X_{23} = -X_{34}, \ X_{ii} = 1 \text{ for } i = 1, \dots, 4\}.$$

In particular, the identity matrix is feasible.

6.4   Elimination method for sparse SOS polynomials

Checking whether a polynomial is always nonnegative is a ubiquitous
task in computational mathematics. This problem is NP-hard, as it
for example encompasses a great variety of hard combinatorial prob-
lems. Instead a common approach utilizes sum of squares formulations.
Indeed, checking whether a polynomial is a sum of squares of poly-
nomials can be modeled as an SDP . A certain hierarchy of sum of
squares problems [78, 103] can then be used to determine the nonneg-
ativity of the original polynomial. The size of the SDP arising from
a sum of squares problem depends on the number of monomials that
must be used in the formulation. In this section, we show how facial
reduction iterations on the cone of sums of squares polynomials can be
used to eliminate monomials yielding a smaller and better conditioned
equivalent SDP formulation. A rigorous explanation of the material in
this section requires some heavier notation; therefore we only outline
the techniques.

Let $\mathbf{R}[x]_{n,2d}$ denote the vector space of polynomials in n variable
with real coefficients of degree at most $2d$. We will write a polynomial
$f \in \mathbf{R}[x]_{n,2d}$ using multi-index notation

$$f(x) = \sum_{\alpha \in N} c_\alpha x^\alpha,$$

where N is some subset of $\mathbb{N}^n$, we set $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, and c_α are
some real coefficients. We will think of $\mathbf{R}[x]_{n,2d}$ as a Euclidean space
with the inner product being the usual dot product between coefficient

vectors. Let $\Sigma_{n,2d} \subseteq \mathbf{R}[x]_{n,2d}$ be the set of polynomials $f \in \mathbf{R}[x]_{n,2d}$ that are sums of squares, meaning that f can be written as $\sum_i f_i^2$ for some polynomials f_i. Clearly $\Sigma_{n,2d} \subseteq \mathbf{R}[x]_{n,2d}$ is a closed convex cone, often called the SOS cone.

A fundamental fact is that membership in the SOS cone $\Sigma_{n,2d}$ can be checked by solving an SDP .

Theorem 6.4.1. Fix a set of monomials $M \subset \mathbb{N}^n$. Then a polynomial $f \in \mathbf{R}[x]_{n,2d}$ is a sum of squares of polynomials over the monomial set M if and only if there exists a matrix $Q \succeq 0$ so that $f(x) = [x]_M^T Q[x]_M$, where $[x]_M$ is a vector of monomials in M.

Proof. If f is a sum of squares $f = \sum_i f_i^2$, then we can form a matrix P whose rows are the coefficient vectors of $\{f_i\}_i$. Then $Q = P^T P$ is the PSD matrix we seek. Conversely, given a PSD matrix Q satisfying $f(x) = [x]_M^T Q[x]_M$, we can form a factorization $Q = P^T P$, and read off the coefficients of each polynomial f_i from the rows of P. $\square$

Notice that the relation $f = [x]_M^T Q[x]_M$ can be easily rewritten as a linear relation on Q by matching coefficient of the left and right-hand-sides. The size of Q is completely dictated by the number of monomials.

More generally, instead of certifying whether a polynomial is SOS, we might be interested in minimizing a linear functional over an affine slice of the SOS cone. More precisely, consider a problem of the form:

$$\min_{u \in \mathbf{R}^m} \sum_{i=1}^m w_i u_i$$
$$\text{s.t. } f = g_0 + \sum_{i=1}^m u_i g_i \tag{6.9}$$
$$f \in \Sigma_{n,2d}$$

where $u \in \mathbf{R}^m$ is the decision variable, $g_i \in \mathbf{R}[x]_{n,2d}$ are specified polynomials and $w \in \mathbf{R}^m$ is a fixed vector. Clearly this problem can be converted to an SDP. The size of the decision matrix X is determined by the number of monomials. Parsing algorithms attempt to choose a (small) set of monomials M so that *every* feasible f for (6.9) can be written as a sum of squares over the monomial set M, thereby

decreasing the size of the SDP. Not surprisingly, some parsing strategies can be interpreted as facial reduction iterations on (6.9).

We next outline such a strategy closely following [111]. To this end, we must first explain which faces of the SOS cone $\Sigma_{n,2d}$ correspond to eliminating monomials. Indeed, there are faces of $\Sigma_{n,2d}$ that do not have such a description.

