PRIMAL–DUAL INTERIOR POINT APPROACH

TO SEMIDEFINITE PROGRAMMING

Cover design: Mirjam Bode

The book is no. 156 of the Tinbergen Institute Research Series. This series is established through cooperation between Thesis Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

PRIMAL–DUAL INTERIOR POINT APPROACH TO SEMIDEFINITE PROGRAMMING

(Primaal–duaal inwendige punt aanpak

voor semidefiniet programmeren)

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
prof. dr P.W.C. Akkermans M.A.
en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 11 september 1997 om 13.30 uur
door

Jos Fredrik Sturm

geboren te Rotterdam, Nederland

Promotiecommissie

Promotor:	Prof. dr ir R. Dekker
Overige leden:	Dr J. Brinkhuis
	Prof. dr Z.-Q. Luo
	Dr C. Roos
	Dr S. Zhang (tevens co-promotor)

Acknowledgments

This is an excellent opportunity to express my gratitude to my promotor, Rommert Dekker. He kindly supported me in the course of my Ph.D. project. Most conveniently, we used to discuss the plans for conference visits while cycling home through the woodland ('Kralingse bos'). Two enthusiastic supervisors, Hans Frenk and Shuzhong Zhang, were an obvious stimulus, and as such they made my position really enviable. My attitude to duality and convexity has been influenced by the course 'Fundamentals of Optimization', that Hans Frenk taught at the Tinbergen Institute.

The cooperation with my daily supervisor Shuzhong Zhang has invariantly been intense. Shuzhong has a thorough, up to date knowledge of linear and nonlinear programming, and he also maintains a large network of scientific friends. What makes him really a stimulating supervisor, is the fact that he is literally always in the mood for discussions. In this respect, I also have to thank Yin, Shuzhong's wife, for her admirable understanding, and her great skills in Chinese cooking.

There is no faculty of mathematics at Erasmus University, and this makes mathematical sources scarce. Initially, I experienced this as a handicap, but my problems were solved after I met Jan Brinkhuis: his capability and willingness in answering my questions in mathematics is invaluable.

The visit of Tom Luo in December 1995 has been an exciting experience. At that time, Shuzhong and I had just finished our first paper on semidefinite programming, and together with Tom we proceeded on some of the most challenging open problems in this area. This month turned out to be extremely productive.

I have greatly benefited from the interior point group in Delft, namely Kees Roos, Tamás Terlaky, Benjamin Jansen and Etienne de Klerk. In particular, I very much appreciate the fact that they invited some well known foreign researchers in the area of interior point methods, and organized the workshop HPOPT96. Altogether, both cooperation and competition with our colleagues in Delft have been very stimulating.

Financial support of NWO, SHELL and the Tinbergen Institute enabled me to visit a number of international conferences. These visits have been essential for keeping track of the progress in my field of research. In this respect, I also like to mention the service that is provided by ‘Interior point methods online’ [29].

When I came to Rotterdam, I immediately enjoyed an exciting, ever–changing international atmosphere, that was largely forged by a blend of cordial foreign colleagues, namely Ana Isabel, Bernd, Bo, Heping, Joaquim, Teresa, Thomas and Vitaly. The friendship with my Portuguese roommates is especially worth mentioning. Teresa has contributed directly to this thesis, by supplying me with the LaTeX commands that comply with the style of the Tinbergen Institute Research Series.

All Ph.D. students in the ‘Mabes’ group are more than colleagues. In fact, the interaction with Arjan, Bart, Gonda, the two Marcels, Patrick, Ralph, Raymond and Richard made a joy of every day. Bart has carefully checked the summary in the back of this thesis.

The truly familiar concern exposed by my parents and my brother has contributed more to this thesis than they will ever know.

Contents

Chapter 1

Introduction

Semidefinite programming is a rich class of optimization problems with (generalized) convex constraints. Constraints in a semidefinite programming problem take the form of linear matrix inequalities. Such inequalities specify bounds on the eigenvalues of linear combinations of symmetric or Hermitian matrices. This way of modeling constraints is common practice in control theory, but it is still unusual in other branches of research. However, semidefinite progamming has a great potential of application outside control theory, since many types of (generalized) convex functions can be modeled by linear matrix inequalities, even if the underlying problem has no obvious relation to matrix theory.

The usual treatment of nonlinear optimization problems involves inequalities of the form

$$g(x) \leq 0, \tag{1.1}$$

where $g(\cdot)$ is some real valued function with a domain in $\Re^n$. Well known packages for solving nonlinear programming problems with constraints in the above form are *Minos* [109], *Lancelot* [25] and *Conopt* [31]. These packages are designed for application on possibly nonconvex problems, since there is no easy way for checking convexity of a function in the form (1.1). Consequently, these packages generally do not warn for nonconvexity, even though such nonconvexity can be caused by a mistake in the input of the model: nonlinear programming models have to be formulated with care. In contrast, the form of semidefinite programs ensures convexity, and this allows practical application of the elegant duality relations that hold for convex programming problems. In particular, dual solutions can (and should) be used to certify the validity of the output of semidefinite programming solvers.

From an algorithmic point of view, the duality structure of semidefinite programs facilitates the development of symmetric primal–dual methods. These type of methods are scale invariant and do not require problem specific parameter tuning. This clearly con-

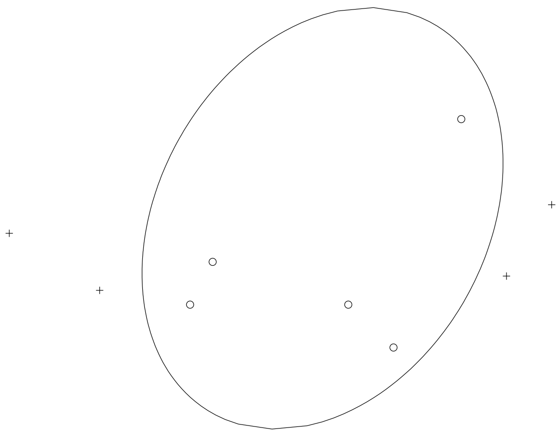

Figure 1.1: A separating ellipsoid

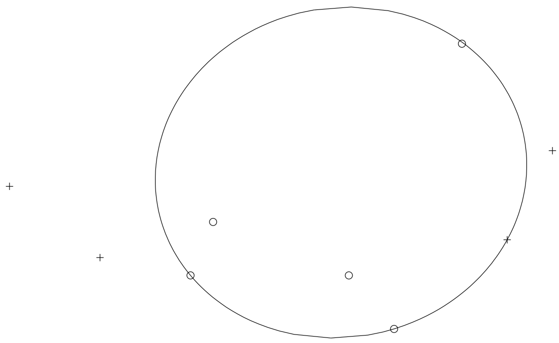

Figure 1.2: A minimal separating ellipsoid

trasts with the behavior of nonlinear programming solvers, which can be very sensitive to small changes in the problem formulation and adjustments to optional parameters [108].

In summary, there are good reasons to try a semidefinite programming formulation, even if a nonlinear programming formulation in the form (1.1) seems more natural. In Appendix B, we discuss semidefinite programming formulations for several classes of (generalized) convex constraints. Of course, there are many cases in which a semidefinite formulation is impossible, for instance if the problem does not allow any convex formulation. (In such cases, semidefinite programming can still be useful for solving semidefinite relaxations.) Similarly, there are applications of semidefinite programming in which traditional nonlinear programming formulations are impossible, or highly unsuitable. We will mention some of those typical applications of semidefinite programming in Section 1.1.

1.1 Applications

The problem of separating two sets of points by a hyperplane can be formulated as a linear programming problem. However, if the two sets cannot be separated by a hyperplane, we may like to separate them by an ellipsoid, such as in Figure 1.1, and this leads to a problem that cannot be handled by linear programming. Instead, we have used semidefinite programming to calculate this separating ellipsoid. Semidefinite programming also allows to optimize the shape of such an ellipsoid. For instance, the ellipsoid in Figure 1.2 has the minimal width. It is also possible to make the ellipsoid as spherical as possible [158], or to minimize the volume of the ellipsoid [159]. If the two sets cannot be separated by any ellipsoid, we can generate a Farkas–type dual solution by semidefinite programming. In general, Farkas–type dual solutions serve as a proof of inconsistency for a system of inequalities. Notice that the graphical interpretation of our example does not lead to an obvious way for demonstrating that two sets cannot be separated by any ellipsoid. However, a rich duality theory is available once the problem is formulated as a semidefinite program. A thorough treatment of this duality theory is given in Chapter 2 of this thesis. For a further discussion of pattern separation by means of semidefinite programming, and other graphical examples, we refer to Vandenberghe and Boyd [158] and Vandenberghe, Boyd and Wu [159]. Other applications of semidefinite programming in cluster and data analysis are considered by Mirkin [92].

Ben-Tal and Bendsoe [8] and Ben-Tal and Nemirovsky [9, 10] use semidefinite programming for computing optimal designs of engineering structures, in particular trusses. A truss is a construction that consists of connected bars, which is typical for bridges, transmission towers and, most notably, the Eiffel tower.

Stability analysis plays an important role in system theory. For linear systems described by differential equations, the stability question can be answered by using the classical Lyapunov theory. However, systems with uncertainty give rise to differential inclusions, for which the classical Lyapunov theory is not applicable. For such systems, the stability can be analyzed by solving a semidefinite programming problem, see [20, 157, 158]. Vandenberghe and Boyd [157] developed a solver that is specialized for these type of control problems.

Another area where semidefinite programming has already proved valuable is signal processing. For instance, Olkin [115] and Olkin and Titterton [116] used semidefinite programming to design an active noise controller. Active noise controllers aim at quieting noisy areas, such as factory floors, by sending out anti–noise signals. The importance of semidefinite programming for designing and analyzing controllers was made clear in the book of Boyd, El Ghaoui, Feron and Balakrishnan [20]. This book has greatly stimulated research in this area, see e.g. [135, 157, 158]. Although the literature on semidefinite programming and linear matrix inequalities addresses mostly the real symmetric case, there are many practical situations that give rise to Hermitian LMIs, especially in signal processing [84, 81].

Related design problems occur in statistical methods for quality improvement. Opportunities for quality improvement arise if we have knowledge about the effect of certain design parameters on the quality of a product. Semidefinite programming can be used to design an experiment for generating the most informative data. Vandenberghe, Boyd and Wu [159] give semidefinite programming formulations for several problems in optimal experiment design.

Semidefinite programming problems also arise from structured estimation problems in statistics. If the structural constraints involve inequalities, it is in general not possible to derive efficient estimators analytically. For instance, in structured covariance estimation, one of the constraints is always semidefiniteness, and this leads to optimization problems in semidefinite matrix variables [159].

The (statistical) moment problem concerns the question what values a generalized moment can attain, given bounds on other moments of the underlying distribution. In these problems, the distribution function is not restricted to any parametric family. Klein Haneveld [68] used duality theory of infinite dimensional linear programming to tackle these type of problems analytically. However, Vandenberghe, Boyd and Wu [159] recently pointed out that some interesting moment problems can also be formulated as semidefinite programming problems, which enables us to solve large scale moment problems numerically. The educational testing problem is a related statistical problem, which has been well studied in the context of semidefinite programming, see [82, 151, 158].

Semidefinite programming can be used within algorithms for nonconvex global optimization. In this context, semidefinite programming relaxations can often provide good lower bound information. In particular, this approach has been advocated for nonconvex quadratic programming, optimization with zero–one variables and, more generally, multi-quadratic programming, see the surveys of Pardalos [119] and Ramana and Pardalos [126].

Combinatorial problems form an important class within global optimization, where connections with semidefinite programming were recognized early. One of the pioneering contributions to this research area is due to Grötschel, Lovász and Schrijver [50]. Semidefinite programming relaxations are provably effective for the max–cut problem, see Goemans and Williamson [40]. Computational experience with max–cut relaxations is reported by Helmberg, Rendl, Vanderbei and Wolkowicz [53]. The semidefinite relaxation for maximum clique problems yields the Lovász number [83]. Semidefinite programming relaxations are also successful for graph bisection [53], and quadratic assignment [166]. Results on semidefinite programming relaxations with the QAPLIB [22] set of quadratic assignment problems are reported by Zhao, Karisch, Rendl and Wolkowicz [166].

Eigenvalue optimization problems have been studied by many authors, among which are [1, 28, 62, 117, 118, 128, 136]. Since these problems involve nonsmooth objective functions, it is natural to apply nonsmooth optimization techniques, as in Cullum, Donath and Wolfe [28]. More recently however, efficient interior point algorithms were proposed by Jarre [62] and Alizadeh [1], based on semidefinite envelope representations of eigenvalue functions. Such representations are discussed in Appendix B of this thesis.

1.2 Notation

The topic of this thesis is on the cutting edge of various branches of applied mathematics. This has unfortunately resulted in a need for a rather elaborate description of notational conventions.

1.2.1 Sets and Linear Spaces

The n–dimensional real Euclidean space is denoted by $\Re^n$. The standard inner product of two vectors $x, y \in \Re^n$ is $y^{\mathrm{T}}x$, where $^{\mathrm{T}}$ denotes the matrix transpose. We let $\Re_+$ ($\Re_{++}$) denote the half–line of nonnegative (positive) real numbers. The n–dimensional complex Euclidean space is denoted by $\mathcal{C}^n$. Given $x \in \mathcal{C}^n$, we let $\mathrm{Re}\, x \in \Re^n$ and $\mathrm{Im}\, x \in \Re^n$ denote the real part and the imaginary part of x respectively, i.e. $x = (\mathrm{Re}\, x) + \mathbf{i}\ \mathrm{Im}\, x$, where $\mathbf{i}^2 = -1$. We let $\mathrm{conj}(x)$ denote the complex conjugate of x, i.e. $\mathrm{conj}(x) = \mathrm{Re}\, x - \mathbf{i}\ \mathrm{Im}\, x$. The standard complex valued inner product of two vectors $x, y \in \mathcal{C}^n$ is $y^{\mathrm{H}}x$, where $^{\mathrm{H}}$ denotes the complex conjugate transpose, or Hermitian adjoint, i.e. $y^{\mathrm{H}} = \mathrm{conj}(y^{\mathrm{T}})$. The Minkowski sum of two sets $\mathcal{T}$ and $\mathcal{T}'$ in $\Re^n$ is

$$\mathcal{T} \oplus \mathcal{T}' = \{z \in \Re^n \mid z = x + y \text{ for some } x \in \mathcal{T},\ y \in \mathcal{T}'\},$$

and the (asymmetric) difference of $\mathcal{T}$ and $\mathcal{T}'$ is

$$\mathcal{T} \setminus \mathcal{T}' := \{x \in \mathcal{T} \mid x \notin \mathcal{T}'\}.$$

If $\mathcal{A}$ is a linear subspace of $\Re^n$, then $P_{\mathcal{A}}$ denotes the orthogonal projection matrix onto $\mathcal{A}$. The dimension of $\mathcal{A}$ is denoted $\dim \mathcal{A}$.

Given a set $\mathcal{T}$, we let $\mathrm{cl}\, \mathcal{T}$, $\mathrm{int}\, \mathcal{T}$ and $\mathrm{rel}\, \mathcal{T}$ denote the closure of $\mathcal{T}$, the interior of $\mathcal{T}$ and the relative interior of $\mathcal{T}$ respectively. If $\mathcal{T}$ is a subset of $\mathcal{C}^n$ and A is an $m \times n$ matrix, then the image of $\mathcal{T}$ under the linear mapping A is denoted by $A\mathcal{T}$, i.e.

$$A\mathcal{T} = \{y \in \mathcal{C}^m \mid y = Ax, \quad \text{for some } x \in \mathcal{T}\}.$$

The kernel, the image and the rank of A are denoted by $\mathrm{Ker}\, A$, $\mathrm{Img}\, A$ and $\mathrm{rank}\, A$ respectively. Notice that $\mathrm{rank}\, A = \dim\ \mathrm{Img}\, A$, $\mathrm{Img}\, A = A\mathcal{C}^n$ and $\mathrm{Ker}\, A = \{x \in \mathcal{C}^n \mid Ax = 0\}$.

1.2.2 Matrices and Vectorization

Given real matrices X and Y in $\Re^{m_1 \times m_2}$, the standard inner product is defined by

$$X \bullet Y = \mathrm{tr}\, X^{\mathrm{T}}Y,$$

where $\mathrm{tr}\,(\cdot)$ denotes the trace of a matrix. We treat the space of $m_1 \times m_2$ matrices with complex entries as a real linear space of dimension $2m_1m_2$. For this space, we define a real valued inner product as

$$X \bullet Y = ((\mathrm{Re}\, X) \bullet \mathrm{Re}\, Y) + ((\mathrm{Im}\, X) \bullet \mathrm{Im}\, Y),$$

or equivalently,

$$X \bullet Y = \text{ Re } \text{ tr } Y^{\mathrm{H}} X.$$

The notation $X \perp Y$ denotes orthogonality in the sense that $X \bullet Y = 0$.

If X is an $\bar{n} \times \bar{n}$ matrix, we say that X is a square matrix of order $\bar{n}$. The linear space of real symmetric $\bar{n} \times \bar{n}$ matrices is denoted by $\mathcal{S}^{(\bar{n})}$. Its orthogonal complement in $\Re^{\bar{n} \times \bar{n}}$, viz. the space of skew symmetric $\bar{n} \times \bar{n}$ matrices, is denoted by $(\mathcal{S}^{(\bar{n})})^{\perp}$. Remark that $X \in \mathcal{S}^{(\bar{n})}$ if and only if $X = X^{\mathrm{T}}$, and $X \in (\mathcal{S}^{(\bar{n})})^{\perp}$ if and only if $X = -X^{\mathrm{T}}$. The real linear space of Hermitian $\bar{n} \times \bar{n}$ matrices is denoted by $\mathcal{H}^{(\bar{n})}$, and the real linear space of skew–Hermitian $\bar{n} \times \bar{n}$ matrices is denoted by $(\mathcal{H}^{(\bar{n})})^{\perp}$. Remark that $X \in \mathcal{H}^{(\bar{n})}$ if and only if $\text{Re } X \in \mathcal{S}^{(\bar{n})}$ and $\text{Im } X \in (\mathcal{S}^{(\bar{n})})^{\perp}$, or equivalently, $X^{\mathrm{H}} = X$. Similarly, $X \in (\mathcal{H}^{(\bar{n})})^{\perp}$ if and only if $\text{Re } X \in (\mathcal{S}^{(\bar{n})})^{\perp}$ and $\text{Im } X \in \mathcal{S}^{(\bar{n})}$, or equivalently, $X^{\mathrm{H}} = -X$. In the space of $\bar{n} \times \bar{n}$ matrices with complex entries, $(\mathcal{H}^{(\bar{n})})^{\perp}$ is the orthogonal complement of $\mathcal{H}^{(\bar{n})}$, with respect to the real valued inner product $X \bullet Y$.

If the value of $\bar{n}$ is clear from the context, we will drop the superscript $^{(\bar{n})}$ and simply write $\mathcal{S}$, $\mathcal{S}^{\perp}$, $\mathcal{H}$ and $\mathcal{H}^{\perp}$, to denote the spaces of symmetric, skew–symmetric, Hermitian and skew–Hermitian matrices respectively. If $X \in \mathcal{H}$ is Hermitian positive (semi–) definite, we write $X \succ 0$ ($X \succeq 0$). The cone of positive semi–definite matrices is denoted by $\mathcal{H}_{+}$ and the cone of Hermitian positive definite matrices is $\mathcal{H}_{++}$. By definition,

$$X \succeq 0 \iff y^{\mathrm{H}} X y \geq 0 \; \forall y \in \mathcal{C}^{\bar{n}}, \quad X \succ 0 \iff y^{\mathrm{H}} X y > 0 \; \forall 0 \neq y \in \mathcal{C}^{\bar{n}}.$$

Furthermore, we write $X \succeq Y$ if $X - Y$ is positive semidefinite. We remark that given an $\bar{n} \times \bar{n}$ matrix X, it holds that $\text{Im } (y^{\mathrm{H}} X y) = 0$ for all $y \in \mathcal{C}^{\bar{n}}$, if and only if $X \in \mathcal{H}^{(\bar{n})}$, see Horn and Johnson [56].

Since $\mathcal{S}^{(\bar{n})}$ is a real linear space of dimension

$$\dim \mathcal{S}^{(\bar{n})} = \bar{n} + (\bar{n} - 1) + \cdots + 1 = \frac{\bar{n}(\bar{n}+1)}{2},$$

it is endowed with an orthonormal basis of cardinality $m := \dim \mathcal{S}^{(\bar{n})}$. Given a matrix $X \in \mathcal{S}^{(\bar{n})}$, we let $x = \text{vec}_{\mathrm{S}}(X)$ denote its coordinate vector in $\Re^{m}$, with respect to a fixed orthonormal basis of $\mathcal{S}^{(\bar{n})}$. Hence, if we denote this fixed basis by the symmetric matrices $\{U_1, U_2, \ldots, U_m\}$, then

$$x_j = U_j \bullet X \quad \text{ for } j = 1, 2, \ldots, m.$$

Remark from the above identity that the inner product $X \bullet Y$ for symmetric matrices X and Y corresponds to the standard Euclidean inner product $x^{\mathrm{T}} y = \sum_{j=1}^{m} x_j y_j$, viz.

$$X \bullet Y = \left(\sum_{j=1}^{m} x_j U_j \right) \left(\sum_{k=1}^{m} y_k U_k \right) = x^{\mathrm{T}} y.$$

Similarly, $(\mathcal{S}^{(\bar{n})})^{\perp}$ is a real linear space of dimension

$$\dim \, (\mathcal{S}^{(\bar{n})})^{\perp} = (\bar{n} - 1) + (\bar{n} - 2) + \cdots + 0 = \frac{\bar{n}(\bar{n}-1)}{2}.$$

Hence, it has an orthonormal basis of skew–symmetric matrices $\{U_{m+1}, U_{m+2}, \ldots, U_{\bar{n}^2}\}$. Since $\mathcal{H}^{(\bar{n})} = \mathcal{S}^{(\bar{n})} + \mathbf{i}\,(\mathcal{S}^{(\bar{n})})^{\perp}$, it follows that

$$U_1, U_2, \ldots, U_m,\ \mathbf{i}\,U_{m+1},\ \mathbf{i}\,U_{m+2}, \ldots,\ \mathbf{i}\,U_{\bar{n}^2}$$

forms an orthonormal basis of the real linear space of $\bar{n} \times \bar{n}$ Hermitian matrices, with respect to the inner product $X \bullet Y = \text{ Re tr}\, Y^{\mathrm{H}} X$. The dimension of this space is $n := \dim \mathcal{H}^{(\bar{n})} = \bar{n}^2$. Given a matrix $X \in \mathcal{H}^{(\bar{n})}$, we let $x = \text{ vec}_{\mathrm{H}}\ (X)$ denote its coordinate vector in $\Re^n$, with respect to the above mentioned orthonormal basis of $\mathcal{H}^{(\bar{n})}$, i.e.

$$x_j = U_j \bullet X = U_j \bullet \text{ Re}\, X \text{ for } j = 1, 2, \ldots, m,$$

$$x_j = (\ \mathbf{i}\,U_j) \bullet X = U_j \bullet \text{ Im}\, X \text{ for } j = (m+1), (m+2), \ldots, n.$$

We will use the convention that lower case letters represent vectors, and upper case letters represent matrices. Once we introduce a Hermitian matrix $X \in \mathcal{H}^{(\bar{n})}$, we implicitly define the lower case symbol x to denote its vector representation $\text{ vec}_{\mathrm{H}}\ (X)$ in $\Re^{\bar{n}^2}$, and vice versa. Moreover, if we introduce a certain set $\mathcal{A}$ of Hermitian matrices, it depends on the context whether we consider $\mathcal{A}$ as a subset of $\mathcal{H}^{(\bar{n})}$ or as a subset of $\Re^{\bar{n}^2}$. Thus, if we write $X \in \mathcal{H}_+^{(\bar{n})}$, then $\mathcal{H}_+^{(\bar{n})}$ (the Hermitian positive semidefinite cone) is treated as a subset of $\mathcal{H}^{(\bar{n})}$, and if we write $x \in \mathcal{H}_+^{(\bar{n})}$, then $\mathcal{H}_+^{(\bar{n})}$ is treated as a subset of $\Re^n$.

Given a square matrix X, we let $P_{\mathcal{H}}(X)$ and $P_{\mathcal{H}^{\perp}}(X)$ denote the projections of X on $\mathcal{H}$ and $\mathcal{H}^{\perp}$ respectively. Namely, we let

$$P_{\mathcal{H}}(X) = \frac{X + X^{\mathrm{H}}}{2}, \quad P_{\mathcal{H}^{\perp}}(X) = \frac{X - X^{\mathrm{H}}}{2}.$$

The identity matrix is denoted by I. For Hermitian X, we let $\lambda_{\min}(X)$ be the smallest eigenvalue of X. (Recall that Hermitian matrices have only real eigenvalues.)

The *Hermitian Kronecker product* notation is convenient when matrix products have to be vectorized. Given two $\bar{n} \times \bar{n}$ matrices X and Z, we consider the linear mapping $f_{X,Z} : \mathcal{H} \to \mathcal{H}$ defined by

$$f_{X,Z}(Y) = P_{\mathcal{H}}(ZYX^{\mathrm{H}}).$$

Since $f_{X,Z}(Y)$ is linear in Y, there exists for given X and Z a matrix $X \otimes_{\mathrm{H}} Z$ such that

$$(X \otimes_{\mathrm{H}} Z)\ \text{vec}_{\mathrm{H}}\ (Y) = \text{ vec}_{\mathrm{H}}\ (f(Y)) \text{ for all } Y \in \mathcal{H}^{(\bar{n})}.$$

This matrix is called the Hermitian Kronecker product of X and Z, and this product is commutative, i.e. $X \otimes_{\mathrm{H}} Z = Z \otimes_{\mathrm{H}} X$. Vectorization and Kronecker products for real symmetric matrices are discussed in detail by Alizadeh, Haeberly and Overton [4] and Todd, Toh and Tütüncü [151].

1.2.3 Norm and Distance

Given a vector $x \in \mathcal{C}^n$, we let $\|x\|$ denote a norm of x, for which the dual norm is $\|x\|^*$, i.e.

$$\|x\|^* = \max_y \{ \operatorname{Re} y^{\mathrm{H}} x \mid \|y\| = 1\}.$$

Moreover, the vector norm $\|\cdot\|$ induces a matrix norm, viz. its operator norm, defined as

$$\|Y\| = \max\{\|Yx\| \mid \|x\| = 1\},$$

where Y is a matrix. The Euclidean norm of a vector x is denoted by $\|x\|_2$; the associated operator norm $\|Y\|_2$ is the maximal singular value of the matrix Y. In particular, if Y is positive semidefinite, then $\|Y\|_2$ is the largest eigenvalue (or spectral radius) of Y. The Frobenius norm of Y is $\|Y\|_F = \sqrt{Y \bullet Y}$. Remark for $Y \in \mathcal{H}$ that

$$\|y\|_2 = \|Y\|_F = \sqrt{\sum_{j=1}^{\bar{n}} \lambda_j(Y)^2},$$

where $\lambda_1(Y), \lambda_2(Y), \ldots, \lambda_{\bar{n}}(Y)$ are the eigenvalues of Y. Some well known properties of the matrix norms $\|\cdot\|_2$ and $\|\cdot\|_F$ are

$$\|X\|_2 \leq \|X\|_F.$$

$$\|XY\|_F \leq \|X\|_2 \|Y\|_F,$$

$$\|X\|_2 = \|X^{\mathrm{H}}\|_2, \quad \|X\|_F = \|X^{\mathrm{H}}\|_F,$$

for arbitrary matrices X and Y.

The distance in Frobenius norm from a Hermitian matrix X to a given convex set $\mathcal{T}$ in the real space of Hermitian matrices is denoted by $\operatorname{dist}_F(X, \mathcal{T})$, i.e.

$$\operatorname{dist}_F(X, \mathcal{T}) = \inf_{Y \in \mathcal{T}} \|X - Y\|_F.$$

The distance from a vector $x \in \Re^n$ to a convex set $\mathcal{T} \subseteq \Re^n$ is

$$\operatorname{dist}(x, \mathcal{T}) = \inf_{y \in \mathcal{T}} \|x - y\|.$$

Similarly, we let

$$\operatorname{dist}(\mathcal{T}, \mathcal{T}') = \inf_{x \in \mathcal{T}} \operatorname{dist}(x, \mathcal{T}')$$

denote the distance between two convex sets $\mathcal{T}$ and $\mathcal{T}'$ in the real Euclidean space.

1.2.4 Big O and Small o

In local convergence analysis, we use the standard Landau notation for sequences. In particular, if $\{u(t) \mid t > 0\}$ and $\{w(t) \mid t > 0\}$ are real sequences with $w(t) > 0$, then $u(t) = O(w(t))$ means that $u(t)/w(t)$ is bounded, independent of t; $u(t) = o(w(t))$ means that $\lim_{t\downarrow 0} u(t)/w(t) = 0$; $u(t) \sim w(t)$ means that $\lim_{t\downarrow 0} u(t)/w(t) = 1$; $u(t) = \Theta(w(t))$ whenever $u(t)/w(t)$ and $w(t)/u(t)$ are both bounded. For a Hermitian positive definite matrix, we use "O" and "Θ" to denote the order of all its eigenvalues. Hence, for $U(t) \in \mathcal{H}_{++}$, the notation $U(t) = \Theta(w(t))$ signifies the existence of $\Gamma > 0$ such that

$$\frac{1}{\Gamma} I \preceq \frac{1}{w(t)} U(t) \preceq \Gamma I, \qquad \text{for all } t > 0.$$

Therefore, the condition $U(t) = \Theta(w(t))$ implies that $\|U(t)\| = \Theta(w(t))$ and the diagonal entries

$$U_{11}(t),\ U_{22}(t),\ \ldots,\ U_{nn}(t)$$

are all $\Theta(w(t))$ too.

In global convergence analysis, the big O is used to denote the existence of a universal bound, instead of a bound for a specific sequence. To prevent confusion, this type of big $\mathbf{O}$ is printed in boldface. In particular, if $u(\cdot)$ and $w(\cdot)$ are positive quantities that depend on the problem data (such as the number of variables and constraints in the problem), then

$$u(\text{problem data}) = \mathbf{O}(w(\text{problem data}))$$

means that there exists a universal constant $\Gamma > 0$ such that

$$u(\text{problem data}) \leq \Gamma w(\text{problem data})$$

for any problem instance.

1.3 Conic Form

To facilitate the theoretical development of semidefinite programming, it is important to agree on a standard formulation of the problem. The *conic form* is the standard that is used throughout this book.

1.3.1 Semidefinite Programming

The optimization problem

$$\inf\{C \bullet X \mid X \in (B + \mathcal{A}) \cap \mathcal{H}_+\} \tag{1.2}$$

is a semidefinite program in conic form. The decision variable is a (Hermitian) positive semidefinite matrix X, which is restricted to an affine space $B+\mathcal{A}$, where B is a given Hermitian matrix and $\mathcal{A}$ is a linear subspace of $\mathcal{H}$. The objective is to minimize the linear function $C \bullet X$, where C is a given Hermitian matrix. Notice that we leave complete freedom regarding the formulation of the linear subspace $\mathcal{A}$. Consequently, any semidefinite program easily fits this framework. It will become clear later why this formulation is addressed to as the conic form.

Many authors treat semidefinite programming in the so–called standard form. The example below shows how a standard form semidefinite programming problem can be transformed into conic form.

Example 1.1 *Consider a standard semidefinite programming problem*

$$\inf\{C \bullet X \mid A^{(i)} \bullet X = \bar{b}_i \text{ for } i=1,\ldots,m,\ X \succeq 0\}. \tag{1.3}$$

Let $B \in \mathcal{H}$ be such that

$$A^{(i)} \bullet B = \bar{b}_i \text{ for } i=1,2,\ldots,m.$$

Notice that if such B does not exist, then the system of linear equalities

$$A^{(i)} \bullet X = \bar{b}_i \text{ for } i=1,2,\ldots,m.$$

is inconsistent, and does not describe any geometrical object. Define further the linear subspace $\mathcal{A} \subseteq \mathcal{H}$ by

$$\mathcal{A} = \{Y \in \mathcal{H} \mid A^{(i)} \bullet Y = 0 \text{ for } i=1,\ldots,m\}.$$

We can now rewrite (1.3) in conic form as

$$\inf\{C \bullet X \mid X \in (B+\mathcal{A}) \cap \mathcal{H}_+\}.$$

At first sight, the standard semidefinite program (1.3) in Example 1.1 may seem somewhat special: it involves only one semidefinite matrix X. In fact, a standard linear programming problem with $\bar{n}$ nonnegativity constraints, say

$$x_1 \in \Re_+,\ x_2 \in \Re_+, \ldots, x_{\bar{n}} \in \Re_+,$$

is a semidefinite program with $\bar{n}$ semidefiniteness constraints,

$$x_1 \in \mathcal{H}_+^{(1)},\ x_2 \in \mathcal{H}_+^{(1)}, \ldots, x_{\bar{n}} \in \mathcal{H}_+^{(1)}.$$

Observe however, that

$$X_1 \succeq 0, X_2 \succeq 0 \iff \begin{bmatrix} X_1 & X_{12} \\ X_{12}^{\mathrm{H}} & X_2 \end{bmatrix} \succeq 0 \quad \text{for some } X_{12},$$

which shows that we can transform two semidefiniteness constraints into a single one, simply by introducing an extra variable X_{12}. (See Appendix A for a review of Hermitian matrices and positive definiteness.) Some readers may get worried here about the number of extra variables that are introduced in this way: a standard linear program with $\bar{n}$ variables for instance, yields a semidefinite program with an $\bar{n} \times \bar{n}$ matrix of decision variables. However, these extra variables are introduced for notational simplicity only; they will not be used in any algorithm.

The fact that the decision variable X is a Hermitian matrix enables us to use such specific features as eigenvalue decomposition, matrix multiplication, matrix square root (if X is positive semidefinite) and matrix inverse (if X is positive definite). However, in those cases where the interpretation of X as Hermitian matrix is not essential, it is more convenient to work with a vector of decision variables. As described in Section 1.2, we use the notational convention that given Hermitian matrices B, C and X in $\mathcal{H}^{(\bar{n})}$, we let b, c and x denote the coordinate vectors $\operatorname{vec}_{\mathrm{H}} B$, $\operatorname{vec}_{\mathrm{H}} C$ and $\operatorname{vec}_{\mathrm{H}} X$ in $\Re^n$, where $n = \dim \mathcal{H}^{(\bar{n})}$. In this way, we can rewrite (1.2) as

$$\inf\{c^{\mathrm{T}} x \mid x \in (b + \mathcal{A}) \cap \mathcal{H}_+\},$$

where the decision variable is now treated as a vector instead of a Hermitian matrix.

1.3.2 Conic Convex Programming

A convex cone in $\Re^n$ is by definition a set $\mathcal{K}$ with the property that $\{0\} \in \mathcal{K}$ and

$$\alpha(\mathcal{K} \oplus \mathcal{K}) = \mathcal{K} \quad \text{for all } \alpha > 0.$$

Instead of 'convex cone', some authors prefer the name 'nonempty convex cone' for the above notion. A linear subspace of $\Re^n$ is a convex cone $\mathcal{A}$ with the property that $\mathcal{A} = -\mathcal{A}$.

The positive semidefinite cone $\mathcal{H}_+$ is an example of a convex cone. Namely, using the characterization

$$X \in \mathcal{H}_+^{(\bar{n})} \iff u^{\mathrm{H}} X u \geq 0 \quad \text{for all } u \in \mathcal{C}^{\bar{n}},$$

it follows that

$$X, Y \in \mathcal{H}_+ \Longrightarrow X + Y \in \mathcal{H}_+, \quad tX \in \mathcal{H}_+ \quad \text{for all } t > 0.$$

A conic convex program is an optimization problem for which the objective is linear and the constraint set is given by the intersection of an affine space with a convex cone. Conic convex programming problems can thus be formulated as follows:

$$\text{(P)} \quad \inf\{c^{\mathrm{T}} x \mid x \in (b + \mathcal{A}) \cap \mathcal{K}\},$$

where $\mathcal{A}$ is a linear subspace of $\Re^n$, $\mathcal{K}$ is a convex cone in $\Re^n$, c and b are vectors in $\Re^n$. Without loss of generality, it is assumed that c and b are in $\mathcal{A}$ and its orthogonal complement respectively. We denote this conic convex program by $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$. In its most general setting, the convex cone $\mathcal{K}$ is not necessarily closed, and consequently, the domain of the conic convex program may not be closed either.

By choosing $\mathcal{K} = \mathcal{H}_+$, we see that semidefinite programming is a special case of conic convex programming. In fact, Nesterov and Nemirovsky [110] pointed out that any convex programming problem can be written in the conic form.

1.4 Organization

Appendix B shows you how to formulate various classes of convex and quasiconvex functions in a semidefinite program. This modeling guide is meant to whet your appetite for the study of numerical algorithms for solving large scale semidefinite programs.

Duality and optimality criteria form the foundation of our treatment. Using duality, it is possible to get upper and lower bounds for the optimal objective value of a convex optimization problem. As such, duality theory offers practical tools for measuring the quality of approximate solutions. As will become clear in Chapter 2, there are some delicate issues in semidefinite programming duality, which demand a cautious treatment. However, in the case that primal and dual interior solutions exist, the same kind of duality relations holds true as in linear programming.

With the duality results in mind, we investigate the possibilities for generating primal and dual solutions such that the associated upper and lower bounds on the optimal value approach each other. We first assume in Chapter 3 that primal and dual interior feasible solutions exist, and even that such a solution pair is known. Under these assumptions, we extend some well known path–following algorithms from linear programming to semidefinite programming. We derive worst–case estimations on the global linear rate of convergence for these algorithms. These estimations generalize the polynomiality results that are known for these algorithms in the case of linear programming.

Of course, there are many situations in which interior feasible solutions are not known in advance, and indeed, it may not even be known whether such solutions exist. In Chapter 4, we investigate the possiblities of solving semidefinite programs without any pre–knowledge, by using the self–dual embedding technique. This technique enables us to embed any semidefinite program in an artificial semidefinite program, for which a pair of interior feasible solutions can be chosen arbitrarily. As a result, the treatment of Chapter 3 is applicable to this artificial program, and this yields efficient solution procedures for solving semidefinite programs without pre–knowledge. However, some nasty cases can arise here if the original semidefinite program does not have interior solutions itself, and we will examine these cases in detail.

Chapters 3 and 4 provide sufficient tools for solving semidefinite programming problems from scratch. In Chapters 5 and 6, we study the local properties of path–following in the vicinity of the optimal solution set. We will first analyze in Chapter 5 the limiting properties of the central path. In Chapter 6, it is shown that path–following algorithms can benefit from these properties by generating the iterates close enough to the central path. In particular, we obtain the remarkable result that superlinear convergence can be achieved, even if there are multiple optimal solutions.

In Chapter 7, we study the possibilities of speeding up the global convergence of primal–dual interior point methods. To this end, we leave the standard path–following framework in favor of the more general central region approach. Whereas path–following methods trace a one–dimensional curve, the iterates of central region methods follow a full dimensional set. This feature typically results in the acceptance of longer steps. Furthermore, we will improve the search directions by applying a second order centrality correction. These modifications lead to the introduction of the so–called centering–predictor–corrector algorithm. This new algorithm is also of interest in the context of linear programming, since it finally may close the notorious gap between theory and practice of interior point methods.

The material in this thesis is largely based on [85, 87, 86, 146, 149, 148, 144]. Additional and related results are contained in [11, 38, 140, 141, 147, 142, 143, 145].