To answer this question, we will need extra notation. Henceforth, fix a set of monomials $M \subseteq \mathbb{N}^n$ and set $d := \max\{\sum_{i=1}^n z_i : z \in M\}$ to be the maximal degree of monomials in M. Let $\Sigma(M)$ be the set of all polynomials that can be written as sums of squares over the monomial set M. Finally, set M^+ to be the set of points in M that are not midpoints of any points in M, namely

$$M^+ := M \setminus \left\{ \frac{\alpha + \beta}{2} : \alpha, \beta \in M \text{ and } \alpha \neq \beta \right\}.$$

Let us now look two types of faces that arise from elimination of monomials.

Theorem 6.4.2 (Type I face)**.** If equality,

$$\mathrm{conv}(M) \cap \mathbb{N}^n = M$$

holds, then $\Sigma(M)$ is a face of $\Sigma_{n,2d}$.

In other words, if the convex hull $\mathrm{conv}(M)$ contains no grid points other than those already in M, then $\Sigma(M)$ is a face of $\Sigma_{n,2d}$.

Theorem 6.4.3 (Type II face)**.** If $\Sigma(M)$ is a face of $\Sigma_{n,2d}$, then $\Sigma(M \setminus \beta)$ is a face of $\Sigma_{n,2d}$ for any $\beta \in M^+$.

Thus given a face $\Sigma(M)$, we can recursively make the face smaller by deleting any $\beta \in M^+$.

Let us now turn to facial reduction. In the first step of facial reduction for (6.9), we must find an exposing vector $v \in \Sigma_{n,2d}^*$ that is orthogonal to all the affine constraints. Doing so in full generality is a difficult proposition. Instead, let us try to replace $\Sigma_{n,2d}^*$ by a polyhedral inner approximation. Then the search for v is a linear program.

Theorem 6.4.4. The polyhedral set

$$\Lambda = \Sigma(M)^{\perp} + \left\{ \sum_{\alpha \in M^+} \lambda_{2\alpha} e_{2\alpha} : \lambda_{2\alpha} \geq 0 \right\}$$

satisfies $\Lambda \subseteq \Sigma(M)^*$.

Thus if we can find $v \in \Lambda$ that is orthogonal to the affine constraints (6.9), then we can use v to expose a face of $\Sigma_{n,2d}$ containing the feasible region. Remarkably, this face can indeed be represented by eliminating monomials from M.

Theorem 6.4.5. Consider a vector

$$v = p + \sum_{\alpha \in M^+} \lambda_{2\alpha} e_{2\alpha},$$

for some $p \in \Sigma(M)^{\perp}$ and nonnegative numbers $\lambda_{2\alpha} \geq 0$. Define the monomial set $\mathcal{I} := \{\alpha \in M^+ : \lambda_{2\alpha} > 0\}$. Then the face $\Sigma(M) \cap v^{\perp}$ coincides with $\Sigma(M \setminus \mathcal{I})$.

Thus we can inductively use this procedure to eliminate monomials. At the end, one would hope that we would be left with a small dimensional SDP to solve. Promising numerical results and further explanations of methods of this type can be found in [74, 110, 111, 140, 141].

6.5 Commentary

Quadratic assignment problem, QAP

Many survey articles and books have appeared on the QAP, e.g., [28, 89, 101, 102]. More recent work on implementation of SDP relaxations include [41, 100, 150]. That the quadratic assignment problem is NP-hard is shown in [125]. The elegant trace formulation we used was introduced in [49].

The classic *Nugent test set* for QAP is given in [99].[4] These problems have proven to be extremely hard to solve to optimality, see e.g., [28].

---

[4]It is maintained within QAPLIB [26] currently online.

The difficulty of these problems is illustrated in the fact that many of them were not solved for 30 odd years, see e.g., [13].

The semi-definite relaxation described here was introduced in [149]. It was derived by using the Lagrangian relaxation after modelling the permutation matrix constraint by various quadratic constraints. The semi-definite relaxation is then the dual of the Lagrangian relaxation, i.e., *the dual of the dual*. Application of FR then results in the surprisingly simplified *gangster operator* formulation.

This gangster formulation along with symmetry for certain QAP models is exploited in [40, 41] to significantly increase the size of QAP problems that can be solved. Other relaxations of QAP based on e.g., eigenvalue bounds are studied in e.g., [14, 52, 53].

Graph partitioning, GP

The *graph partitioning, GP*, problem is very similar to the QAP in that it involves a trace quadratic objective with a $0, 1$ matrix variable X, i.e. the matrix whose columns are the incidence vectors of the sets for the partition. A similar successful SDP relaxation can be found [147]. More recently, successful bounding results have been found in [38, 113, 130].

Second lift of Max-Cut

The second lifting from Section 6.2 is derived in [10, 11, 12] but in a different way, i.e. using the nullspace of the barycenter approach. The bounds found were extremely successful and, in fact, found the optimal solution of the **MC** in almost all but very special cases. The algorithm used for the SDP relaxation was the spectral bundle approach [65] and only problems of limited size could be solved. More recently an ADMM approach was much more successful in solving larger problems in [131].

Lifts of combinatorial problems

The SDP lifting of combinatorial regions described in Section 6.3 is standard; see for [137] for many examples and references. The material on the minimal face of the SDP lift follows [135], though our explanation here is stated in dual terms, i.e. using exposing vectors.

Monomial elimination from SOS problems

The topic of eliminating monomials from sum of squares problems has a rich history. The section in the current text follows entirely the exposition in [111]. The technique of solving linear programs in order to approximate exposing vectors was extensively studied in [110]. Important earlier references on monomial elimination include [74, 140, 141]. For an exposition of how to use SOS hierarchies to solve polynomial optimization problems see the monograph [79].

Acknowledgements

Index

References

[1] A. Alfakih. Graph rigidity via Euclidean distance matrices. *Linear Algebra Appl.*, 310(1-3):49–165, 2000.

[2] A. Alfakih, M.F. Anjos, V. Piccialli, and H. Wolkowicz. Euclidean distance matrices, semidefinite programming, and sensor network localization. *Portug. Math.*, 68(1):53–102, 2011.

[3] A. Alfakih, A. Khandani, and H. Wolkowicz. Solving Euclidean distance matrix completion problems via semidefinite programming. *Comput. Optim. Appl.*, 12(1-3):13–30, 1999. A tribute to Olvi Mangasarian.

[4] A. Alfakih and H. Wolkowicz. Matrix completion problems. In *Handbook of semidefinite programming*, volume 27 of *Internat. Ser. Oper. Res. Management Sci.*, pages 533–545. Kluwer Acad. Publ., Boston, MA, 2000.

[5] A.Y. Alfakih. A remark on the faces of the cone of Euclidean distance matrices. *Linear Algebra Appl.*, 414(1):266–270, 2006.

[6] B. Alipanahi, N. Krislock, A. Ghodsi, H. Wolkowicz, L. Donaldson, and M. Li. Determining protein structures from NOESY distance constraints by semidefinite programming. *J. Comput. Biol.*, 20(4):296–310, 2013.

[7] B. Alipanahi, N. Krislock, A. Ghodsi, H. Wolkowicz, L. Donaldson, and M. Li. Determining protein structures from NOESY distance constraints by semidefinite programming. *J. Comput. Biol.*, 20(4):296–310, 2013.

[8] E.D. Andersen and K.D. Andersen. The Mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In *High performance optimization*, volume 33 of *Appl. Optim.*, pages 197–232. Kluwer Acad. Publ., Dordrecht, 2000.

[9] A.F. Anjos and J.B. Lasserre, editors. *Handbook on Semidefinite, Conic and Polynomial Optimization*. International Series in Operations Research & Management Science. Springer-Verlag, 2011.

[10] M. F. Anjos and H. Wolkowicz. Strengthened semidefinite programming relaxations for the max-cut problem. In *Advances in convex analysis and global optimization (Pythagorion, 2000)*, volume 54 of *Nonconvex Optim. Appl.*, pages 409–420. Kluwer Acad. Publ., Dordrecht, 2001.

[11] M.F. Anjos. *New Convex Relaxations for the Maximum Cut and VLSI Layout Problems*. PhD thesis, University of Waterloo, 2001.

[12] M.F. Anjos and H. Wolkowicz. Strengthened semidefinite relaxations via a second lifting for the Max-Cut problem. *Discrete Appl. Math.*, 119(1-2):79–106, 2002. Foundations of heuristics in combinatorial optimization.

[13] K.M. Anstreicher. Recent advances in the solution of quadratic assignment problems. *Math. Program.*, 97(1-2, Ser. B):27–42, 2003. ISMP, 2003 (Copenhagen).