Semidefinite programming is a young branch in optimization, and many of its interesting features are hardly explored yet. In Chapter 8, we give directions for further research in this challenging area.

Chapter 2

Duality

For a given solution to a semidefinite programming problem, it is easy to check its feasibility and to calculate the amount of constraint violation. But how do you check optimality or near optimality in terms of the objective value? Duality theory provides an answer to these questions. The idea is to associate the semidefinite program with a *dual* semidefinite program in such a way that upper and lower bounds on the optimal value can be calculated, once feasible solutions to the SDP problem and its dual are known.

If the optimal value p^* of (P) is finite and the infimum is attained, then an optimal solution of (P) should consist of a feasible solution x^* for which $c^{\mathrm{T}}x^* = p^*$ as well as a certificate proving the claim that p^* is really the infimum. Dual solutions can serve as certificates of this type. Similarly, if one claims that (P) is infeasible, one has to prove this claim with a certificate which can be a Farkas–type dual solution. In fact, duality theory is an indispensable tool for checking the validity of solutions to optimization problems.

To a large extent, duality results for linear programming can be generalized to the setting of conic convex programming, as was pointed out by Duffin [32]. However, certificates in the context of conic convex programming, such as those proposed in [32], can be infinitely long. More recently, Borwein and Wolkowicz [19] proposed a regularization scheme which results in certificates of finite length. We propose a dual version of their regularization scheme in Section 2.6. However, as we will see in Example 2.6, checking the feasibility (correctness) of regularized certificates can be a nontrivial task. Fortunately, the structure of regularized certificates is now well understood in the special case of semidefinite programming, due to the recent results of Ramana [124].

On account of the finite precision in computational algorithms, we are led naturally to study the properties of approximate dual solutions for $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$. We will see that while exact dual solutions provide a lower bound on the optimal value of a conic convex optimization problem, approximate dual solutions provide a lower bound for the optimal value of 'reasonably' sized (primal) solutions. Our analysis of approximate solutions follows the approach initiated by Todd and Ye [153].

We present a unified treatment of duality theory for finite dimensional conic convex programming. The results apply to the conic convex programming problem in its most general form, in the sense that there are no such restrictions as closedness, pointedness, solidness, or constraint qualifications. We carefully survey some existing results known for $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$, and show that various duality results that were previously known only for the case of closed, pointed and / or solid convex cones can be extended to this general setting. Many new and interesting proofs and examples make the survey self–contained.

2.1 Terminology and Preliminaries

Given a convex cone $\mathcal{K}$, we let $\mathcal{K}^*$ denote the corresponding dual cone, i.e.

$$\mathcal{K}^* := \{z \in \Re^n \mid z^{\mathrm{T}}\mathcal{K} \subseteq \Re_+\}.$$

The cone $-\mathcal{K}^*$ is also known as the polar cone of $\mathcal{K}$ [132]. It is easily verified that the dual cone $\mathcal{K}^*$ is convex and closed. We let $\operatorname{sub}\mathcal{K}$ denote the largest linear subspace that is contained in $\mathcal{K}$, i.e.

$$\operatorname{sub}\mathcal{K} := \mathcal{K} \cap (-\mathcal{K}).$$

Notice that

$$\operatorname{sub}\mathcal{K}^* = \{z \in \Re^n \mid z^{\mathrm{T}}\mathcal{K} = \{0\}\};$$

we define $\mathcal{K}^\perp := \operatorname{sub}\mathcal{K}^*$. A convex cone $\mathcal{K}$ is said to be *pointed* if $\operatorname{sub}\mathcal{K} = \{0\}$; $\mathcal{K}$ is said to be *solid* if $\operatorname{int}\mathcal{K} \neq \emptyset$. The linear subspace that is spanned by elements of $\mathcal{K}$ is

$$\operatorname{span}\mathcal{K} := \mathcal{K} \oplus -\mathcal{K}.$$

Example 2.1 *Let* $\mathcal{K} = \Re \times \Re_+ \times \{0\}$, *then* $\mathcal{K}^* = \{0\} \times \Re_+ \times \Re$, $\operatorname{span}\mathcal{K} = \Re^2 \times \{0\}$, $\operatorname{sub}\mathcal{K} = \Re \times \{0\} \times \{0\}$ *and* $\mathcal{K}^\perp = \{0\} \times \{0\} \times \Re$.

Consider now problem (P), i.e. the conic convex program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$. If $\mathcal{K}$ is closed, we say that (P) is a *closed conic convex program.* The set of feasible solutions of (P) is

$$\mathcal{F}_P := (b + \mathcal{A}) \cap \mathcal{K}.$$

It is easy to see that

$$\mathcal{F}_P = \mathcal{F}_P \oplus (\mathcal{A} \cap \mathcal{K}). \tag{2.1}$$

If $x \in \mathcal{A} \cap \mathcal{K}$, then x is called a *direction* (or a recession direction in the terminology of Rockafellar [132]); x is an *interior direction* if it belongs to $\mathcal{A} \cap \operatorname{rel}\mathcal{K}$. If x is a direction and $-x$ is not, i.e. $x \in (\mathcal{A} \cap \mathcal{K}) \setminus \text{ sub } (\mathcal{A} \cap \mathcal{K})$, then x is a *one–sided direction.* If $x \in \mathcal{A} \cap \mathcal{K}$ is such that $c^{\mathrm{T}} x \leq 0$, then x is a *lower level direction*; such x is a *one–sided lower level direction* if $-x$ is not a lower level direction. An *improving direction* is a direction x with $c^{\mathrm{T}} x < 0$. An *improving direction sequence* is a sequence $x^{(1)}, x^{(2)}, \ldots$ in $\mathcal{K}$ such that

$$\limsup_{i \to \infty} c^{\mathrm{T}} x^{(i)} < 0, \quad \lim_{i \to \infty} \text{ dist}(x^{(i)}, \mathcal{A}) = 0.$$

Notice that if there exists an improving direction, then there certainly exists an improving direction sequence. We will see in Section 2.4 that the converse is in general not true.

If sub $(\mathcal{A} \cap \mathcal{K}) \neq \{0\}$, it is often convenient to consider only solutions in $(\text{ sub } (\mathcal{A} \cap \mathcal{K}))^{\perp}$. Namely, it follows from (2.1) that

$$\mathcal{F}_P = \left(\mathcal{F}_P \cap (\text{ sub } (\mathcal{A} \cap \mathcal{K}))^{\perp} \right) \oplus \text{ sub } (\mathcal{A} \cap \mathcal{K}).$$

Based on this observation, we say that $x \in \Re^n$ is a *normalized feasible solution* of (P) if

$$x \in \mathcal{F}_P \cap (\text{ sub } (\mathcal{A} \cap \mathcal{K}))^{\perp}.$$

Obviously, if $\mathcal{A} \cap \mathcal{K}$ is pointed, then any feasible solution is a normalized feasible solution.

Example 2.2 *The standard LP problem*

$$\min\{\tilde{c}^T y \mid A^T y + z = \tilde{b},\, y \in \Re^m,\, z \geq 0\},$$

where A is an $m \times n$ matrix, can be cast as a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$ in $\Re^{m+n}$ by letting

$$b^T := \left[\begin{array}{cc} 0, & \tilde{b}^T \end{array}\right]^T, \quad c^T = \left[\begin{array}{cc} \tilde{c}^T, & 0 \end{array}\right]^T,$$

$$\mathcal{A} := \text{ Ker } \left[\begin{array}{cc} A^T, & I \end{array}\right], \quad \mathcal{K} := \Re^m \times \Re^n_+,$$

where I denotes the identity matrix of order n. Since in this case there holds

$$\text{sub } (\mathcal{A} \cap \mathcal{K}) = (\text{ Ker } A^T) \times \{0\}^n,$$

the normalized feasible set is

$$\mathcal{F}_P \cap (\text{ sub } (\mathcal{A} \cap \mathcal{K}))^{\perp} = \{(y, z) \in \mathcal{F}_P \mid y \in \text{ Img } A\}.$$

It is customary in linear programming theory to assume that A has full row rank, i.e. Img $A = \Re^m$*, which implies that $\mathcal{F}_P$ consists only of normalized feasible solutions.*

The set of *interior solutions* is defined as

$$\overset{o}{\mathcal{F}}_P := \mathcal{F}_P \cap \text{ rel } \mathcal{K}.$$

We say that (P) is *feasible* (or consistent) if $\mathcal{F}_P \neq \emptyset$ and (P) is *strongly feasible* (or super–consistent in the terminology of Duffin [32]) if $\overset{o}{\mathcal{F}}_P \neq \emptyset$. If (P) is feasible but not strongly feasible, then (P) is said to be *weakly feasible*.

Strong feasibility as defined above is also known as the generalized Slater's constraint qualification.

Obviously, if (P) is feasible, i.e. if $(b+\mathcal{A})\cap\mathcal{K}\neq\emptyset$, then $\operatorname{dist}(b+\mathcal{A},\mathcal{K})=0$. The converse is in general not true, even if $\mathcal{K}$ is closed; see Example 2.3 at the end of this section. This observation gives rise to the definition of *weak infeasibility*, which is sometimes referred to as sub–consistency [32] or asymptotic consistency [12]. Problem (P) is said to be *weakly infeasible* if

$$\operatorname{dist}(b+\mathcal{A},\mathcal{K})=0 \quad \text{but} \quad \mathcal{F}_P=\emptyset.$$

If

$$\operatorname{dist}(b+\mathcal{A},\mathcal{K})>0,$$

then (P) is called *strongly infeasible*.

Let

$$p^* := \inf c^{\mathrm{T}}\mathcal{F}_P$$

denote the optimal value of (P). The set of feasible solutions for which the optimal value is attained is

$$\mathcal{F}_P^* := \{x\in\mathcal{F}_P \mid c^{\mathrm{T}}x=p^*\},$$

and the *normalized optimal set* is

$$\mathcal{F}_P^* \cap (\text{ sub } (\mathcal{A}\cap\mathcal{K}))^{\perp}.$$

Problem (P) is said to be *solvable* (or convergent in the terminology of Duffin [32]) if $\mathcal{F}_P^*\neq\emptyset$. A special case of unsolvability occurs when $p^*=-\infty$. In this case, we say that (P) is *unbounded*. Notice that if (P) is feasible and there exists an improving direction, then (P) is unbounded.

Associated with (P) is a *dual program* (D), viz.

$$\begin{aligned} \inf \quad & b^{\mathrm{T}}z \\ \text{s.t.} \quad & z-c\in\mathcal{A}^{\perp} \\ & z\in\mathcal{K}^*. \end{aligned} \tag{D}$$

In other words, the dual of the conic convex program $\mathrm{CP}(b,c,\mathcal{A},\mathcal{K})$ is by definition the closed conic convex program $\mathrm{CP}(c,b,\mathcal{A}^{\perp},\mathcal{K}^*)$. In analogy to the definitions of $\mathcal{F}_P$, $\overset{o}{\mathcal{F}}_P$, p^* and $\mathcal{F}_P^*$ for the primal program, we define

$$\mathcal{F}_D := (c+\mathcal{A}^{\perp})\cap\mathcal{K}^*, \quad \overset{o}{\mathcal{F}}_D := (c+\mathcal{A}^{\perp})\cap \text{ rel } \mathcal{K}^*,$$

and

$$d^* := \inf b^{\mathrm{T}}\mathcal{F}_D, \quad \mathcal{F}_D^* := \{z \in \mathcal{F}_D \mid b^{\mathrm{T}} z = d^*\},$$

for the dual program. If (D) is weakly (in)feasible, strongly (in)feasible or solvable, then (P) is said to be *dual* weakly (in)feasible, dual strongly (in)feasible or dual solvable, respectively. Similarly, (P) is said to have a dual level direction, dual improving direction, etc., if (D) has a level direction, improving direction, and so on. When we discuss the bipolar theorem (Theorem 2.1), we will see that if $\mathcal{K}$ is closed, then the dual of (D) is again (P).

Notice that if $x \in \mathcal{F}_P$ and $z \in \mathcal{F}_D$, then

$$x^{\mathrm{T}} z = b^{\mathrm{T}} z + (x-b)^{\mathrm{T}} z = b^{\mathrm{T}} z + c^{\mathrm{T}} x, \tag{2.2}$$

where we used the fact that $b \in \mathcal{A}^{\perp}$ and $c \in \mathcal{A}$. The quantity $x^{\mathrm{T}} z$ is called the *duality gap* at the feasible solution pair (x, z).

The following example, a semidefinite programming problem, illustrates some of the terminology introduced above.

Example 2.3 *Consider the program CP$(b, c, \mathcal{A}, \mathcal{K})$ in $\Re^3$, with*

$$b = \begin{bmatrix} 0, & 0, & 1 \end{bmatrix}^T, \quad c = \begin{bmatrix} 0, & c_2, & 0 \end{bmatrix}^T,$$

$$\mathcal{A} = \{x \in \Re^3 \mid x_1 = 0, x_3 = 0\},$$

and

$$\mathcal{K} = \left\{ x \in \Re^3 \;\middle|\; \begin{bmatrix} x_1 & x_3/\sqrt{2} \\ x_3/\sqrt{2} & x_2 \end{bmatrix} \succeq 0 \right\}.$$

Then $\mathcal{K}^* = \mathcal{K}$ and $\mathcal{A}^{\perp} = \{z \in \Re^3 \mid z_2 = 0\}$. The primal is weakly infeasible,

$$p^* = \inf \left\{ c_2 x_2 \;\middle|\; \begin{bmatrix} 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & x_2 \end{bmatrix} \succeq 0 \right\} = \infty.$$

The dual is

$$d^* = \inf \left\{ z_3 \;\middle|\; \begin{bmatrix} z_1 & z_3/\sqrt{2} \\ z_3/\sqrt{2} & c_2 \end{bmatrix} \succeq 0 \right\}.$$

Hence, the dual is strongly infeasible if $c_2 < 0$, weakly feasible and solvable with optimal value $d^* = 0$ if $c_2 = 0$, and strongly feasible and unbounded if $c_2 > 0$.

2.2 Basic Properties of Convex Cones

The result below is quoted from Corollary 16.4.2 of Rockafellar [132]. We give a direct proof for completeness.

Lemma 2.1 *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be two convex cones, then*

$$\mathcal{K}_1^* \cap \mathcal{K}_2^* = (\mathcal{K}_1 \oplus \mathcal{K}_2)^*.$$

Proof. By definition, we have

$$x \in (\mathcal{K}_1 \oplus \mathcal{K}_2)^*$$

if and only if

$$x^{\mathrm{T}}(\mathcal{K}_1 \oplus \mathcal{K}_2) \subseteq \Re_+.$$

Since $0 \in \mathcal{K}_1 \cap \mathcal{K}_2$, the above relation is equivalent with

$$x^{\mathrm{T}}\mathcal{K}_1 \subseteq \Re_+, \quad x^{\mathrm{T}}\mathcal{K}_2 \subseteq \Re_+,$$

i.e. $x \in \mathcal{K}_1^* \cap \mathcal{K}_2^*$. **Q.E.D.**

Based on Lemma 2.1, one may guess that the cones $\mathcal{K}_1^* \oplus \mathcal{K}_2^*$ and $(\mathcal{K}_1 \cap \mathcal{K}_2)^*$ are identical. However, this is in general not true even if $\mathcal{K}_1$ and $\mathcal{K}_2$ are both closed, since the Minkowski sum $\mathcal{K}_1^* \oplus \mathcal{K}_2^*$ may not be closed. For instance in Example 2.3, we have $\begin{bmatrix} 0, & 0, & 1 \end{bmatrix}^{\mathrm{T}} \in \mathrm{cl}\,(\mathcal{A} \oplus \mathcal{K})$, but $\begin{bmatrix} 0, & 0, & 1 \end{bmatrix}^{\mathrm{T}} \notin \mathcal{A} \oplus \mathcal{K}$.

Corollary 2.1 *Let $\mathcal{K}$ be a convex cone. Then* $\mathrm{span}\,\mathcal{K} = (\,\mathrm{sub}\,\mathcal{K}^*)^\perp = \mathcal{K}^{\perp\perp}$, *i.e.* $\mathrm{span}\,\mathcal{K}$ *is the smallest linear subspace containing $\mathcal{K}$.*

Proof. Apply Lemma 2.1 with $\mathcal{K}_1 = \mathcal{K}$ and $\mathcal{K}_2 = -\mathcal{K}$. **Q.E.D.**

Recall that a solid convex cone in $\Re^n$ is by definition a convex cone $\mathcal{K}$ for which $\mathrm{int}\,\mathcal{K} \neq \emptyset$, or equivalently, for which the smallest subspace containing $\mathcal{K}$ is $\Re^n$. Hence, we obtain from Corollary 2.1 that $\mathcal{K}$ is solid if and only if $\mathcal{K}^*$ is pointed. Notice however, that if $\mathcal{K}$ is not closed then $\mathcal{K}$ may be pointed whereas $\mathcal{K}^*$ is not solid. (For instance, consider $\mathcal{K} = (\Re \times \Re_{++}) \cup \{0\}$.)

Notice from Lemma 2.1 and Corollary 2.1 that

$$\begin{aligned} \left(\,\mathrm{sub}\,(\mathcal{A}^\perp \cap \mathcal{K}^*)\right)^\perp &= (\,\mathrm{sub}\,(\mathcal{A}^* \cap \mathcal{K}^*))^\perp \\ &= (\,\mathrm{sub}\,(\mathcal{A} \oplus \mathcal{K})^*)^\perp \\ &= \mathrm{span}\,(\mathcal{A} \oplus \mathcal{K}). \end{aligned}$$

Hence, it follows that

$$\mathcal{F}_D \cap \mathrm{span}\,(\mathcal{A} \oplus \mathcal{K})$$

is the normalized dual feasible set.

A well known result is the *bipolar theorem* (see Duffin [32], Ben-Israel [7] and Rockafellar [132], among others).

Theorem 2.1 (bipolar theorem) *Let $\mathcal{K}$ be a convex cone in $\Re^n$. There holds*

$$\mathrm{cl}\,\mathcal{K} = \mathcal{K}^{**}.$$

The bipolar theorem shows the nice symmetry between the dual pair (P) and (D): if $\mathcal{K}$ is closed, then $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ is the dual of the conic convex program $\mathrm{CP}(c, b, \mathcal{A}^\perp, \mathcal{K}^*)$.

The bipolar theorem gives a dual characterization of $\mathrm{cl}\,\mathcal{K}$. Theorem 2.2 below gives a dual characterization of $\mathrm{rel}\,\mathcal{K}$. To the best of the author's knowledge, this characterization is new.

Theorem 2.2 *Let $\mathcal{K}$ be a convex cone in $\Re^n$. Then*

$$x \in \mathrm{rel}\,\mathcal{K}$$

if and only if

$$x \in \mathrm{span}\,\mathcal{K}, \quad x^T(\mathcal{K}^* \setminus \mathcal{K}^\perp) \subseteq \Re_{++}.$$

Proof. Let z be an arbitrary nonzero vector in $(\mathcal{K}^* \setminus \mathcal{K}^\perp)$, and let $\hat{z}$ denote the nonzero orthogonal projection of z onto the subspace $K^\perp$. By definition, $x \in \mathrm{rel}\,\mathcal{K}$ implies that there exists a positive number $\delta(z)$ such that $x - \delta(z)\hat{z} \in \mathcal{K}$. This yields

$$0 \le z^{\mathrm{T}}(x - \delta(z)\hat{z}) = z^{\mathrm{T}}x - \delta(z)\|\hat{z}\|_2^2 < z^{\mathrm{T}}x.$$

Moreover, since $\mathrm{rel}\,\mathcal{K} \subseteq \mathrm{span}\,\mathcal{K}$, we have $x \in \mathrm{span}\,\mathcal{K}$.

Conversely, suppose that $x \in \mathrm{span}\,\mathcal{K}$ is such that $x^{\mathrm{T}}(\mathcal{K}^* \setminus \mathcal{K}^\perp) \subseteq \Re_{++}$. Since $x^{\mathrm{T}}\mathcal{K}^\perp = \{0\}$ (see Corollary 2.1), it follows that $x^{\mathrm{T}}\mathcal{K}^* \subseteq \Re_+$, i.e. $x \in \mathcal{K}^{**}$. Let

$$\epsilon := \inf_z \{x^{\mathrm{T}}z \mid z \in \mathcal{K}^* \cap \mathrm{span}\,\mathcal{K},\ \|z\| = 1\}.$$

Then $\epsilon > 0$, because $\mathcal{K}^*$ and $\mathrm{span}\,\mathcal{K}$ are closed. By construction, we have for all $y \in \mathrm{span}\,\mathcal{K}$, $y \neq 0$, and $z \in \mathcal{K}^*$ that

$$z^{\mathrm{T}}\left(x + \frac{\epsilon}{\|y\|_*}y\right) \ge 0,$$

which implies that $x \in \mathrm{rel}\,\mathcal{K}^{**}$, Using the bipolar theorem, it follows that $x \in \mathrm{rel}\,\mathcal{K}$. **Q.E.D.**

The following lemma gives a formula for the relative interior of a Minkowski sum of cones. It follows from Corollary 6.1.1 in Rockafellar [132], but we give a direct proof for completeness.

Lemma 2.2 *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be convex cones in $\Re^n$. Then*

$$\text{rel}\,(\mathcal{K}_1 \oplus \mathcal{K}_2) = (\text{ rel}\,\mathcal{K}_1) \oplus \text{ rel}\,\mathcal{K}_2.$$

Proof. From Corollary 2.1, we know that span $(\mathcal{K}_1 \oplus \mathcal{K}_2)$ is the smallest linear subspace containing $\mathcal{K}_1 \oplus \mathcal{K}_2$. Since

$$\text{span}\,(\mathcal{K}_1 \oplus \mathcal{K}_2) = (\text{ span}\,\mathcal{K}_1) \oplus \text{ span}\,\mathcal{K}_2,$$

it follows that

$$x \in (\text{ rel}\,\mathcal{K}_1) \oplus \text{ rel}\,\mathcal{K}_2 \;\Rightarrow\; x \in \text{ rel}\,(\mathcal{K}_1 \oplus \mathcal{K}_2). \tag{2.3}$$

On the other hand, we have cl rel $\mathcal{K}_1 =$ cl $\mathcal{K}_1$ and cl rel $\mathcal{K}_2 =$ cl $\mathcal{K}_2$ because $\mathcal{K}_1$ and $\mathcal{K}_2$ are convex, and hence

$$\mathcal{K}_1 \oplus \mathcal{K}_2 \subseteq \text{ cl}\,((\text{ rel}\,\mathcal{K}_1) \oplus \text{ rel}\,\mathcal{K}_2), \tag{2.4}$$

where we used the fact that the closure of a set is the union of that set with its limit points. Relation (2.4) implies that

$$\text{rel}\,(\mathcal{K}_1 \oplus \mathcal{K}_2) \subseteq (\text{ rel}\,\mathcal{K}_1) \oplus \text{ rel}\,\mathcal{K}_2. \tag{2.5}$$

Combining (2.3) and (2.5) yields

$$(\text{ rel}\,\mathcal{K}_1) \oplus \text{ rel}\,\mathcal{K}_2 = \text{ rel}\,(\mathcal{K}_1 \oplus \mathcal{K}_2).$$

Q.E.D.

In fact, Lemma 2.2 holds not only for convex cones but also for more general sets known as *robust sets*. (A set $\mathcal{S}$ is said to be robust if it satisfies cl rel $\mathcal{S} =$ cl $\mathcal{S}$.) This fact can be shown by the same proof as used in Lemma 2.2.

The following lemma shows how an invertible linear transformation of a cone affects its dual.

Lemma 2.3 *Let $\mathcal{K}$ be a convex cone in $\Re^n$ and let $M \in \Re^{n\times n}$ be an invertible matrix. Then*

$$(M^T\mathcal{K})^* = M^{-1}\mathcal{K}^*.$$

Proof. We note the following relations

$$\begin{aligned} y \in (M^{\mathrm{T}}\mathcal{K})^* &\iff y^{\mathrm{T}}M^{\mathrm{T}}\mathcal{K} \subseteq \Re_+ \\ &\iff My \in \mathcal{K}^* \\ &\iff y \in M^{-1}\mathcal{K}^*. \end{aligned}$$

Q.E.D.

2.3 Characterization of Strong Feasibility

Combining Lemma 2.1, Theorem 2.2, and Lemma 2.2, we obtain the following result.

Theorem 2.3 *There exists a primal interior solution if and only if there exists no one–sided dual level direction.*

Proof. By definition, a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$ has an interior solution if and only if $(b + \mathcal{A}) \cap \text{ rel } \mathcal{K} \neq \emptyset$, i.e.

$$b \in \mathcal{A} \oplus \text{ rel } \mathcal{K} = \text{ rel } (\mathcal{A} \oplus \mathcal{K}),$$

where we used Lemma 2.2. From Theorem 2.2, we know that the above relation holds if and only if

$$b \in \text{ span } (\mathcal{A} \oplus \mathcal{K}), \quad b^{\mathrm{T}}((\mathcal{A} \oplus \mathcal{K})^* \setminus (\mathcal{A} \oplus \mathcal{K})^{\perp}) \subseteq \Re_{++},$$

which, using Lemma 2.1, is equivalent with

$$b^{\mathrm{T}}(\mathcal{A}^{\perp} \cap \mathcal{K}^*) = \{0\}, \quad b^{\mathrm{T}}((\mathcal{A}^{\perp} \cap \mathcal{K}^*) \setminus \text{ sub } (\mathcal{A}^{\perp} \cap \mathcal{K}^*)) \subseteq \Re_{++},$$

i.e. there exist no one–sided dual level directions. **Q.E.D.**

The above characterization of strong feasibility was established by Carver [23] for the case that $\mathcal{K} = \Re^n_+$. For general solid closed convex cones, the result can be found in Fan [36], Duffin [32], and Berman and Ben-Israel [13]. Notice however, that Theorem 2.3 above is applicable also if $\mathcal{K}$ is not solid.

Special cases of Theorem 2.3 are the arbitrage and pricing result in the theory of financial markets [57] and well known theorems of Lyapunov, Stein and Taussky in matrix theory (see the discussion in Berman and Ben-Israel [13] and Berman [12]). Applying Theorem 2.3 to the conic convex program CP$(0, 0, \mathcal{A}, \mathcal{K})$ yields a characterization of the existence of primal interior directions:

Corollary 2.2 *Consider a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$. There exists a primal interior direction, i.e.*

$$\mathcal{A} \cap \text{ rel } \mathcal{K} \neq \emptyset$$

if and only if there is no one–sided dual direction.

If $\mathcal{K} = \Re^n_+$ (the polyhedral case), Corollary 2.2 reduces to a classical result of Gordan [49] and Stiemke [139].

Combining Theorem 2.3 and Corollary 2.2, it follows that if there exists an interior direction ($\mathcal{A} \cap \text{ rel } \mathcal{K} \neq \emptyset$), then there must also exist an interior solution ($\overset{o}{\mathcal{F}}_P \neq \emptyset$).

Based on Theorem 2.3, we derive a characterization of the existence of *improving* interior directions:

Corollary 2.3 *Consider a conic convex program CP($b, c, \mathcal{A}, \mathcal{K}$). There exists a primal improving interior direction, i.e.*

$$c^T(\mathcal{A} \cap \text{ rel } \mathcal{K}) \not\subseteq \Re_+$$

if and only if the dual is infeasible and there is no one–sided dual direction.

Proof. Notice that if $c = 0$ then there exist no primal improving directions, and the dual has a feasible solution, viz. $0 \in \mathcal{A}^\perp \cap \mathcal{K}^*$.

Suppose now that $c \neq 0$. Below, we will construct an artificial conic convex program, for which the interior solutions correspond to primal improving directions of the original program. The corollary will then follow as an application of Theorem 2.3. First, since $c \in \mathcal{A}$, there holds

$$-\frac{c}{\|c\|_2^2} + (\mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}) = \{x \in \mathcal{A} \mid c^{\mathrm{T}} x = -1\}.$$

Hence, x is a primal improving interior direction if and only if there exists some $\alpha > 0$ such that

$$\alpha x \in \left(-\frac{c}{\|c\|_2^2} + (\mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}) \right) \cap \text{ rel } \mathcal{K}.$$

Applying Theorem 2.3, it follows that there exist primal improving interior directions if and only if the conic convex program $\mathrm{CP}(-c/\|c\|_2^2, 0, \mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}, \mathcal{K})$ has no one–sided dual level directions. The dual of $\mathrm{CP}(-c/\|c\|_2^2, 0, \mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}, \mathcal{K})$ is $\mathrm{CP}(0, -c/\|c\|_2^2, \mathcal{A}^\perp \oplus \text{ Img } c, \mathcal{K}^*)$, and if it has a one–sided level direction s, it must be contained in $(\mathcal{A}^\perp \oplus \text{ Img } c) \cap \mathcal{K}^*$. Since $c \in \mathcal{A}$, it follows that either $c^{\mathrm{T}} z > 0$ and there is $\alpha > 0$ such that $\alpha z \in (c + \mathcal{A}^\perp) \cap \mathcal{K}$, or $c^{\mathrm{T}} z = 0$ and

$$z \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus -\mathcal{K}^*.$$

Hence, $\mathrm{CP}(-c/\|c\|_2^2, 0, \mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}, \mathcal{K})$ has no one–sided dual level directions if and only if $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ is dual infeasible and has no one–sided dual directions. **Q.E.D.**

We remark that Nesterov, Todd and Ye [114] called a closed conic convex program *strictly infeasible* if it has a dual improving interior direction. Corollary 2.3 shows that a program is strictly infeasible in the sense of [114] if and only if it is infeasible and has no one–sided directions. We will see in Corollary 2.4 that strict infeasibility implies strong infeasibility.

The relation between dual directions and primal strong feasibility has now been fully investigated. We now proceed to study the relationships between dual directions and boundedness of the dual feasible set, the dual lower level sets and the dual optimal set.

Lemma 2.4 *Consider a conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$ for which the primal is strongly feasible. Let $z^{(1)}, z^{(2)}, \dots$ in $\mathcal{K}^* \cap \text{ span } (\mathcal{A} \oplus \mathcal{K})$ be a sequence with*

$$\lim_{i\to\infty} \text{ dist}(z^{(i)}, c + \mathcal{A}^\perp) = 0, \quad \limsup_{i\to\infty} b^T z^{(i)} < \infty.$$

Then $z^{(i)}$, $i = 1, 2, \dots$, is a bounded sequence.

Proof. Suppose to the contrary that $\mathcal{K}^* \cap \text{ span } (\mathcal{A} \oplus \mathcal{K})$ contains some sequence $z^{(1)}, z^{(2)}, \dots$ such that

$$\lim_{i\to\infty} \left\| z^{(i)} \right\| = \infty, \tag{2.6}$$

whereas $\lim_{i\to\infty} \text{ dist}(z^{(i)}, c + \mathcal{A}^\perp) = 0$ and $\limsup_{i\to\infty} b^{\mathrm{T}} z^{(i)} < \infty$.

Without loss of generality, we assume that $\left\| z^{(i)} \right\| > 0$ for all i and that the limit

$$y := \lim_{i\to\infty} \frac{z^{(i)}}{\|z^{(i)}\|}$$

exists. Since the sequence $z^{(i)} / \left\| z^{(i)} \right\|$, $i = 1, 2, \dots$, is contained in the closed cone $\mathcal{K}^* \cap \text{ span } (\mathcal{A} \oplus \mathcal{K})$, it follows that

$$y \in \mathcal{K}^* \cap \text{ span } (\mathcal{A} \oplus \mathcal{K}) = \mathcal{K}^* \cap (\text{ sub } (\mathcal{A}^\perp \cap \mathcal{K}^*))^\perp, \tag{2.7}$$

where we used Lemma 2.1. Moreover, using (2.6) we have

$$y = y - \lim_{i\to\infty} \frac{1}{\|z^{(i)}\|} c = \lim_{i\to\infty} \frac{z^{(i)} - c}{\|z^{(i)}\|} \in \mathcal{A}^\perp, \tag{2.8}$$

and, since $\limsup_{i\to\infty} b^{\mathrm{T}} z^{(i)} < \infty$,

$$b^{\mathrm{T}} y = \lim_{i\to\infty} \frac{b^{\mathrm{T}} z^{(i)}}{\|z^{(i)}\|} = 0. \tag{2.9}$$

By construction, $\|y\| = 1$, so that (2.7)—(2.9) implies

$$y \in (\mathcal{A}^\perp \cap \mathcal{K}^*) \setminus -\mathcal{K}^*, \quad b^{\mathrm{T}} y = 0,$$

i.e. y is a one–sided lower level direction, which contradicts the primal strong feasibility (see Theorem 2.3). **Q.E.D.**

Theorem 2.4 *Consider a conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$. If the dual is weakly infeasible then the primal has no interior direction.*

Proof. Apply Lemma 2.4 to the dual feasibility problem $\mathrm{CP}(0, c, \mathcal{A}, \mathcal{K})$. **Q.E.D.**

Combining Theorem 2.4 with the dual characterization of (primal) interior directions (Corollary 2.2) yields the following result.

Corollary 2.4 *Consider a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$. If the dual is weakly infeasible then the dual has a one–sided direction.*

Notice that Corollary 2.4 is stated in terms of the dual program, since the closedness of $\mathcal{K}^*$ is essential for this result.

Theorem 2.5 *A dual feasible conic convex program is primal strongly feasible if and only if the normalized dual optimal set is nonempty and bounded.*

Proof. If the normalized dual optimal set is nonempty and bounded, then there is obviously no one–sided dual level direction. Using Theorem 2.3, this implies that $\overset{o}{\mathcal{F}}_P \neq \emptyset$. Conversely, we know from Lemma 2.4 that if $\overset{o}{\mathcal{F}}_P \neq \emptyset$ then any sequence $z^{(1)}, z^{(2)}, \ldots$ of normalized dual feasible solutions with $\lim_{i\to\infty} b^{\mathrm{T}} z^{(i)} = d^*$ is bounded. Since $\mathcal{F}_D$ is nonempty and closed, it follows that the normalized dual optimal set is nonempty and bounded. **Q.E.D.**

Corollary 2.5 *Consider a dual feasible conic convex program. The normalized dual feasible set is bounded if and only if there exists a primal interior direction.*

Proof. Apply Theorem 2.5 to the dual feasibility problem $\mathrm{CP}(0, c, \mathcal{A}, \mathcal{K})$. **Q.E.D.**

2.4 Farkas–Type Lemmas

In the previous section, we have discussed a dual characterization of strong feasibility. We will now give a characterization of strong infeasibility.

Lemma 2.5 (First Farkas–type lemma) *Consider a conic convex programming problem CP$(b, c, \mathcal{A}, \mathcal{K})$. The primal is strongly infeasible if and only if there exists a dual improving direction.*

Proof. By definition, the primal is not strongly infeasible if and only if $\mathrm{dist}(b+\mathcal{A},\mathcal{K}) = 0$. Since $\mathrm{dist}(b+\mathcal{A},\mathcal{K}) = 0$ if and only if there exists a sequence $x^{(1)}, x^{(2)}, \ldots$ in $\mathcal{K}$ such that

$$\lim_{i\to\infty} P_{\mathcal{A}^\perp} x^{(i)} = b,$$

we obtain the relation

$$\mathrm{dist}(b+\mathcal{A},\mathcal{K}) = 0 \Longleftrightarrow b \in \ \mathrm{cl}\ P_{\mathcal{A}^\perp}\mathcal{K}. \tag{2.10}$$

It is easy to see that a linearly transformed convex cone is also a convex cone. Therefore, we can apply the bipolar theorem, which states that

$$\mathrm{cl}\ P_{\mathcal{A}^\perp}\mathcal{K} = (P_{\mathcal{A}^\perp}\mathcal{K})^{**}. \tag{2.11}$$

Combining the relation (2.10)–(2.11) yields

$$\mathrm{dist}(b+\mathcal{A},\mathcal{K}) = 0 \Longleftrightarrow b^{\mathrm{T}}(P_{\mathcal{A}^\perp}\mathcal{K})^* \subseteq \Re_+, \tag{2.12}$$

where

$$\begin{aligned}(P_{\mathcal{A}^\perp}\mathcal{K})^* &= \{\sigma \in \Re^n \mid \sigma^{\mathrm{T}} P_{\mathcal{A}^\perp}\mathcal{K} \subseteq \Re_+\}\\ &= \{\sigma \in \Re^n \mid P_{\mathcal{A}^\perp}\sigma \in \mathcal{K}^*\}\\ &= (\mathcal{K}^* \cap \mathcal{A}^\perp) + \mathcal{A}.\end{aligned}$$

Combining the above relation with (2.12) and noting $b \in \mathcal{A}^\perp$, we obtain

$$\mathrm{dist}(b+\mathcal{A},\mathcal{K}) = 0 \Longleftrightarrow b^{\mathrm{T}}(\mathcal{K}^* \cap \mathcal{A}^\perp) \subseteq \Re_+.$$

Q.E.D.

For the case that $\mathcal{K} = \Re^n_+$, Lemma 2.5 reduces to the famous lemma of Farkas [37]. For general closed convex programming, the result has been established by Duffin [32] and Berman [12].

Applying Lemma 2.5 to the conic convex program $\mathrm{CP}(c, b, \mathcal{A}^\perp, \mathcal{K}^*)$, we see that

$$\mathrm{dist}(c+\mathcal{A}^\perp,\mathcal{K}^*) = 0 \iff c^{\mathrm{T}}(\mathcal{A} \cap \mathcal{K}^{**}) \subseteq \Re_+. \tag{2.13}$$

However, from the bipolar theorem we have $\mathcal{K}^{**} = \ \mathrm{cl}\ \mathcal{K}$. This together with Theorem 2.4 leads to the following characterization of feasibility for conic convex programs satisfying a generalized Slater condition.

Corollary 2.6 *Consider a conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$ with $\mathcal{A} \cap \ \mathrm{rel}\,\mathcal{K} \neq \emptyset$. There holds*

$$\mathcal{F}_D \neq \emptyset$$

if and only if

$$c^T(\mathcal{A} \cap \mathcal{K}) \subseteq \Re_+.$$

Proof. Since $\mathcal{A} \cap \text{ rel } \mathcal{K} \neq \emptyset$, we have

$$\mathcal{A} \cap \text{ cl } \mathcal{K} = \text{ cl } (\mathcal{A} \cap \mathcal{K}).$$

Hence, we can replace $\mathcal{K}^{**}$ with $\mathcal{K}$ in relation (2.13). Moreover, we know from Theorem 2.4 that (D) cannot be weakly infeasible. The corollary thus follows from relation (2.13). **Q.E.D.**

The result of Corollary 2.6 is due to Wolkowicz [160]. For the special case that $\mathcal{K}$ is closed and pointed, Corollary 2.6 reduces to a generalization of Farkas' lemma as it can be found in many papers, including [1, 7, 12, 13, 27, 156].

Naturally, we are also interested in a characterization of feasibility without a Slater-type condition. We can easily obtain such a characterization from Lemma 2.5.

Lemma 2.6 (Second Farkas–type lemma) *A conic convex program is dual feasible if and only if there does not exist any primal improving direction sequence.*

Proof. Consider a conic convex programming problem $\text{CP}(b, c, \mathcal{A}, \mathcal{K})$. Notice that if $c = 0$ then $0 \in \mathcal{F}_D$ and $c^{\mathrm{T}}\mathcal{K} = \{0\}$. In other words, we have dual feasibility and no primal improving direction sequence if $c = 0$. It remains to consider the case that $c \neq 0$.