[14] K.M. Anstreicher and N.W. Brixius. A new bound for the quadratic assignment problem based on convex quadratic programming. *Math. Program.*, 89(3, Ser. A):341–357, 2001.

[15] M. Bakonyi and C.R. Johnson. The Euclidean distance matrix completion problem. *SIAM J. Matrix Anal. Appl.*, 16(2):646–654, 1995.

[16] M. Bakonyi and H.J. Woerdeman. *Matrix completions, moments, and sums of Hermitian squares*. Princeton University Press, Princeton, NJ, 2011.

[17] A. Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[18] A. Ben-Israel. Theorems of the alternative for complex linear inequalities. *Israel J. Math.*, 7:129–136, 1969.

[19] A. Ben-Israel, A. Ben-Tal, and S. Zlobec. *Optimality in nonlinear programming: a feasible directions approach*. John Wiley & Sons, Inc., New York, 1981. A Wiley-Interscience Publication.

[20] A. Ben-Tal and A.S. Nemirovski. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, algorithms, and engineering applications.

[21] J.M. Borwein and A.S. Lewis. *Convex analysis and nonlinear optimization.* Springer-Verlag, New York, 2000. Theory and examples.

[22] J.M. Borwein and H. Wolkowicz. Characterization of optimality for the abstract convex program with finite-dimensional range. *J. Austral. Math. Soc. Ser. A*, 30(4):390–411, 1980/81.

[23] J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81.

[24] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. Appl.*, 83(2):495–530, 1981.

[25] J.M. Borwein and H. Wolkowicz. Characterizations of optimality without constraint qualification for the abstract convex program. *Math. Programming Stud.*, 19:77–100, 1982. Optimality and stability in mathematical programming.

[26] R.E. Burkard, S. Karisch, and F. Rendl. QAPLIB – a quadratic assignment problem library. *European J. Oper. Res.*, 55:115–119, 1991. anjos.mgi.polymtl.ca/qaplib/.

[27] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[28] E. Çela. *The quadratic assignment problem*, volume 1 of *Combinatorial Optimization*. Kluwer Academic Publishers, Dordrecht, 1998. Theory and algorithms.

[29] Y.-L. Cheung. *Preprocessing and Reduction for Semidefinite Programming via Facial Reduction: Theory and Practice*. PhD thesis, University of Waterloo, 2013.

[30] Y-L. Cheung, S. Schurr, and H. Wolkowicz. Preprocessing and regularization for degenerate semidefinite programs. In D.H. Bailey, H.H. Bauschke, P. Borwein, F. Garvan, M. Thera, J. Vanderwerff, and H. Wolkowicz, editors, *Computational and Analytical Mathematics, In Honor of Jonathan Borwein's 60th Birthday*, volume 50 of *Springer Proceedings in Mathematics & Statistics*, pages 225–276. Springer, 2013.

[31] R. Connelly and S.J. Gortler. Iterative universal rigidity. *Discrete Comput. Geom.*, 53(4):847–877, 2015.

[32] G. Cravo. Recent progress on matrix completion problems. *J. Math. Sci. Adv. Appl.*, 1(1):69–90, 2008.

[33] G. Cravo. Matrix completion problems. *Linear Algebra Appl.*, 430(8-9):2511–2540, 2009.

[34] G.M. Crippen and T.F. Havel. *Distance geometry and molecular conformation*, volume 15 of *Chemometrics Series*. Research Studies Press Ltd., Chichester, 1988.

[35] G. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey, 1963.

[36] G. Dantzig. Linear programming. In *History of Mathematical Programming: A Collection of Personal Reminiscences*. CWI North-Holland, Amsterdam, 1991.

[37] J. Dattorro. Convex optimization & Euclidean distance geometry. `https://ccrma.stanford.edu/~dattorro/mybook.html`, Meboo Publishing, 2005.

[38] E. de Klerk, M. E.-Nagy, and R. Sotirov. On semidefinite programming bounds for graph bandwidth. *Optim. Methods Softw.*, 28(3):485–500, 2013.

[39] E. de Klerk, C. Roos, and T. Terlaky. Infeasible–start semidefinite programming algorithms via self–dual embeddings. In *Topics in Semidefinite and Interior-Point Methods*, volume 18 of *The Fields Institute for Research in Mathematical Sciences, Communications Series*, pages 215–236. American Mathematical Society, 1998.

[40] E. de Klerk and R. Sotirov. Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem. *Math. Program.*, 122(2, Ser. A):225–246, 2010.

[41] E. de Klerk, R. Sotirov, and U. Truetsch. A new semidefinite programming relaxation for the quadratic assignment problem and its computational perspectives. *INFORMS J. Comput.*, 27(2):378–391, 2015.

[42] Y. Ding, N. Krislock, J. Qian, and H. Wolkowicz. Sensor network localization, Euclidean distance matrix completions, and graph realization. *Optim. Eng.*, 11(1):45–66, 2010.

[43] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli. Euclidean distance matrices: A short walk through theory, algorithms and applications. *CoRR*, abs/1502.07541, 2015.

[44] A.L. Dontchev, A.S. Lewis, and R.T. Rockafellar. The radius of metric regularity. *Trans. Amer. Math. Soc.*, 355(2):493–517 (electronic), 2003.

[45] A.W.M. Dress and T.F. Havel. The fundamental theory of distance geometry. In *Computer aided geometric reasoning, Vol. I, II (Sophia-Antipolis, 1987)*, pages 127–169. INRIA, Rocquencourt, 1987.

[46] D. Drusvyatskiy, N. Krislock, Y-L. Cheung Voronin, and H. Wolkowicz. Noisy sensor network localization: robust facial reduction and the Pareto frontier. *SIAM Journal on Optimization*, pages 1–30, 2017. arXiv:1410.6852, 20 pages, to appear.

[47] D. Drusvyatskiy, G. Pataki, and H. Wolkowicz. Coordinate shadows of semidefinite and Euclidean distance matrices. *SIAM J. Optim.*, 25(2):1160–1178, 2015.

[48] H. Dym and I. Gohberg. Extensions of band matrices with band inverses. *Linear Algebra Appl.*, 36:1–24, 1981.

[49] C.S. Edwards. The derivation of a greedy approximator for the koopmans-beckmann quadratic assignment problem. *Proc. CP77 Combinatorial Prog. Conf.*, pages 55–86, 1977.

[50] M. Fazel. *Matrix Rank Minimization with Applications.* PhD thesis, Stanford University, Stanford, CA, 2001.

[51] M. Fazel, H. Hindi, and S.P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings American Control Conference*, pages 4734–4739, 2001.

[52] G. Finke, R.E. Burkard, and F. Rendl. Quadratic assignment problems. *Ann. Discrete Math.*, 31:61–82, 1987.

[53] G. Finke, R.E. Burkard, and F. Rendl. Quadratic assignment problems. In *Surveys in combinatorial optimization (Rio de Janeiro, 1985)*, volume 132 of *North-Holland Math. Stud.*, pages 61–82. North-Holland, Amsterdam, 1987.

[54] J. Gauvin and J.W. Tolle. Differential stability in nonlinear programming. *SIAM J. Control Optimization*, 15(2):294–311, 1977.

[55] S.J. Gortler and D.P. Thurston. Characterizing the universal rigidity of generic frameworks. *Discrete Comput Geom*, 51:1017–1036, 2014.

[56] S.J. Gortler and D.P. Thurston. Characterizing the universal rigidity of generic frameworks. *Discrete Comput. Geom.*, 51(4):1017–1036, 2014.

[57] F.J. Gould and J.W. Tolle. Geometry of optimality conditions and constraint qualifications. *Math. Programming*, 2(1):1–18, 1972.

[58] J.C. Gower. Euclidean distance geometry. *Math. Sci.*, 7(1):1–14, 1982.

[59] J.C. Gower. Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra Appl.*, 67:81–97, 1985.

[60] H. Grassmann. *Extension theory*, volume 19 of *History of Mathematics*. American Mathematical Society, Providence, RI, 2000. Translated from the 1896 German original and with a foreword, editorial notes and supplementary notes by Lloyd C. Kannenberg.

[61] B. Grone, C.R. Johnson, E. Marques de Sa, and H. Wolkowicz. Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.*, 58:109–124, 1984.

[62] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM J. of Control*, 7(2):232–241, 1969.

[63] K.J. Harrison. Matrix completions and chordal graphs. *Acta Math. Sin. (Engl. Ser.)*, 19(3):577–590, 2003. International Workshop on Operator Algebra and Operator Theory (Linfen, 2001).