Suppose $c \neq 0$. We use a similar technique as in the proof of Corollary 2.3, namely we will construct an artificial conic convex program, for which the dual improving directions correspond to dual feasible solutions of the original program. The corollary will then follow from Lemma 2.5. First, since $c \in \mathcal{A}$, there holds

$$c^{\mathrm{T}} z > 0 \quad \text{and} \quad z \in (\mathcal{A}^{\perp} \oplus \text{ Img } c) \cap \mathcal{K}^{*}$$

if and only if

$$\alpha z \in (c + \mathcal{A}^{\perp}) \cap \mathcal{K}^{*} \quad \text{ for some } \alpha > 0.$$

We conclude that the dual feasible set $(c + \mathcal{A}^{\perp}) \cap \mathcal{K}^{*}$ is nonempty if and only if the conic convex program $\text{CP}(-c/\|c\|_2^2, 0, \left(\mathcal{A}^{\perp} \oplus \text{ Img } c\right)^{\perp}, \mathcal{K})$ has a dual improving direction. Applying Lemma 2.5, it follows that $\mathcal{F}_D \neq \emptyset$ if and only if

$$\text{dist}\left(\frac{-c}{\|c\|_2^2} + \left(\mathcal{A}^{\perp} \oplus \text{ Img } c\right)^{\perp}, \mathcal{K}\right) > 0. \tag{2.14}$$

From Lemma 2.1, we have

$$\begin{aligned} \frac{-c}{\|c\|_2^2} + \left(\mathcal{A}^{\perp} \oplus \text{ Img } c\right)^{\perp} &= \frac{-c}{\|c\|_2^2} + (\mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}) \\ &= \{x \in \mathcal{A} \mid c^{\mathrm{T}} x = -1\}. \end{aligned}$$

primal	dual
Interior direction	Bounded normalized feasible set
Strongly feasible	Nonempty and bounded normalized optimal set
Weakly feasible	One–sided level direction, but no improving direction sequence
Weakly infeasible	Improving direction sequence, but no improving direction
Strongly infeasible	Improving direction
Strongly infeasible and no one–sided direction	Interior improving direction

Table 2.1: Feasibility characterizations ('if–and–only–if') for dual feasible closed conic convex programs

This implies that (2.14) holds if and only if there exists no primal improving direction sequence. **Q.E.D.**

The result of Lemma 2.6 can also be found in Duffin [32], with a different proof. In Section 2.1, we already remarked that improving directions may not exist, even if know an improving direction sequence. Combining Lemma 2.5 and Lemma 2.6, we see that this is exactly what happens in the case of dual weak infeasibility, see also Example 2.3.

Table 2.1 summarizes the feasibility characterizations for dual feasible closed conic convex programs. Since duality is completely symmetric for closed conic convex programs, we can make an analogous table of dual (in)feasibility characterizations for primal feasible programs. The characterizations that are listed in Table 2.1 are direct applications of Corollary 2.5, Theorem 2.5, Theorem 2.3, Lemma 2.6, Lemma 2.5 and Corollary 2.3.

2.5 Strong Duality

It is well known that if (P) is a linear program and p^* is finite, then strong duality holds, i.e. $p^* + d^* = 0$. Our objective is to generalize the strong duality result for linear programming to conic convex programming. Notice that, for a general conic convex program, it is possible that d^* is finite but (P) is weakly infeasible (see e.g. Example 2.3 with $c_2 = 0$). This means that we should allow an arbitrarily small constraint violation for the primal and define its *subvalue* as:

$$p^- := \lim_{\epsilon \downarrow 0} \inf_x \{c^{\mathrm{T}} x \mid x \in \mathcal{K}, \ \ \mathrm{dist}(x, b + \mathcal{A}) < \epsilon\}.$$

If (P) is strongly infeasible, then $p^- = \infty$, but for weakly infeasible programs, the subvalue is possibly finite. We also define a matrix M_c,

$$M_c := \begin{bmatrix} I & -c \\ 0 & 1 \end{bmatrix}, \tag{2.15}$$

where I denotes the identity matrix of order n.

Lemma 2.7 *Let $\gamma \in \Re$, and consider a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$. Then there holds*

$$\operatorname{dist}\left(\begin{bmatrix} b \\ \gamma \end{bmatrix} + M_c^T(\mathcal{A} \times \{0\}), \mathcal{K} \times \Re_+\right) = 0$$

if and only if $p^- \leq \gamma$.

Proof. By definition, we have $p^- \leq \gamma$ if and only if there exists a sequence $x^{(1)}, x^{(2)}, \ldots$ in $\mathcal{K}$ such that

$$\lim_{i\to\infty} \operatorname{dist}(x^{(i)}, b + \mathcal{A}) = 0, \quad \lim_{i\to\infty} c^{\mathrm{T}} x^{(i)} \leq \gamma. \tag{2.16}$$

Letting

$$x_{n+1}^{(i)} := \max\{0,\ \gamma - c^{\mathrm{T}} x^{(i)}\} \quad \text{for } i = 1, 2, \ldots,$$

we obtain a sequence $(x^{(1)}, x_{n+1}^{(1)}), (x^{(2)}, x_{n+1}^{(2)}), \ldots$ in $\mathcal{K} \times \Re_+$ with

$$\lim_{i\to\infty} \operatorname{dist}\left(\begin{bmatrix} x^{(i)} \\ x_{n+1}^{(i)} \end{bmatrix}, \begin{bmatrix} b \\ \gamma \end{bmatrix} + M_c^{\mathrm{T}}(\mathcal{A} \times \{0\})\right) = 0. \tag{2.17}$$

Conversely, if $(x^{(i)}, x_{n+1}^{(i)}) \in \mathcal{K} \times \Re_+$, $i = 1, 2, \ldots$, is a sequence satisfying (2.17), then $x^{(1)}, x^{(2)}, \ldots$ is a sequence satisfying (2.16). **Q.E.D.**

Lemma 2.8 *Let $\gamma \in \Re$, and consider a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$ with* $\operatorname{dist}(b + \mathcal{A}, \mathcal{K}) = 0$. *Then there holds*

$$\begin{bmatrix} b^T & \gamma \end{bmatrix} \left((\mathcal{K} \times \Re_+) \cap M_c^{-1}(\mathcal{A}^\perp \times \Re)\right) \subseteq \Re_+$$

if and only if $d^* \geq -\gamma$.

Proof. Notice that

$$M_c^{-1} = \begin{bmatrix} I & c \\ 0 & 1 \end{bmatrix}. \tag{2.18}$$

By definition, we have $d^* < -\gamma$ if and only if there exists a vector $z \in \mathcal{F}_D$ such that

$$b^{\mathrm{T}} z + \gamma < 0.$$

Letting $z_{n+1} := 1$, we see that

$$\begin{bmatrix} z \\ z_{n+1} \end{bmatrix} \in (\mathcal{K}^* \times \Re_+) \cap M_c^{-1}(\mathcal{A}^\perp \times \Re), \quad b^{\mathrm{T}} z + \gamma z_{n+1} < 0. \tag{2.19}$$

Conversely, suppose that there exists (z, z_{n+1}) satisfying (2.19). Notice that if $z_{n+1} = 0$ then s is a dual improving direction which contradicts the assumption that $\operatorname{dist}(b + \mathcal{A}, \mathcal{K}) = 0$ (see Lemma 2.5). Hence, $z_{n+1} > 0$ and $z/z_{n+1} \in \mathcal{F}_D$. Moreover, we have

$$d^* \leq c^{\mathrm{T}} z / z_{n+1} < -\gamma,$$

which completes the proof. **Q.E.D.**

Combining Lemmas 2.7 and 2.8 with the extended Farkas' lemma, we obtain a strong duality theorem:

Theorem 2.6 (Strong duality) *Consider a conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$. If the dual is infeasible and the primal is strongly infeasible, then*

$$p^- = d^* = \infty.$$

Otherwise, there holds

$$p^- = -d^*.$$

Proof. First consider the case of primal strong infeasibility, i.e. $\operatorname{dist}(b + \mathcal{A}, \mathcal{K}) > 0$. In that we case, $p^- = \infty$. From Lemma 2.5 it then follows that there exists a dual improving direction. Hence, we have $d^* = -\infty = -p^-$ if there exists a dual solution, and $d^* = \infty = p^-$ otherwise.

It remains to consider the case $\operatorname{dist}(b + \mathcal{A}, \mathcal{K}) = 0$. Given $\gamma \in \Re$, we know from Lemma 2.7 that

$$p^- \leq \gamma \iff \operatorname{dist}\left(\begin{bmatrix} b \\ \gamma \end{bmatrix} + M_c^{\mathrm{T}}(\mathcal{A} \times \{0\}), \mathcal{K} \times \Re_+ \right) = 0.$$

The above relation implies, using Lemma 2.5 and Lemma 2.3, that

$$p^- \leq \gamma \iff \begin{bmatrix} b^{\mathrm{T}}, & \gamma \end{bmatrix} \left((\mathcal{K} \times \Re_+) \cap M_c^{-1}(\mathcal{A}^\perp \times \Re) \right) \subseteq \Re_+.$$

Applying now Lemma 2.8 yields

$$p^- \leq \gamma \iff d^* \geq -\gamma.$$

Since γ is arbitrary, it follows that

$$p^{-} = -d^{*}.$$

Q.E.D.

Since $p^* \geq p^-$, we obtain from Theorem 2.6 the *weak duality relation*

$$p^* \geq -d^*. \tag{2.20}$$

For the case that the primal and the dual are not both infeasible, we see from the above theorem that $p^- = -d^*$. We will now show that if the primal has an interior solution (generalized Slater condition), then the subvalue coincides with the optimal value, i.e. $p^* = p^-$. Hence, we can strengthen the duality result for the case in which the generalized Slater condition holds. We thus arrive at the following strong duality theorem.

Theorem 2.7 (Slater duality) *Suppose that $\overset{o}{\mathcal{F}}_P \neq \emptyset$. Then*

$$p^* = p^- = -d^*.$$

Moreover, if $p^* > -\infty$ then

$$\mathcal{F}_D^* \neq \emptyset,$$

and the normalized dual optimal solution set is bounded.

Proof. Observe that since $\overset{o}{\mathcal{F}}_P \neq \emptyset$, there holds

$$\operatorname{cl} \mathcal{F}_P = (b + \mathcal{A}) \cap \operatorname{cl} \mathcal{K}.$$

Therefore, for the purpose of proving the theorem, we can assume without loss of generality that $\mathcal{K}$ is closed, i.e. $\mathcal{K} = \mathcal{K}^{**}$.

Analogous to the definition of p^-, we define the *dual subvalue* d^- as

$$d^- := \lim_{\epsilon \downarrow 0} \inf_z \{ b^{\mathrm{T}} z \mid z \in \mathcal{K}^*, \ \ \operatorname{dist}(z, c + \mathcal{A}^{\perp}) < \epsilon \}.$$

Applying Theorem 2.6 to the conic convex program $\mathrm{CP}(c, b, \mathcal{A}^{\perp}, \mathcal{K}^*)$, we obtain $p^* = -d^- \geq -d^*$. Hence, if $p^* = -\infty$ then $d^* = \infty$ and the theorem holds true. It remains to consider the case that $p^* = -d^- > -\infty$. By definition, the condition

$$d^- = -p^* < \infty$$

means that there exists a sequence $z^{(1)}, z^{(2)}, \ldots$ in $\mathcal{K}^* \cap \operatorname{span}(\mathcal{A} \oplus \mathcal{K})$ with

$$\lim_{i \to \infty} \operatorname{dist}(z^{(i)}, c + \mathcal{A}^{\perp}) = 0 \quad \text{and} \quad \limsup_{i \to \infty} b^{\mathrm{T}} z^{(i)} = -p^* < \infty.$$

It follows from Lemma 2.4 that this sequence has a cluster point, say $z^{(\infty)}$. Obviously

$$z^{(\infty)} \in \mathcal{F}_D, \quad b^{\mathrm{T}} z^{(\infty)} = -p^*.$$

It follows from the relation

$$-p^* = d^* \leq b^{\mathrm{T}} z^{(\infty)} = -p^*,$$

that

$$p^* + d^* = 0, \quad \mathcal{F}_D^* \neq \emptyset.$$

Finally, the boundedness of the normalized dual optimal solution set follows from Theorem 2.5. **Q.E.D.**

For the case that $\mathcal{K}$ is closed and solid, the strong duality theorem with Slater condition is well known; see for example [1, 32, 110], among others. The result of Theorem 2.7, which holds for general convex cones, is due to Borwein and Wolkowicz [18].

Theorem 2.7 implies the following well known fact: if $\mathcal{K}$ is closed and $\overset{o}{\mathcal{F}}_P \times \overset{o}{\mathcal{F}}_D \neq \emptyset$, then $\mathcal{F}_P^* \times \mathcal{F}_D^* \neq \emptyset$ and

$$(x^*)^{\mathrm{T}} z^* = c^{\mathrm{T}} x^* + b^{\mathrm{T}} z^* = 0, \quad \text{for all } (x^*, z^*) \in \mathcal{F}_P^* \times \mathcal{F}_D^*.$$

For the above case, we say that a conic convex programming problem has a *complementary solution.* A complementary solution is a pair $(x, z) \in \mathcal{F}_P \times \mathcal{F}_D$ such that

$$x^{\mathrm{T}} z = 0.$$

A *face* of a cone $\mathcal{K}$ is a set

$$\text{face}\,(\mathcal{K}, z) := \{x \in \mathcal{K} \mid x^{\mathrm{T}} z = 0\},$$

where $z \in \mathcal{K}^*$. Notice for $x, y \in \mathcal{K}$ that

$$x + y \in \text{ face}\,(\mathcal{K}, z) \implies x, y \in \text{ face}\,(\mathcal{K}, z),$$

which explains why "face $(\mathcal{K}, z)$" is called a face of $\mathcal{K}$. We remark that Theorem 2.2 implies that $\mathcal{K} \cap$ sub $\mathcal{K}$ is the smallest face of $\mathcal{K}$. If (x, z) is a complementary solution, then

$$\mathcal{F}_P^* = \mathcal{F}_P \cap \text{ face}\,(\mathcal{K}, z), \quad \mathcal{F}_D^* = \mathcal{F}_D \cap \text{ face}\,(\mathcal{K}^*, x).$$

Therefore, $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ are also known as the *optimal faces* of (P) and (D) respectively. A *strictly complementary solution pair* of (P) is a pair $(x, z) \in \mathcal{F}_P \times \mathcal{F}_D$ such that

$$x \in \text{ rel (face}\,(\mathcal{K}, z)), \quad z \in \text{ rel (face}\,(\mathcal{K}^*, x)).$$

By definition, such a solution pair is also a complementary solution pair. It was shown by Tucker [155] and Goldman and Tucker [44] that any solvable linear programming problem has a strictly complementary solution pair. Unfortunately, a conic convex program

		primal feasible		primal infeasible	
		strong	weak	weak	strong
dual feasible	strong	$p^* = -d^*$ (P)+(D) solvable	$p^* = -d^*$ (P) solvable	(D) unbounded	(D) unbounded
	weak		possible	possible	(D) unbounded
dual infeasible	weak			possible	possible
	strong				possible

Table 2.2: Duality for closed conic convex programs

may not have any strictly complementary solution pair, even if it satisfies primal and dual Slater conditions. One therefore also encounters the term *maximal complementary solution pair*, which is a complementary solution pair $(x, z) \in (\text{ rel } \mathcal{F}_P^*) \times \text{ rel } (\mathcal{F}_D^*)$.

Obviously, any complementary solution pair is an optimal solution pair. The converse however, is in general not true unless $\overset{o}{\mathcal{F}}_P \neq \emptyset$ or $\overset{o}{\mathcal{F}}_D \neq \emptyset$. This is because without the latter condition, there may exist a positive duality gap and as a result there cannot exist a complementary solution. Moreover, strong duality is necessary, but not sufficient for the existence of a complementary solution. The duality relations for conic convex programs CP$(b, c, \mathcal{A}, \mathcal{K})$ with $\mathcal{K}$ closed are summarized in Table 2.2.

All entries in the table represent possible combinations of the status of the primal and dual problem. Only if we cannot conclude anything more, we explicitly mention that the entry represents a possible state. Due to the complete symmetry of the closed conic convex programming duality, the table is symmetric, so we only need to consider the upper-right block. The entries in the first row of the table are denoted by 'A1', 'A2', 'A3' and 'A4', in the second row by 'B1', 'B2', and so on. The entries 'A1', 'A2', 'A3' and 'A4', are due to Theorem 2.7. Lemma 2.5 implies entry 'B4'. The possibility of states 'A3' and 'B3' and 'D3' (and hence 'C4') is demonstrated by Example 2.3, while the entry 'C3' is illustrated by Example 2.4, a semidefinite programming problem.

Example 2.4 *Consider the program CP$(b, c, \mathcal{A}, \mathcal{K})$ in $\Re^6$ with*

$$b = \begin{bmatrix} 0, & 0, & 0, & 0, & 1, & 0 \end{bmatrix}^T, \quad c = \begin{bmatrix} 0, & 0, & 0, & 1, & 0, & 0 \end{bmatrix}^T,$$

$$\mathcal{A} = \{x \in \Re^6 \mid x_2 = 0, x_5 = 0\},$$

and

$$\mathcal{K} = \left\{ x \in \Re^6 \;\middle|\; \begin{bmatrix} x_1 & x_4/\sqrt{2} & x_6/\sqrt{2} \\ x_4/\sqrt{2} & x_2 & x_5/\sqrt{2} \\ x_6/\sqrt{2} & x_5/\sqrt{2} & x_3 \end{bmatrix} \succeq 0 \right\}.$$

Then $\mathcal{K}^* = \mathcal{K}$ *and* $\mathcal{A}^\perp = \{z \in \Re^6 \mid z_1 = z_3 = z_4 = z_6 = 0\}$. *The primal is weakly infeasible,*

$$p^* = \inf\left\{ x_4 \;\middle|\; \begin{bmatrix} x_1 & x_4/\sqrt{2} & x_6/\sqrt{2} \\ x_4/\sqrt{2} & 0 & 1 \\ x_6/\sqrt{2} & 1 & x_3 \end{bmatrix} \succeq 0 \right\} = \infty,$$

and the dual is also weakly infeasible:

$$d^* = \inf\left\{ z_5 \;\middle|\; \begin{bmatrix} 0 & 1 & 0 \\ 1 & z_2 & z_5/\sqrt{2} \\ 0 & z_5/\sqrt{2} & 0 \end{bmatrix} \succeq 0 \right\} = \infty.$$

Finally, the possibility of the entries in Table 2.2 where weak infeasibility is not involved, can be demonstrated by a 2-dimensional linear programming problem:

Example 2.5 *Let* $n = 2$, $c \in \Re^2$, $\mathcal{K} = \mathcal{K}^* = \Re^2_+$ *and*

$$\mathcal{A} = \{(x_1, x_2) \mid x_1 = 0\}, \quad \mathcal{A}^\perp = \{(z_1, z_2) \mid z_2 = 0\}.$$

We see that (P) is strongly feasible if $c_1 > 0$*, weakly feasible if* $c_1 = 0$ *and strongly infeasible if* $c_1 < 0$*. Similarly, (D) is strongly feasible if* $c_2 > 0$*, weakly feasible if* $c_2 = 0$ *and strongly infeasible if* $c_2 < 0$*.*

Weak infeasibility does not exist in linear programming. However, Examples 2.3 and 2.4 illustrate that weakly infeasible problems do exist in semidefinite programming.

2.6 Regularization

In Theorem 2.6, we have shown that $z \in \mathcal{F}_D$ is an optimal solution of (D) if and only if there exists a sequence $x^{(i)} \in \mathcal{K}$, $i = 1, 2, \ldots$, with

$$\lim_{i\to\infty} \operatorname{dist}(b + \mathcal{A}, x^{(i)}) = 0, \quad \lim_{i\to\infty} c^{\mathrm{T}} x^{(i)} = -b^{\mathrm{T}} z. \tag{2.21}$$

Such a sequence is called a *certificate* of the optimality of the dual solution s. Since this certificate is a sequence, it has the rather inconvenient property that it has an infinite length. We will see in this section that finite certificates can be obtained by means of *regularization.*

If $\overset{o}{\mathcal{F}}_P \neq \emptyset$ (a generalized Slater condition), then $p^* = -d^*$ and no regularization is needed, see Theorem 2.7. For a weakly feasible conic convex program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ however, we may try to replace $\mathcal{K}$ by a lower dimensional face, say face $(\mathcal{K}, z)$ for a certain $z \in \mathcal{K}^*$, such that

$$(b+\mathcal{A}) \cap \text{ face } (\mathcal{K}, z) = (b+\mathcal{A}) \cap \mathcal{K}, \quad (b+\mathcal{A}) \cap \text{ rel face } (\mathcal{K}, z) \neq \emptyset.$$

If we succeed in finding such a face (which is then known as the *minimal cone* [18, 19, 160]), then we can regularize $\text{CP}(b, c, \mathcal{A}, \mathcal{K})$ to $\text{CP}(b, c, \mathcal{A}, \text{ face } (\mathcal{K}, z))$, which satisfies the generalized Slater condition. Such a regularization approach, which we call *primal regularization*, was proposed by Borwein and Wolkowicz [18, 19] and Wolkowicz [160].

In this section, we propose a *dual regularization* approach, which is based on the dual characterization of strong feasibility, Theorem 2.3. In this approach, we transform all one–sided, non–improving, dual level directions into two–sided directions (lines), thus enlarging the dimension of sub $\mathcal{K}^*$. For notational convenience, we will now interchange the role of the primal and the dual: we assume that we want to solve the dual program $\text{CP}(c, b, \mathcal{A}^\perp, \mathcal{K}^*)$, and to this end we transform one–sided primal level directions into lines, thus enlarging $\mathcal{K}$.

Let $\mathcal{K}$ be a convex cone in $\Re^n$, and let $\mathcal{A}$ be a linear subspace of $\Re^n$. We define an operator $\Gamma_\mathcal{A}$ on $\mathcal{K}$ as follows:

$$\Gamma_\mathcal{A}\mathcal{K} := \text{ cl } (\mathcal{K} \oplus \text{ span } (\mathcal{A} \cap \text{ cl } \mathcal{K})). \tag{2.22}$$

Observe from this definition that

$$\mathcal{A} \cap \text{ cl } \mathcal{K} \subseteq \text{ sub } (\Gamma_\mathcal{A}\mathcal{K}),$$

i.e. if $x \in \mathcal{A} \cap \mathcal{K} \setminus -\mathcal{K}$ is a one–sided direction with respect to $\mathcal{K}$, then this direction x is not one–sided with respect to $\Gamma_\mathcal{A}\mathcal{K}$. Observe also that $\Gamma_\mathcal{A}\mathcal{K} = \text{ cl } \mathcal{K}$ if and only if the convex program $\text{CP}(b, c, \mathcal{A}, \mathcal{K})$ has no primal one–sided directions, i.e.

$$\Gamma_\mathcal{A}\mathcal{K} = \text{ cl } \mathcal{K} \iff \text{ span } (\mathcal{A} \cap \text{ cl } \mathcal{K}) \subseteq \text{ cl } \mathcal{K}. \tag{2.23}$$

Although $\text{CP}(b, c, \mathcal{A}, \mathcal{K} \oplus \text{ span } (\mathcal{A} \cap \mathcal{K}))$ has no primal one–sided directions, it is quite possible for the *closed* conic convex program $\text{CP}(b, c, \mathcal{A}, \Gamma_\mathcal{A}\mathcal{K})$ to have primal one–sided directions (see Example 2.6 below). Therefore, it makes sense to apply the operator $\Gamma_\mathcal{A}$ k times in succession, resulting in an operator $\Gamma_\mathcal{A}^k$. More precisely, we let

$$\begin{cases} \Gamma_\mathcal{A}^0\mathcal{K} := \mathcal{K}, \\ \Gamma_\mathcal{A}^k\mathcal{K} := \Gamma_\mathcal{A}\Gamma_\mathcal{A}^{k-1}\mathcal{K}, \quad \text{for } k = 1, 2, \ldots. \end{cases} \tag{2.24}$$

In addition, we define

$$\Gamma_\mathcal{A}^\infty\mathcal{K} := \Gamma_\mathcal{A}^{\dim \mathcal{A}}\mathcal{K}. \tag{2.25}$$

Each time that we apply the operator $\Gamma_\mathcal{A}$ to a cone $\Gamma_\mathcal{A}^k\mathcal{K}$, we move any one–sided direction in $\mathcal{A} \cap \Gamma_\mathcal{A}^k\mathcal{K}$ into sub $\Gamma_\mathcal{A}^{k+1}\mathcal{K}$, so that it is not one–sided with respect to the larger cone $\Gamma_\mathcal{A}^{k+1}\mathcal{K}$. After applying the $\Gamma_\mathcal{A}$ operator dim $\mathcal{A}$ times in succesion, there will be no one–sided directions in $\mathcal{A} \cap \Gamma_\mathcal{A}^\infty\mathcal{K}$, as the following lemma shows.

Lemma 2.9 *Let $\mathcal{K}$ be a convex cone in $\Re^n$ and let $\mathcal{A}$ be a linear subspace of $\Re^n$. Then*

$$\Gamma^k_{\mathcal{A}}\mathcal{K} = \Gamma^\infty_{\mathcal{A}}\mathcal{K} \quad \textit{for all } k \geq \dim \mathcal{A}.$$

Proof. Since the sets $\{\Gamma^k_{\mathcal{A}}\mathcal{K} : k = 0, 1, 2, \ldots\}$ are nested, we only need to show that $\Gamma^k_{\mathcal{A}}\mathcal{K} = \Gamma^{k+1}_{\mathcal{A}}\mathcal{K}$ for some finite k. Suppose $\Gamma^k_{\mathcal{A}}\mathcal{K} \neq \Gamma^{k+1}_{\mathcal{A}}\mathcal{K}$ for some k so that

$$\text{span}\,(\mathcal{A} \cap \Gamma^k_{\mathcal{A}}\mathcal{K}) \not\subseteq \Gamma^k_{\mathcal{A}}\mathcal{K}. \tag{2.26}$$

From the definition (2.22), we have

$$\text{sub}\,(\mathcal{A} \cap \Gamma^k_{\mathcal{A}}\mathcal{K}) \subset \text{span}\,(\mathcal{A} \cap \Gamma^k_{\mathcal{A}}\mathcal{K}) \subseteq \text{sub}\,(\mathcal{A} \cap \Gamma^{k+1}_{\mathcal{A}}\mathcal{K}),$$

so that

$$\dim\ \text{sub}\left(\mathcal{A} \cap \Gamma^k_{\mathcal{A}}\mathcal{K}\right) < \dim\ \text{sub}\left(\mathcal{A} \cap \Gamma^{k+1}_{\mathcal{A}}\mathcal{K}\right) \leq \dim \mathcal{A}.$$

Thus, the dimension of $\text{sub}\left(\mathcal{A} \cap \Gamma^k_{\mathcal{A}}\mathcal{K}\right)$ is increased by one whenever (2.26) holds. Using an inductive argument, it follows that $k + 1 \leq \dim \mathcal{A}$. Consequently, there will be some $k \leq \dim \mathcal{A}$ for which (2.26) does not hold. Together with (2.23), this implies the lemma. **Q.E.D.**

We will now show that the property of strong infeasibility is invariant under the operator $\Gamma_{\mathcal{A}}$.

Lemma 2.10 *Consider a conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$ and let $\mathcal{A}' \subseteq \mathcal{A}$ be a linear subspace. There holds*

$$\text{dist}\left(b + \mathcal{A}, \Gamma^{k+1}_{\mathcal{A}'}\mathcal{K}\right) = \text{dist}\left(b + \mathcal{A}, \Gamma^k_{\mathcal{A}'}\mathcal{K}\right)$$

for all $k = 0, 1, \ldots$.

Proof. Since $\Gamma^k_{\mathcal{A}'}\mathcal{K} \subseteq \Gamma^{k+1}_{\mathcal{A}'}\mathcal{K}$, we obviously have

$$\text{dist}\left(b + \mathcal{A}, \Gamma^{k+1}_{\mathcal{A}'}\mathcal{K}\right) \leq \text{dist}\left(b + \mathcal{A}, \Gamma^k_{\mathcal{A}'}\mathcal{K}\right). \tag{2.27}$$

To prove the converse, we fix any vector x in $\Gamma^{k+1}_{\mathcal{A}'}\mathcal{K}$. It follows from the definition (2.22) that there exists a sequence $\{(u^{(i)}, v^{(i)})\}$ with

$$u^{(i)} \in \Gamma^k_{\mathcal{A}'}\mathcal{K}, \quad v^{(i)} \in \text{span}\left(\mathcal{A}' \cap \Gamma^k_{\mathcal{A}'}\mathcal{K}\right), \qquad i = 1, 2, \ldots,$$

such that

$$x = \lim_{i\to\infty}(u^{(i)} + v^{(i)}).$$

As $v^{(i)} \in \text{ span } \left(\mathcal{A}' \cap \Gamma^k_{\mathcal{A}'}\mathcal{K}\right) \subseteq \mathcal{A}' \subseteq \mathcal{A}$, we have

$$\text{dist}(u^{(i)} + v^{(i)}, b + \mathcal{A}) = \text{ dist}(u^{(i)}, b + \mathcal{A}) \geq \text{ dist}\left(b + \mathcal{A}, \Gamma^k_{\mathcal{A}'}\mathcal{K}\right),$$

where the last step is due to $u^{(i)} \in \Gamma^k_{\mathcal{A}'}\mathcal{K}$. Letting $i \to \infty$ yields

$$\text{dist}(x, b + \mathcal{A}) \geq \text{ dist}\left(b + \mathcal{A}, \Gamma^k_{\mathcal{A}'}\mathcal{K}\right).$$

Since x is an arbitrary element of $\Gamma^{k+1}_{\mathcal{A}'}\mathcal{K}$, we obtain

$$\text{dist}\left(b + \mathcal{A}, \Gamma^{k+1}_{\mathcal{A}'}\mathcal{K}\right) \geq \text{ dist}\left(b + \mathcal{A}, \Gamma^k_{\mathcal{A}'}\mathcal{K}\right).$$

Combining this with (2.27) proves the lemma. **Q.E.D.**

The following lemma shows that regularization with the subspace ($\mathcal{A} \cap$ Ker c^{T}) does not change the dual feasible set.

Lemma 2.11 *Consider a conic convex program CP*$(b, c, \mathcal{A}, \mathcal{K})$. *There holds*

$$\mathcal{F}_D = (c + \mathcal{A}^{\perp}) \cap \left(\Gamma^k_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K}\right)^*$$

for all $k \in \{0, 1, 2, \ldots\}$.

Proof. Let $z \in \mathcal{F}_D$. It suffices to prove that this implies

$$z \in \left(\Gamma^k_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}}\mathcal{K}\right)^*, \quad k = 0, 1, 2, ... \tag{2.28}$$

Since $(\Gamma^0_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}}\mathcal{K})^* = \mathcal{K}^*$, relation (2.28) holds trivially for $k = 0$. Now assume that (2.28) holds for some $k \in \{0, 1, 2, \ldots\}$. We need to show that (2.28) holds for $k + 1$ in the sense that $x^{\mathrm{T}} z \geq 0$ for any $x \in \Gamma^{k+1}_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}}\mathcal{K}$. By definition, $x \in \Gamma^{k+1}_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}}\mathcal{K}$ means that there exists some sequence $(u^{(i)}, v^{(i)})$, $i = 1, 2, \ldots$, satisfying

$$u^{(i)} \in \Gamma^k_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}}\mathcal{K}, \quad v^{(i)} \in \text{ span }\left(\mathcal{A} \cap \text{ Ker } c^{\mathrm{T}} \cap \Gamma^k_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}}\mathcal{K}\right),$$

such that

$$x = \lim_{i\to\infty}(u^{(i)} + v^{(i)}).$$

However, since $z \in c + \mathcal{A}^{\perp}$ we have $z^{\mathrm{T}} v^{(i)} = 0$, whereas (2.28) implies $z^{\mathrm{T}} u(i) \geq 0$, for all i. Consequently, there holds $z^{\mathrm{T}} x \geq 0$. **Q.E.D.**

Although regularization does not affect the dual feasible set, it can change the nature of dual (in)feasibility.

Lemma 2.12 *For a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$, the regularized program*

$$\mathrm{CP}\,(b, c, \mathcal{A}, \Gamma^{\infty}_{\mathcal{A}\cap\ \mathrm{Ker}\ c^T}\mathcal{K})$$

is either dual strongly infeasible or dual strongly feasible.

Proof. Recall from Lemma 2.9 that $\Gamma_{\mathcal{A}\cap\ \mathrm{Ker}\ c^{\mathrm{T}}}\Gamma^{\infty}_{\mathcal{A}\cap\ \mathrm{Ker}\ c^{\mathrm{T}}}\mathcal{K} = \Gamma^{\infty}_{\mathcal{A}\cap\ \mathrm{Ker}\ c^{\mathrm{T}}}\mathcal{K}$. Hence, if the regularized primal has a one–sided level direction, it must be an improving direction. It thus follows from Lemma 2.5 that the regularized dual is strongly infeasible if and only if the regularized primal has one–sided level directions. Using Theorem 2.3, we conclude that the regularized dual is either dual strongly infeasible or dual strongly feasible. **Q.E.D.**

Together, Lemma 2.11 and Lemma 2.12 imply that the regularization of a dual weakly feasible problem results in a dual strongly feasible problem. Similarly, the regularization of a dual weakly infeasible problem results in a dual strongly infeasible problem. However, the *set* of dual feasible solutions is not affected by regularization. These conclusions are summarized in the following theorem.

Theorem 2.8 *Consider a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$ and let*

$$\mathcal{K}' := \Gamma^{\infty}_{\mathcal{A}\cap\ \mathrm{Ker}\ c^T}\mathcal{K}.$$

There holds

- *The dual feasible sets of CP$(b, c, \mathcal{A}, \mathcal{K})$ and its regularization CP$(b, c, \mathcal{A}, \mathcal{K}')$ coincide, i.e.*
$$\mathcal{F}_D = (c + \mathcal{A}^{\perp}) \cap (\mathcal{K}')^*.$$
- *The regularized program CP$(b, c, \mathcal{A}, \mathcal{K}')$ is dual strongly feasible if and only if $\mathcal{F}_D \neq \emptyset$.*
- *The regularized program CP$(b, c, \mathcal{A}, \mathcal{K}')$ is dual strongly infeasible if and only if $\mathcal{F}_D = \emptyset$.*

Combining Theorem 2.8 with Table 2.2, we see that the regularized conic convex program is in perfect duality:

Corollary 2.7 *Assume the same setting as in Theorem 2.8. Then there holds*

- *If $d^* = \infty$, then the regularized primal CP$(b, c, \mathcal{A}, \mathcal{K}')$ is either infeasible or unbounded.*

- *If $-\infty < d^* < \infty$, then the regularized primal is solvable with optimal value equal to $-d^*$, i.e.*
$$d^* = -\min c^T\left((b+\mathcal{A}) \cap \mathcal{K}'\right).$$
- *If $d^* = -\infty$ then the regularized primal CP$(b, c, \mathcal{A}, \mathcal{K}')$ is infeasible.*

Applying Corollary 2.7 to the conic convex program CP$(0, c, \mathcal{A}, \mathcal{K})$, we obtain a generalization of Farkas' lemma:

Corollary 2.8 *A conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$ is dual feasible if and only if*

$$c^T\left(\mathcal{A} \cap \Gamma^{\infty}_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K})\right) \subseteq \Re_+.$$

We have seen that the regularization of a dual feasible conic program results in a conic program with strong dual feasibility. Due to this property, the dual regularized cone is called the *minimal cone* for (D). The regularization scheme presented here is a dual version of the minimal cone duality of Borwein and Wolkowicz [18, 19] and Wolkowicz [160]. Namely, the dual conic convex program (D) is regularized in [18, 19, 160] by replacing the original cone $\mathcal{K}^*$ by a smaller cone, such that the resulting program will be strongly feasible whenever (D) is feasible. In the preceding, we regularized (P) by transforming all its one–sided, non-improving, level directions into two-sided directions (lines), thus enlarging the cone $\mathcal{K}$. In this way, new primal solutions are created that play the role of sequences that approach feasibility for the original problem (P), as can be seen from Lemma 2.9 and Lemma 2.10.

An illustration of regularization for a semidefinite programming problem is given in Example 2.6.

Example 2.6 *Consider the program CP$(b, c, \mathcal{A}, \mathcal{K})$ in $\Re^6$ with*

$$\mathcal{K} = \mathcal{S}_+ \times \mathcal{S}_+,$$

where we let

$$\mathcal{S}_+ := \left\{ x \in \Re^3 \;\middle|\; \begin{bmatrix} x_1 & x_3/\sqrt{2} \\ x_3/\sqrt{2} & x_2 \end{bmatrix} \succeq 0 \right\}.$$

Moreover, we let

$$b = \begin{bmatrix} 0, & 0, & 0, & 0, & 0, & \sqrt{2} \end{bmatrix}^T, \quad c = \begin{bmatrix} 0, & 0, & c_4, & c_4, & 0, & 0 \end{bmatrix}^T,$$

$$\mathcal{A} = \{x \in \Re^6 \mid x_2 = 0, x_3 = x_4, x_5 = x_6 = 0\}.$$

Then $\mathcal{K}^* = \mathcal{K}$ *and* $\mathcal{A}^\perp = \{z \in \Re^6 \mid z_1 = 0, z_3 = -z_4\}$. *In other words, the primal is*

$$p^* = \inf\left\{2c_4x_4 \;\middle|\; \begin{bmatrix} x_1 & x_4/\sqrt{2} \\ x_4/\sqrt{2} & 0 \end{bmatrix} \succeq 0, \begin{bmatrix} x_4 & 1 \\ 1 & 0 \end{bmatrix} \succeq 0 \right\} = \infty$$

with dual

$$d^* = \inf\left\{\sqrt{2}z_6 \;\middle|\; \begin{bmatrix} 0 & z_3/\sqrt{2} \\ z_3/\sqrt{2} & z_2 \end{bmatrix} \succeq 0, \begin{bmatrix} 2c_4 - z_3 & z_6/\sqrt{2} \\ z_6/\sqrt{2} & z_5 \end{bmatrix} \succeq 0 \right\}.$$

Notice that the primal is weakly infeasible, whereas (D) is

- *weakly infeasible if* $c_4 < 0$,
- *weakly feasible and solvable with optimal value* $d^* = 0$ *if* $c_4 = 0$, *and*
- *weakly feasible and unbounded if* $c_4 > 0$.

Since $\mathcal{A} \cap \mathcal{K} = \Re_+ \times \{0\}^5 \subset \text{ Ker } c^T$, *we obtain*

$$\Gamma_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K} = \Re \times \Re_+ \times \Re \times \mathcal{S}_+, \quad (\Gamma_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K})^* = \{0\} \times \Re_+ \times \{0\} \times \mathcal{S}_+,$$

for all c_4. *Notice that if* $c_4 \neq 0$, *then* $\dim(\mathcal{A} \cap \text{ Ker } c^T) = 1$ *and* $CP(b, c, \mathcal{A}, \Gamma_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K})$ *is the regularized program. Indeed, the regularized dual is strongly infeasible for* $c_4 < 0$ *and strongly feasible for* $c_4 > 0$.