[64] T.F. Havel, I.D. Kuntz, and B. Crippen. Errata: "The theory and practice of distance geometry" [Bull. Math. Biol. **45** (1983), no. 5, 665–720; MR0718540 (84k:92037)]. *Bull. Math. Biol.*, 47(1):157, 1985.

[65] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM J. Optim.*, 10(3):673 – 696, 2000.

[66] S. Huang and H. Wolkowicz. Low-rank matrix completion using nuclear norm with facial reduction. *J. Global Optim.*, 2017. 23 pages, to appear.

[67] X. Huang. *Preprocessing and postprocessing in linear optimization*. PhD thesis, McMaster University, 2004.

[68] C.R. Johnson. Matrix completion problems: a survey. In *Matrix theory and applications (Phoenix, AZ, 1989)*, pages 171–198. Amer. Math. Soc., Providence, RI, 1990.

[69] C.R. Johnson. Matrix completion problems: a survey. *Proceedings of Symposium in Applied Mathematics*, 40:171–198, 1990.

[70] C.R. Johnson, editor. *Matrix theory and applications*, volume 40 of *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, Providence, RI, 1990. Lecture notes prepared for the American Mathematical Society Short Course held in Phoenix, Arizona, January 10–11, 1989, AMS Short Course Lecture Notes.

[71] C.R. Johnson. Positive definite completions: a guide to selected literature. In *Signal processing, Part I*, volume 22 of *IMA Vol. Math. Appl.*, pages 169–188. Springer, New York, 1990.

[72] C.R. Johnson, B. Kroschel, and H. Wolkowicz. An interior-point method for approximate positive semidefinite completions. *Comput. Optim. Appl.*, 9(2):175–190, 1998.

[73] C.R. Johnson and P. Tarazaga. Connections between the real positive semidefinite and distance matrix completion problems. *Linear Algebra Appl.*, 223/224:375–391, 1995. Special issue honoring Miroslav Fiedler and Vlastimil Pták.

[74] M. Kojima, S. Kim, and H. Waki. Sparsity in sums of squares of polynomials. *Math. Program.*, 103(1, Ser. A):45–62, 2005.

[75] N. Krislock. *Semidefinite Facial Reduction for Low-Rank Euclidean Distance Matrix Completion.* PhD thesis, University of Waterloo, 2010.

[76] N. Krislock and H. Wolkowicz. Explicit sensor network localization using semidefinite representations and facial reductions. *SIAM Journal on Optimization*, 20(5):2679–2708, 2010.

[77] N. Krislock and H. Wolkowicz. Euclidean distance matrices and applications. In *Handbook on Semidefinite, Cone and Polynomial Optimization*, number 2009-06 in International Series in Operations Research & Management Science, pages 879–914. Springer-Verlag, 2011.

[78] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817 (electronic), 2000/01.

[79] J.B. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series.* Imperial College Press, London, 2010.

[80] M. Laurent. A connection between positive semidefinite and Euclidean distance matrix completion problems. *Linear Algebra Appl.*, 273:9–22, 1998.

[81] M. Laurent. A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In *Topics in semidefinite and interior-point methods (Toronto, ON, 1996)*, volume 18 of *Fields Inst. Commun.*, pages 51–76. Amer. Math. Soc., Providence, RI, 1998.

[82] M. Laurent. A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In *Topics in semidefinite and interior-point methods (Toronto, ON, 1996)*, pages 51–76. Amer. Math. Soc., Providence, RI, 1998.

[83] M. Laurent. Polynomial instances of the positive semidefinite and Euclidean distance matrix completion problems. *SIAM J. Matrix Anal. Appl.*, 22:874–894, 2000.

[84] M. Laurent. Matrix completion problems. In *Encyclopedia of Optimization*, pages 1311–1319. Springer US, 2001.

[85] M. Laurent and F. Vallentin. Semidefinite Optimization. `http://homepages.cwi.nl/~monique/master_SDP_2016.pdf`, 2016. [Online; accessed 25-April-2016].

[86] L. Liberti and C. Lavor. Open research areas in distance geometry. *ArXiv e-prints*, October 2016.

[87] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino. Euclidean distance geometry and applications. *SIAM Review*, 56(1):3–69, 2014.

[88] M. Liu and G. Pataki. Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. Technical report, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 2015.

[89] E.M. Loiola, Nair M. Maia d A., P.O. Boaventura-Netto, P. Hahn, and T. Querido. A survey for the quadratic assignment problem. *European J. Oper. Res.*, 176(2):657–690, 2007.

[90] Z-Q. Luo, J.F. Sturm, and S. Zhang. Duality results for conic convex programming. Technical Report Report 9719/A, April, Erasmus University Rotterdam, Econometric Institute EUR, P.O. Box 1738, 3000 DR, The Netherlands, 1997.

[91] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.*, 17:37–47, 1967.

[92] O.L. Mangasarian. *Nonlinear Programming*. McGraw-Hill, New York, NY, 1969.

[93] H. Massam. Optimality conditions for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 27(2):141–162, 1979.

[94] H. Massam. Optimality conditions using sum-convex approximations. *J. Optim. Theory Appl.*, 35(4):475–495, 1981.

[95] J. Maybee and J. Quirk. Qualitative problems in matrix theory. *SIAM Rev.*, 11:30–51, 1969.

[96] C. Mészáros and U.H. Suhl. Advanced preprocessing techniques for linear and quadratic programming. *OR Spectrum*, 25:575–595, 2003. 10.1007/s00291-003-0130-x.

[97] A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, editors. *Distance geometry*. Springer, New York, 2013. Theory, methods, and applications.

[98] Y.E. Nesterov and A.S. Nemirovski. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[99] C.E. Nugent, T.E. Vollman, and J. Ruml. An experimental comparison of techniques for the assignment of facilities to locations. *Operations Research*, 16:150–173, 1968.

[100] D.E. Oliveira, H. Wolkowicz, and Y. Xu. ADMM for the SDP relaxation of the QAP. Technical report, University of Waterloo, Waterloo, Ontario, 2015. arXiv:1512.05448, under revision Oct. 2016, 12 pages.

[101] P. Pardalos, F. Rendl, and H. Wolkowicz. The quadratic assignment problem: a survey and recent developments. In P.M. Pardalos and H. Wolkowicz, editors, *Quadratic assignment and related problems (New Brunswick, NJ, 1993)*, pages 1–42. Amer. Math. Soc., Providence, RI, 1994.

[102] P. Pardalos and H. Wolkowicz, editors. *Quadratic assignment and related problems*. American Mathematical Society, Providence, RI, 1994. Papers from the workshop held at Rutgers University, New Brunswick, New Jersey, May 20–21, 1993.

[103] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003. Algebraic and geometric methods in discrete optimization.

[104] G. Pataki. A simple derivation of a facial reduction algorithm and extended dual systems. Technical report, Columbia University, New York, 2000.

[105] G. Pataki. Bad semidefinite programs: they all look the same. *ArXiv e-prints*, December 2011.

[106] G. Pataki. On the connection of facially exposed and nice cones. *J. Math. Anal. Appl.*, 400(1):211–221, 2013.

[107] G. Pataki. Strong duality in conic linear programming: facial reduction and extended duals. In David Bailey, Heinz H. Bauschke, Frank Garvan, Michel Thera, Jon D. Vanderwerff, and Henry Wolkowicz, editors, *Computational and analytical mathematics*, volume 50 of *Springer Proc. Math. Stat.*, pages 613–634. Springer, New York, 2013.

[108] J. Peña and J. Renegar. Computing approximate solutions for convex conic systems of constraints. *Math. Program.*, 87(3, Ser. A):351–383, 2000.

[109] F. Permenter, H. Friberg, and E. Andersen. Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach. Technical report, MIT, Boston, MA, 2015.

[110] F. Permenter and P. Parrilo. Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone. Technical Report Preprint arXiv:1408.4685, MIT, Boston, MA, 2014.

[111] F. Permenter and P. A. Parrilo. Basis selection for sos programs via facial reduction and polyhedral approximations. In *53rd IEEE Conference on Decision and Control*, pages 6615–6620, Dec 2014.

[112] F. Permenter and P. A. Parrilo. Dimension reduction for semidefinite programs via Jordan algebras. *ArXiv e-prints*, August 2016.

[113] T.K. Pong, H. Sun, N. Wang, and H. Wolkowicz. Eigenvalue, quadratic programming, and semidefinite programming relaxations for a cut minimization problem. *Comput. Optim. Appl.*, 63(2):333–364, 2016.

[114] M.V. Ramana. *An Algorithmic Analysis of Multiquadratic and Semidefinite Programming Problems*. PhD thesis, Johns Hopkins University, Baltimore, Md, 1993.

[115] M.V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Programming*, 77(2):129–162, 1997.