However, if $c_4 = 0$ *then* $\dim(\mathcal{A} \cap \text{ Ker } c^T) = \dim \mathcal{A} = 2$, *and we have to take one step more. It can easily be verified that*

$$\Gamma^2_{\mathcal{A}}\mathcal{K} = (\Re \times \Re_+ \times \Re) \times (\Re \times \Re_+ \times \Re),$$

and

$$\left(\Gamma^2_{\mathcal{A}}\mathcal{K}\right)^* = (\{0\} \times \Re_+ \times \{0\}) \times (\{0\} \times \Re_+ \times \{0\}).$$

Consequently, the regularization makes the dual strongly feasible, and makes the primal solvable with optimal value 0.

Example 2.6 reveals a drawback of the regularization scheme: although the regularized certificates are finite, it may not be easy to check their feasibility, since this involves the cone $\Gamma^\infty_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K}$. For semidefinite programming (as in Example 2.6) however, we will see in Section 2.7 that $\Gamma^\infty_{\mathcal{A}\cap \text{ Ker } c^T}\mathcal{K}$ can be completely described by semidefiniteness constraints, after adding artificial variables. The resulting regularized semidefinite program coincides with the regularized dual of Ramana [124], which was originally derived in a very different way. The relation between *primal* regularization and the so-called extended Lagrange–Slater dual of Ramana [124] was already recognized by Ramana, Tunçel and

Wolkowicz [127]. The way in which Zhao, Karisch, Rendl and Wolkowicz [166] make the regularization explicit, is more or less the opposite of the technique of Ramana. Namely, in [166], the regularized semidefinite relaxation of a quadratic assignment problem is transformed into a strongly feasible semidefinite programming problem by *eliminating* variables, instead of adding variables.

2.7 Regularization of Semidefinite Programs

We will now further analyze the structure of the regularized conic convex program, for the special case that $\mathcal{K}$ is the cone of positive semidefinite matrices, i.e. $\mathcal{K} = \mathcal{H}_+$. We consider the real linear space $\mathcal{H}^{(\bar{n})}$ of $\bar{n} \times \bar{n}$ Hermitian matrices, with the real valued inner product $X \bullet Y$, see Section 1.2.2. Given two Hermitian matrices $X, Y \in \mathcal{H}^{(\bar{n})}$, it holds that

$$X \bullet Y = \operatorname{tr} XY.$$

In terms of this inner product, we can define an orthonormal basis of $\mathcal{H}^{(\bar{n})}$. In this way, we obtain a one–to–one correspondence between Hermitian matrices in $\mathcal{H}^{(\bar{n})}$ and their coordinate vectors in $\Re^n$, where $n = \dim \mathcal{H}^{(\bar{n})}$. Recall from Section 1.2.2 that if $x, y \in \Re^n$ are the coordinate vectors of $X, Y \in \mathcal{H}^{(\bar{n})}$, then $X \bullet Y = x^{\mathrm{T}} y$. We can therefore treat elements of $\mathcal{H}^{(\bar{n})}$ both as $\bar{n} \times \bar{n}$ Hermitian matrices, and as real vectors of order n.

For a semidefinite program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{H}_+^{(\bar{n})})$, we let $B \in \mathcal{H}^{(\bar{n})}$ and $C \in \mathcal{H}^{(\bar{n})}$ denote the matrix representations of the coordinate vectors $b \in \Re^n$ and $c \in \Re^n$, respectively.

Consider an $l \times \bar{n}$ matrix R satisfying $RR^{\mathrm{H}} = I$, where $l \in \{1, 2, \ldots, \bar{n}\}$. Remark that for such R, there must exist a $(\bar{n} - l) \times \bar{n}$ matrix Q such that $\begin{bmatrix} Q^{\mathrm{H}}, & R^{\mathrm{H}} \end{bmatrix}$ is a unitary matrix. We define the following linear subspace of $\mathcal{H}^{(\bar{n})}$,

$$\mathrm{HKer}\, R := \{X \in \mathcal{H}^{(\bar{n})} \mid RXR^{\mathrm{H}} = 0\},$$

where R^{H} denotes the complex conjugate transpose (or adjoint) of R. (In terms of the Hermitian Kronecker product, $\mathrm{HKer}\, R$ corresponds to $\mathrm{Ker}\, R \otimes_{\mathrm{H}} R$ in $\Re^n$.) There holds

$$\mathcal{H}_+^{(\bar{n})} \oplus \mathrm{HKer}\, R = \{X \in \mathcal{H}^{(\bar{n})} \mid RXR^{\mathrm{H}} \succeq 0\}, \tag{2.29}$$

which is a closed convex cone in $\mathcal{H}^{(\bar{n})}$. Here, it is convenient to interpret the unitary matrix $\begin{bmatrix} Q^{\mathrm{H}}, & R^{\mathrm{H}} \end{bmatrix}$ as a basis of the complex Euclidean space $\mathcal{C}^n$. In this way,

$$\mathrm{HKer}\, R = \left\{ X \,\middle|\middle|\, \begin{bmatrix} Q \\ R \end{bmatrix} X \begin{bmatrix} Q^{\mathrm{H}}, & R^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^{\mathrm{H}} & 0 \end{bmatrix} \text{ for some } X_{11},\, X_{12} \right\},$$

and

$$\mathcal{H}_+^{(\bar{n})} \oplus \mathrm{HKer}\, R = \left\{ X \,\middle|\middle|\, \begin{bmatrix} Q \\ R \end{bmatrix} X \begin{bmatrix} Q^{\mathrm{H}}, & R^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^{\mathrm{H}} & X_{22} \end{bmatrix},\ X_{22} \succeq 0 \right\}.$$

We will derive in this section that the regularized cones $\Gamma^k_{\mathcal{A}}\mathcal{H}^{(\bar{n})}_+$ are of the form (2.29). First of all, we notice that this is indeed the case for $k = 0$, viz.

$$\mathcal{H}^{(\bar{n})}_+ = \mathcal{H}^{(\bar{n})}_+ \oplus \text{ HKer } I,$$

where I is the $\bar{n} \times \bar{n}$ identity matrix.

We will see below that $\text{Img } R^{\mathrm{H}} = R^{\mathrm{H}}\mathcal{C}^l$ plays a crucial role. We want to make clear that $\text{Img } R^{\mathrm{H}}$ is a *complex* linear subspace of $\mathcal{C}^{\bar{n}}$, where we use the standard complex valued inner product $y^{\mathrm{H}}x$ for $x, y \in \mathcal{C}^{\bar{n}}$. This is in contrast with the space of Hermitian matrices which is real: although the off–diagonal entries of Hermitian matrices are complex, the inner product $X \bullet Y$ is *real* valued for $X,\ Y \in \mathcal{H}^{(\bar{n})}$.

Lemma 2.13 *Let $\mathcal{A}$ be a linear subspace of $\mathcal{H}^{(\bar{n})}$, and let $\mathcal{K} = \mathcal{H}^{(\bar{n})}_+ \oplus$ HKer (R). If $Y \in$ rel $(\mathcal{A} \cap \mathcal{K})$, then*

$$(\text{ Ker } Y) \cap \text{ Img } R^H \subseteq (\text{ Ker } \tilde{Y}) \cap \text{ Img } R^H \quad \forall \tilde{Y} \in \mathcal{A} \cap \mathcal{K}.$$

Proof. Suppose to the contrary that there exists a $\tilde{Y} \in \mathcal{A} \cap \mathcal{K}$ such that $\tilde{Y}u \neq 0$ for some $u \in (\text{ Ker } Y) \cap \text{ Img } R^{\mathrm{H}}$. Since $\tilde{Y} \in \mathcal{K}$, it holds

$$u^{\mathrm{H}}\tilde{Y}u > 0.$$

This implies that for any $\epsilon > 0$,

$$u^{\mathrm{H}}(Y - \epsilon\tilde{Y})u = -\epsilon\, u^{\mathrm{H}}\tilde{Y}u < 0,$$

which contradicts the fact that $Y \in$ rel $(\mathcal{A} \cap \mathcal{K})$. **Q.E.D.**

Lemma 2.14 *Let R and $\mathcal{K}$ be as in Lemma 2.13, and let $Y \in \mathcal{A} \cap \mathcal{K}$. If*

$$W \in (Y + \text{ HKer } (R)) \cap \mathcal{H}^{(\bar{n})}_+$$

then

$$\text{Ker } W \supseteq (\text{ Ker } Y) \cap \text{ Img } R^H, \tag{2.30}$$

with equality holding if and only if $W \in$ rel $((Y + \text{ HKer } (R)) \cap \mathcal{H}^{(\bar{n})}_+)$.

Proof. First, we remark that since W is positive semidefinite, we have

$$u \in \text{ Ker } W \iff u^{\mathrm{H}}Wu = 0.$$

Since $W \in Y +$ HKer (R) and $Y \in \mathcal{K}$, it holds

$$u^{\mathrm{H}} W u = u^{\mathrm{H}} Y u \geq 0 \quad \forall u \in \text{ Img } R^{\mathrm{H}}, \tag{2.31}$$

with $u^{\mathrm{H}} Y u = 0$ if and only if $u \in (\text{ Ker } Y) \cap \text{ Img } R^{\mathrm{H}}$. This proves the conclusion (2.30). Now consider $v \in (\text{ Img } R^{\mathrm{H}})^{\perp} = \text{ Ker } R$, $v \neq 0$, and notice that

$$vv^{\mathrm{H}} \in \text{ HKer } (R).$$

Hence,

$$W + \lambda vv^{\mathrm{H}} \in (Y + \text{ HKer } (R)) \cap \mathcal{H}_+^{(\bar{n})} \text{ for all } \lambda \geq 0.$$

For $W \in$ rel $((Y +$ HKer $(R)) \cap \mathcal{H}_+^{(\bar{n})})$, it thus follows that $v^{\mathrm{H}} W v > 0$. Consequently, Ker $W \subseteq$ Img R^{H}. Together with (2.31), we obtain

$$\text{Ker } W = (\text{ Ker } Y) \cap \text{ Img } R^{\mathrm{H}}.$$

Q.E.D.

We arrive now at the central result of this section, viz., if $\mathcal{K}$ is of the form (2.29), then so is $\Gamma_{\mathcal{A}} \mathcal{K}$.

Theorem 2.9 *If* $\mathcal{K} = \mathcal{H}_+^{(\bar{n})} \oplus$ HKer (R), *then*

$$\Gamma_{\mathcal{A}} \mathcal{K} = \mathcal{H}_+^{(\bar{n})} \oplus \text{ HKer } (\tilde{R}),$$

where $\tilde{R}$ *is any matrix satisfying*

$$\begin{cases} \text{Img } \tilde{R}^H = (\text{ Img } R^H) \cap \text{ Ker } Y, \textit{ for some } Y \in \text{ rel } (\mathcal{A} \cap \mathcal{K}) \\ \tilde{R}\tilde{R}^H = I. \end{cases}$$

Proof. Suppose that $X \in \Gamma_{\mathcal{A}} \mathcal{K}$, i.e. $X = \lim_{i \to \infty} (X^{(i)} - Y^{(i)})$ for some sequences $X^{(1)}, X^{(2)}, \ldots$ and $Y^{(1)}, Y^{(2)}, \ldots$ in $\mathcal{K}$ and $\mathcal{A} \cap \mathcal{K}$ respectively. We know from Lemma 2.13 and the definition of $\tilde{R}$ that Img $\tilde{R}^{\mathrm{H}} \subseteq$ (Img $R^{\mathrm{H}}) \cap$ Ker $Y^{(i)}$, so that

$$\tilde{R}(X^{(i)} + Y^{(i)})\tilde{R}^{\mathrm{H}} = \tilde{R} X^{(i)} \tilde{R}^{\mathrm{H}} \succeq 0$$

for all $i \in \{1, 2, \ldots\}$. The above relation shows that $X \in \mathcal{H}_+^{(\bar{n})} \oplus$ HKer $(\tilde{R})$, from which we conclude that $\Gamma_{\mathcal{A}} \mathcal{K} \subseteq \mathcal{H}_+^{(\bar{n})} \oplus$ HKer $(\tilde{R})$. To prove the converse inclusion, consider a matrix $X \in \mathcal{H}_+^{(\bar{n})} \oplus$ HKer $(\tilde{R})$. Without loss of generality, we may assume that there exists a matrix Q such that

$$R^{\mathrm{H}} = \begin{bmatrix} Q^{\mathrm{H}}, & \tilde{R}^{\mathrm{H}} \end{bmatrix}.$$

We partition RYR^{H}, $Y \in$ rel $(\mathcal{A} \cap \mathcal{K})$ as follows:

$$RYR^{\mathrm{H}} = \begin{bmatrix} Q \\ \tilde{R} \end{bmatrix} Y \begin{bmatrix} Q^{\mathrm{H}}, & \tilde{R}^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{12}^{\mathrm{H}} & Y_{22} \end{bmatrix}.$$

By definition of $\tilde{R}$ and using the fact that $RYR^{\mathrm{H}} \succeq 0$, it follows that

$$Y_{11} \succ 0, \quad Y_{12} = 0, \quad Y_{22} = 0.$$

Similarly, we partition the matrix RXR^{H} as follows:

$$RXR^{\mathrm{H}} = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^{\mathrm{H}} & X_{22} \end{bmatrix}.$$

Since $X \in \mathcal{H}_+^{(\bar{n})} \oplus \text{ HKer } (\tilde{R})$, we have $X_{22} \succeq 0$. Let α_1 and α_2 be positive numbers such that

$$\alpha_1 Y_{11} + X_{11} \succeq 0, \quad \alpha_2 Y_{11} \succeq X_{12} X_{12}^{\mathrm{H}}.$$

(Such numbers exist, because Y_{11} is positive definite.) Then for any $\epsilon > 0$ there holds

$$R(X + \epsilon I + (\alpha_1 + \alpha_2/\epsilon)Y)R^{\mathrm{H}} \succeq 0.$$

Letting

$$X(\epsilon) := X + \epsilon I + (\alpha_1 + \alpha_2/\epsilon)Y,$$

it follows that $X(\epsilon) \in \mathcal{K}$ and

$$X = \lim_{\epsilon \downarrow 0} (X(\epsilon) - (\alpha_1 + \alpha_2/\epsilon)Y),$$

so that $X \in \Gamma_{\mathcal{A}} \mathcal{K}$. **Q.E.D.**

We already observed that

$$\Gamma_{\mathcal{A}}^0 \mathcal{H}_+^{(\bar{n})} = \mathcal{H}_+^{(\bar{n})} = \mathcal{H}_+^{(\bar{n})} \oplus \text{ HKer } (I).$$

With an inductive argument, it thus follows from Theorem 2.9 that there exist matrices $R^{(1)},\ R^{(2)}, \ldots$ such that

$$\Gamma_{\mathcal{A}}^k \mathcal{H}_+^{(\bar{n})} = \mathcal{H}_+^{(\bar{n})} \oplus \text{ HKer } (R^{(k)}),$$

for $k = 1, 2, \ldots$. Notice also from Theorem 2.9 that $\Gamma_{\mathcal{A}} \mathcal{K} \neq \mathcal{K}$ if and only if $\text{rank } \tilde{R} < \text{rank } R$. Together with Lemma 2.9, this implies that

$$\Gamma_{\mathcal{A}}^k \mathcal{K} = \Gamma_{\mathcal{A}}^\infty \mathcal{K} \quad \text{ for all } k \geq \min\{\bar{n},\ \dim \mathcal{A}\}. \tag{2.32}$$

It should be noted that $\bar{n}$ can be considerably smaller than $\dim \mathcal{A}$.

The following lemma shows the interesting fact that the linear subspace $\text{HKer } (\tilde{R})$, where $\tilde{R}$ is defined as in Theorem 2.9, can be modeled by semidefinite constraints.

Lemma 2.15 *Let R and $\tilde{R}$ as in Theorem 2.9. There holds*

$$\text{HKer } (\tilde{R}) = \left\{ W_{12} + W_{12}^H \;\middle|\; \begin{bmatrix} W_{11} & W_{12} \\ W_{12}^H & I \end{bmatrix} \succeq 0, W_{11} + U \in \mathcal{A}, U \in \text{ HKer } (R) \right\}.$$

Proof. We first notice that

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{12}^{\mathrm{H}} & I \end{bmatrix} \succeq 0 \iff W_{12}W_{12}^{\mathrm{H}} \preceq W_{11}, \tag{2.33}$$

see Section A.4. Consider W_{11}, W_{12}, W_{22} and U satisfying

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{12}^{\mathrm{H}} & I \end{bmatrix} \succeq 0, \quad W_{11} + U \in \mathcal{A}, \quad U \in \mathrm{\ HKer\ }(R). \tag{2.34}$$

Let $Y = W_{11} + U$, then

$$W_{11} \in (Y + \mathrm{\ HKer\ }(R)) \cap \mathcal{H}_+^{(\bar{n})}, \quad Y \in \mathcal{A} \cap \left(\mathcal{H}_+^{(\bar{n})} \oplus \mathrm{\ HKer\ }(R)\right).$$

Using respectively Lemma 2.13, Lemma 2.14 and (2.33), we obtain

$$\mathrm{Img\ } \tilde{R}^{\mathrm{H}} \subseteq (\mathrm{\ Img\ } R^{\mathrm{H}}) \cap \mathrm{\ Ker\ } Y \subseteq \mathrm{\ Ker\ } W_{11} \subseteq \mathrm{\ Ker\ } W_{12}^{\mathrm{H}},$$

so that $W_{12}^{\mathrm{H}}\tilde{R}^{\mathrm{H}} = (\tilde{R}W_{12})^{\mathrm{H}} = 0$, and

$$W_{12} + W_{12}^{\mathrm{H}} \in \mathrm{\ HKer\ }(\tilde{R}).$$

Conversely, suppose that $X \in \mathrm{\ HKer\ }(\tilde{R})$. Since $\tilde{R}\tilde{R}^{\mathrm{H}} = I$ there exists some matrix Q such that $\begin{bmatrix} Q^{\mathrm{H}}, & \tilde{R}^{\mathrm{H}} \end{bmatrix}$ is unitary. By definition, $X \in \mathrm{\ HKer\ }(\tilde{R})$ means that

$$\begin{bmatrix} Q \\ \tilde{R} \end{bmatrix} X \begin{bmatrix} Q^{\mathrm{H}}, & \tilde{R}^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^{\mathrm{H}} & 0 \end{bmatrix},$$

for some X_{11} and X_{12}. Letting

$$W_{12} := \begin{bmatrix} Q^{\mathrm{H}}, & \tilde{R}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} X_{11}/2 & X_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q \\ \tilde{R} \end{bmatrix},$$

it follows that

$$X = W_{12} + W_{12}^{\mathrm{H}}, \quad \mathrm{Img\ } \tilde{R}^{\mathrm{H}} \subseteq \mathrm{\ Ker\ } W_{12}^{\mathrm{H}}.$$

Since $\mathrm{Img\ } \tilde{R}^{\mathrm{H}} = (\mathrm{\ Img\ } R) \cap \mathrm{\ Ker\ } Y$ for some $Y \in \mathrm{\ rel\ } (\mathcal{A} \cap (\mathcal{H}_+^{\bar{n}} \oplus \mathrm{\ HKer\ }(R)))$, we know from Lemma 2.14 and the above inclusion that

$$\mathrm{Ker\ } \bar{W}_{11} = \mathrm{\ Img\ } \tilde{R}^{\mathrm{H}} \subseteq \mathrm{\ Ker\ } W_{12}^{\mathrm{H}}$$

for any $\bar{W}_{11} \in \mathrm{\ rel\ } ((Y + \mathrm{\ HKer\ }(R)) \cap \mathcal{H}_+^{(\bar{n})})$. Letting $\bar{U} = Y - \bar{W}_{11}$, it follows that

$$\begin{bmatrix} \alpha\bar{W}_{11} & W_{12} \\ W_{12}^{\mathrm{H}} & I \end{bmatrix} \succeq 0, \quad \bar{W}_{11} + \bar{U} \in \mathcal{A}, \quad \bar{U} \in \mathrm{\ HKer\ }(R)$$

for sufficiently large $\alpha > 0$. Letting $W_{11} := \alpha \bar{W}_{11}$ and $U := \alpha \bar{U}$, we see that (2.34) is satisfied. **Q.E.D.**

Consider the kth regularized semidefinite program

$$\inf\{C \bullet X \mid X \in (B + \mathcal{A}) \cap \Gamma^k_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}} \mathcal{H}_+^{(\bar{n})}\}. \tag{2.35}$$

We already know from (2.32) that for all $k = 0, 1, 2, \ldots$, there exist $R^{(k)}$ such that

$$\Gamma^k_{\mathcal{A}\cap \text{ Ker } c^{\mathrm{T}}} \mathcal{H}_+^{(\bar{n})} = \mathcal{H}_+^{(\bar{n})} \oplus \text{ HKer } (R^{(k)}).$$

Using Lemma 2.15, it follows that (2.35) is equivalent to

$$\begin{aligned}
\inf \quad & C \bullet (X + W_{12}^{(k)} + (W_{12}^{(k)})^{\mathrm{H}}) \\
\text{s.t.} \quad & X + W_{12}^{(k)} + (W_{12}^{(k)})^{\mathrm{H}} \in B + \mathcal{A} \\
& \begin{bmatrix} W_{11}^{(k)} & W_{12}^{(k)} \\ (W_{12}^{(k)})^{\mathrm{H}} & I \end{bmatrix} \succeq 0, \quad X \succeq 0, \\
& W_{11}^{(k)} + U \in \mathcal{A} \cap \text{ Ker } c^{\mathrm{T}}, \quad U \in \text{ HKer } (R^{(k-1)}).
\end{aligned}$$

With a recursive argument, we obtain

$$\begin{aligned}
\inf \quad & C \bullet (X + W_{12}^{(k)} + (W_{12}^{(k)})^{\mathrm{H}}) \\
\text{s.t.} \quad & X + W_{12}^{(k)} + (W_{12}^{(k)})^{\mathrm{H}} \in B + \mathcal{A} \\
& X \succeq 0, \\
& \begin{bmatrix} W_{11}^{(i)} & W_{12}^{(i)} \\ (W_{12}^{(i)})^{\mathrm{H}} & I \end{bmatrix} \succeq 0 \quad \text{for } i = 1, 2, \ldots, k \\
& W_{11}^{(i)} + W_{12}^{(i-1)} + (W_{12}^{(i-1)})^{\mathrm{H}} \in \mathcal{A} \cap \text{ Ker } c^{\mathrm{T}} \quad \text{for } i = 2, 3, \ldots, k \\
& W_{11}^{(1)} \in \mathcal{A} \cap \text{ Ker } c^{\mathrm{T}},
\end{aligned} \tag{PRAM}$$

which is again a semidefinite program. This regularized program was proposed by Ramana [124] for the real symmetric case, see also [126, 127]. Moreover, Ramana uses $k = \dim \mathcal{A}^{\perp} = \bar{n}^2 - \dim \mathcal{A}$, whereas we show that $k = \min(\bar{n}, \dim \mathcal{A})$ is sufficient. It is important to note that checking the feasibility of a solution for Ramana's regularized semidefinite program is easy, because it involves only linear and positive semidefiniteness constraints.

Remark 2.1 *The introduction of auxiliary variables $W^{(k)}$ into the regularized program has also disadvantages. In particular, the duality relation of (D) and its Ramana dual*

(PRAM) is asymmetric, since Ramana's dualization scheme increases the dimension of the problem. In order to regain symmetricity, Ramana and Freund [125] propose to consider the primal–dual pair of (PRAM) and its standard dual semidefinite programming problem. Since the subvalue of (PRAM) is equal to its optimal value, it follows from (2.6) that this primal–dual pair is again in perfect duality, and this fact is known from Ramana and Freund [125]. However, as recently noticed by De Klerk, Roos and Terlaky [71], it is possible that the dual of (PRAM) is weakly (in)feasible, and we can therefore not obtain results as in Theorem 2.8 for the primal–dual pair of Ramana and Freund.

2.8 Inexact Dual Solutions

As pointed out by Nesterov and Nemirovsky [110], interior point methods are well suited for solving conic convex programs. Although interior point methods typically require the existence of primal and dual interior solutions, it is possible to solve conic programs that are not strongly feasible by using the self–dual embedding technique, see Chapter 4. With (P) being a nonlinear program, it is not surprising that the interior point methods (or indeed any other methods) require an infinite number of iterations to obtain an exact solution. Within a finite number of iterations these iterative methods can only compute an approximate solution of (P). Naturally such an approximate solution of (P) can be interpreted as an exact solution of a perturbed problem (backward error analysis). However, this interpretation is of little practical use. In what follows, we show that an approximate solution of (P) can be used to infer many useful properties of the original conic program (P) such as 'approximate infeasibility'.

In the analysis of approximate solutions, it is convenient to add a variable which measures the constraint violation. A good way to construct such a variable is by making use of the *norm* cone, which is defined as follows:

$$\mathcal{K}_{\text{norm}} := \{(x_0, x) \in \Re_+ \times \Re^n \mid x_0 \geq \|x\|\}.$$

Using the basic properties of norms, it is easily seen that $\mathcal{K}_{\text{norm}}$ is a closed, pointed and solid convex cone. Moreover, it follows from the definition of dual norms that

$$\mathcal{K}^*_{\text{norm}} = \{(x_0, x) \in \Re_+ \times \Re^n \mid x_0 \geq \|x\|^*\}.$$

The theorem below shows that if we have an approximate primal improving direction, viz. some $x \in \mathcal{A}$ such that $c^{\mathrm{T}}x = -1$ and x 'almost' in $\mathcal{K}$, then the dual cannot have any 'reasonably' sized feasible solution.

Theorem 2.10 *Consider a conic convex program CP$(b, c, \mathcal{A}, \mathcal{K})$. There holds*

$$\inf_{s \in \mathcal{F}_D} \|s\|^* = \sup\{-c^T x \mid x \in \mathcal{A},\ \operatorname{dist}(x, \mathcal{K}) \leq 1\}.$$

Proof. Construct the conic convex program

$$\mathrm{CP}\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\c\end{bmatrix},\{0\}\times\mathcal{A},(\{0\}\times\mathcal{K})\oplus\mathcal{K}_{\mathrm{norm}}\right), \tag{2.36}$$

which can be written as

$$\inf\{c^{\mathrm{T}}x \mid x\in\mathcal{A},\ x-u\in\mathcal{K},\ \|u\|\le x_0=1\}=\inf\{c^{\mathrm{T}}x \mid x\in\mathcal{A},\ \operatorname{dist}(x,\mathcal{K})\le 1\}.$$

Using Lemma 2.1, it follows that the dual of the conic convex program (2.36) is

$$\inf\{z_0 \mid z\in(c+\mathcal{A}^{\perp})\cap\mathcal{K}^*,\ \|z\|^*\le z_0\}=\inf\{\|z\|^* \mid s\in\mathcal{F}_D\}.$$

Notice now that (2.36) is primal strongly feasible, because it has a trivial interior solution $\begin{bmatrix}1, & 0\end{bmatrix}^{\mathrm{T}}\in\operatorname{int}\mathcal{K}_{\mathrm{norm}}$. Theorem 2.7 is therefore applicable, and it yields

$$\inf_{z\in\mathcal{F}_D}\|z\|^*=-\inf\{c^{\mathrm{T}}x \mid x\in\mathcal{A},\ \operatorname{dist}(x,\mathcal{K})\le 1\}.$$

Q.E.D.

Remark 2.2 *For the case of ℓ_p norms and $\mathcal{K}=\Re^n_+$, the statement of Theorem 2.10 is known from Todd and Ye [153].*

Remark 2.3 *Suppose that we have an approximate solution $\hat{x}$, with*

$$\operatorname{dist}(\hat{x},\mathcal{K})<\epsilon_1,\qquad \operatorname{dist}(\hat{x},\mathcal{A})<\epsilon_2.$$

Let Δx be such that $\hat{x}+\Delta x\in\mathcal{A}$ and $\|\Delta x\|<\epsilon_2$. Then, we can invoke Theorem 2.10 with $x:=(\hat{x}+\Delta x)/(\epsilon_1+\epsilon_2)$ to conclude that

$$\inf_{s\in\mathcal{F}_D}\|s\|^*\ge -c^Tx\ge\frac{-c^T\hat{x}}{\epsilon_1+\epsilon_2}-\frac{\epsilon_2}{\epsilon_1+\epsilon_2}\|c\|^*,$$

With an approximate Farkas–type dual solution, we can now deduce a lower bound on the norm of primal feasible solutions. However, we cannot conclude primal infeasibility, unless the Farkas–type dual solution is exact (which actually implies primal strong infeasibility). This is not surprising, since the distinction between weakly feasible problems and weakly infeasible problems is a delicate issue. In the terminology of Renegar [131], such problems are ill–posed. Renegar [130, 131] also defined a condition number for conic convex programs, that is based on the distance to ill-posedness, see also Epelman and Freund [34].

We will now show that based on an approximate primal solution, viz. some $x\in b+\mathcal{A}$ such that x is 'almost' in $\mathcal{K}$, we obtain a lower bound on the objective value of any 'reasonably' sized dual feasible solution. To the best of our knowledge, this result (Theorem 2.11) is new.

Theorem 2.11 *Consider a conic convex program $CP(b,c,\mathcal{A},\mathcal{K})$. For all $\gamma \in \Re$, there holds*

$$\begin{aligned}&\inf\{\|z\|^* \mid z \in \mathcal{F}_D,\, b^T z \le \gamma\}\\ &= \sup\{-(c^T x + \gamma x_{n+1}) \mid x \in x_{n+1} b + \mathcal{A},\ \operatorname{dist}(x,\mathcal{K}) \le 1,\, x_{n+1} \ge 0\}.\end{aligned}$$

Proof. Recall from (2.15) and (2.18) that

$$M_b := \begin{bmatrix} I & -b \\ 0 & 1 \end{bmatrix}, \quad M_b^{-1} = \begin{bmatrix} I & b \\ 0 & 1 \end{bmatrix}. \tag{2.37}$$

Now, we use a similar argumentation as in the proof of Theorem 2.10. First, construct the conic convex program

$$\mathrm{CP}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ c \\ \gamma \end{bmatrix}, \{0\} \times M_b^{-1}(\mathcal{A} \times \Re), ((\{0\} \times \mathcal{K}) \oplus \mathcal{K}_{\mathrm{norm}}) \times \Re_+ \right), \tag{2.38}$$

which can be written as

$$\begin{aligned}&\inf\{c^{\mathrm{T}} x + \gamma x_{n+1} \mid x - x_{n+1} b \in \mathcal{A},\, x - u \in \mathcal{K},\, \|u\| \le x_0 = 1,\, x_{n+1} \ge 0\}\\ &= \inf\{c^{\mathrm{T}} x + \gamma x_{n+1} \mid x \in x_{n+1} b + \mathcal{A},\ \operatorname{dist}(x,\mathcal{K}) \le 1,\, x_{n+1} \ge 0\}.\end{aligned}$$

Using Lemma 2.1, it follows that the dual of the conic convex program (2.36) is

$$\inf\{z_0 \mid z \in (c+\mathcal{A}^\perp)\cap\mathcal{K}^*, z_{n+1} = \gamma - b^{\mathrm{T}} s \ge 0, z_0 \ge \|z\|^*\} = \inf\{\|z\|^* \mid z \in \mathcal{F}_D,\, b^{\mathrm{T}} z \le \gamma\}.$$

If $b = 0$, then (2.38) has the trivial interior solution $\begin{bmatrix} 1, & 0, & 1 \end{bmatrix}^{\mathrm{T}} \in (\operatorname{int} \mathcal{K}_{\mathrm{norm}}) \times \Re_{++}$. And if $b \ne 0$, then $\begin{bmatrix} 1, & b^{\mathrm{T}}/\|b\|, & 1/\|b\| \end{bmatrix}^{\mathrm{T}} \in (\operatorname{int} \mathcal{K}_{\mathrm{norm}}) \times \Re_{++}$ is a primal interior feasible solution. Hence, (2.38) is primal strongly feasible. Theorem 2.7 is therefore applicable, and it yields

$$\begin{aligned}&\inf\{\|z\|^* \mid z \in \mathcal{F}_D,\, b^{\mathrm{T}} z \le \gamma\} =\\ &\quad -\inf\{c^{\mathrm{T}} x + \gamma x_{n+1} \mid x \in x_{n+1} b + \mathcal{A},\ \operatorname{dist}(x,\mathcal{K}) \le 1,\, x_{n+1} \ge 0\}.\end{aligned}$$

Q.E.D.

Suppose that $x \in b + \mathcal{A}$. If $x \in \mathcal{K}$, then $d^* \ge -c^{\mathrm{T}} x$, as we already knew from (2.20). If $\operatorname{dist}(x,\mathcal{K}) > 0$, then x is an approximate solution, and we obtain from Theorem 2.11 that

$$\inf\{\|z\|^* \mid z \in \mathcal{F}_D,\, b^{\mathrm{T}} z \le \gamma\} \ge \frac{-c^{\mathrm{T}} x - \gamma}{\operatorname{dist}(x,\mathcal{K})}.$$

Remark that Theorem 2.10 follows from Theorem 2.11 by letting $\gamma \to \infty$. Theorem 2.6 can also be seen as an application of Theorem 2.11 (the converse is true as well, as has just been demonstrated).

In semidefinite programming, a natural norm is the matrix 2-norm, i.e. the maximal singular value of a matrix. In order to apply Theorem 2.10 and 2.11, we should be able to compute the dual of this matrix norm. Fortunately, this norm can easily be computed for Hermitian matrices, using the eigenvalue decomposition. In particular, if $X \in \mathcal{H}$ and $X = Q^{\mathrm{H}}\Lambda_X Q$ is the eigenvalue decomposition of X, then

$$\|X\|_2^* = \max_{\|Y\|_2 \le 1} X \bullet Y = \max_{\|Y\|_2 \le 1} \operatorname{tr} \Lambda_X Q^{\mathrm{H}} Y Q.$$

Notice now that $\|Y\|_2 = \|Q^{\mathrm{H}} Y Q\|_2$, since Q is unitary. Hence,

$$\|X\|_2^* = \max_{\|Y\|_2 \le 1} \operatorname{tr} \Lambda_X Y = \max_{\|Y\|_2 \le 1} \sum_{j=1}^{\bar{n}} (\Lambda_X)_{jj} Y_{jj} = \sum_{j=1}^{\bar{n}} |(\Lambda_X)_{jj}|,$$

which is the ℓ_1 norm of the eigenvalues of X. Remark also that the Frobenius norm in $\mathcal{H}^{(\bar{n})}$ coincides with the Euclidean vector norm in $\Re^n$, so that $\|X\|_F^* = \|X\|_F$.

2.9 Discussion

We have treated conic convex programming duality in a unified fashion. Special attention has been given to conic convex programs that do not satisfy constraint qualifications. It has also been shown how recent duality approaches of [18, 19, 124, 127, 160] fit into the framework. Elaborating on the results of [153], we have also discussed the value of approximate dual solutions.

We believe that duality results without constraint qualifications have not received enough attention in the past. It is our hope that this thesis will help popularize these results in future. In Chapter 4, we show that this type of duality relation can be used fruitfully in the design of algorithms whose convergence is guaranteed even in the absence of constraint qualifications.

Our survey is restricted to conic convex programming in finite dimensional real linear spaces. As such, it includes conic convex programming with complex numbers, if a real inner product is used. For instance, we can treat $\mathcal{C}^n$ as a $2n$-dimensional *real* linear space by using the real valued inner product Re $z^{\mathrm{H}}x$. However, due to the lack of ordering of complex numbers, there is no obvious way to generalize duality results to *complex* linear spaces (ordering is crucial in the definition of convex cones, among others). Duality results for convex programming in infinite dimensional real linear spaces have not been discussed in this chapter. The strong duality result of Theorem 2.6 can be generalized to semi–infinite linear and convex programming, see the collective work of [16, 17, 32, 64, 66, 67, 160]. Results for conic convex programming with infinitely many variables and a bounded feasible set are given in [26]; see also the books [5, 39].

Chapter 3

Polynomiality of Path–following Methods

Modern solvers for linear and quadratic programming often incorporate some variant of the primal–dual path–following method, which is a class of interior point methods. This method turns out to be suited for solving semidefinite programming problems as well. In this chapter, we propose a framework for analyzing primal–dual path–following algorithms for semidefinite programming. Our framework is a direct extension of the v–space approach, which was developed by Kojima et al. [73] in the context of complementarity problems.

Four basic variants of the path–following method are described in this chapter. All these variants share the strong property that the maximum number of main iterations needed for computing an approximate solution grows at most polynomially in the problem size.

Interior point algorithms for semidefinite programming have been studied intensively in recent years. We give an overview of the literature in Section 3.6.

3.1 Primal–Dual Transformations

In this section, we give an interpretation of the so–called V–space solution that is associated with a pair of primal and dual matrix variables. A good understanding of V–space solutions is crucial for developing and analyzing efficient primal–dual interior point methods for semidefinite programming.

In conic form, semidefinite programming problems are stated as

$$(P) \quad \inf\{C \bullet X \mid X \in (B + \mathcal{A}) \cap \mathcal{H}_+\},$$

where $B \in \mathcal{H}$ and $C \in \mathcal{H}$ are given Hermitian matrices, $\mathcal{A} \subseteq \mathcal{H}$ is a linear subspace of $\mathcal{H}$, and $X \in \mathcal{H}$ is the decision variable. Let $\bar{n}$ denote the order of B, C and X, i.e. B, C and X are $\bar{n} \times \bar{n}$ matrices.

A well known property of the semidefinite cone $\mathcal{H}_+$ is that for any invertible $\bar{n} \times \bar{n}$ matrix L, it holds

$$L^{-1} X L^{-\mathrm{H}} \in \mathcal{H}_+ \iff X \in \mathcal{H}_+.$$

The above relation can be written more concisely by using Hermitian Kronecker product notation, as follows:

$$(L^{-1} \otimes_{\mathrm{H}} L^{-1})\mathcal{H}_+ = \mathcal{H}_+.$$

Whereas the linear transformation $L^{-1} \otimes_{\mathrm{H}} L^{-1}$ does not affect the the cone $\mathcal{H}_+$, it transforms the linear subspace $\mathcal{A}$ into a different linear subspace, denoted by $\mathcal{A}(L) := (L^{-1} \otimes_{\mathrm{H}} L^{-1})\mathcal{A}$, i.e.

$$\mathcal{A}(L) = \{X \mid L X L^{\mathrm{H}} \in \mathcal{A}\}.$$

Obviously, X is a feasible (optimal) solution of (P) if and only if $\bar{X} := L^{-1} X L^{-\mathrm{H}}$ is a feasible (optimal) solution of the transformed semidefinite program

$$(\mathrm{P}_L) \quad \inf\{(L^{\mathrm{H}} C L) \bullet \bar{X} \mid \bar{X} \in (L^{-1} B L^{-\mathrm{H}} + \mathcal{A}(L)) \cap \mathcal{H}_+\}.$$

We conclude that the programs $\mathrm{CP}(B, C, \mathcal{A}, \mathcal{H}_+)$ and $\mathrm{CP}(L^{-1} B L^{-\mathrm{H}}, L^{\mathrm{H}} C L, \mathcal{A}, \mathcal{H}_+)$ are equivalent.