[116] M.V. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM J. Optim.*, 7(3):641–662, 1997.

[117] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.

[118] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.

[119] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.

[120] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Programming*, 70(3, Ser. A):279–351, 1995.

[121] S.M. Robinson. Stability theorems for systems of inequalities, part i: linear systems. *SIAM J. Numerical Analysis*, 12:754–769, 1975.

[122] S.M. Robinson. Stability theorems for systems of inequalities, part ii: differentiable nonlinear systems. *SIAM J. Numerical Analysis*, 13:497–513, 1976.

[123] R.T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[124] B. Rosgen and L. Stewart. Complexity results on graphs with few cliques. *Discrete Math. Theor. Comput. Sci.*, 9(1):127–135 (electronic), 2007.

[125] S. Sahni and T. Gonzales. P-complete approximation problems. *Journal of ACM*, 23:555–565, 1976.

[126] A. Singer. A remark on global positioning from local distances. *Proc. Natl. Acad. Sci. USA*, 105(28):9507–9511, 2008.

[127] F.J. Sturm. *Primal-dual interior point approach Semidefinite programming.* PhD thesis, Erasmus University, Rotterdam, Netherlands, 1997.

[128] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999. Interior point methods.

[129] J.F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10(4):1228–1248 (electronic), 2000.

[130] H. Sun. Admm for sdp relaxation of gp. Master's thesis, University of Waterloo, 2016.

[131] Z. Tang and H. Wolkowicz. ADMM for the second lifting SDP relaxation of MC. Technical report, University of Waterloo, Waterloo, Ontario, 2017. in progress.

[132] S. Tanigawa. Singularity degree of the positive semidefinite matrix completion problem. Technical Report arXiv:1603.09586, Research Institute for Mathematical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan, 2016.

[133] P. Tarazaga. Faces of the cone of Euclidean distance matrices: characterizations, structure and induced geometry. *Linear Algebra Appl.*, 408:1–13, 2005.

[134] P. Tarazaga, T.L. Hayden, and J. Wells. Circum-Euclidean distance matrices and faces. *Linear Algebra Appl.*, 232:77–96, 1996.

[135] L. Tunçel. On the Slater condition for the SDP relaxations of nonconvex sets. *Oper. Res. Lett.*, 29(4):181–186, 2001.

[136] L. Tunçel. *Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization*, volume 27 of *Fields Institute Monographs.* American Mathematical Society, Providence, RI, 2010.

[137] L. Tunçel. *Polyhedral and semidefinite programming methods in combinatorial optimization*, volume 27 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2010.

[138] L. Tunçel and H. Wolkowicz. Strong duality and minimal representations for cone optimization. *Comput. Optim. Appl.*, 53(2):619–648, 2012.

[139] L. Vandenberghe and M.S. Andersen. Chordal graphs and semidefinite optimization. *Found. Trends Opt.*, 1(4):241–433, 2015.

[140] H. Waki, S. Kim, M. Kojima, and M. Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.*, 17(1):218–242, 2006.

[141] H. Waki and M. Muramatsu. A facial reduction algorithm for finding sparse SOS representations. *Oper. Res. Lett.*, 38(5):361–365, 2010.

[142] H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013.

[143] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Trans. Amer. Math. Soc.*, 36(1):63–89, 1934.

[144] H. Wolkowicz. Geometry of optimality conditions and constraint qualifications: the convex case. *Math. Programming*, 19(1):32–60, 1980.

[145] H. Wolkowicz. Some applications of optimization in matrix theory. *Linear Algebra Appl.*, 40:101–118, 1981.

[146] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of semidefinite programming*. International Series in Operations Research & Management Science, 27. Kluwer Academic Publishers, Boston, MA, 2000. Theory, algorithms, and applications.

[147] H. Wolkowicz and Q. Zhao. Semidefinite programming relaxations for the graph partitioning problem. *Discrete Appl. Math.*, 96/97:461–479, 1999. Selected for the special Editors' Choice, Edition 1999.

[148] S. Wright. *Primal-Dual Interior-Point Methods.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, 1996.

[149] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *J. Comb. Optim.*, 2(1):71–109, 1998. Semidefinite programming and interior-point approaches for combinatorial optimization problems (Fields Institute, Toronto, ON, 1996).

[150] Y-R. Zhu. *Recent Advances and Challenges in Quadratic Assignment and Related Problems.* PhD thesis, University of Pennsylvania, 2007. PhD Thesis.