Associated with (P_L) is the dual semidefinite program

$$(\mathrm{D}_L) \quad \inf\{(L^{-1} B L^{-\mathrm{H}}) \bullet \bar{Z} \mid \bar{Z} \in (L^{\mathrm{H}} C L + \mathcal{A}^{\perp}(L)) \cap \mathcal{H}_+\},$$

where, using Lemma 2.3,

$$\mathcal{A}^{\perp}(L) = (L^{\mathrm{H}} \otimes_{\mathrm{H}} L^{\mathrm{H}})\mathcal{A}^{\perp} = \{\bar{Z} \mid L^{-\mathrm{H}} \bar{Z} L^{-1} \in \mathcal{A}^{\perp}\}.$$

The feasible solution sets of (P_L) and (D_L) are denoted by

$$\mathcal{F}_P(L) := (L^{-1} B L^{-\mathrm{H}} + \mathcal{A}(L)) \cap \mathcal{H}_+$$

and

$$\mathcal{F}_D(L) := (L^{\mathrm{H}} C L + \mathcal{A}^{\perp}(L)) \cap \mathcal{H}_+,$$

respectively. Similarly, $\overset{o}{\mathcal{F}}_P(L)$ denotes the set of primal interior solutions,

$$\overset{o}{\mathcal{F}}_P(L) := (L^{-1} B L^{-\mathrm{H}} + \mathcal{A}(L)) \cap \mathcal{H}_{++},$$

and $\overset{o}{\mathcal{F}}_D(L)$ denotes the set of dual interior solutions,

$$\overset{o}{\mathcal{F}}_D(L) := (L^{\mathrm{H}} C L + \mathcal{A}^{\perp}(L)) \cap \mathcal{H}_{++}.$$

Notice that $\mathcal{F}_P = \mathcal{F}_P(I)$ and $\mathcal{F}_D = \mathcal{F}_D(I)$.

We assume that $\mathrm{CP}(B, C, \mathcal{A}, \mathcal{H}_+)$ is both primal and dual strongly feasible, which means that $\overset{o}{\mathcal{F}}_P(L) \neq \emptyset$ and $\overset{o}{\mathcal{F}}_D(L) \neq \emptyset$ for all invertible L. We have seen in Chapter 2 that this assumption guarantees the existence of a complementary solution, which is by definition a solution $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ such that $X \bullet Z = 0$. Recall that for given $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$, the quantity $X \bullet Z$ is called the duality gap at the solution pair (X, Z), see (2.2). Solving (P) is now equivalent to the duality gap minimization problem,

$$\min\{X \bullet Z \mid X \in \mathcal{F}_P,\ Z \in \mathcal{F}_D\},$$

for which the optimal value is zero. We remark that the duality gap $X \bullet Z$ is invariant under the transformation $L^{-1} \otimes_{\mathrm{H}} L^{-1}$, viz. if $\bar{X} = L^{-1}XL^{-\mathrm{H}}$ and $\bar{Z} = L^{\mathrm{H}}ZL$, then

$$\bar{X} \bullet \bar{Z} = \operatorname{tr} \bar{X}\bar{Z} = \operatorname{tr} L^{-1}XL^{-\mathrm{H}}L^{\mathrm{H}}ZL = X \bullet Z. \tag{3.1}$$

Now consider $X \in \overset{o}{\mathcal{F}}_P$ and $Z \in \overset{o}{\mathcal{F}}_D$. From a primal point of view, an interesting choice for L is $L = X^{1/2}$, yielding

$$\bar{X} = L^{-1}XL^{-\mathrm{H}} = I.$$

Hence,

$$\bar{X} + \{\Delta\bar{X} \in \mathcal{H} \mid \left\|\Delta\bar{X}\right\|_F < 1\} \subset \mathcal{H}_{++}. \tag{3.2}$$

This means that we can take a large step in the transformed primal space without leaving the positive definite cone. The transformed dual solution $\bar{Z} = X^{1/2}ZX^{1/2}$ however, may be close to the boundary of $\mathcal{H}_{++}$. Based on this transformation, it is therefore not clear how the duality gap can be reduced substantially. One would rather like to use a transformation that is neither preoccupied with the primal problem, nor with the dual problem. Such a transformation will be called a symmetric primal-dual transformation.

A *symmetric primal-dual transformation* is by definition a transformation L_d such that

$$L_d^{-1}XL_d^{-\mathrm{H}} = L_d^{\mathrm{H}}ZL_d.$$

(The subscript ‘d’ is reminiscent to the standard notation used for primal–dual interior point methods in linear programming, where d denotes a primal–dual scaling vector.) The following lemma characterizes L_d.

Lemma 3.1 *Let $X \in \mathcal{H}_{++}$ and $Z \in \mathcal{H}_{++}$. An invertible $\bar{n} \times \bar{n}$ matrix L_d is a symmetric primal-dual transformation in the sense that*

$$L_d^{-1}XL_d^{-H} = L_d^HZL_d,$$

if and only if

$$L_dL_d^H = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}.$$

Proof. We have $L_d^{-1}XL_d^{-\mathrm{H}} = L_d^{\mathrm{H}}ZL_d$ if and only if

$$X = L_d L_d^{\mathrm{H}} Z L_d L_d^{\mathrm{H}} = Z^{-1/2}(Z^{1/2}L_d L_d^{\mathrm{H}} Z^{1/2})^2 Z^{-1/2},$$

or equivalently,

$$(Z^{1/2}L_d L_d^{\mathrm{H}} Z^{1/2})^2 = Z^{1/2}XZ^{1/2},$$

from which it follows that

$$L_d L_d^{\mathrm{H}} = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}.$$

Q.E.D.

Based on Lemma 3.1, we let

$$D(X,Z) := Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}, \tag{3.3}$$

for $X, Z \in \mathcal{H}_{++}$. It is now clear that L_d is a symmetric primal-dual transformation if and only if $L_d L_d^{\mathrm{H}} = D(X,Z)$. Obviously, L_d is not uniquely defined: possible choices for L_d include the Cholesky factor of $D(X,Z)$, in which case L_d is lower triangular, and the square root of $D(X,Z)$, in which case L_d is Hermitian. In fact, if L_d is a symmetric primal–dual transformation, then so is L_dQ for any unitary matrix Q of order $\bar{n}$. Since the transformed solution $L_d^{\mathrm{H}}ZL_d$ is Hermitian positive definite, we can choose Q in such a way that $Q^{\mathrm{H}}L_d^{\mathrm{H}}ZL_dQ$ is a positive diagonal matrix. Hence, there exists a symmetric primal–dual transformation L_d that transforms the primal and dual solutions X and Z into a positive diagonal matrix V, i.e.

$$V = L_d^{-1}XL_d^{-\mathrm{H}} = L_d^{\mathrm{H}}ZL_d. \tag{3.4}$$

Remark that the matrix XZ is similar to V^2, viz.

$$XZ = L_d(L_d^{-1}XL_d^{-\mathrm{H}})(L_d^{\mathrm{H}}ZL_d)L_d^{-1} = L_dV^2L_d^{-1}.$$

This implies that the positive entries of the diagonal matrix V^2 are exactly the eigenvalues of the matrix XZ, see Appendix A. The order in which these eigenvalues appear on the diagonal of V^2 obviously depends on the choice of L_d.

The $\bar{n}\times\bar{n}$ matrices L_d and V can be computed numerically, with the following procedure:

Algorithm 3.1 (Symmetric primal–dual transformation)

1. *Compute Cholesky factor L_X of $X \in \mathcal{H}_{++}$,*

$$X = L_X L_X^H.$$

2. *Compute spectral decomposition* (Q, Λ_{XZ}) *of* $L_X^H Z L_X \in \mathcal{H}_{++}$,

$$L_X^H Z L_X = Q^H \Lambda_{XZ} Q.$$

3. *Let*

$$L_d = L_X Q^H \Lambda_{XZ}^{-1/4}, \quad V = \Lambda_{XZ}^{1/2}.$$

The above procedure consists of a Cholesky factorization of an $\bar{n} \times \bar{n}$ Hermitian matrix, and a spectral decomposition of an $\bar{n} \times \bar{n}$ Hermitian matrix; both decompositions involve $\mathbf{O}(\bar{n}^3)$ floating point operations.

Unlike the primal transformation $L_X L_X^{\mathrm{H}} = X$, the symmetric primal–dual transformation $L_d L_d^{\mathrm{H}} = D(X, Z)$ does not discriminate between the primal and the dual. However, the advantage of the primal transformation, viz. that it maps to the central solution I (see (3.2)), is also lost in general. Fortunately, a property that is similar to (3.2) does hold with the symmetric primal–dual transformation L_d, if the pair (X, Z) is on the *central path.*

Definition 3.1 *The primal-dual central path is the set*

$$\mathrm{CPATH} := \left\{ (X, Z) \in \mathcal{F}_P \times \mathcal{F}_D \,\middle|\, XZ = \mu I,\ \mu = \frac{\mathrm{tr}\ XZ}{\bar{n}} \right\}.$$

We remark that the central path is invariant under transformations L, viz.

$$XZ = \mu I \iff \bar{X}\bar{Z} = L^{-1} XZL = \mu I,$$

where as before $\bar{X} := L^{-1} X L^{-\mathrm{H}}$ and $\bar{Z} = L^{\mathrm{H}} Z L$. In particular, for the symmetric primal–dual transformation L_d we have

$$V = L_d^{-1} X L_d^{-\mathrm{H}} = L_d^{\mathrm{H}} Z L_d = \sqrt{\mu} I,$$

which shows that V is a multiple of the identity matrix, if (X, Z) is on the central path. Similar to (3.2), we therefore have

$$V + \{\Delta V \in \mathcal{H} \mid \|\Delta V\|_F < \sqrt{\mu}\} \subset \mathcal{H}_{++} \quad \text{if } (X, Z) \in \text{ CPATH}.$$

Hence, a large step can be taken in the transformed primal space as well as in the transformed dual space if the current iterate is on the central path.

3.2 The Newton Direction

In the V-space treatment, the central object is the V–solution, where V^2 is a diagonal matrix of the eigenvalues of XZ. However, we steer the V-solution by updating the primal and dual solutions X and Z respectively. In order to determine sensible updates for the X and Z variables, it is of primary interest to know how changes in X and Z affect the associated V-space solution. In this section, we obtain such insight by deriving the relation between the derivatives of $(X(t), Z(t))$ trajectories, and the derivatives of corresponding $V(t)$ trajectories, where t is a step length parameter. In system theoretic terms, we may think of X and Z as inputs with output V, and our next step is to derive the differential equations that link the inputs with the output.

As before, we consider an interior solution pair (X, Z),

$$X \in (B + \mathcal{A}) \cap \mathcal{H}_{++}, \quad Z \in (C + \mathcal{A}^{\perp}) \cap \mathcal{H}_{++},$$

and we let V denote a corresponding diagonal V–space solution. In other words, V is a positive diagonal matrix satisfying

$$V = L_d^{-1} X L_d^{-\mathrm{H}} = L_d^{\mathrm{H}} Z L_d,$$

for some invertible matrix L_d. Suppose that $(X(t), Z(t))$ is a smooth trajectory of interior feasible primal–dual pairs, for t in some neighborhood of zero, and $X(0) = X$, $Z(0) = Z$. To this trajectory, we apply the linear transformation

$$\bar{X}(t) := L_d^{-1} X(t) L_d^{-\mathrm{H}}, \quad \bar{Z}(t) := L_d^{\mathrm{H}} Z(t) L_d,$$

which is the symmetric primal–dual transformation at $t = 0$. Obviously, it holds that

$$\bar{X}(0) = \bar{Z}(0) = V.$$

However, $\bar{X}(t)$ and $\bar{Z}(t)$ are in general not identical for $t \neq 0$. We will therefore apply a symmetric primal–dual transformation to the pair $(\bar{X}(t), \bar{Z}(t))$. Namely, we let $G(t)$ be a smooth trajectory such that $G(t)G(t)^{\mathrm{H}} = D(\bar{X}(t), \bar{Z}(t))$ and $G(0) = I$. Then $V(t) := G(t)^{\mathrm{H}} \bar{Z}(t) G(t)$ is a V-space trajectory for $(X(t), Z(t))$ with symmetric primal–dual transformation $L_d G(t)$. In particular,

$$G(t)^{-1} \bar{X}(t) G(t)^{-\mathrm{H}} = G(t)^{\mathrm{H}} \bar{Z}(t) G(t).$$

Remark that without further restrictions on $G(t)$, the V–space solution $V(t)$ is in general not diagonal for $t \neq 0$. Notice also that a feasible choice for $G(t)$ is $G(t) = D(\bar{X}(t), \bar{Z}(t))^{1/2}$. Implicit differentiation of the identity

$$V(t) = \frac{1}{2} G(t)^{\mathrm{H}} \bar{Z}(t) G(t) + \frac{1}{2} G(t)^{-1} \bar{X}(t) G(t)^{-\mathrm{H}}$$

yields

$$\begin{aligned} V^{[1]}(t) &= P_{\mathcal{H}}\left(G^{[1]}(t)^{\mathrm{H}}\bar{Z}(t)G(t) - G(t)^{-1}G^{[1]}(t)G(t)^{-1}\bar{X}(t)G(t)^{-\mathrm{H}}\right) \\ &\quad + \frac{1}{2}G(t)^{\mathrm{H}}\bar{Z}^{[1]}(t)G(t) + \frac{1}{2}G(t)^{-1}\bar{X}^{[1]}(t)G(t)^{-\mathrm{H}} \\ &= P_{\mathcal{H}}\left([G^{[1]}(t)^{\mathrm{H}}G(t)^{-\mathrm{H}} - G(t)^{-1}G^{[1]}(t)]V(t)\right) \\ &\quad + \frac{1}{2}G(t)^{\mathrm{H}}\bar{Z}^{[1]}(t)G(t) + \frac{1}{2}G(t)^{-1}\bar{X}^{[1]}(t)G(t)^{-\mathrm{H}}. \end{aligned} \tag{3.5}$$

Here, we use superscript $^{[1]}$ to denote first order derivatives, i.e.

$$V^{[1]}(t) = \frac{\mathrm{d}\, V(t)}{\mathrm{d}\, t},\ \ \bar{X}^{[1]}(t) = \frac{\mathrm{d}\, \bar{X}(t)}{\mathrm{d}\, t},\ \text{etc.}$$

(In (3.5), we used the identity

$$0 = \frac{\mathrm{d}\, G(t)G(t)^{-1}}{\mathrm{d}\, t} = G^{[1]}(t)G(t)^{-1} + G(t)\,\frac{\mathrm{d}\, G(t)^{-1}}{\mathrm{d}\, t},$$

i.e. $(\,\mathrm{d}\, G(t)^{-1}/\,\mathrm{d}\, t) = -G(t)^{-1}G^{[1]}(t)G(t)^{-1}$.)

Using the fact that $G(0) = I$, we obtain from (3.5) that

$$V^{[1]}(0) = \frac{1}{2}(\bar{Z}^{[1]}(0) + \bar{X}^{[1]}(0)) + P_{\mathcal{H}}([G^{[1]}(0)^{\mathrm{H}} - G^{[1]}(0)]V). \tag{3.6}$$

Moreover, differentiating the feasibility constraints

$$\bar{X}(t) \in V + \mathcal{A}(L_d), \quad \bar{Z}(t) \in V + \mathcal{A}^{\perp}(L_d),$$

we have

$$\bar{X}^{[1]}(0) \in \mathcal{A}(L_d), \quad \bar{Z}^{[1]}(0) \in \mathcal{A}^{\perp}(L_d).$$

In the sequel, we restrict to trajectories $\bar{X}(t)$ and $\bar{Z}(t)$ that are affine in the step length t, i.e.

$$\bar{X}(t) = V + tD_X, \quad \bar{Z}(t) = V + tD_Z,$$

where we let

$$D_X := \bar{X}^{[1]}(0), \quad D_Z := \bar{Z}^{[1]}(0)$$

denote the primal and dual search directions in the transformed linear spaces $\mathcal{A}(L_D)$ and $\mathcal{A}^{\perp}(L_d)$ respectively. These directions can be transformed back into directions $\Delta X \in \mathcal{A}$ and $\Delta Z \in \mathcal{A}^{\perp}$ by the relations

$$\Delta X = L_d D_X L_d^{\mathrm{H}}, \quad \Delta Z = L_d^{-\mathrm{H}} D_Z L_d^{-1}. \tag{3.7}$$

Furthermore, we choose a specific $G(t)$ trajectory, viz. $G(t) = D(\bar{X}(t), \bar{Z}(t))^{1/2}$. Since $G(t)$ is now Hermitian for all t, we have $G^{[1]}(0) \in \mathcal{H}$, and we can simplify (3.6) to

$$V^{[1]}(0) = \frac{1}{2}(\bar{Z}^{[1]}(0) + \bar{X}^{[1]}(0)). \tag{3.8}$$

We comment on this choice of $G(t)$ in Section 3.6.

Based on (3.8), we can compute the Newton direction (D_X, D_Z) towards a solution on the central path with a duality gap of $\bar{n}\gamma\mu$ for given $\gamma \in [0,1]$. Namely, we propose to linearize the nonlinear system

$$\begin{cases} V(1)^2 = \gamma\mu I \\ D_X \in \mathcal{A}(L_d), \quad D_Z \in \mathcal{A}^{\perp}(L_d). \end{cases}$$

Using (3.8) and the relation

$$\frac{\mathrm{d}\, V(t)^2}{\mathrm{d}\, t} = 2P_{\mathcal{H}}\left(V(t)V^{[1]}(t)\right),$$

this yields

$$P_{\mathcal{H}}(V(D_X + D_Z)) = \gamma\mu I - V^2, \tag{3.9}$$

$$D_X \in \mathcal{A}(L_d), \quad D_Z \in \mathcal{A}^{\perp}(L_d). \tag{3.10}$$

We remark that (3.9) is a Sylvester equation in $D_X + D_Z$; see e.g. Higham [54] for a discussion on Sylvester equations. Since V is a positive diagonal matrix, the equation is here particularly easy to solve. The solution is made explicit in the following lemma.

Lemma 3.2 *Let V be a positive diagonal matrix, and Y a Hermitian matrix of the same order. Then*

$$P_{\mathcal{H}}(VX) = Y \iff X_{ij} = \frac{2Y_{ij}}{V_{ii} + V_{jj}} \text{ for all } i, j.$$

The proof of Lemma 3.2 is straightforward, and hence omited. Here, the unique solution of (3.9) is

$$D_X + D_Z = \gamma\mu V^{-1} - V, \tag{3.11}$$

which is an $\bar{n} \times \bar{n}$ real diagonal matrix. From (3.10), the Newton direction (D_X, D_Z) now follows as an orthogonal decomposition of $D_X + D_Z$ onto $\mathcal{A}(L_d)$ and $\mathcal{A}^{\perp}(L_d)$ respectively. Therefore,

$$\|D_X\|_F^2 + \|D_Z\|_F^2 = \|D_X + D_Z\|_F^2. \tag{3.12}$$

From $V \perp (\mu V^{-1} - V)$, we obtain

$$\|D_X + D_Z\|_F^2 = \left\|\gamma(\mu V^{-1} - V) - (1-\gamma)V\right\|_F^2 = \gamma^2 \left\|\mu V^{-1} - V\right\|_F^2 + (1-\gamma)^2 \bar{n}\mu. \tag{3.13}$$

For the duality gap, we have

$$\|V(t)\|_F^2 = \operatorname{tr}\,(V + tD_X)(V + tD_Z),$$

so that using $D_X \perp D_Z$ and (3.11) it follows that

$$\|V(t)\|_F^2 = \|V\|_F^2 + tV \bullet (D_X + D_Z) = (1 - t + \gamma t)\,\|V\|_F^2,$$

i.e.

$$\mu(t) := \|V(t)\|_F^2 / \bar{n} = (1 - t + \gamma t)\mu. \tag{3.14}$$

The Newton equations (3.9)-(3.10) can be stated in terms of the untransformed variables, using (3.4) and (3.7) as follows. Pre-multiplying (3.11) with L_d and post-multiplying with L_d^{H} yields, using (3.7),

$$\Delta X + D(X, Z)\Delta Z D(X, Z) = \gamma \mu L_d V^{-1} L_d^{\mathrm{H}} - L_d V L_d^{\mathrm{H}}.$$

Applying (3.4) to the above relation, we obtain

$$\Delta X + D(X, Z)\Delta Z D(X, Z) = \gamma \mu Z^{-1} - X. \tag{3.15}$$

At first sight, the above relation is more appealing than (3.11), since it does not involve L_d and V. However, the transformation to the V–space as in (3.11) turns out to have many computational advantages, especially if the algorithms get more involved such as in Chapter 7.

It is important to note that equation (3.15) also follows (implicitly) from the primal-dual path-following scheme of Nesterov and Todd [112, 113] for self-scaled cones. Namely, in the treatment of Nesterov and Todd one may choose the logarithmically homogeneous barrier function $\log(\det X)$, and use $D(X, Z)$ as a so–called scaling point, to derive the direction (3.15). However, its motivation is then very different. See also Section 3.6.

3.3 Neighborhoods of the Central Path

In path–following methods, the choice of the parameter γ and the step length t has to be made in such a way that, starting from a well centered solution (X, Z), the resulting solution pair $(X(t), Z(t))$ is again close to the central path. In V–space terminology, we demand that the transformed solution $V(t)$ is close to a multiple of the identity matrix I. We introduce $\delta : \mathcal{H} \to \Re_+$ as a measure of distance to the central path as follows:

$$\delta(Y) := \left\| I - \frac{\bar{n}}{I \bullet Y^2} Y^2 \right\|_F .$$

In particular, we have

$$\delta(V(t)) = \left\| I - \frac{1}{\mu(t)} V(t)^2 \right\|_F ,$$

and

$$\mu(t)^2 \delta(V(t))^2 = \left\| V(t)^2 \right\|_F^2 - \bar{n}\mu(t)^2 = \operatorname{tr}\left[(X(t)Z(t))^2 \right] - \bar{n}\mu(t)^2. \tag{3.16}$$

It follows that the $\delta(V(t))$ measure is not influenced by the way in which we factorize the matrix $D(X(t), Z(t))$. In fact, $V(t)^2 - \mu(t)I$ is similar to $X(t)Z(t) - \mu(t)I$, viz.

$$L_d G(t)(V(t)^2 - \mu(t)I)G(t)^{-1} L_d^{-1} = X(t)Z(t) - \mu(t)I.$$

Hence, if we denote the eigenvalues of $X(t)Z(t)$ by $\lambda_1(t), \ldots, \lambda_{\bar{n}}(t) \in \Re$, then $\delta(V(t)) = \sum_{i=1}^{\bar{n}}(1 - \lambda_i(t)/\mu(t))^2$.

Based on the $\delta(Y)$ measure, we define a neighborhood of the central path as follows:

$$\mathcal{N}_2(\beta) := \{Y \in \mathcal{H}_{++} \mid \delta(Y) \leq \beta\},$$

where $0 < \beta < 1$ is a given constant. The name $\mathcal{N}_2$ is reminiscent of the classification of neighborhoods of the central path for linear programming by Mizuno, Todd and Ye [95]. If $V(t) \in \mathcal{N}_2(\beta)$, then obviously

$$(1+\beta)\mu(t)I \succeq V(t)^2 \succeq (1-\beta)\mu(t)I. \tag{3.17}$$

We also define a *wide neighborhood* of the central path, viz. the $\mathcal{N}_\infty^-$–neighborhood:

$$\mathcal{N}_\infty^-(\beta) := \left\{ Y \in \mathcal{H}_{++} \,\middle|\, Y^2 \succeq (1-\beta)\frac{I \bullet Y^2}{\bar{n}} I \right\}, \tag{3.18}$$

where $0 < \beta < 1$ is a given constant. Notice from (3.17) that

$$V(t) \in \mathcal{N}_2(\beta) \Longrightarrow V(t)^2 \succeq (1-\beta)\mu(t)I \Longrightarrow V(t) \in \mathcal{N}_\infty^-(\beta),$$

i.e. $\mathcal{N}_2(\beta) \subseteq \mathcal{N}_\infty^-(\beta)$, which explains why $\mathcal{N}_\infty^-(\beta)$ is called a *wide* neighborhood. For a discussion of neighborhoods of the central path in the context of linear programming, we refer to Mizuno, Todd and Ye [95] and Sturm and Zhang [146].

3.4 Technical Results

In this section, we derive some technical lemmas that will be used in proving the main results of this chapter.

We define

$$U(t) := P_{\mathcal{H}^\perp}((V + tD_X)(V + tD_Z))$$

and

$$W(t) := P_{\mathcal{H}}((V + tD_X)(V + tD_Z)).$$

Notice that $U(t)$ is skew-Hermitian, whereas $W(t)$ is Hermitian. This implies that

$$U(t) \perp W(t). \tag{3.19}$$

Lemma 3.3 *Suppose that $\delta(V) < 1$ and $0 \leq t < t^*$, where we let*

$$t^* = \max\{t \mid V + tD_X \succeq 0,\ V + tD_Z \succeq 0\}.$$

There holds

$$\delta(V(t))^2 = \left\| \frac{1}{\mu(t)} W(t) - I \right\|_F^2 - \frac{\|U(t)\|_F^2}{\mu(t)^2}.$$

Proof. As in (3.16), we remark that

$$\mu(t)^2\delta(V(t))^2 = \left\|V(t)^2\right\|_F^2 - \bar{n}\mu(t)^2, \tag{3.20}$$

where

$$\left\|V(t)^2\right\|_F^2 = \text{ tr } ((V+tD_X)(V+tD_Z))^2 = \text{ tr } (W(t)+U(t))^2.$$

Now using the fact that $U(t)$ is skew-Hermitian, and using (3.19),

$$\left\|V(t)^2\right\|_F^2 = (W(t)-U(t)) \bullet (W(t)+U(t)) = \|W(t)\|_F^2 - \|U(t)\|_F^2. \tag{3.21}$$

As

$$I \bullet W(t) = \bar{n}\mu(t),$$

it now follows together with (3.20) and (3.21) that

$$\begin{aligned}\delta(V(t))^2 &= \frac{\|W(t)\|_F^2 - \bar{n}\mu(t)^2}{\mu(t)^2} - \frac{\|U(t)\|_F^2}{\mu(t)^2} \\ &= \left\|\frac{1}{\mu(t)}W(t) - I\right\|_F^2 - \frac{\|U(t)\|_F^2}{\mu(t)^2}.\end{aligned}$$

Q.E.D.

It follows from (3.9) that

$$\begin{aligned}W(t) &= V^2 + tP_{\mathcal{H}}(V(D_X+D_Z)) + t^2 P_{\mathcal{H}}(D_X D_Z) \\ &= (1-t)V^2 + t\gamma\mu I + t^2 P_{\mathcal{H}}(D_X D_Z). \end{aligned} \tag{3.22}$$

Combining this relation with (3.14) and Lemma 3.3, it follows that

$$\begin{aligned}\mu(t)^2\delta(V(t))^2 &= \left\|(1-t)(V^2-\mu I) + t^2 P_{\mathcal{H}}(D_X D_Z)\right\|_F^2 - \|U(t)\|_F^2 \\ &\leq \left\|(1-t)(V^2-\mu I) + t^2 P_{\mathcal{H}}(D_X D_Z)\right\|_F^2.\end{aligned}$$

Applying the triangle inequality to the above relation, we obtain

$$\mu(t)\delta(V(t)) \leq (1-t)\mu\delta(V) + t^2 \|P_{\mathcal{H}}(D_X D_Z)\|_F. \tag{3.23}$$

Lemma 3.4 *There holds*

$$\|P_{\mathcal{H}}(D_X D_Z)\|_F \leq \frac{1}{2}\|D_X + D_Z\|_F^2.$$

Proof. Using the triangle inequality and the geometric–arithmetic mean inequality respectively, we have

$$\begin{aligned} \|D_X D_Z + D_Z D_X\|_F &\leq \|D_X D_Z\|_F + \|D_Z D_X\|_F \\ &\leq 2\|D_X\|_F \|D_Z\|_F \\ &\leq \|D_X\|_F^2 + \|D_Z\|_F^2, \end{aligned}$$

which together with (3.12) implies that

$$\|P_{\mathcal{H}}(D_X D_Z)\|_F = \frac{1}{2}\|D_X D_Z + D_Z D_X\|_F \leq \frac{1}{2}\|D_X + D_Z\|_F^2.$$

Q.E.D.

Lemma 3.5 *There holds:*

1. *If* $\delta(V) < 1$ *then*
$$\frac{1}{\mu}\|D_X + D_Z\|_F^2 \leq \frac{\gamma^2 \delta(V)^2}{1-\delta(V)} + \bar{n}(1-\gamma)^2.$$
2. *If* $V \in \mathcal{N}_\infty^-(\beta)$ *for some* $\beta \in (0,1)$ *then*
$$\frac{1}{\mu}\|D_X + D_Z\|_F^2 \leq \frac{\bar{n}\gamma^2\beta}{1-\beta} + \bar{n}(1-\gamma)^2.$$

Proof. From (3.13) we know that

$$\|D_X + D_Z\|_F^2 = \gamma^2 \left\|\mu V^{-1} - V\right\|_F^2 + (1-\gamma)^2 \bar{n}\mu. \tag{3.24}$$

Based on the above identity, we will now establish the two bounds.

1. For the case that $\delta(V) < 1$, we use the inequality
$$\left\|\mu V^{-1} - V\right\|_F \leq \left\|\mu V^{-1}\right\|_2 \left\|I - \frac{1}{\mu}V^2\right\|_F = \delta(V)\left\|\mu V^{-1}\right\|_2. \tag{3.25}$$
As $V^2 \succeq (1-\delta(V))\mu I$, see (3.17), it follows that $I \succeq (1-\delta(V))\mu V^{-2}$. Therefore, using also that $\delta(V) < 1$,
$$\left\|\mu V^{-1}\right\|_2^2 \leq \frac{\mu}{1-\delta(V)}.$$
Combining the above relation with (3.24)–(3.25), it follows that
$$\frac{1}{\mu}\|D_X + D_Z\|_F^2 \leq \frac{\gamma^2\delta(V)^2}{1-\delta(V)} + \bar{n}(1-\gamma)^2.$$

2. Suppose that $V \in \mathcal{N}_\infty^-(\beta)$ for some $\beta \in (0,1)$. Using the definition of the Frobenius norm,
$$\|\mu V^{-1} - V\|_F^2 = \|\mu V^{-1}\|_F^2 - \bar{n}\mu \leq \bar{n}\|\mu V^{-1}\|_2^2 - \bar{n}\mu,$$
from which we obtain for $V \in \mathcal{N}_\infty^-(\beta)$ that
$$\|\mu V^{-1} - V\|_F^2 \leq \frac{\bar{n}\mu}{1-\beta} - \bar{n}\mu = \frac{\beta}{1-\beta}\bar{n}\mu.$$
Together with (3.24), the above relation yields
$$\frac{1}{\mu}\|D_X + D_Z\|_F^2 \leq \frac{\bar{n}\gamma^2\beta}{1-\beta} + \bar{n}(1-\gamma)^2.$$

Q.E.D.

The next result follows by combining Lemma 3.4 and Lemma 3.5 with the inequality (3.23).

Lemma 3.6 *Suppose $\delta(V) < 1$. For $0 \leq t < t^*$ there holds*
$$(1 - t + \gamma t)\delta(V(t)) \leq (1-t)\delta(V) + \frac{t^2}{2}\left(\frac{\gamma^2\delta(V)^2}{1-\delta(V)} + \bar{n}(1-\gamma)^2\right).$$

We will now investigate how we can choose the parameter γ to guarantee feasibility of the full Newton step. Our basic observation is that if $\delta(V(t)) < 1$ for $0 \leq t \leq 1$ then, by continuity, $V(1) \succ 0$, implying that the full Newton step is feasible.

Suppose $V \in \mathcal{N}_2(\beta)$. From Lemma 3.6, it follows for $\gamma \in (0,1]$ that
$$(1 - t + \gamma t)\delta(V(t)) \leq (1-t)\beta + \frac{1}{2}(\gamma t)^2\left(\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2\right) \tag{3.26}$$
yielding the following lemma.

Lemma 3.7 *Let $\gamma \in (0,1]$ and $\beta \in (0,1)$. If $V \in \mathcal{N}_2(\beta)$ and*
$$\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2 \leq 1 \tag{3.27}$$
then $t^* > 1$ and $V(1) \succ 0$.

Proof. Let $t \in [0, t^*)$. Based on (3.26) and (3.27) we obtain
$$(1 - t + \gamma t)\delta(V(t)) \leq (1-t)\beta + \frac{1}{2}(\gamma t)^2,$$

so that

$$\delta(V(t)) < 1 \quad \text{if } 0 \le t \le 1.$$

By continuity this implies that $t^* > 1$ and

$$X(t) \succ 0, \quad Z(t) \succ 0 \quad \text{for } 0 \le t \le 1.$$

The lemma is proved. **Q.E.D.**

If we set $\gamma = 1$, we obtain a so–called *pure centering* step. If we are close enough to the central path, the pure centering step reduces the distance $\delta(V)$ quadratically fast to zero, as the following lemma shows.

Lemma 3.8 *If $\gamma = 1$ and $\delta(V) \le 1/2$, then*

$$\delta(V(1)) \le \delta(V)^2.$$

Proof. It follows from Lemma 3.7 that $t^* > 1$ if $\gamma = 1$ and $\delta(V) \le 1/2$. Hence, the desired result is an immediate consequence of Lemma 3.6. **Q.E.D.**

The following lemma is crucial for algorithms that use the wide $\mathcal{N}_\infty^-$–neighborhood.

Lemma 3.9 *Suppose that $0 < \gamma < 1$ and $V \in \mathcal{N}_\infty^-(\beta)$ for some $\beta \in (0,1)$. Then $V(t) \in \mathcal{N}_\infty^-(\beta)$ for any step length t satisfying*

$$0 \le \bar{n}t \le \frac{2\beta\gamma}{\beta\gamma^2/(1-\beta) + (1-\gamma^2)}.$$

Proof. Remark from definition (3.18) that $V(t) \in \mathcal{N}_\infty^-(\beta)$ if and only if

$$\lambda_{\min}(V(t)^2) \succeq (1-\beta)\mu(t),$$

where $\lambda_{\min}(V(t)^2)$ is the smallest eigenvalue of $V(t)^2$. Since the Hermitian matrix $V(t)^2$ is similar to $(V + tD_X)(V + tD_Z)$, we know from Lemma A.2 that

$$\lambda_{\min}(V(t)^2) = \lambda_{\min}((V + tD_X)(V + tD_Z)) \ge \lambda_{\min}(W(t)). \tag{3.28}$$

However, we know from (3.22) that

$$W(t) = (1-t)V^2 + t\gamma\mu I + t^2 P_{\mathcal{H}}(D_X D_Z),$$

which, using also $V \in \mathcal{N}_\infty^-(\beta)$ and (3.14), implies

$$\begin{aligned}\lambda_{\min}(W(t)) &\ge (1-t)(1-\beta)\mu + t\gamma\mu - t^2 \left\|P_{\mathcal{H}}(D_X D_Z)\right\|_2 \\ &= (1-\beta)\mu(t) + t(\beta\gamma\mu - t\left\|P_{\mathcal{H}}(D_X D_Z)\right\|_2).\end{aligned}$$

Combining the above relation with (3.28), it follows that

$$\lambda_{\min}(V(t)^2) \geq (1-\beta)\mu(t) \quad \text{if } \mu(t) \geq 0 \text{ and } 0 \leq \|P_{\mathcal{H}}(D_X D_Z)\|_2\, t \leq \beta\gamma\mu.$$

Since we know from Lemmas 3.4 and 3.5 that

$$\|P_{\mathcal{H}}(D_X D_Z)\|_2 \leq \|P_{\mathcal{H}}(D_X D_Z)\|_F \leq \frac{\bar{n}\mu}{2}\left(\frac{\gamma^2\beta}{1-\beta} + (1-\gamma)^2\right),$$

it follows that

$$V(t) \in \mathcal{N}_\infty^-(\beta) \quad \text{for } 0 \leq \bar{n}t \leq \frac{2\beta\gamma}{\beta\gamma^2/(1-\beta) + (1-\gamma^2)}.$$

Q.E.D.

3.5 Four Path–Following Algorithms

In [46], Gonzaga provides a clear survey of interior point methods for linear programming and linear complementarity problems. He discussed three algorithms, viz. (1) the short step algorithm, (2) the predictor-corrector algorithm and (3) the largest step algorithm. In this section, we generalize these algorithms to semidefinite programming. In addition, we will generalize the long step algorithm of Kojima, Mizuno and Yoshise [75].

We first present a generic path-following algorithm with unspecified parameters β, $\gamma^{(k)}$ and $t^{(k)}$, for $k = 0, 1, \ldots$, and a certain neighborhood $\mathcal{N}(\beta)$ of the central path. Specific choices for these parameters and the neighborhood will then lead to the four above mentioned algorithms.

Algorithm 3.2 (Path–following)

Data: $\epsilon > 0$, $(X^{(0)}, Z^{(0)})$ *with* $V^{(0)} \in \mathcal{N}(\beta)$.

Step 0 *Set* $k = 0$.

Step 1 *If* $X^{(k)} \bullet Z^{(k)} < \epsilon$ *then stop.*

Step 2 *Choose* $\gamma^{(k)} \in [0, 1]$ *and solve* $(\Delta X^{(k)}, \Delta Z^{(k)})$ *from*

$$\begin{cases} \Delta X^{(k)} + D(X^{(k)}, Z^{(k)})\Delta Z^{(k)} D(X^{(k)}, Z^{(k)}) = \gamma^{(k)}\mu^{(k)}(Z^{(k)})^{-1} - X^{(k)}, \\ \Delta X^{(k)} \in \mathcal{A}, \quad \Delta Z^{(k)} \in \mathcal{A}^\perp, \end{cases}$$

with $\mu^{(k)} = X^{(k)} \bullet Z^{(k)}/\bar{n}$.

Step 3 *Choose* $t^{(k)}$ *and let* $X^{(k+1)} = X^{(k)} + t^{(k)}\Delta X^{(k)}$, $Z^{(k+1)} = Z^{(k)} + t^{(k)}\Delta Z^{(k)}$.

Step 4 *Set* $k = k + 1$ *and return to Step 1.*

We will choose the parameters such that $V^{(k)} \in \mathcal{N}(\beta)$ for all k.

3.5.1 The Short Step Algorithm

In the generic algorithm, the choice

$$\begin{cases} \mathcal{N}(\cdot) = \mathcal{N}_2(\cdot) \\ \beta = \frac{1}{2} \\ \gamma^{(k)} = \frac{1}{1+1/\sqrt{2\bar{n}}} & \text{for } k = 0, 1, 2, \ldots \\ t^{(k)} = 1 & \text{for } k = 0, 1, 2, \ldots \end{cases}$$

leads to the so-called short step algorithm. This type of algorithm was studied for linear programming by Monteiro and Adler [98] and Kojima, Mizuno and Yoshise [74]. We let

$$\gamma := \frac{1}{1 + 1/\sqrt{2\bar{n}}}$$

so that

$$\gamma^{(k)} = \gamma \text{ for } k = 0, 1, 2, \ldots.$$

Based on the technical results of the previous section, we obtain the following result for the short step algorithm.

Lemma 3.10 *For the short step algorithm, we have*

$$V^{(k)} \in \mathcal{N}_2(\beta)$$

for $k = 0, 1, \ldots$.

Proof. Let $k \in \{0, 1, 2, \ldots\}$. For the given choice of parameters, we have

$$\frac{\beta^2}{1-\beta} + \bar{n}(\frac{1-\gamma}{\gamma})^2 = \frac{1}{2} + \frac{1}{2} = 1.$$

Hence, we have from Lemma 3.7 that the maximum step length t^* is greater than 1. Using (3.26) with $t = 1$ it thus follows that

$$\gamma\delta(V^{(k+1)}) \leq \frac{\gamma^2}{2}$$

which implies

$$\delta(V^{(k+1)}) < \frac{1}{2}.$$

As $V^{(0)} \in \mathcal{N}_2(\beta)$ by hypothesis, the lemma follows by induction. **Q.E.D.**

We are now in a position to prove the polynomiality of the algorithm for obtaining an ϵ-optimal solution.

Theorem 3.1 *The short step algorithm computes an ϵ-optimal solution in*

$$\mathbf{O}(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

Proof. We know from Lemma 3.10 that all iterates of the short step algorithm are contained in $\mathcal{N}_2(\beta)$. Therefore, the algorithm is well defined. For given $k \in \{0, 1, \ldots\}$ we know from (3.14) that

$$X^{(k+1)} \bullet Z^{(k+1)} = \bar{n}\mu^{(k)}(1) = \gamma X^{(k)} \bullet Z^{(k)}.$$

Taking the logarithm on both sides and using the definiton of γ, we arrive at the relation

$$\begin{aligned} \log\left(X^{(k+1)} \bullet Z^{(k+1)}\right) &= \log\gamma + \log\left(X^{(k)} \bullet Z^{(k)}\right) \\ &= (k+1)\log\gamma + \log\left(X^{(0)} \bullet Z^{(0)}\right) \\ &= \log\left(X^{(0)} \bullet Z^{(0)}\right) - (k+1)\log\left(1 + \frac{1}{\sqrt{2\bar{n}}}\right) \\ &\leq \log\left(X^{(0)} \bullet Z^{(0)}\right) - \frac{k+1}{2\sqrt{2\bar{n}}}. \end{aligned}$$

The theorem is proved. **Q.E.D.**

3.5.2 The Predictor–Corrector Algorithm

The short step path following algorithm is of little practical value, because it has a fixed rate of convergence, viz. $\gamma = 1/(1+1/\sqrt{2\bar{n}})$. In order to perform better than the worst case behavior, an adaptive rate of convergence is necessary. The predictor-corrector algorithm of Mizuno, Todd and Ye [95] is such an adaptive algorithm. This predictor-corrector algorithm is obtained by the following choice of parameters:

$$\mathcal{N}(\cdot) = \mathcal{N}_2(\cdot), \quad \beta = 1/2,$$

$$\gamma^{(k)} = \begin{cases} 1 & \text{for } k = 0, 2, 4, \ldots, \\ 0 & \text{for } k = 1, 3, 5, \ldots, \end{cases}$$

$$t^{(k)} = \begin{cases} 1 & \text{for } k = 0, 2, 4, \ldots, \\ \max\{\bar{t} \mid V^{(k)}(t) \in \mathcal{N}_2(\beta) \text{ for } 0 \leq t \leq \bar{t}\} & \text{for } k = 1, 3, 5, \ldots. \end{cases}$$

The even numbered iterations are called *corrector* steps, the odd numbered iterations are called *predictor* steps.

At the start of a corrector iteration, we have $V \in \mathcal{N}_2(\frac{1}{2})$. From Lemma 3.8, we have

$$\delta(V(1)) \leq \delta(V)^2 \leq \frac{1}{4}.$$

Predictor iterations therefore always start with $V \in \mathcal{N}_2(\frac{1}{4})$. Using Lemma 3.6, it follows with $\gamma = 0$ that

$$\delta(V(t)) \leq \frac{1}{4} + \frac{\bar{n}t^2}{2(1-t)}.$$

This means that if

$$0 \leq t \leq \frac{2}{1+\sqrt{1+8\bar{n}}}$$

then

$$\delta(V(t)) \leq \frac{1}{2},$$

implying that the step lengths in the predictor iterations are never shorter than $2/(1+\sqrt{1+8\bar{n}})$. Similar to Theorem 3.1, this yields:

Theorem 3.2 *The predictor-corrector algorithm computes an ϵ-optimal solution in*

$$\mathbf{O}(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

3.5.3 The Largest Step Algorithm

The practical performance of the predictor-corrector algorithm can be enhanced by combining the predictor and corrector steps in a single iteration. This idea leads to the largest step algorithm of Gonzaga [47]. In each iteration of the largest step algorithm, we determine the smallest $\gamma \in [0, 1]$ such that $V(1) \in \mathcal{N}_2(\beta)$. In particular, we compute a centering direction $(\Delta X^c, \Delta Z^c)$ from

$$\begin{cases} \Delta X^c + D(X,Z)\Delta Z^c D(X,Z) = \mu Z^{-1} - X \\ \Delta X^c \in \mathcal{A}, \quad \Delta Z^c \in \mathcal{A}^\perp \end{cases}$$

and the so-called affine scaling direction $(\Delta X^a, \Delta Z^a)$ from

$$\begin{cases} \Delta X^a + D(X,Z)\Delta Z^a D(X,Z) = -X \\ \Delta X^a \in \mathcal{A}, \quad \Delta Z^a \in \mathcal{A}^\perp. \end{cases}$$

We can now rewrite (3.15) as

$$\Delta X_\gamma = \gamma \Delta X^c + (1-\gamma)\Delta X^a, \quad \Delta Z_\gamma = \gamma \Delta Z^c + (1-\gamma)\Delta Z^a,$$

where we added a subscript γ to stress the dependence of ΔX and ΔZ on γ.

In the largest step algorithm we set $\gamma = \gamma^*$ where γ^* is obtained by computing

$$\gamma^* := \min\{\bar{\gamma} \mid \delta(V_\gamma(1)) \leq \beta \text{ for } \bar{\gamma} \leq \gamma \leq 1\},$$

which amounts to solving a quartic equation, as can be seen from Lemma 3.3. We choose $\beta = 1/2$. Our analysis of the short step algorithm implies $\gamma^* < 1/(1 + 1/\sqrt{2\bar{n}})$, and this gives the following result:

Theorem 3.3 *The largest step algorithm computes an ϵ-optimal solution in*

$$\mathbf{O}(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

As discussed above, the search direction in the largest step algorithm can be interpreted as a path–following direction (3.15), where the parameter γ is determined adaptively. However, there is a different interpretation of the largest step search direction, which is sometimes more convenient. Namely, we can view the largest step algorithm as a predictor–corrector algorithm with an inexact (or simplified) computation of the predictor. This simplified predictor direction is defined as the solution $(\Delta X^p, \Delta Z^p)$ of

$$\Delta X^p + D(X, Z)\Delta Z^p D(X, Z) = -\mu Z^{-1}$$

$$\Delta X^p \in \mathcal{A}, \quad \Delta Z^p \in \mathcal{A}^\perp.$$

The search direction $(\Delta X_\gamma, \Delta Z_\gamma)$ can be written in terms of the corrector and simplified predictor as

$$\Delta X_\gamma = \Delta X^c + (1-\gamma)\Delta X^p, \quad \Delta Z_\gamma = \Delta Z^c + (1-\gamma)\Delta Z^p.$$

In this interpretation, we first take a unit step along the centering direction $(\Delta X^c, \Delta Z^c)$, and next a step of length $t = 1 - \gamma$ along the simplified predictor direction $(\Delta X^p, \Delta Z^p)$, until we reach the boundary of the $\mathcal{N}_2$–neighborhood. This amounts to the same scheme as in the predictor–corrector algorithm of Section 3.5.2. However, in Section 3.5.2 the predictor (affine scaling) direction was computed with the updated primal–dual transformation *after* completing the corrector (centering) step, whereas we define the simplified predictor direction in terms of the outdated transformation, i.e. prior to the centering step. The simplified approach has computational advantages, since it allows the reuse of matrix factorizations in the orthogonal decomposition of $D_X^p + D_Z^p$.

Obviously, it is also possible to update the primal–dual transformation only at predictor steps and use a simplified corrector step. This approach leads to the simplified predictor–corrector algorithm as studied for linear programming by Gonzaga and Tapia [48]. However, this scheme does not lead to a polynomial algorithm, unless a safeguard is used at the corrector step. Problems in the simplified corrector arise due to the adaptiveness of the predictor step length. Namely, after a long predictor step, the old primal–dual transformation is in general no longer adequate, see [48].

3.5.4 The Long Step Algorithm

A considerable amount of research in linear programming [73, 75, 95, 165] concerns long step primal–dual interior point algorithms that generate iterates in the wide neighborhood $\mathcal{N}_\infty^-(\beta)$. Such algorithms have been extended to semidefinite programming by Monteiro [97], Monteiro and Zhang [105] and Sturm and Zhang [148]. A main iteration of this method can be described as follows.

First, a search direction is obtained for a target on the central path with a considerably smaller duality gap, i.e. $\gamma^{(k)} \in (0,1)$ is chosen such that

$$\frac{1}{1-\gamma^{(k)}} = \mathbf{O}(1), \quad \frac{1}{\gamma^{(k)}} = \mathbf{O}(1),$$

for all $k = 0, 1, 2, \ldots$. Recall that in the short step algorithm, $1/(1-\gamma^{(k)}) = 1 + \sqrt{2\bar{n}}$, which is not universally bounded. This explains the terminology 'short step' and 'long step'.

Second, the method takes the largest possible step in the $(\Delta X^{(k)}, \Delta Z^{(k)})$ direction, without leaving the $\mathcal{N}_\infty^-(\beta)$–neighborhood, i.e. $t^{(k)}$ is maximal with respect to

$$V^{(k)}(t) \in \mathcal{N}_\infty^-(\beta) \text{ for } 0 \le t \le t^{(k)}. \tag{3.29}$$

The parameter β is a universal constant in $(0,1)$, i.e.

$$\frac{1}{\beta(1-\beta)} = \mathbf{O}(1).$$

From Lemma 3.9, we know that

$$\bar{n}t^{(k)} \ge \frac{2\beta\gamma^{(k)}}{\beta(\gamma^{(k)})^2/(1-\beta) + (1-(\gamma^{(k)})^2)}, \tag{3.30}$$

for each iteration $k = 1, 2, \ldots$ of the long step algorithm. This gives the following result:

Theorem 3.4 *The long step algorithm computes an ϵ-optimal solution in*

$$\mathbf{O}(\bar{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

Remark that the worst case estimation for the long step algorithm is an order $\sqrt{\bar{n}}$ worse then the estimates for the three $\mathcal{N}_2$–neighborhood methods. However, since the $\mathcal{N}_\infty^-$–neighborhood is wider than the $\mathcal{N}_2$–neighborhood, the iterates of the long step method trace the central path more loosely than the iterates of $\mathcal{N}_2$–neighborhood methods. In typical iterations, the long step algorithm can use this freedom to take longer steps. In the special case of linear programming, it is widely known that the worst case estimate (3.30), viz. $1/t^{(k)} = \mathbf{O}(\bar{n})$, can be highly pessimistic. Using probabilistic arguments, Mizuno, Todd and Ye [95] argue that $t^{(k)}$ is typically $1/\mathbf{O}(\log \bar{n})$ for the long step path–following algorithm, and this behavior is also observed in practice. A variant of the long step algorithm was implemented in OB1, which is one of the first interior point codes for large scale linear programming, see Lustig, Marsten and Shanno [88]. For variants of the long step algorithm that do not use the $\mathcal{N}_\infty^-$–neighborhood, we refer to Roos, Terlaky and Vial [134].

3.6 Discussion

In recent years, the generalization of interior point methods to semidefinite programming has been intensively investigated. However, in the case of primal–dual search directions the topic still appears to be rather controversial.

Alizadeh [1], Jarre [62], Nesterov and Nemirovsky [110], Nesterov and Todd [112] and Vandenberghe and Boyd [157] generalized some interior point algorithms that are based on optimization of a potential function or a barrier function. These kind of interior point algorithms were proposed for linear programming by Karmarkar [65], Renegar [129], Todd and Ye [152] and Gonzaga [45]. Another classical interior point method for linear programming is the affine scaling method of Dikin [30], and generalization of this method to semidefinite programming is straightforward. However, Muramatsu [106] gives an example in semidefinite programming for which this algorithm fails to converge to an optimal solution, if the starting point is poorly chosen. In linear programming, the interest in barrier, potential reduction and affine scaling methods decreased a couple of years ago, in favor of (pure) primal–dual methods, which were proposed by Kojima, Mizuno and Yoshise [75]. The excellent numerical results for primal–dual methods as obtained by Lustig, Marsten and Shanno [88, 89] helped the primal–dual method to become the central issue in interior point methods.

It appears that there are several ways in which primal–dual methods can be extended to semidefinite programming. In particular, primal–dual directions can be derived on the basis of

- self–scaled barriers,
- matrix targets, and

- eigenvalue targets.

3.6.1 Self-Scaled Barriers

The symmetric primal–dual approach with self–scaled barriers was proposed by Nesterov and Todd [112, 113], in a slightly more general setting than semidefinite programming. The central object in this approach is the self–scaled barrier function $F(X) := -\log\det X$. (It is not important here to know the definition of self–scaled barriers. However, the precise definition is given in [112, 113].) In Definition 3.1, we have defined the primal–dual central path as the set of feasible primal–dual pairs (X, Z) for which $XZ = \mu I$ for some μ. We then derived a search direction by setting a μ–target on the central path. A pair (X, Z) on the central path corresponding to a parameter μ is called a μ–center. A μ–center can also be characterized as the minimizer of the function

$$C \bullet X + B \bullet Z + \mu(F(X) + F(Z)).$$

A Newton–type direction for minimizing the above function is naturally based on the Hessian of $F(\cdot)$. Instead of evaluating the Hessian at the current iterate, Nesterov and Todd [113] propose to evaluate it at the so–called primal–dual scaling point. This primal–dual scaling point is exactly the matrix $D(X, Z)$ as defined in (3.3). In this way, Nesterov and Todd were the first who obtained the path–following direction $(\Delta X, \Delta Z)$ as defined in (3.15), which is therefore called the Nesterov–Todd direction. The interpretation of this Nesterov–Todd direction as a Newton direction towards an eigenvalue target (as in this chapter) is due to Sturm and Zhang [149]. More recently, Todd, Toh and Tütüncü [151] and Kojima, Shida and Shindoh [77] interpreted the Nesterov–Todd direction as a Newton direction towards a matrix target.

3.6.2 Matrix Targets

A very popular approach for deriving primal–dual path–following directions is based on linearization of the system

$$\begin{cases} (X + \Delta X)(Z + \Delta Z) = \gamma\mu I \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^\perp, \end{cases} \tag{3.31}$$

yielding

$$\begin{cases} X\Delta Z + (\Delta X)Z = \gamma\mu I - XZ \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^\perp. \end{cases} \tag{3.32}$$

However, (3.32) is over-determined and therefore in general not solvable. So, we should either add a number of artificial variables, or relax some constraints. Helmberg, Rendl, Vanderbei and Wolkowicz [53] proposed to replace the restriction $\Delta X \in \mathcal{A}$ by $\Delta X \in \mathcal{A} \oplus \mathcal{H}^\perp$, which makes the resulting system solvable. In order to get primal feasible iterates, [53] further proposes to take the Hermitian part of the solution ΔX in the actual steps. This technique is also known as the XZ–method.

Independently of [53], Kojima, Shindoh and Hara [80] developed a more general framework for linearizing (3.31). Namely, they propose to add skew–Hermitian variables U_X and U_Z to the system (3.32), yielding

$$\begin{cases} X(\Delta Z + U_Z) + (\Delta X + U_X)Z = \gamma\mu I - XZ \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^\perp \\ (U_X, U_Z) \in \mathcal{M}, \end{cases} \tag{3.33}$$

where $\mathcal{M}$ is an arbitrary maximal monotone linear subspace of $\mathcal{H}^\perp \times \mathcal{H}^\perp$. By definition, this is a linear subspace $\mathcal{M}$ of dimension $\dim \mathcal{H}^\perp = \bar{n}^2$, with the property that

$$(U_X, U_Z) \in \mathcal{M} \Longrightarrow U_X \bullet U_Z \geq 0.$$

For given $\mathcal{M}$, the system (3.33) has a unique solution $(\Delta X, \Delta Z, U_X, U_Z)$, see [80]. We remark that if $\mathcal{M}_P$ is a linear subspace of $\mathcal{H}^\perp$ then $\mathcal{M} = \mathcal{M}_P \times \mathcal{M}_P^\perp$ is maximal monotone, where $\mathcal{M}_P^\perp$ is the orthogonal complement of $\mathcal{M}_P$ in $\mathcal{H}^\perp$. For example, if we let $\mathcal{M}_P = \mathcal{H}^\perp$, then $\mathcal{M}_P^\perp = \{0\}$, and (3.33) describes the Helmberg–Rendl–Vanderbei-Wolkowicz direction. Moreover, if we let $\mathcal{M} = \{(U_X, U_Z) \mid L_d^{-1} U_X L_d^{-\mathrm{H}} = L_d^{\mathrm{H}} U_Z L_d\}$, then (3.33) describes the Nesterov–Todd direction, see Kojima, Shida and Shindoh [77]. We will see in Section 3.6.3 how the Kojima–Shindoh–Hara family of path–following directions can be interpreted in the V–space framework.

Instead of adding artificial variables, it is also possible to reduce the number of constraints in (3.32) in order to make it solvable, and this leads us to the similarity–symmetrization techniques. The basic observation here is that

$$(X + \Delta X)(Z + \Delta Z) = \mu I \iff P_{\mathcal{H}}((X + \Delta X)(Z + \Delta Z)) = \mu I.$$

The Hermitian system $P_{\mathcal{H}}((X + \Delta X)(Z + \Delta Z)) = \gamma\mu I$ describes only $\dim \mathcal{H}$ nonlinear equations, instead of the former $\dim \mathcal{H} + \dim \mathcal{H}^\perp$ equations. Linearization of this Hermitian system yields

$$\begin{cases} P_{\mathcal{H}}(X\Delta Z + (\Delta X)Z) = P_{\mathcal{H}}(\gamma\mu I - XZ) \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^\perp. \end{cases} \tag{3.34}$$

This approach, which was proposed by Alizadeh, Haeberly and Overton [4], is sometimes called the $XZ + ZX$ method. Shida, Shindoh and Kojima [138] showed that the system (3.34) is solvable if $P_{\mathcal{H}}(XZ) \succ 0$, which holds true if X and Z are well centered.

A generalization of this technique was proposed by Zhang [164], inspired by the work of Monteiro [97]. Zhang proposed to let a similarity transformation precede the symmetrization in (3.34), namely,

$$\begin{cases} P_{\mathcal{H}}\{L^{-1}(X\Delta Z + (\Delta X)Z)L\} = P_{\mathcal{H}}\{L^{-1}(\gamma\mu I - XZ)L\} \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^{\perp}, \end{cases} \tag{3.35}$$

where L is a given invertible matrix. For $L = I$, we obtain the Alizadeh–Haeberly–Overton direction, for $L = L_d$ the Nesterov–Todd direction and for $L = Z^{-1/2}$ the Helmberg–Rendl–Vanderbei–Wolkowicz direction. The class of directions (3.35) was further analysed by Monteiro and Zhang [105] and Monteiro [96]. A disadvantage of the Alizadeh–Haeberly–Overton and Monteiro–Zhang approach is that it is in general relatively hard to solve the search direction $(\Delta X, \Delta Z)$. In fact, the system (3.35) may not have a unique solution and can even be inconsistent, see Todd, Toh and Tütüncü [151] for an example. However, it can be show that there exists a unique solution to (3.35) if $P_{\mathcal{H}}(L^{-1}XZL) \succ 0$ or $\|I - V^2/\mu\|_2 \leq 1/2$, see Shida, Shindoh and Kojima [138], and Monteiro and Zanjácomo [104], respectively.

Recall that the Nesterov–Todd direction satisfies relation (3.15), i.e.

$$\Delta X + D(X,Z)\Delta Z D(X,Z) = \gamma\mu Z^{-1} - X.$$

where $D(X,Z) := Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}$, see definition (3.3). Tseng [154] analyzed path–following methods in a more general framework, namely he considers search directions that satisfy the system

$$\begin{cases} P_{\mathcal{H}}\left(\Delta X + D_{\tau}(X,Z)\Delta Z D_{1-\tau}(X,Z)\right) = \gamma\mu Z^{-1} - X \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^{\perp}, \end{cases} \tag{3.36}$$

with

$$D_{\tau}(X,Z) = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{\tau}Z^{-1/2}.$$

As special cases, we have $D_0(X,Z) = Z^{-1}$, $D_{1/2}(X,Z) = D(X,Z)$ and $D_1(X,Z) = X$. Therefore, (3.36) yields the Helmberg–Rendl–Vanderbei–Wolkowicz direction for $\tau = 0$, and the Nesterov–Todd direction for $\tau = 1/2$.

In the V–space approach of Section 3.2, we considered a positive diagonal matrix V^2, with on the diagonal the eigenvalues of XZ. In particular, if L_X is the Cholesky factor of X, i.e. $X = L_X L_X^{\mathrm{H}}$, then $L_X^{\mathrm{H}} Z L_X = Q^{\mathrm{H}} V^2 Q$ for some orthogonal matrix Q, see also Algorithm 3.1 in Section 3.1. Letting $\hat{L}_X := L_X Q^{\mathrm{H}}$, we have

$$\hat{L}_X^{\mathrm{H}} Z \hat{L}_X = V^2, \quad \hat{L}_X \hat{L}_X^{\mathrm{H}} = X.$$

We derived a Newton direction for steering $V(t)^2$, by finding out how the derivative of $V(t)^2$ is related to the derivatives of $X(t)$ and $Z(t)$. A very similar, and in fact equivalent, idea was recently worked out by Monteiro and Tsuchiya [102]. However, the motivation of their direction fits more in the framework of matrix targets, and the relation with the V–space directions of Sturm and Zhang [149] has not been realized before.

Monteiro and Tsuchiya [102] propose to linearize the system

$$\begin{cases} R_X(1)L^{-1}(Z+\Delta Z)L^{-\mathrm{H}}R_X(1) = \gamma\mu I \\ R_X(t)^2 = L^{\mathrm{H}}(X+t\Delta X)L \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^{\perp}, \end{cases} \tag{3.37}$$

where L is a given invertible matrix. Notice that $R_X(t) = \left(L^{\mathrm{H}}(X+t\Delta X)L\right)^{1/2}$. We introduced the auxiliary function $R_X(t)$ in order to simplify the presentation.

Letting $L_X(t) := L^{-\mathrm{H}}R_X(t)$, we can rewrite (3.37) as

$$\begin{cases} L_X(1)^{\mathrm{H}}(Z+\Delta Z)L_X(1) = \gamma\mu I \\ L_X(t)L_X(t)^{\mathrm{H}} = X + t\Delta X, \quad L^{\mathrm{H}}L_X(t) \in \mathcal{H} \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^{\perp}. \end{cases} \tag{3.38}$$

Now, define the target function

$$W(t) := L_X(t)^{\mathrm{H}}(Z+t\Delta Z)L_X(t).$$

Implicit differentiation of the identity

$$(X+t\Delta X)(Z+t\Delta Z) = L_X(t)W(t)L_X(t)^{-1}$$

yields

$$\begin{aligned} X\Delta Z + (\Delta X)Z + 2t\Delta X\Delta Z \quad &= \quad L_X(t)W^{[1]}(t)L_X(t)^{-1} + L_X^{[1]}(t)W(t)L_X(t)^{-1} \\ &\quad\; - L_X(t)W(t)L_X(t)^{-1}L_X^{[1]}(t)L_X(t)^{-1}. \end{aligned} \tag{3.39}$$

Linearization of the target equation $W(1) = \gamma\mu I$ yields the identity $W^{[1]}(0) = \gamma\mu I - W(0)$, where $W(0) = L_X^{\mathrm{H}}ZL_X$, $L_X := L_X(0)$. Substitute this into (3.39), to obtain for $t = 0$ that

$$X\Delta Z + (\Delta X)Z = (\gamma\mu I - XZ) + (L_X^{[1]}(0)L_X^{-1})XZ - XZ(L_X^{[1]}(0)L_X^{-1}).$$

It follows that the Monteiro–Tsuchiya directions are solutions of the system

$$\begin{cases} X\Delta Z + (\Delta X)Z = (\gamma\mu I - XZ) + (L_X^{[1]}(0)L_X^{-1})XZ - XZ(L_X^{[1]}(0)L_X^{-1}) \\ 2P_{\mathcal{H}}\left(L_X^{[1]}(0)L_X^{\mathrm{H}}\right) = \Delta X \\ L^{\mathrm{H}}L_X,\ L^{\mathrm{H}}L_X^{[1]}(0) \in \mathcal{H} \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^{\perp}. \end{cases} \tag{3.40}$$

We will see in Section 3.6.3 how the Monteiro–Tsuchiya family of path–following directions can be interpreted in the V–space framework.

3.6.3 Eigenvalue Targets

The third main approach, after the self–scaled barrier approach and the many variants of the matrix target approach, is the eigenvalue target approach. In this approach, we set a target for the eigenvalues of the matrix XZ, instead of its entries.

An iterate in the Q–method consists of a primal–dual pair of diagonal matrices (Λ_X, Λ_Z) and a common eigenvector basis Q. This method was proposed by Alizadeh, Haeberly and Overton [2, 4]. A Newton direction is obtained by linearizing the diagonal system

$$(\Lambda_X + \Delta\Lambda_X)(\Lambda_Z + \Delta\Lambda_Z) = \gamma\mu I.$$

However, the choice of variables in this method makes the feasibility restrictions nonlinear, viz. $Q\Lambda_X Q^{\mathrm{H}} \in B + \mathcal{A}$ and $Q\Lambda_Z Q^{\mathrm{H}} \in C + \mathcal{A}^{\perp}$. Another weak point of the Q–method is its lack of global convergence.

A different method with eigenvalue targets was proposed by Sturm and Zhang [149, 144], as a way to extend the v–space approach of Kojima, Megiddo, Noma and Yoshise [73] to semidefinite programming. In the semidefinite V–space approach, we derive a Newton direction by setting a target for V^2, the diagonal matrix of eigenvalues of XZ, as discussed in this chapter. For path–following methods, we arrived at the Newton equation (3.15), which corresponds with the Nesterov–Todd direction [113].

In order to obtain (3.15) in the framework of Nesterov and Todd [113], it is important to use a modified Newton direction, in which the Hessian of the barrier function is evaluated at the scaling point $D(X, Z)$, instead of the solution pair (X, Z) itself. This scaling point gives the Nesterov–Todd direction an artificial flavor, and this is probably one reason why the Nesterov–Todd direction is not as popular as it should be.

However, Sturm and Zhang [149] derived the Nesterov–Todd direction as a pure Newton direction for the central path equation $V^2 = \gamma\mu I$, exactly as we did in this chapter. The concept of a symmetric primal–dual transformation is used to explain our interest in V–space solutions. The V–space solutions themselves are scale invariant, and the Newton direction (3.15) is pure. Unfortunately, many people somehow got the wrong impression that the direction (3.15) is some kind of modified or scaled Newton direction.

The V–space approach seems to be more versatile than self–scaled barrier and matrix target techniques, since it is not restricted to targets on the central path. In fact, (3.8) shows that any diagonal search direction $D_X + D_Z$ can be interpreted as a Newton direction for V by choosing an appropriate target, see also Chapter 7. We will now show that off-diagonal entries $(D_X + D_Z)_{ij}$, $i \neq j$ also have an interpretation in the V–space, for those (i, j) where $V_{ii} \neq V_{jj}$.

To see this, recall that the direction (3.15) was derived for a specific choice of the transformation $G(t)$, namely for Hermitian positive definite $G(t)$. If we consider factorizations

$D(\bar{X}(t), \bar{Z}(t)) = G(t)G(t)^{\mathrm{H}}$ with non–Hermitian $G(t)$, we obtain a family of Newton directions for a given target in the V–space. Namely, if we set a target $\tilde{D}_G \in \mathcal{H}^{\perp}$ for the skew–Hermitian part of $G(1)$, by demanding

$$P_{\mathcal{H}^{\perp}} G^{[1]}(0) = \tilde{D}_G,$$

we obtain from (3.6) that

$$V^{[1]}(0) = \frac{1}{2}(D_Z + D_X) - 2P_{\mathcal{H}}(\tilde{D}_G V).$$

Linearizing the nonlinear system $V(1)^2 = \gamma\mu I$ yields

$$\begin{cases} P_{\mathcal{H}}(V(D_X + D_Z)) = (\gamma\mu I - V^2) + 4P_{\mathcal{H}}(V P_{\mathcal{H}}(\tilde{D}_G V)) \\ \qquad D_X \in \mathcal{A}(L_d), \quad D_Z \in \mathcal{A}^{\perp}(L_d), \quad \tilde{D}_G \in \mathcal{H}^{\perp}, \end{cases} \tag{3.41}$$

from which we obtain

$$D_X + D_Z = \gamma\mu V^{-1} - V + 4P_{\mathcal{H}}(\tilde{D}_G V).$$

Using the fact that $\tilde{D}_G$ is skew–Hermitian, we have

$$4P_{\mathcal{H}}(\tilde{D}_G V) = 2(\tilde{D}_G V - V\tilde{D}_G),$$

which is a Hermitian matrix with only off-diagonal entries, whereas $(\gamma\mu V^{-1} - V)$ is a diagonal matrix. This implies that

$$\|D_X + D_Z\|_F^2 = \|\gamma\mu V^{-1} - V\|_F^2 + 4\|\tilde{D}_G V - V\tilde{D}_G\|_F^2. \tag{3.42}$$

Notice that on the i-th row and j-th column of $(D_X + D_Z)$, we have

$$(D_X + D_Z)_{ij} = (\gamma\mu V^{-1} - V)_{ij} + 2(V_{jj} - V_{ii})(\tilde{D}_G)_{ij}. \tag{3.43}$$

Therefore, (D_X, D_Z) is a V–space direction if

$$V_{ii} = V_{jj},\ i \neq j \Longrightarrow (D_X + D_Z)_{ij} = 0 \quad \text{for } i, j \in \{1, 2, \ldots, \bar{n}\}.$$

In this way, we can interpret the Kojima–Shindoh–Hara family of primal–dual path–following directions (3.33) in the V–space framework. Namely, if we let

$$\bar{U}_X := L_d^{-1} U_X L_d^{-\mathrm{H}}, \quad \bar{U}_Z := L_d^{\mathrm{H}} U_Z L_d,$$

we can rewrite (3.33) as

$$\begin{cases} V(D_Z + \bar{U}_Z) + (D_X + \bar{U}_X)V = \gamma\mu I - V^2 \\ \qquad D_X \in \mathcal{A}(L_d), \quad D_Z \in \mathcal{A}^{\perp}(L_d) \\ \qquad (L_d \bar{U}_X L_d^{\mathrm{H}}, L_d^{-\mathrm{H}} \bar{U}_Z L_d^{-1}) \in \mathcal{M}. \end{cases} \tag{3.44}$$

Hence, the Kojima–Shindoh–Hara directions satisfy

$$\begin{aligned}\gamma\mu I - V^2 &= P_{\mathcal{H}}(\gamma\mu I - V^2)\\ &= P_{\mathcal{H}}(V(D_Z + \bar{U}_Z) + (D_X + \bar{U}_X)V)\\ &= P_{\mathcal{H}}(V(D_X + D_Z)) + P_{\mathcal{H}}(V\bar{U}_Z + \bar{U}_X V).\end{aligned}$$

Since $\bar{U}_X, \bar{U}_Z \in \mathcal{H}^\perp$, it holds that $P_{\mathcal{H}}(V\bar{U}_Z + \bar{U}_X V) = P_{\mathcal{H}}(V(\bar{U}_Z - \bar{U}_X))$, and

$$(D_X + D_Z)_{ij} = (\gamma\mu V^{-1} - V)_{ij} + \frac{V_{jj} - V_{ii}}{V_{ii} + V_{jj}}(\bar{U}_Z - \bar{U}_X)_{ij}, \tag{3.45}$$

for all rows i and columns j.

Letting

$$(\tilde{D}_G)_{ij} := \frac{(\bar{U}_Z)_{ij} - (\bar{U}_X)_{ij}}{2(V_{ii} + V_{jj})} \quad \forall\, i, j, \tag{3.46}$$

it follows that any Kojima–Shindoh–Hara direction (3.45) can be interpreted as a V–space direction (3.43). Hence, our remarks on the V–space search direction are applicable, such as the observation that the diagonal of $D_X + D_Z$ is equal to $\gamma\mu V^{-1} - V$, for any path–following direction in this family. In this chapter, we have made the choice $\tilde{D}_G = 0$ in the V–space framework, which corresponds to $\bar{U}_X = \bar{U}_Z$ in the Kojima–Shindoh–Hara framework, and this yields the Nesterov–Todd direction. As can be seen from (3.42), this choice results in the shortest Newton step $D_X + D_Z$. Considering the fact that the region of quadratic convergence of Newton's method is closely tied to the length of the Newton step, this explains why we prefer to choose $\tilde{D}_G = 0$.

In a similar fashion, we can interpret the Monteiro–Tsuchiya [102] family of primal–dual path–following directions (3.40) in the V–space framework. Namely, if we let

$$\hat{U}_X := L_d^{-1}(L_X^{[1]}(0)L_X^{-1})L_d,$$

then we see from (3.40) that the Monteiro–Tsuchiya directions satisfy

$$VD_Z + D_X V = (\gamma\mu I - V^2) + (\hat{U}_X V^2 - V^2\hat{U}_X).$$

By taking the Hermitian part, we have

$$P_{\mathcal{H}}(V(D_X + D_Z)) = (\gamma\mu I - V^2) + P_{\mathcal{H}}\left((\hat{U}_X - \hat{U}_X^{\mathrm{H}})V^2\right),$$

which implies

$$(D_X + D_Z)_{ij} = (\gamma\mu V^{-1} - V)_{ij} + (V_{jj} - V_{ii})(\hat{U}_X - \hat{U}_X^{\mathrm{H}})_{ij}. \tag{3.47}$$

for all rows i and columns j. Referring to (3.43), it follows that Monteiro–Tsuchiya directions are V–space directions with $\tilde{D}_G = P_{\mathcal{H}^\perp}(\hat{U}_X)$. In the Monteiro–Tsuchiya framework, we can set any desired value for $P_{\mathcal{H}^\perp}(\hat{U}_X)$, by choosing an appropriate invertible matrix

L in (3.40). Therefore, the Monteiro–Tsuchiya framework is equivalent to the V–space framework, and all our remarks concerning the V–space directions are again applicable.

Interestingly, we obtain the Nesterov–Todd direction ($\tilde{D}_G = 0$) by choosing $L = L_Z$ with $L_Z L_Z^{\mathrm{H}} = Z$ in the Monteiro–Tsuchiya framework. Namely, from (3.40) we have then $L_Z^{\mathrm{H}} L_X \in \mathcal{H}$ and consequently $L_X L_Z^{-1} = L_Z^{-\mathrm{H}}(L_Z^{\mathrm{H}} L_X) L_Z^{-1} \in \mathcal{H}$. This further implies the identity

$$X = (L_X L_Z^{-1}) Z (L_X L_Z^{-1}),$$

so that $L_X L_Z^{-1} = D(X, Z)$. Now, we have

$$L_Z^{\mathrm{H}} L_X^{[1]}(0) \in \mathcal{H} \Longrightarrow L_Z^{-\mathrm{H}}(L_Z^{\mathrm{H}} L_X^{[1]}(0)) L_Z^{-1} = (L_X^{[1]}(0) L_X^{-1})(L_X L_Z^{-1}) = L_d \hat{U}_X L_d^{\mathrm{H}} \in \mathcal{H},$$

and hence $\tilde{D}_G = P_{\mathcal{H}^\perp}(\hat{U}_X) = 0$.

3.6.4 Final Remarks on the Search Direction

As pointed out by Kojima, Shida and Shindoh [79], the difference between the various primal–dual path–following directions is minor if XZ is close to a multiple of the identity matrix, which is the case in the $\mathcal{N}_2$–neighborhood. However, for iterates that are only loosely centered, such as in the $\mathcal{N}_\infty^-$–neighborhood, the difference can be considerable. In particular, the long step path–following method achieves an $\mathbf{O}(\bar{n} \log(X^{(0)} \bullet Z^{(0)}/\epsilon))$ iteration bound with the Nesterov–Todd direction (see Monteiro and Zhang [105], Sturm and Zhang [148] and Theorem 3.4), whereas the iteration bound is $\mathbf{O}(\bar{n}^{3/2} \log(X^{(0)} \bullet Z^{(0)}/\epsilon))$ with the Helmberg–Rendl–Vanderbei–Wolkowicz direction (see Monteiro [97] and Monteiro and Zhang [105]). Another interior point algorithm that does not follow the central path closely is the primal–dual affine scaling algorithm, which was proposed for linear programming by Monteiro, Adler and Resende [99]. De Klerk, Roos and Terlaky [72] proved polynomial convergence for an extension of this affine scaling algorithm to semidefinite programming, using the V–space framework (with $\tilde{D}_G = 0$). However, Muramatsu and Vanderbei [107] show that the algorithm may fail if Helmberg–Rendl–Vanderbei–Wolkowicz directions are used. Some non–path–following methods are only defined in the V–space framework, such as the primal–dual Dikin–type algorithm of Jansen, Roos and Terlaky [60] and the primal–dual cone affine scaling algorithm of Sturm and Zhang [145], which were extended to semidefinite programming by De Klerk Roos and Terlaky [72] and Berkelaar, Sturm and Zhang [11], respectively.

Finally, it should be noticed that in this book, semidefinite programming is treated in Hermitian matrices, whereas all cited references study the real symmetric case.

Chapter 4

Self–Dual Embedding Technique

In Chapter 3, we have analyzed the iteration complexity of finding an ϵ-optimal solution, if an interior, sufficiently centered pair of primal and dual solutions is known beforehand. We will see in this chapter how we can adapt the algorithms of Chapter 3 to solve semidefinite programming problems without any pre–knowledge. To this end, we use the self–dual embedding technique. This technique will also be used to tackle semidefinite programming problems that may be unbounded, unsolvable, or infeasible.

Self-duality will be discussed within the framework of conic convex programming. In particular, we consider the conic convex program

$$\text{(P)} \quad \inf\{c^{\mathrm{T}}x \mid x \in (b + \mathcal{A}) \cap \mathcal{K}\}$$

and its dual

$$\text{(D)} \quad \inf\{b^{\mathrm{T}}z \mid z \in (c + \mathcal{A}^{\perp}) \cap \mathcal{K}^{*}\},$$

where $\mathcal{K} \subset \Re^n$ is a convex cone, $\mathcal{A}$ is a linear subspace of $\Re^n$ and b and c are given vectors in $\mathcal{A}^{\perp}$ and $\mathcal{A}$ respectively. Throughout this chapter, we make the following assumption.

Assumption 4.1 *The convex cone $\mathcal{K}$ is closed, solid ($\operatorname{int} \mathcal{K} \neq \emptyset$) and pointed ($\mathcal{K} \cap -\mathcal{K} = \{0\}$).*

Notice that under Assumption 4.1, any nonzero direction (for (P) or (D)) is a one–sided direction and vice versa.

4.1 Self–Duality

Self–duality has been defined by Duffin [32] for conic convex programs that are formulated in the so–called symmetric form. More recently, Ye, Todd and Mizuno [163] formulated a linear program in a different form, and argued that their program is self–dual since "the dual of the problem is equivalent to the primal". Below, we propose a definition of self–duality that does not depend on the specific form in which the program is formulated.

Definition 4.1 *A conic convex program CP*$(b, c, \mathcal{A}, \mathcal{K})$ *is* Π self–dual *if* Π *is a symmetric permutation matrix such that*

$$c = \Pi b,\ \mathcal{A}^{\perp} = \Pi\mathcal{A},\ \mathcal{K}^* = \Pi\mathcal{K}.$$

If, in addition, $b = c = 0$*, then the program is said to be* homogeneous*.*

Notice from the above definition that a self–dual program is indeed its own dual, after a simple reordering of the variables. More precisely, if CP$(b, c, \mathcal{A}, \mathcal{K})$ is a Π self–dual conic convex program, then

$$\Pi(b + \mathcal{A}) = c + \mathcal{A}^{\perp}, \quad \Pi\mathcal{K} = \mathcal{K}^*,$$

so that $x \in \Re^n$ is primal feasible if and only if Πx is dual feasible, and, using the symmetry of Π,

$$c^{\mathrm{T}} = b^{\mathrm{T}}\Pi,$$

so that the primal objective value $c^{\mathrm{T}}x$ is identical to the dual objective value $b^{\mathrm{T}}(\Pi x)$ for all $x \in \Re^n$. Notice also that $\mathcal{K}^* = \Pi\mathcal{K}$ implies that $\mathcal{K}$ is closed; self–dual conic convex programs are therefore always closed.

Remark 4.1 *The requirement that* Π *is symmetric, i.e.* $\Pi = \Pi^T$*, is natural. Namely, for any permutation matrix* Π*, there holds* $\Pi^T = \Pi^{-1}$*. From the relation* $\mathcal{A}^{\perp} = \Pi\mathcal{A}$*, it therefore follows using Lemma 2.3 that*

$$\Pi\mathcal{A} = \mathcal{A}^{\perp} = (\Pi^T\Pi\mathcal{A})^{\perp} = \Pi^{-1}(\Pi\mathcal{A})^{\perp} = \Pi^T\mathcal{A},$$

i.e. $\Pi\mathcal{A} = \Pi^T\mathcal{A}$ *and similarly,* $\Pi\mathcal{A}^{\perp} = \Pi^T\mathcal{A}^{\perp}$*. Hence, the relation* $\mathcal{A}^{\perp} = \Pi\mathcal{A}$ *already implies some symmetry.*

Below are two elementary results for self–dual conic convex programs.

Lemma 4.1 *Let* $\mathcal{A}$ *be a linear subspace of* $\Re^n$ *such that* $\mathcal{A}^{\perp} = \Pi\mathcal{A}$ *for some symmetric permutation matrix* Π*. Then*

$$\Pi P_{\mathcal{A}} = P_{\mathcal{A}^{\perp}}\Pi, \quad \Pi P_{\mathcal{A}^{\perp}} = P_{\mathcal{A}}\Pi.$$

Proof. Since Π is a symmetric permutation matrix, we know that $\Pi\mathcal{A}^\perp = \Pi^2\mathcal{A} = \mathcal{A}$. Hence, the orthogonal decomposition

$$x = y + z, \quad y \in \mathcal{A},\ z \in \mathcal{A}^\perp$$

is equivalent to

$$\Pi x = \Pi y + \Pi z, \quad \Pi y \in \mathcal{A}^\perp,\ \Pi z \in \mathcal{A}.$$

From the above decompositions, it follows for arbitrary x that $\Pi P_{\mathcal{A}} x = \Pi y$ and $P_{\mathcal{A}^\perp}\Pi x = \Pi y$ respectively. **Q.E.D.**

Lemma 4.2 *Let $CP(b, c, \mathcal{A}, \mathcal{K})$ be a Π self–dual program where $\mathcal{K}$ is a solid convex cone. Then*

$$y^T P_{\mathcal{A}}\Pi y = y^T P_{\mathcal{A}^\perp}\Pi y = \frac{1}{2} y^T \Pi y > 0 \text{ for all } y \in \text{ int } \mathcal{K}.$$

Proof. Let $y \in \text{ int } \mathcal{K}$, then

$$\Pi y \in \Pi \text{ int } \mathcal{K} = \text{ int } \Pi\mathcal{K} = \text{ int } \mathcal{K}^*,$$

where in the first identity, we used the fact that Π is an invertible matrix. Since $\mathcal{K}^*$ is pointed (cf. Corollary 2.1), and $0 \neq \Pi y \in \text{ int } \mathcal{K}^*$, it follows from Theorem 2.2 that

$$0 < y^{\mathrm{T}}\Pi y = y^{\mathrm{T}}\Pi P_{\mathcal{A}} y + y^{\mathrm{T}}\Pi P_{\mathcal{A}^\perp} y.$$

Using now Lemma 4.1 and the symmetricity of Π, the lemma follows. **Q.E.D.**

The properties of self–dual conic convex programs result in a particularly nice form of weak–duality:

Theorem 4.1 *If $CP(b, c, \mathcal{A}, \mathcal{K})$ is a Π self–dual conic convex program, then*

$$c^T x = \frac{1}{2} x^T \Pi x \geq 0$$

for all $x \in (b + \mathcal{A}) \cap \mathcal{K}$.

Proof. Since $x - b \in \mathcal{A}$, $\Pi x - c \in \mathcal{A}^\perp$ and $b \perp c$, we have

$$0 = (x - b)^{\mathrm{T}}(\Pi x - c) = x^{\mathrm{T}}\Pi x - b^{\mathrm{T}}\Pi x - c^{\mathrm{T}} x = x^{\mathrm{T}}\Pi x - 2c^{\mathrm{T}} x.$$

Moreover, $x^{\mathrm{T}}\Pi x \geq 0$ because $x \in \mathcal{K}$ and $\Pi x \in \mathcal{K}^*$. **Q.E.D.**

If x^* is a solution to a Π self–dual program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ and $(x^*)^{\mathrm{T}}\Pi x^* = 0$, then x^* is called a *self–complementary* solution. It follows from Theorem 4.1 that if x^* is a self–complementary solution, then it is also an optimal solution.

As before, we denote the optimal value and the subvalue of the conic convex program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ by p^* and p^-, respectively. That is,

$$p^* := \inf_x \{c^{\mathrm{T}}x \mid x \in (b + \mathcal{A}) \cap \mathcal{K}\}, \quad p^- := \lim_{\epsilon \downarrow 0} \inf_x \{c^{\mathrm{T}}x \mid x \in \mathcal{K},\ \operatorname{dist}(x, b + \mathcal{A}) < \epsilon\}.$$

If $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ is self–dual, then

$$p^- = -p^*, \tag{4.1}$$

as follows from Theorem 2.6. Since $p^* \geq p^-$, we know that the optimal value p^* is nonnegative, and the subvalue p^- is non-positive. This implies in particular that a self–dual program cannot be unbounded. In fact, only the diagonal entries in Table 2.2 (which summarizes duality relations) are applicable to self–dual programs, because for such programs, the primal and the dual are essentially identical.

4.2 Self–Dual Embedding

There are basically two different types of dual variables involved in closed conic convex programming, viz.

- Dual feasible solutions, and
- Nonzero dual directions.

The former yield lower bounds on the optimal value, and the latter concern the feasibility of the problem, see Chapter 2. The homogeneous self–dual embedding, to be discussed in Section 4.2.1, combines both types of dual variables into a single self–dual program. In Section 4.2.2, we will treat the extended self–dual model, which is a strongly feasible self–dual model for which the optimal solution set corresponds to feasible solutions of the homogeneous self–dual model.

4.2.1 The Homogeneous Self–Dual Model

Consider a Π_{SD} self–dual program (SD),

$$\text{(SD)} \qquad \inf_{x_{\mathrm{SD}}} \{c_{\mathrm{SD}}^{\mathrm{T}} x_{\mathrm{SD}} \mid x_{\mathrm{SD}} \in (b_{\mathrm{SD}} + \mathcal{A}_{\mathrm{SD}}) \cap \mathcal{K}_{\mathrm{SD}}\},$$

i.e. (SD) is the conic convex program $\mathrm{CP}(b_{\mathrm{SD}}, c_{\mathrm{SD}}, \mathcal{A}_{\mathrm{SD}}, \mathcal{K}_{\mathrm{SD}})$. We assume that $\mathcal{K}_{\mathrm{SD}}$ is solid. The optimal value of (SD) is denoted by p^*_{SD}. We introduce an invertible matrix $M(b_{\mathrm{SD}}, c_{\mathrm{SD}})$,

$$M(b_{\mathrm{SD}}, c_{\mathrm{SD}}) := \begin{bmatrix} I & b_{\mathrm{SD}} & 0 \\ 0 & 1 & 0 \\ -c_{\mathrm{SD}}^{\mathrm{T}} & 0 & 1 \end{bmatrix},$$

and we let

$$\mathcal{A}_{\mathrm{H}} := M(b_{\mathrm{SD}}, c_{\mathrm{SD}})(\mathcal{A}_{\mathrm{SD}} \times \Re \times \{0\}), \quad \mathcal{K}_{\mathrm{H}} := \mathcal{K}_{\mathrm{SD}} \times \Re_+ \times \Re_+.$$

Applying Lemma 2.3, we have

$$\mathcal{A}_{\mathrm{H}}^{\perp} = M(b_{\mathrm{SD}}, c_{\mathrm{SD}})^{-\mathrm{T}}(\mathcal{A}_{\mathrm{SD}}^{\perp} \times \{0\} \times \Re),$$

where

$$M(b_{\mathrm{SD}}, c_{\mathrm{SD}})^{-\mathrm{T}} = \begin{bmatrix} I & 0 & c_{\mathrm{SD}} \\ -b_{\mathrm{SD}}^{\mathrm{T}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Using the self–duality of (SD), it follows that $\mathrm{CP}(0, 0, \mathcal{A}_{\mathrm{H}}, \mathcal{K}_{\mathrm{H}})$ is a homogeneous Π_{H} self–dual program with

$$\Pi_{\mathrm{H}} = \begin{bmatrix} \Pi_{\mathrm{SD}} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

This type of homogeneous self–dual program was proposed by Goldman and Tucker [43, 155] in the context of linear programming. The convex cone $\mathcal{A}_{\mathrm{H}} \cap \mathcal{K}_{\mathrm{H}}$ consists of those solutions $(x_{\mathrm{SD}}, x_0, z_0)$ for which

$$(\mathrm{H}) \qquad \begin{cases} x_{\mathrm{SD}} \in (x_0 b_{\mathrm{SD}} + \mathcal{A}_{\mathrm{SD}}) \cap \mathcal{K}_{\mathrm{SD}}, \\ x_0 \geq 0, \\ z_0 = -c_{\mathrm{SD}}^{\mathrm{T}} x_{\mathrm{SD}} \geq 0. \end{cases}$$

The concepts of complementary solution, improving direction and nonzero lower level direction for (SD) can be characterized in terms of the homogeneous program $(0, 0, \mathcal{A}_{\mathrm{H}}, \mathcal{K}_{\mathrm{H}})$ as follows:

- x_{SD} is a *self–complementary solution* to (SD) if and only if

$$(x_{\mathrm{SD}}, 1, 0) \in \mathcal{A}_{\mathrm{H}} \cap \mathcal{K}_{\mathrm{H}}.$$

- x_{SD} is an *improving direction* for (SD) if and only if

$$(x_{\mathrm{SD}}, 0, z_0) \in \mathcal{A}_{\mathrm{H}} \cap \mathcal{K}_{\mathrm{H}} \text{ for some } z_0 > 0.$$

- x_{SD} is a *nonzero lower level direction* for (SD) if and only if

$$(x_{\mathrm{SD}}, 0, z_0) \in \mathcal{A}_{\mathrm{H}} \cap \mathcal{K}_{\mathrm{H}} \setminus \{0\} \text{ for some } z_0 \geq 0.$$

With the above observation, it is straightforward to prove the following theorem.

Theorem 4.2 *If $(x_{SD}, x_0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H$, then*

$$x_{SD}^T \Pi_{SD} x_{SD} = 0, \quad x_0 z_0 = 0.$$

Moreover, if $(x_{SD}, x_0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H \setminus \{0\}$ and (SD) is strongly feasible, then $x_0 > 0$ and x_{SD}/x_0 is a self–complementary solution of (SD).

Remark 4.2 *It follows from Theorem 4.2 that if*

$$(x_{SD}, 0, z_0) \in \mathcal{A}_H \cap \mathcal{K}_H \setminus \{0\},$$

then (SD) is not strongly feasible, i.e. it is either weakly feasible, or weakly infeasible, or strongly infeasible. However, if it is strongly infeasible, then it must have an improving direction x'_{SD} and $(x'_{SD}, 0, -c_{SD}^T x'_{SD}) \in \mathcal{A}_H \cap \mathcal{K}_H$.

Remark 4.3 *Even if (SD) is not strongly feasible, it may have a self–complementary solution x_{SD}, in which case $(x_{SD}, 1, 0) \in \mathcal{A}_H \cap \mathcal{K}_H$.*

4.2.2 The Extended Self–Dual Model

Given a Π_{SD} self–dual program (SD), we have constructed a Π_{H} homogeneous self–dual model $(0, 0, \mathcal{A}_{\mathrm{H}}, \mathcal{K}_{\mathrm{H}})$, generalizing the Goldman–Tucker [43] model to conic convex programming. We will now add a normalization constraint to the model, and we will make the program strongly dual. This results in an extension of Ye, Todd and Mizuno's self–dual formulation [163] for linear programming, to the context of conic convex programming.

Choose $\iota \in \operatorname{int} \mathcal{K}_{\mathrm{H}}$, and let

$$\rho := \frac{\iota^{\mathrm{T}} P_{\mathcal{A}_{\mathrm{H}}} \Pi_{\mathrm{H}} \iota}{\left\| P_{\mathcal{A}_{\mathrm{H}}^{\perp}} \iota \right\|_2^2}.$$

By construction, $\mathcal{K}_{\mathrm{H}}$ is solid, and we can apply Lemma 4.2 to conclude that ρ is a well defined positive quantity. Define

$$b_{\mathrm{E}} := \rho\, P_{\mathcal{A}_{\mathrm{H}}} \Pi_{\mathrm{H}} \iota, \quad c_{\mathrm{E}} := \rho\, P_{\mathcal{A}_{\mathrm{H}}^{\perp}} \iota, \tag{4.2}$$

and notice from Lemma 4.1 that

$$c_{\mathrm{E}} = \Pi_{\mathrm{H}} b_{\mathrm{E}}. \tag{4.3}$$

Since c_{E} is simply a permutation of b_{E} and $\rho > 0$, we have

$$\|b_{\mathrm{E}}\|_2 = \|c_{\mathrm{E}}\|_2 > 0.$$

Using the definitions of c_{E} and ρ, and then applying Lemma 4.2, we further have

$$\frac{\|b_{\mathrm{E}}\|_2^2}{\rho} = \frac{\|c_{\mathrm{E}}\|_2^2}{\rho} = \rho \|P_{\mathcal{A}_{\mathrm{H}}^{\perp}} \iota\|_2^2 = \iota^{\mathrm{T}} P_{\mathcal{A}_{\mathrm{H}}} \Pi_{\mathrm{H}} \iota = \frac{\iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota}{2}. \tag{4.4}$$

We shall now study the conic convex program $\mathrm{CP}(b_{\mathrm{E}}, c_{\mathrm{E}}, \mathcal{A}_{\mathrm{E}}, \mathcal{K}_{\mathrm{H}})$, where

$$\mathcal{A}_{\mathrm{E}} := (\mathcal{A}_{\mathrm{H}} \cap \text{ Ker } b_{\mathrm{E}}^{\mathrm{T}}) \oplus \text{ Img } c_{\mathrm{E}}.$$

Just as in the homogeneous self–dual model, we partition the decision variable as

$$x_{\mathrm{E}} = (x_{\mathrm{SD}}, x_0, z_0),$$

with $x_{\mathrm{SD}} \in \mathcal{K}_{\mathrm{SD}}$, $x_0 \in \Re_+$ and $z_0 \in \Re_+$. Introducing an auxiliary variable

$$y_0 := \frac{\rho}{\|c_{\mathrm{E}}\|_2^2}\, c_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}}, \tag{4.5}$$

we can reformulate $\mathrm{CP}(b_{\mathrm{E}}, c_{\mathrm{E}}, \mathcal{A}_{\mathrm{E}}, \mathcal{K}_{\mathrm{H}})$ as follows:

$$(\mathrm{E}) \qquad \begin{array}{ll} \min & \left(\iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota / 2\right) y_0 \\ s.t. & x_{\mathrm{E}} - y_0 \iota \in \mathcal{A}_{\mathrm{H}} \\ & b_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}} = \|b_{\mathrm{E}}\|_2^2 \\ & x_{\mathrm{E}} \in \mathcal{K}_{\mathrm{H}},\ y_0 \in \Re. \end{array}$$

To see this, remark that by definition of c_{E},

$$x_{\mathrm{E}} - y_0 \iota \in \mathcal{A}_{\mathrm{H}} \iff x_{\mathrm{E}} - \frac{y_0}{\rho} c_{\mathrm{E}} \in \mathcal{A}_{\mathrm{H}}.$$

Since obviously,

$$b_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}} = \|b_{\mathrm{E}}\|_2^2 \iff x_{\mathrm{E}} - b_{\mathrm{E}} \in \text{ Ker } b_{\mathrm{E}}^{\mathrm{T}},$$

we obtain from the above two relations and the fact that $b_{\mathrm{E}} \in \mathcal{A}_{\mathrm{H}}$, $c_{\mathrm{E}} \in \mathcal{A}_{\mathrm{H}}^{\perp}$, that

$$x_{\mathrm{E}} - b_{\mathrm{E}} \in \mathcal{A}_{\mathrm{E}} \iff \begin{cases} x_{\mathrm{E}} - y_0 \iota \in \mathcal{A}_{\mathrm{H}}, \\ b_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}} = \|b_{\mathrm{E}}\|_2^2 \end{cases}$$

From the relation $x_{\mathrm{E}} \in y_0 \iota + \mathcal{A}_{\mathrm{H}}$, we obtain

$$\iota^{\mathrm{T}} \Pi_{\mathrm{H}} \left(P_{\mathcal{A}_{\mathrm{H}}^{\perp}} x_{\mathrm{E}} \right) = \iota^{\mathrm{T}} \Pi_{\mathrm{H}} \left(y_0 P_{\mathcal{A}_{\mathrm{H}}^{\perp}} \iota \right) = \frac{y_0}{2} \iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota,$$

where we used Lemma 4.2 in the last identity. Using the definition of b_{E}, it thus follows that

$$\frac{b_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}}}{\rho} = \iota^{\mathrm{T}} \Pi_{\mathrm{H}} x_{\mathrm{E}} - \frac{y_0}{2} \iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota.$$

Combining this with (4.4), we obtain an alternative form for the normalization constraint of (E), viz.

$$b_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}} = \|b_{\mathrm{E}}\|_2^2 \iff \iota^{\mathrm{T}} \Pi_{\mathrm{H}} x_{\mathrm{E}} = \frac{1 + y_0}{2} \iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota. \tag{4.6}$$

It is obvious from (4.6) that the lower level sets of (E) are bounded. Moreover, it is now easily verified that $\iota \in (b_{\mathrm{E}} + \mathcal{A}_{\mathrm{E}}) \cap \operatorname{int} \mathcal{K}_{\mathrm{H}}$, i.e. ι is an interior solution, which can serve as an initial solution in interior point methods (remark that we can choose any $\iota \in \operatorname{int} \mathcal{K}_{\mathrm{H}}$). The feasible solutions x_{E} of (E) for which $y_0 = 0$ correspond to those solutions of the homogeneous model (H) that are normalized by the constraint $\iota^{\mathrm{T}} \Pi_{\mathrm{H}} x_{\mathrm{E}} = \iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota / 2$. The normalization guarantees that if $y_0 = 0$, then x_{E} is a nonzero direction of the homogeneous model (H).

Theorem 4.3 *The conic convex program CP*$(b_E, c_E, \mathcal{A}_E, \mathcal{K}_H)$ *has the following properties:*

1. *It is self–dual,*
2. *It has an interior solution, viz.* $\iota \in (b_E + \mathcal{A}_E) \cap \operatorname{int} \mathcal{K}_H$, *and hence*
3. *It has a self–complementary solution,*
4. *Any self–complementary solution is a nonzero direction of CP*$(0, 0, \mathcal{A}_H, \mathcal{K}_H)$, *and*
5. *For any nonzero direction* x_E *of the homogeneous program CP*$(0, 0, \mathcal{A}_H, \mathcal{K}_H)$, *there exists* $\alpha > 0$ *such that* αx_{SD} *is a self–complementary solution for CP*$(b_E, c_E, \mathcal{A}_E, \mathcal{K}_H)$.

Proof.

1. It is already known from (4.3) that $c_{\mathrm{E}} = \Pi_{\mathrm{H}} b_{\mathrm{E}}$. Moreover, $\mathcal{K}_{\mathrm{H}}^{*} = \Pi_{\mathrm{H}}\mathcal{K}_{\mathrm{H}}$, since $\mathrm{CP}(0, 0, \mathcal{A}_{\mathrm{H}}, \mathcal{K}_{\mathrm{H}})$ is self–dual. It remains to show that $\mathcal{A}_{\mathrm{E}}^{\perp} = \Pi_{\mathrm{H}}\mathcal{A}_{\mathrm{E}}$. To this end, we remark using Lemma 2.1 that

$$\begin{aligned}\mathcal{A}_{\mathrm{E}}^{\perp} &= \left[\ (\mathcal{A}_{\mathrm{H}} \cap \ \mathrm{Ker}\ b_{\mathrm{E}}^{\mathrm{T}}) \oplus \ \mathrm{Img}\ c_{\mathrm{E}}\ \right]^{\perp} \\ &= (\mathcal{A}_{\mathrm{H}} \cap \ \mathrm{Ker}\ b_{\mathrm{E}}^{\mathrm{T}})^{\perp} \cap \ \mathrm{Ker}\ c_{\mathrm{E}}^{\mathrm{T}} \\ &= (\mathcal{A}_{\mathrm{H}}^{\perp} \oplus \ \mathrm{Img}\ b_{\mathrm{E}}) \cap \ \mathrm{Ker}\ c_{\mathrm{E}}^{\mathrm{T}} \\ &= (\mathcal{A}_{\mathrm{H}}^{\perp} \cap \ \mathrm{Ker}\ c_{\mathrm{E}}^{\mathrm{T}}) \oplus \ \mathrm{Img}\ b_{\mathrm{E}},\end{aligned}$$

   where we used $b_{\mathrm{E}} \perp c_{\mathrm{E}}$ in the last identity. Using (4.3) and the fact that $\mathcal{A}_{\mathrm{H}}^{\perp} = \Pi_{\mathrm{H}}\mathcal{A}_{\mathrm{H}}$, it follows that $\mathcal{A}_{\mathrm{E}}^{\perp} = \Pi_{\mathrm{H}}\mathcal{A}_{\mathrm{E}}$.

2. Using (4.6), it is easily verified that the solution $x_{\mathrm{E}} = \iota$, $y_0 = 1$ satisfies all the constraints of (E).

3. The self–dual program $\mathrm{CP}(b_{\mathrm{E}}, c_{\mathrm{E}}, \mathcal{A}_{\mathrm{E}}, \mathcal{K}_{\mathrm{H}})$ has a self–complementary solution because it is strongly feasible, see Theorem 2.7.

4. Let x_{E} be a self-complementary solution of (E). By definition, this means that

$$0 = x_{\mathrm{E}}^{\mathrm{T}}\Pi_{\mathrm{H}} x_{\mathrm{E}} = 2\left\|c_{\mathrm{E}}\right\|_2^2 y_0.$$

   Consequently, $y_0 = 0$ and $x_{\mathrm{E}} \in \mathcal{A}_{\mathrm{H}} \cap \mathcal{K}_{\mathrm{H}}$. Moreover, $x_{\mathrm{E}} \neq 0$ due to the normalization constraint (4.6).

5. Let $x_{\mathrm{E}} \in \mathcal{A}_{\mathrm{H}} \cap \mathcal{K}_{\mathrm{H}} \setminus \{0\}$. Since $\Pi_{\mathrm{H}}\iota \in \ \mathrm{int}\ \mathcal{K}_{\mathrm{H}}^{*}$, we obtain using Theorem 2.2 that

$$\iota^{\mathrm{T}}\Pi_{\mathrm{H}} x_{\mathrm{E}} > 0,$$

   so that

$$\frac{\iota^{\mathrm{T}}\Pi_{\mathrm{H}}\iota}{\iota^{\mathrm{T}}\Pi_{\mathrm{H}} x_{\mathrm{E}}} x_{\mathrm{E}} \in \left[b_{\mathrm{E}} + (\mathcal{A}_{\mathrm{H}} \cap \ \mathrm{Ker}\ b_{\mathrm{E}}^{\mathrm{T}})\right] \cap \mathcal{K}_{\mathrm{H}}.$$

Q.E.D.

Remark from Theorem 4.2 and Theorem 4.3 that if (SD) is strongly feasible and (x, x_0, z_0) is an optimal solution of (E), then x/x_0 is a self–complementary solution of (SD). Using the interior point method [110], we can thus obtain an optimal solution to (SD) by solving the artificial program (E), for which we can *choose* an initial feasible solution $\iota \in \ \mathrm{int}\ \mathcal{K}_{\mathrm{SD}}$. We will see in the next section that even if (SD) is not strongly feasible (in which case it may not be solvable), it is still a good idea to solve the embedding (E), if the solution method generates a so–called *weakly centered sequence.*

4.3 Weakly Centered Sequences

Up to now, we did not use the special structure of the homogeneous program (H) in our study of the extended self–dual program (E). In this section however, we will focus on the full structure of (E), and we partition the decision variable as $x_{\mathrm{E}} = (x_{\mathrm{SD}}, x_0, z_0)$, just like we did in the homogeneous model previously. Similarly, we write

$$\iota^{\mathrm{T}} = \left[\begin{array}{ccc} u_{\mathrm{SD}}^{\mathrm{T}}, & u_0, & v_0 \end{array}\right].$$

Throughout this section, we make the following assumption:

Assumption 4.2 *The cone $\mathcal{K}_{SD}$ is solid, i.e.* $\operatorname{int} \mathcal{K}_{SD} \neq \emptyset$.

Since $\iota \in \operatorname{int} \mathcal{K}_{\mathrm{H}} = (\operatorname{int} \mathcal{K}_{\mathrm{SD}}) \times \Re_{++} \times \Re_{++}$, there holds

$$u_{\mathrm{SD}} \in \operatorname{int} \mathcal{K}_{\mathrm{SD}}, \quad u_0 > 0, \quad v_0 > 0.$$

We can now formulate the model (E) as follows:

$$\begin{array}{lll} \min & (\iota^{\mathrm{T}}\Pi_{\mathrm{H}}\iota/2)y_0 & \\ \text{s.t.} & x_{\mathrm{SD}} - y_0 u_{\mathrm{SD}} \in (x_0 - y_0 u_0) b_{\mathrm{SD}} + \mathcal{A}_{\mathrm{SD}} & (4.7) \\ & z_0 - y_0 v_0 = -c_{\mathrm{SD}}^{\mathrm{T}}(x_{\mathrm{SD}} - y_0 u_{\mathrm{SD}}) & (4.8) \\ & \iota^{\mathrm{T}}\Pi_{\mathrm{H}} x_{\mathrm{E}} = (1+y_0)\iota^{\mathrm{T}}\Pi_{\mathrm{H}}\iota/2 & (4.9) \\ & x_{\mathrm{SD}} \in \mathcal{K}_{\mathrm{SD}},\ x_0 \in \Re_+,\ z_0 \in \Re_+,\ y_0 \in \Re. & \end{array}$$

Remark from (4.5) and Theorem 4.1 that

$$(\iota^{\mathrm{T}}\Pi_{\mathrm{H}}\iota/2)y_0 = c_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}} = x_0 z_0 + (x^{\mathrm{T}}\Pi_{\mathrm{SD}} x)/2. \tag{4.10}$$

In the sequel, we will analyze the behavior of weakly centered sequences for (E); the existence of such sequences will be demonstrated in Section 4.5.

Definition 4.2 *A sequence $x_E^{(k)} = (x_{SD}^{(k)}, x_0^{(k)}, z_0^{(k)}) \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H$, $k = 1, 2, \ldots,$ is* weakly centered *if and only if there exists some constant $\omega \in (0,1)$ such that*

$$x_0^{(k)} z_0^{(k)} \geq \omega c_E^T x_E^{(k)} > 0 \text{ for all } k = 1, 2, \ldots, \tag{4.11}$$

and $\lim_{k\to\infty} c_E^T x_E^{(k)} = 0$.

Condition (4.11) is also known as the minimal centrality condition [153]. This condition holds true for all path–following algorithms, and for some potential reduction methods. In particular, Nesterov and Todd [112] developed a framework of primal–dual interior point algorithms for solving self–scaled conic convex programming, which is a subclass of conic convex programming that includes linear programming and semidefinite programming, among others. All their algorithms generate a sequence of weakly centered iterates. We remark here that if (SD) is a linear (semidefinite, self–scaled) programming problem, then (E) is also a linear (semidefinite, self–scaled) programming problem, since $\mathcal{K}_{\mathrm{H}} = \mathcal{K}_{\mathrm{SD}} \times \Re_+ \times \Re_+$.

Since $c_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}} = (\iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota/2) y_0$, see (4.4) and (4.5), it follows by definition that weakly centered sequences satisfy

$$x_0^{(k)} z_0^{(k)} \geq \omega(\iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota/2) y_0^{(k)} > 0 \text{ for all } k = 1, 2, \ldots.$$

This immediately implies the following result.

Lemma 4.3 *Let $x_E^{(1)}, x_E^{(2)}, \ldots$ be a weakly centered sequence, then*

$$\lim_{k\to\infty} \frac{y_0^{(k)}}{x_0^{(k)}} = 0 \iff \lim_{k\to\infty} z_0^{(k)} = 0,$$

and

$$\lim_{k\to\infty} \frac{y_0^{(k)}}{z_0^{(k)}} = 0 \iff \lim_{k\to\infty} x_0^{(k)} = 0.$$

The lemma below shows a crucial property of weakly centered sequences: the components $x_0^{(k)}$ and $z_0^{(k)}$ avoid the boundary of the cone $\Re_+$ essentially as much as possible.

Lemma 4.4 *Let $x_E = (x, x_0, z_0) \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H$ and $\omega \in (0, 1)$ be such that $x_0 z_0 \geq \omega c_E^T x_E > 0$. For any $x'_E = (x', x'_0, z'_0) \in (b_E + \mathcal{A}_E) \cap \mathcal{K}_H$ there holds*

$$x_0 \geq \frac{\omega}{1 + (c_E^T x'_E / c_E^T x_E)} x'_0$$

and

$$z_0 \geq \frac{\omega}{1 + (c_E^T x'_E / c_E^T x_E)} z'_0.$$

Proof. Because $x'_{\mathrm{E}} - x_{\mathrm{E}} \in \mathcal{A}_{\mathrm{E}}$ and $\Pi_{\mathrm{H}}(x'_{\mathrm{E}} - x_{\mathrm{E}}) \in \mathcal{A}_{\mathrm{E}}^{\perp}$, there holds

$$\begin{aligned} 0 &= (x'_{\mathrm{E}} - x_{\mathrm{E}})^{\mathrm{T}} \Pi_{\mathrm{H}} (x'_{\mathrm{E}} - x_{\mathrm{E}}) \\ &= (x'_{\mathrm{E}})^{\mathrm{T}} \Pi_{\mathrm{H}} x'_{\mathrm{E}} + x_{\mathrm{E}}^{\mathrm{T}} \Pi_{\mathrm{H}} x_{\mathrm{E}} - 2 x_{\mathrm{E}}^{\mathrm{T}} \Pi_{\mathrm{H}} x'_{\mathrm{E}} \\ &= 2 c_{\mathrm{E}}^{\mathrm{T}} (x'_{\mathrm{E}} + x_{\mathrm{E}}) - 2 x^{\mathrm{T}} \Pi x' - 2(x_0 z'_0 + z'_0 x_0). \end{aligned} \tag{4.12}$$

Since $x \in \mathcal{K}$ and $\Pi x' \in \mathcal{K}^*$, we have $x^{\mathrm{T}} \Pi x' \geq 0$. We thus obtain from (4.12) that

$$c_{\mathrm{E}}^{\mathrm{T}} (x'_{\mathrm{E}} + x_{\mathrm{E}}) \geq x_0 z'_0 + z_0 x'_0 \tag{4.13}$$

The lemma follows by multiplying (4.13) with $x_0/c_{\mathrm{E}}^{\mathrm{T}}(x'_{\mathrm{E}}+x_{\mathrm{E}})$ and $z_0/c_{\mathrm{E}}^{\mathrm{T}}(x'_{\mathrm{E}}+x_{\mathrm{E}})$ respectively. **Q.E.D.**

The argumentation that is used in the proof of Lemma 4.4 is due to Güler and Ye [52].

Theorem 4.4 below shows why weakly centered sequences are so interesting in the context of self–dual embeddings. Namely, if we can generate a weakly centered sequence for (E) then we can also solve (SD), whenever it has a complementary solution or an improving direction. In other cases, (SD) must be either weakly feasible or weakly infeasible, and we can generate a sequence of solutions for (SD), for which the amount of constraint violation converges to zero and the corresponding objective values are in the limit contained in the interval $[p_{\mathrm{SD}}^{-}, p_{\mathrm{SD}}^{*}]$.

Theorem 4.4 *Let $x_E^{(k)} = (x^{(k)}, x_0^{(k)}, z_0^{(k)})$, $k = 1, 2, \ldots$, be a weakly centered sequence for (E). There holds*

1. $\liminf_{k\to\infty} x_0^{(k)} > 0$ *if and only if (SD) has a self–complementary solution. Moreover, if (SD) has a self–complementary solution then $x_{SD}^{(k)}/x_0^{(k)}$, $k = 1, 2, \ldots$ is a bounded sequence and therefore it has a cluster point $x_{SD}^{(\infty)}$. Any such cluster point $x_{SD}^{(\infty)}$ is a self–complementary solution of (SD).*

2. $\liminf_{k\to\infty} z_0^{(k)} > 0$ *if and only if (SD) is strongly infeasible. Moreover, if (SD) is strongly infeasible, then $x_{SD}^{(k)}/z_0^{(k)}$, $k = 1, 2, \ldots$ is a bounded sequence and therefore it has a cluster point $x_{SD}^{(\infty)}$. Any such cluster point $x_{SD}^{(\infty)}$ is an improving direction of (SD).*

3. *If $\lim_{k\to\infty} z_0^{(k)} = 0$, then $x_{SD}^{(k)}/x_0^{(k)}$ is a sequence in $\mathcal{K}_{SD}$ for which*
$$\lim_{k\to\infty} \operatorname{dist}\left(\frac{x_{SD}^{(k)}}{x_0^{(k)}}, b_{SD} + \mathcal{A}_{SD}\right) = 0, \quad \lim_{k\to\infty} \frac{c_{SD}^T x_{SD}^{(k)} + z_0^{(k)}}{x_0^{(k)}} = 0 \le p_{SD}^*.$$

4. *If $\lim_{k\to\infty} z_0^{(k)}/x_0^{(k)} = \infty$, then (SD) is infeasible, and $x_{SD}^{(k)}/z_0^{(k)}$ is an improving direction sequence in $\mathcal{K}_{SD}$, viz.*
$$\lim_{k\to\infty} \operatorname{dist}\left(\frac{x_{SD}^{(k)}}{z_0^{(k)}}, \mathcal{A}_{SD}\right) = 0, \quad \lim_{k\to\infty} \frac{c_{SD}^T x_{SD}^{(k)}}{z_0^{(k)}} = -1.$$

Proof.

1. From Theorem 4.3 and the discussion in Section 4.2.1, it follows that (SD) has a self–complementary solution if and only if (E) has an optimal solution $(x_{\mathrm{SD}}^*, x_0^*, 0)$ with

$x_0^* > 0$. From Lemma 4.4, we know that if (E) has an optimal solution $(x_{\mathrm{SD}}^*, x_0^*, 0)$ with $x_0^* > 0$, then

$$\liminf_{k\to\infty} x_0^{(k)} \geq \omega x_0^* > 0.$$

The converse is also true, because the sequence $(x_{\mathrm{SD}}^{(k)}, x_0^{(k)}, z_0^{(k)})$ is bounded (and hence it has a cluster point, which must be an optimal solution to (E)).

2. Similarly, (SD) is strongly infeasible if and only if the embedding (E) has an optimal solution $(x_{\mathrm{SD}}^*, 0, z_0^*)$ with $z_0^* > 0$, which is equivalent with the relation

$$\liminf_{k\to\infty} z_0^{(k)} \geq \omega z_0^* > 0.$$

3. Suppose that $\lim_{k\to\infty} z_0^{(k)} = 0$, so that, using Lemma 4.3,

$$\lim_{k\to\infty} \frac{y_0^{(k)}}{x_0^{(k)}} = 0.$$

Combining this with (4.7)-(4.8), we obtain

$$\lim_{k\to\infty} \operatorname{dist}(\frac{x_{\mathrm{SD}}^{(k)}}{x_0^{(k)}}, b_{\mathrm{SD}} + \mathcal{A}_{\mathrm{SD}}) = 0, \quad \lim_{k\to\infty} \frac{c_{\mathrm{SD}}^{\mathrm{T}} x_{\mathrm{SD}}^{(k)} + z_0^{(k)}}{x_0^{(k)}} = 0.$$

4. Suppose that $\lim_{k\to\infty} z_0^{(k)}/x_0^{(k)} = \infty$. Then $x_0^{(k)} \to 0$ so that using Lemma 4.3,

$$\lim_{k\to\infty} \frac{y_0^{(k)}}{z_0^{(k)}} = 0.$$

Combining this with (4.7)-(4.8), we obtain

$$\lim_{k\to\infty} \operatorname{dist}(\frac{x_{\mathrm{SD}}^{(k)}}{z_0^{(k)}}, \mathcal{A}_{\mathrm{SD}}) = 0, \quad \lim_{k\to\infty} \frac{c_{\mathrm{SD}}^{\mathrm{T}} x_{\mathrm{SD}}^{(k)}}{z_0^{(k)}} = -1.$$

By definition, $x_{\mathrm{SD}}^{(k)}/z_0^{(k)}$ is then an improving direction sequence, which implies that (SD) is infeasible (see Lemma 2.6).

Q.E.D.

Theorem 4.5 *Let $x_E^{(k)} = (x^{(k)}, x_0^{(k)}, z_0^{(k)})$, $k = 1, 2, \ldots$, be a weakly centered sequence for (E). There holds*

$$p_{SD}^* \geq \limsup_{k\to\infty} \frac{z_0^{(k)}}{x_0^{(k)}}.$$

Proof. It is known from Theorem 4.4 that $\liminf_{k\to\infty} z_0^{(k)} > 0$ if (SD) is strongly infeasible. Using e.g. Lemma 4.3, it follows that $\liminf_{k\to\infty} x_0^{(k)} = 0$, and hence

$$p^*_{\mathrm{SD}} = \lim_{k\to\infty} \frac{z_0^{(k)}}{x_0^{(k)}} = \infty.$$

Now suppose that (SD) is not strongly infeasible, or equivalently, $\liminf_{k\to\infty} z_0^{(k)} = 0$. For this case, we know from Theorem 4.4 that $x_{\mathrm{SD}}^{(k)}/x_0^{(k)}$ is a sequence in $\mathcal{K}_{\mathrm{SD}}$, with

$$\lim_{k\to\infty} \operatorname{dist}(\frac{x_{\mathrm{SD}}^{(k)}}{x_0^{(k)}}, b_{\mathrm{SD}} + \mathcal{A}) = 0.$$

Therefore, we have the following inequality for the subvalue p^-_{SD} of (SD):

$$p^-_{\mathrm{SD}} \le \liminf_{k\to\infty} \frac{c_{\mathrm{SD}}^{\mathrm{T}} x_{\mathrm{SD}}^{(k)}}{x_0^{(k)}} = -\limsup_{k\to\infty} \frac{z_0^{(k)}}{x_0^{(k)}}.$$

Using (4.1), the theorem follows. **Q.E.D.**

Remark 4.4 *It follows from Theorem 4.5 that if* $\limsup_{k\to\infty}(z_0^{(k)}/x_0^{(k)}) = \infty$, *then (SD) is infeasible.*

Remark 4.5 *Theorem 4.5 also shows that if there is no duality gap, i.e.* $p^*_{SD} = 0$, *then* $\lim_{k\to\infty}(z_0^{(k)}/x_0^{(k)}) = 0$.

4.4 The Primal–Dual Model

Up to now, we have only considered self–dual embeddings for self–dual programs. However, the self–dual embedding technique is applicable to general closed conic convex programs, simply by combining the original primal and dual programs into a single, self–dual program. To be more specific, consider a closed conic convex program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$, and let

$$b_{\mathrm{SD}} := \left[\begin{array}{cc} b^{\mathrm{T}}, & c^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}, \quad c_{\mathrm{SD}} := \left[\begin{array}{cc} c^{\mathrm{T}}, & b^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}, \tag{4.14}$$

and

$$\mathcal{A}_{\mathrm{SD}} := \mathcal{A} \times \mathcal{A}^{\perp}, \quad \mathcal{K}_{\mathrm{SD}} := \mathcal{K} \times \mathcal{K}^*. \tag{4.15}$$

The program (SD) is easily seen to be Π_{SD} self–dual, with

$$\Pi_{\mathrm{SD}} = \left[\begin{array}{cc} 0 & I \\ I & 0 \end{array}\right].$$

If $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ is both primal and dual feasible, then $(b, c, \mathcal{A}, \mathcal{K})$ and (SD) are equivalent, as follows from the weak duality relation for conic convex programming. Therefore, it is interesting to study the self–dual embedding (E) of the above constructed self-dual model (SD). Since our basic interest lies in the connection with the original program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$, we partition x_{SD} and u_{SD} as follows:

$$x_{\mathrm{SD}}^{\mathrm{T}} = \left[\ x^{\mathrm{T}},\quad z^{\mathrm{T}}\ \right], \quad u_{\mathrm{SD}}^{\mathrm{T}} = \left[\ u_p^{\mathrm{T}},\quad u_d^{\mathrm{T}}\ \right].$$

Since (SD) is self–dual, the results of Theorems 4.4 and 4.5 are applicable. However, if (SD) has no self–complementary solution, it is not fully equivalent with the original conic convex program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$. Below, we will therefore use the special structure of the primal–dual model, to deduce as much information as possible for $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{K})$ and its dual.

Theorem 4.6 *Consider a closed conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$ and let $x_E^{(k)}$, $k = 1, 2, \ldots$, be a weakly centered sequence for the self–dual embedding (E), where*

$$CP(b_{SD}, c_{SD}, \mathcal{A}_{SD}, \mathcal{K}_{SD})$$

is defined as in (4.14)–(4.15). Then

1. *(SD) has a self–complementary solution if and only if* $\limsup_{k\to\infty} x_0^{(k)} > 0$.
2. *(SD) is strongly infeasible if and only if* $\limsup_{k\to\infty} z_0^{(k)} > 0$.
3. *If* $\lim_{k\to\infty} z_0^{(k)} = 0$ *then*

$$\lim_{k\to\infty} \mathrm{dist}\left(\frac{x^{(k)}}{x_0^{(k)}}, b + \mathcal{A}\right) = 0, \quad \limsup_{k\to\infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} \le p^*$$

and

$$d^* \ge -\liminf_{k\to\infty} \frac{c^T x^{(k)}}{x_0^{(k)}} \ge -\limsup_{k\to\infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}}.$$

Proof. The cases of self–complementarity and strong infeasibility are known from Theorem 4.4.

If $\lim_{k\to\infty} z_0^{(k)} = 0$ then neither (P) nor (D) is strongly infeasible, and it follows from Theorem 2.6 that

$$p^* = -d^-, \quad p^- = -d^*. \tag{4.16}$$

Moreover, we know from Theorem 4.4 that

$$\lim_{k\to\infty} \mathrm{dist}\left(\frac{x^{(k)}}{x_0^{(k)}}, b + \mathcal{A}\right) = 0,$$

so that by definition of the subvalue,

$$p^- \leq \liminf_{k\to\infty} \frac{c^{\mathrm{T}} x^{(k)}}{x_0^{(k)}}, \quad d^- \leq \liminf_{k\to\infty} \frac{b^{\mathrm{T}} z^{(k)}}{x_0^{(k)}}. \tag{4.17}$$

Moreover, using (4.8) and Lemma 4.3, it follows that

$$\lim_{k\to\infty} \frac{c^{\mathrm{T}} x^{(k)} + b^{\mathrm{T}} z^{(k)} + z_0^{(k)}}{x_0^{(k)}} = 0. \tag{4.18}$$

Combining (4.16)–(4.18), the theorem follows. **Q.E.D.**

Some remarks concerning Theorem 4.6 have to be made:

Remark 4.6 *The case of self–complementarity was already known from Theorem 4.4, which also states that self–complementarity will be demonstrated by a self–complementary solution, say $x^*_{SD} = (x^*, z^*)$. It is obvious that (x^*, z^*) is then a complementary solution pair for (P) and (D).*

Remark 4.7 *Similarly, we know from Theorem 4.4 that strong infeasibility will be demonstrated by an improving direction, say $x^*_{SD} = (x^*, z^*)$. In this case, we have $c^T x^* + b^T z^* < 0$, so that either $c^T x^*$ and $b^T z^*$ are both negative, or exactly one of the quantities $c^T x^*$ and $b^T z^*$ is negative, say $c^T x^* < 0$ and $b^T z^* \geq 0$. In the former case, z^* and x^* demonstrate primal* and *dual strong infeasibility respectively. In the latter case, it follows that (D) is strongly infeasible, but we do not have complete information about (P): (P) can be either unbounded or infeasible.*

Remark that by definition, (P) is unbounded if and only if $p^* = -\infty$, and (D) is infeasible if and only if $d^* = \infty$. The following is therefore a consequence of Theorem 4.6.

Corollary 4.1 *If (P) is unbounded and (D) is weakly infeasible, then*

$$\lim_{k\to\infty} z_0^{(k)} = 0, \quad \lim_{k\to\infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} = -\infty.$$

Conversely, if

$$\lim_{k\to\infty} z_0^{(k)} = 0, \quad \lim_{k\to\infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} = -\infty, \tag{4.19}$$

then (D) is weakly infeasible and either (P) is unbounded or $p^* > -d^*$.

Remark 4.8 *If (P) is strongly feasible, then $p^* = -d^*$ (see Theorem 2.7), and Corollary 4.1 characterizes the case of primal unboundedness. In general however, we cannot conclude unboundedness from (4.19), as is illustrated later in this chapter by Example 4.7.*

There are still some cases that are not described by Theorem 4.6 and Corollary 4.1. Namely, it can happen that

- p^* is finite, and $p^* = -d^*$, but (P) is not solvable,
- p^* and d^* are finite, but $p^* + d^* > 0$,
- (P) is weakly feasible, but (D) is weakly infeasible,
- (P) and (D) are both weakly infeasible.

In all these remaining cases, we will obtain some partial information, based on the value of $\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)}$.

Theorem 4.7 *Consider a closed conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$ and let $x_E^{(k)}$, $k = 1, 2, \ldots$, be a weakly centered sequence for the self–dual embedding (E), where we use the primal–dual model $CP(b_{SD}, c_{SD}, \mathcal{A}_{SD}, \mathcal{K}_{SD})$ as defined in (4.14)–(4.15). Suppose that $\lim_{k\to\infty} x_0^{(k)} = \lim_{k\to\infty} z_0^{(k)} = 0$ and*

$$\liminf_{k\to\infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} > -\infty, \quad \liminf_{k\to\infty} \frac{b^T z^{(k)} + z_0^{(k)}}{x_0^{(k)}} > -\infty.$$

Then

1. *If $\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} = \infty$, then (SD) is weakly infeasible, and $p^* \neq -d^*$. Hence, $p^* + d^* = \infty$. Moreover, $x_{SD}^{(k)}/z_0^{(k)}$, $k = 1, 2, \ldots$, is an improving direction sequence.*
2. *If $0 < \limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} < \infty$, then $p^* \neq -d^*$ and neither (P) nor (D) is strongly feasible. Moreover, $\left\|x^{(k)}\right\| > 0$ for all sufficiently large k, and any cluster point of the sequence $x^{(k)}/\left\|x^{(k)}\right\|$ is a nonzero lower level direction, demonstrating the fact that (D) is not strongly feasible.*
3. *If $\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} = 0$, we have the following:*
   - *If $\lim_{k\to\infty} \left\|x^{(k)}\right\|/x_0^{(k)} = \infty$, then (D) is not strongly feasible. Moreover, $\left\|x^{(k)}\right\|$ is positive for all sufficiently large k, and any cluster point of the sequence $x^{(k)}/\left\|x^{(k)}\right\|$ is a nonzero lower level direction, demonstrating the fact that (D) is not strongly feasible.*
   - *Otherwise, i.e. if $\liminf_{k\to\infty} \left\|x^{(k)}\right\|/x_0^{(k)} < \infty$, then (P) is solvable and weakly feasible. Moreover, any cluster point of the sequence $x^{(k)}/x_0^{(k)}$ is an optimal solution for (P) and $\liminf_{k\to\infty} c^T z^{(k)}/x_0^{(k)} = -p^*$.*

Proof.

1. It is already known from Theorem 4.4 that if $\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} = \infty$, then (SD) is weakly infeasible, and $x_{\mathrm{SD}}^{(k)}/z_0^{(k)}$, $k = 1, 2, \ldots$, is an improving direction sequence. Using Corollary 4.1, we have $p^* > -\infty$ and $d^* > -\infty$, and it follows that $p^* + d^* = \infty$ and $p^* \neq -d^*$.

2. We use that $p^* > -\infty$ and $d^* > -\infty$ to conclude from Theorem 4.5 that if

$$\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} > 0,$$

then $p^* + d^* > 0$. Together with Theorem 2.7, this implies that neither (P) nor (D) is strongly feasible. Using Theorem 4.6, we know that the sequence $x^{(k)}/x_0^{(k)}$ cannot have any cluster point. Namely, if x is such a cluster point then $x \in (b + \mathcal{A}) \cap \mathcal{K}$ and $c^{\mathrm{T}}x < p^*$, a contradiction. Consequently, $\lim_{k\to\infty} \left\|x^{(k)}\right\| / x_0 = \infty$ and we obtain from (4.7) that any cluster point x^* of the sequence $x^{(k)}/\left\|x^{(k)}\right\|$ is a nonzero direction, i.e. $0 \neq x^* \in \mathcal{A} \cap \mathcal{K}$. Dividing (4.8) by x_0, and using Lemma 4.3, we have

$$\limsup_{k\to\infty} \frac{c^{\mathrm{T}}x^{(k)}}{x_0^{(k)}} = -\liminf_{k\to\infty} \frac{b^{\mathrm{T}}z^{(k)} + z_0^{(k)}}{x_0^{(k)}} < \infty,$$

where the inequality is an assumption of the lemma. Since $x_0^{(k)} = o(\|x^{(k)}\|)$, this inequality implies that $c^{\mathrm{T}}x^* \leq 0$, i.e. x^* is a lower–level direction.

3. Suppose that $\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} = 0$. The case that $\lim_{k\to\infty} \left\|x^{(k)}\right\| / x_0^{(k)} = \infty$ is completely analogous to the case $0 < \limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} < \infty$, which has been treated above. If $\liminf_{k\to\infty} \left\|x^{(k)}\right\| / x_0^{(k)} < \infty$, then the sequence $x^{(k)}/x_0^{(k)}$ must have a cluster point, and it follows from Theorem 4.6 that such a cluster point is an optimal solution for (P). We also know from Theorem 4.6 that there is no complementary solution pair, and hence $\lim_{k\to\infty} \left\|z^{(k)}\right\| / x_0^{(k)} = \infty$. We have already seen above that this implies that any cluster point of $z^{(k)}/\left\|z^{(k)}\right\|$ is a dual lower level direction, demonstrating weak feasibility of (P). From (4.8) and $\limsup_{k\to\infty} z_0^{(k)}/x_0^{(k)} = 0$, it follows that $\liminf_{k\to\infty} c^{\mathrm{T}}z^{(k)}/x_0^{(k)} = -p^*$.

Q.E.D.

Solving the self–dual embedding (E) is really equivalent to solving (P) if (P) is strongly feasible, as the following theorem shows.

Theorem 4.8 *Consider a closed conic convex program $CP(b, c, \mathcal{A}, \mathcal{K})$, and suppose that it is primal strongly feasible. Let $x_E^{(k)}$, $k = 1, 2, \ldots$, be a weakly centered sequence for the self–dual embedding (E), where $CP(b_{SD}, c_{SD}, \mathcal{A}_{SD}, \mathcal{K}_{SD})$ is defined as in (4.14)–(4.15). Then $CP(b, c, \mathcal{A}, \mathcal{K})$ is*

1. *Solvable if and only if*

$$\liminf_{k\to\infty} x_0^{(k)} > 0.$$

Moreover, if (P) is solvable, then any cluster point of the (bounded) sequence

$$\left\{ \left(\frac{x^{(k)}}{x_0^{(k)}}, \frac{z^{(k)}}{x_0^{(k)}}\right) \; |k = 1, 2, \ldots \right\},$$

is a complementary solution pair. (Cf. Theorem 4.6.)

2. *Unbounded and dual strongly infeasible if and only if*

$$\liminf_{k\to\infty} z_0^{(k)} > 0.$$

Moreover, if (D) is strongly infeasible then any cluster point of the (bounded) sequence $x^{(1)}, x^{(2)}, \ldots$ *is a primal improving direction, certifying the dual strong infeasibility. (Cf. Theorem 4.6 and Remark 4.7.)*

3. *Unbounded and dual weakly infeasible if and only if*

$$\lim_{k\to\infty} z_0^{(k)} = 0, \quad \lim_{k\to\infty} \frac{c^T x^{(k)} + z_0^{(k)}}{x_0^{(k)}} = -\infty. \tag{4.20}$$

Moreover, if (D) is weakly infeasible then $x^{(k)}/z_0^{(k)}$, $k = 1, 2, \ldots$ *is a primal improving direction sequence, certifying the dual weak infeasibility. (Cf. Corollary 4.1.)*

4. *Dual weakly feasible and not primal solvable, if and only if*

$$\lim_{k\to\infty} x_0^{(k)} = 0, \quad \lim_{k\to\infty} z_0^{(k)}/x_0^{(k)} = 0.$$

Moreover, if (D) is feasible and (P) is not solvable, then

- $z^{(k)}/x_0^{(k)}$, $k = 1, 2, \ldots$, *is a bounded sequence, and any cluster point of this sequence is a dual optimal solution.*
- *Any cluster point of the sequence* $x^{(k)}$, $k = 1, 2, \ldots$ *is a nonzero lower level direction, certifying that (D) is not strongly feasible.*
- $x^{(k)}/x_0^{(k)}$, $k = 1, 2, \ldots$ *is a sequence of approximate primal solutions, with*

$$\lim_{k\to\infty} \frac{c^T x^{(k)}}{x_0^{(k)}} = p^*, \quad \lim_{k\to\infty} \operatorname{dist}\left(\frac{x^{(k)}}{x_0^{(k)}}, b + \mathcal{A}\right) = 0.$$

(Cf. Theorem 4.7.)

Applying Theorem 4.6, Corollary 4.1 and Theorem 4.7, it is straightforward to prove Theorem 4.8.

Only slightly weaker results than those of Theorem 4.8 hold under the condition that $p^* = -d^*$, without requiring primal strong feasibility. Such results are then applicable to Ramana's regularized semidefinite programs, see Section 2.7. See De Klerk, Roos and Terlaky [71] for a discussion of the self–dual embedding for regularized semidefinite programs.

4.4.1 Examples in Semidefinite Programming

Several primal-dual interior point algorithms were recently extended from linear to semidefinite programming, see Chapter 3 and [80, 82, 97, 113, 149], among others. All these algorithms generate a sequence of weakly centered iterates, so that all results of Section 4.4 are applicable.

We will illustrate the theory of weakly centered sequences for (E) with some semidefinite programming problems, i.e. $\mathcal{K} = \mathcal{K}^* = \mathcal{H}_+$. We continue with our convention that given a Hermitian matrix $Y \in \mathcal{H}^{(\bar{n})}$, the lower case symbol y denotes $\mathrm{vec}_{\mathrm{H}}\ Y$, which is the coordinate vector of Y with respect to a fixed orthonormal basis of the real linear space $\mathcal{H}^{(\bar{n})}$ of Hermitian $\bar{n} \times \bar{n}$ matrices. Letting n denote the dimension of $\mathcal{H}^{(\bar{n})}$, i.e. $n = \bar{n}^2$, it follows that $y \in \Re^n$. The pair of primal and dual semidefinite programming problems is

$$\text{(P)} \quad \inf\{C \bullet X \mid X \in (B + \mathcal{A}) \cap \mathcal{H}_+\},$$

and

$$\text{(D)} \quad \inf\{B \bullet Z \mid Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_+\}.$$

Since $I \in \mathcal{H}_{++}$, we can choose $\iota = \begin{bmatrix} u_p^{\mathrm{T}}, & u_d^{\mathrm{T}}, & u_0^{\mathrm{T}}, & v_0^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ as follows:

$$U_p = U_d = I, \quad u_0 = v_0 = 1.$$

With this choice, there holds

$$\frac{\iota^{\mathrm{T}} \Pi_{\mathrm{H}} \iota}{2} = \bar{n} + 1.$$

We obtain the following formulation of the extended self–dual model (E) from (4.7)–(4.9), by specializing it to semidefinite programming.

$$\begin{array}{ll} \min & (\bar{n}+1)y_0 \\ \text{s.t.} & X - y_0 I \in (x_0 - y_0)B + \mathcal{A} \\ & Z - y_0 I \in (x_0 - y_0)C + \mathcal{A}^\perp \\ & z_0 - y_0 = -C \bullet (X - y_0 I) - B \bullet (Z - y_0 I) \\ & \mathrm{tr}\ X + \mathrm{tr}\ Z + x_0 + z_0 = (1 + y_0)(\bar{n}+1) \\ & X \succeq 0,\ Z \succeq 0,\ x_0 \geq 0.\ z_0 \geq 0,\ y_0 \in \Re. \end{array}$$

Weakly centered sequences will now be parameterized by a continuous parameter $\epsilon > 0$ such that

$$\lim_{\epsilon \downarrow 0} y_0(\epsilon) = 0.$$

We will only discuss those difficult cases where $\lim_{\epsilon\downarrow 0}(x_0(\epsilon) + z_0(\epsilon)) = 0$.

First, we consider a weakly infeasible problem.

Example 4.1 (Weakly infeasible) *Let* $\bar{n} = 2$ *and*

$$B = \begin{bmatrix} 0 & 1 \\ & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ & 1 \end{bmatrix}, \quad \mathcal{A} = \left\{ X \,\middle|\, X = \begin{bmatrix} 0 & 0 \\ & x_{22} \end{bmatrix} \right\}.$$

The primal is weakly infeasible,

$$p^* = \inf \left\{ x_{22} \,\middle|\, X = \begin{bmatrix} 0 & 1 \\ & x_{22} \end{bmatrix} \succeq 0 \right\} = \infty$$

and the dual is strongly feasible and unbounded,

$$d^* = \inf \left\{ 2z_{12} \,\middle|\, Z = \begin{bmatrix} z_{11} & z_{12} \\ & 1 \end{bmatrix} \succeq 0 \right\} = -\infty.$$

We construct a weakly centered sequence for $0 < \epsilon \leq 1/3$ *as follows:*

$$X(\epsilon) = \epsilon^3 I + \begin{bmatrix} 0 & \epsilon^2 \\ & \epsilon \end{bmatrix}, \quad Z(\epsilon) = \epsilon^3 I + \begin{bmatrix} 3 - (2\epsilon + 2\epsilon^2 + 3\epsilon^3) & -\epsilon \\ & \epsilon^2 \end{bmatrix},$$

$$y_0(\epsilon) = \epsilon^3, \quad x_0(\epsilon) = \epsilon^2 + \epsilon^3, \quad z_0(\epsilon) = \epsilon + \epsilon^3.$$

We see that

$$\lim_{\epsilon\downarrow 0} z_0(\epsilon) = 0, \quad \lim_{\epsilon\downarrow 0} \frac{B \bullet Z(\epsilon) + z_0(\epsilon)}{x_0(\epsilon)} = -\infty,$$

which indeed implies that the dual is unbounded and the primal is weakly infeasible, see Theorem 4.8. Finally, notice that

$$\lim_{\epsilon\downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)} = \infty.$$

In order to be able to solve the self–dual embedding (E), we specialized the predictor-corrector algorithm for semidefinite programming (see Chapter 3) to the special structure of (E). The plots below show the numerical results for the examples in this section. The solid lines represent the primal objective values

$$\frac{C \bullet X^{(k)}}{x_0^{(k)} - y_0^{(k)}}, \quad \frac{C \bullet X^{(k)} + z_0^{(k)}}{x_0^{(k)} - y_0^{(k)}},$$

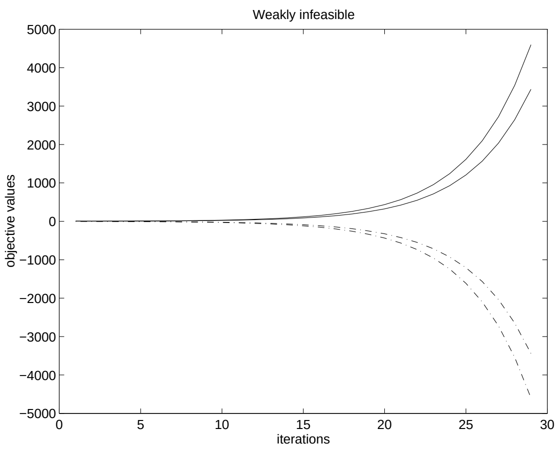

Figure 4.1: Example 4.1

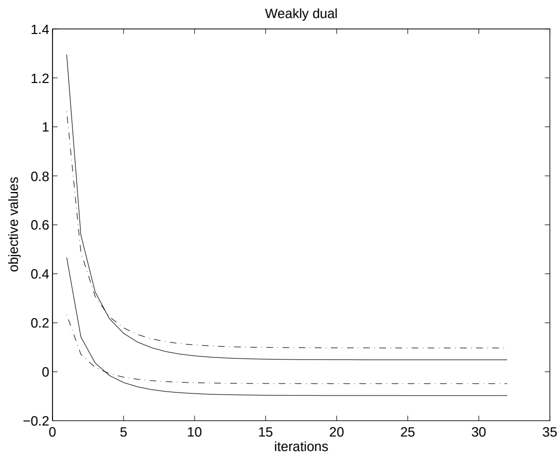

Figure 4.2: Example 4.2

whereas the dual objective values

$$\frac{B \bullet Z^{(k)}}{x_0^{(k)} - y_0^{(k)}}, \quad \frac{B \bullet Z^{(k)} + z_0^{(k)}}{x_0^{(k)} - y_0^{(k)}}$$

are represented by dashed lines. Recall from Theorem 4.6 that $(C \bullet X^{(k)} + z_0^{(k)})/(x_0^{(k)} - y_0^{(k)})$ and $(B \bullet Z^{(k)} + z_0^{(k)})/(x_0^{(k)} - y_0^{(k)})$ are lower bounds for p^* and d^* respectively.

The next example, which is from Vandenberghe and Boyd [158], gives a feasible problem, where strong duality fails to hold.

Example 4.2 (Weakly dual) *Let* $\bar{n} = 3$ *and*

$$B = \begin{bmatrix} 0 & 1/3 & 0 \\ & 0 & 0 \\ & & 1/3 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & -1/3 & 0 \\ & 0 & 0 \\ & & 2/3 \end{bmatrix},$$

$$\mathcal{A} = \left\{ X \;\middle|\; X = \begin{bmatrix} x_{11} & -x_{33}/2 & x_{13} \\ & 0 & x_{23} \\ & & x_{33} \end{bmatrix} \right\},$$

so that the primal is solvable and weakly feasible,

$$p^* = \inf \left\{ \frac{2}{3}x_{33} - \frac{2}{3}x_{12} \;\middle|\; X = \begin{bmatrix} x_{11} & (1-x_{33})/2 & x_{13} \\ & 0 & x_{23} \\ & & x_{33} \end{bmatrix} \succeq 0 \right\} = \frac{2}{3}$$

and the dual is also solvable and weakly feasible,

$$d^* = \inf \left\{ \frac{1}{3} z_{33} + \frac{2}{3} z_{12} \;\middle|\; Z = \begin{bmatrix} 0 & z_{33}-1 & 0 \\ & z_{22} & 0 \\ & & z_{33} \end{bmatrix} \succeq 0 \right\} = \frac{1}{3}.$$

Remark that $p^* + d^* = 1 > 0$ *so that strong duality fails. A weakly centered sequence for* $0 < \epsilon \leq 1/2$ *is given by*

$$X(\epsilon) = \epsilon^2 I + \begin{bmatrix} 1 & \epsilon/2 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix}, \quad Z(\epsilon) = \epsilon^2 I + \begin{bmatrix} 0 & -\epsilon & 0 \\ & 3-(2\epsilon+4\epsilon^2) & 0 \\ & & 0 \end{bmatrix},$$

$$y_0(\epsilon) = \epsilon^2, \quad x_0(\epsilon) = \epsilon + \epsilon^2, \quad z_0(\epsilon) = \epsilon + \epsilon^2$$

so that

$$\lim_{\epsilon \downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)} = 1,$$

which indeed implies that $p^* \neq -d^*$*, see Theorem 4.7.*

The third case is a problem where strong duality holds, but there exists no complementary solution pair (see Vandenberghe and Boyd [158]).

Example 4.3 (Strongly dual) *Let* $\bar{n} = 2$ *and*

$$B = \begin{bmatrix} 0 & 0 \\ & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 \\ & 0 \end{bmatrix}, \quad \mathcal{A} = \left\{ X \;\middle|\; X = \begin{bmatrix} 0 & x_{12} \\ & 0 \end{bmatrix} \right\},$$

so that the primal is solvable and weakly feasible,

$$p^* = \inf \left\{ 2x_{12} \;\middle|\; X = \begin{bmatrix} 0 & x_{12} \\ & 1 \end{bmatrix} \succeq 0 \right\} = 0$$

and the dual is strongly feasible but not solvable,

$$d^* = \inf \left\{ z_{22} \;\middle|\; Z = \begin{bmatrix} z_{11} & 1 \\ & z_{22} \end{bmatrix} \succeq 0 \right\} = 0.$$

Notice that $p^* + d^* = 0$*, but the dual has no optimal solution. A weakly centered sequence for* $0 < \epsilon \leq 1/3$ *is*

$$X(\epsilon) = \epsilon^3 I + \begin{bmatrix} 0 & -\epsilon^2 \\ & \epsilon \end{bmatrix}, \quad Z(\epsilon) = \epsilon^3 I + \begin{bmatrix} 3-(2\epsilon+3\epsilon^2+3\epsilon^3) & \epsilon \\ & \epsilon^2 \end{bmatrix},$$

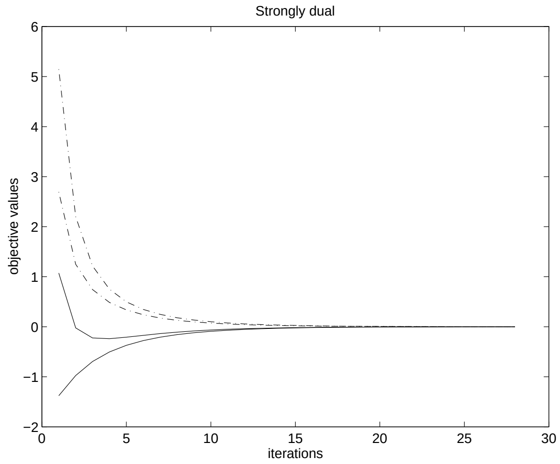

Figure 4.3: Example 4.3

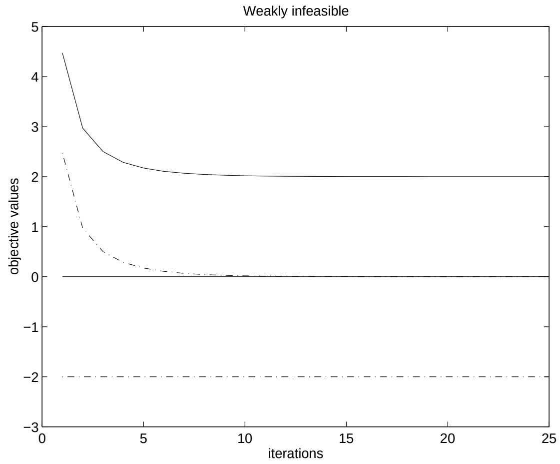

Figure 4.4: Example 4.4

$$y_0(\epsilon) = \epsilon^3, \quad x_0(\epsilon) = \epsilon + \epsilon^3, \quad z_0(\epsilon) = 2\epsilon^2 + \epsilon^3.$$

Hence,

$$\lim_{\epsilon \downarrow 0} x_0(\epsilon) = 0, \quad \lim_{\epsilon \downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)} = 0,$$

which indeed holds if and only if the primal is weakly feasible and the dual is not solvable, see Theorem 4.8.

So far, we have seen a weakly infeasible problem with $\lim_{\epsilon \downarrow 0} z_0(\epsilon)/x_0(\epsilon) = \infty$, a feasible problem with only weak duality and $\lim_{\epsilon \downarrow 0} z_0(\epsilon)/x_0(\epsilon) \in (0, \infty)$ and a strongly dual problem with $\lim_{\epsilon \downarrow 0} z_0(\epsilon)/x_0(\epsilon) = 0$. The reader may wonder whether the asymptotic behavior of the indicator $z_0(\epsilon)/x_0(\epsilon)$ completely characterizes the three cases that we consider. Unfortunately, this is not the case, as the next example shows.

Example 4.4 (Weakly infeasible) *Let $\bar{n} = 2$, and consider*

$$B = \begin{bmatrix} 0 & 1 \\ & 0 \end{bmatrix}, \quad C = 0, \quad \mathcal{A} = \left\{ X \;\middle|\; X = \begin{bmatrix} 0 & 0 \\ & x_{22} \end{bmatrix} \succeq 0 \right\}.$$

The primal is weakly infeasible,

$$p^* = \inf \left\{ 0 \;\middle|\; X = \begin{bmatrix} 0 & 1 \\ & x_{22} \end{bmatrix} \succeq 0 \right\} = \infty,$$

and the dual is solvable and weakly feasible,

$$d^* = \inf \left\{ 2z_{12} \;\middle|\; Z = \begin{bmatrix} z_{11} & z_{12} \\ & 0 \end{bmatrix} \succeq 0 \right\} = 0.$$

We construct a weakly centered sequence for $0 < \epsilon \leq 1/3$ *as*

$$X(\epsilon) = \epsilon^2 I + \begin{bmatrix} 0 & \epsilon \\ & 1 \end{bmatrix}, \quad Z(\epsilon) = \epsilon^2 I + \begin{bmatrix} 2 - 3(\epsilon + \epsilon^2) & -\epsilon \\ & 0 \end{bmatrix},$$

$$y_0(\epsilon) = \epsilon^2, \quad x_0(\epsilon) = \epsilon + \epsilon^2, \quad z_0(\epsilon) = 2\epsilon + \epsilon^2.$$

We see that

$$\lim_{\epsilon \downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)} = 2.$$

The reader may still wonder whether we can distinguish weak infeasibility from strong duality. It appears somewhat difficult indeed, to construct an example of an infeasible problem where $z_0(\epsilon)/x_0(\epsilon) \to 0$, but it does exist.

Example 4.5 (Weakly infeasible) *Let* $\bar{n} = 3$ *and*

$$B = 0, \quad C = \begin{bmatrix} 0 & -1 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix}, \quad \mathcal{A} = \left\{ X \;\middle|\; X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ & 0 & -x_{11}/2 \\ & & x_{33} \end{bmatrix} \right\}.$$

The primal is solvable and weakly feasible,

$$p^* = \inf \left\{ -2x_{12} \;\middle|\; X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ & 0 & -x_{11}/2 \\ & & x_{33} \end{bmatrix} \succeq 0 \right\} = 0$$

and the dual is weakly infeasible,

$$d^* = \inf \left\{ 0 \;\middle|\; Z = \begin{bmatrix} z_{11} & -1 & 0 \\ & z_{22} & z_{11} \\ & & 0 \end{bmatrix} \succeq 0 \right\} = \infty.$$

We construct a weakly centered sequence for $0 < \epsilon \leq 1/3$ *by*

$$X(\epsilon) = \epsilon^4 I + \begin{bmatrix} \epsilon^2 & \epsilon^3/2 & 0 \\ & 0 & -\epsilon^2/2 \\ & & 1 \end{bmatrix},$$

$$Z(\epsilon) = \epsilon^4 I + \begin{bmatrix} \epsilon^2 & -\epsilon & 0 \\ & 3 - (\epsilon + 2\epsilon^2 + \epsilon^3 + 4\epsilon^4) & \epsilon^2 \\ & & 0 \end{bmatrix},$$

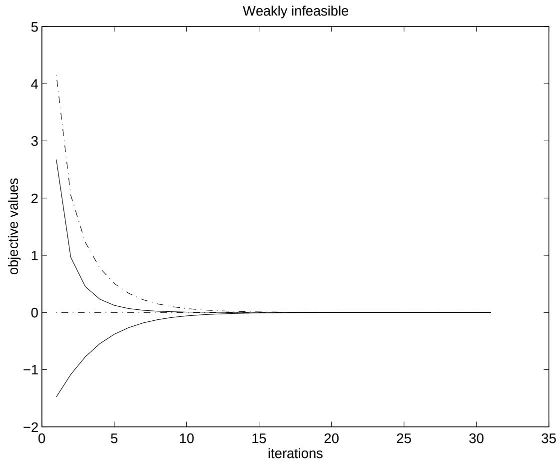

Figure 4.5: Example 4.5

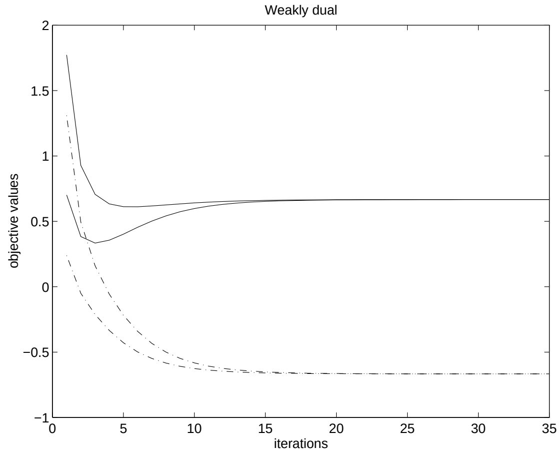

Figure 4.6: Example 4.6

$$y_0(\epsilon) = \epsilon^4, \quad x_0(\epsilon) = \epsilon + \epsilon^4, \quad z_0(\epsilon) = \epsilon^3 + \epsilon^4,$$

so that

$$\lim_{\epsilon \downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)} = 0.$$

After Example 4.5 there is little hope that feasibility with lack of strong duality would imply $\limsup_{\epsilon \downarrow 0} z_0(\epsilon)/x_0(\epsilon) > 0$. Indeed, we can construct a feasible problem with only weak duality but $\lim_{\epsilon \downarrow 0} z_0(\epsilon)/x_0(\epsilon) = 0$.

Example 4.6 (Weakly dual) *Let* $\bar{n} = 4$ *and*

$$B = \begin{bmatrix} 0 & 1/3 & 0 & 0 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 1/3 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & -1/3 & 0 & 0 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 2/3 \end{bmatrix},$$

$$\mathcal{A} = \left\{ X \;\middle|\; X = \begin{bmatrix} x_{11} & -x_{44}/2 & x_{13} & x_{14} \\ & 0 & -x_{11}/2 & x_{24} \\ & & x_{33} & x_{34} \\ & & & x_{44} \end{bmatrix} \right\}.$$

The primal is solvable and weakly feasible,

$$p^* = \inf \left\{ \frac{2}{3}x_{44} - \frac{2}{3}x_{12} \;\middle|\; X = \begin{bmatrix} x_{11} & (1-x_{44})/2 & x_{13} & x_{14} \\ & 0 & -x_{11}/2 & x_{24} \\ & & x_{33} & x_{34} \\ & & & x_{44} \end{bmatrix} \succeq 0 \right\} = \frac{2}{3},$$

and the dual is also solvable and weakly feasible,

$$d^* = \inf\left\{ \frac{2}{3}z_{12} + \frac{1}{3}z_{44} \;\middle|\; Z = \begin{bmatrix} z_{11} & z_{44}-1 & 0 & 0 \\ & z_{22} & z_{11} & 0 \\ & & 0 & 0 \\ & & & z_{44} \end{bmatrix} \succeq 0 \right\} = \frac{1}{3}.$$

Hence, $p^* + d^* = 1 > 0$, so that strong duality fails. Remark that the first three rows and columns of this program are the same as in Example 4.5, with the additional constraints

$$x_{12} = \frac{1 - x_{44}}{2} \leq \frac{1}{2}, \quad z_{12} = z_{44} - 1 \geq -1.$$

As a consequence, we can construct a weakly centered sequence that is very similar to the one in Example 4.5, viz.

$$X(\epsilon) = \epsilon^4 I + \begin{bmatrix} \epsilon^2 & \epsilon^3/2 & 0 & 0 \\ & 0 & -\epsilon^2/2 & 0 \\ & & 1 & 0 \\ & & & \epsilon - \epsilon^3 \end{bmatrix},$$

$$Z(\epsilon) = \epsilon^4 I + \begin{bmatrix} \epsilon^2 & -\epsilon & 0 & 0 \\ & 4 - (2\epsilon + 2\epsilon^2 + 5\epsilon^4) & \epsilon^2 & 0 \\ & & 0 & 0 \\ & & & 0 \end{bmatrix},$$

$$y_0(\epsilon) = \epsilon^4, \quad x_0(\epsilon) = \epsilon + \epsilon^4, \quad z_0(\epsilon) = \epsilon^3 + \epsilon^4,$$

and

$$\lim_{\epsilon \downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)} = 0.$$

Our results on the indicator

$$\limsup_{\epsilon \downarrow 0} \frac{z_0(\epsilon)}{x_0(\epsilon)}$$

are summarized in Table 4.1. The possible combinations in the table are illustrated by Examples 4.1–4.6. The impossibility of the remaining combinations follows from Theorem 4.5; see also Remark 4.4 and Remark 4.5.

As promised in Remark 4.8, we will now give an example where $(c^{\mathrm{T}}x(\epsilon) + s_0(\epsilon))/x_0(\epsilon) \to -\infty$, but (P) is not unbounded. We consider this as an extremely nasty case, since it implies $b^{\mathrm{T}}s(\epsilon)/x_0(\epsilon) \to \infty$, even though $d^- = -p^* < \infty$.

	$\limsup_{\epsilon\downarrow 0} z_0(\epsilon)/x_0(\epsilon)$		
	0	$(0,\infty)$	∞
$p^*_{\mathrm{SD}} = \infty$	Example 4.5	Example 4.4	Example 4.1
$p^*_{\mathrm{SD}} \in (0,\infty)$	Example 4.6	Example 4.2	impossible
$p^*_{\mathrm{SD}} = 0$	Example 4.3	impossible	impossible

Table 4.1: Distinguishing the difficult cases where $x_0^{(k)} + s_0^{(k)} \to 0$

Example 4.7 *We consider a semidefinite program, which has some similarity to Example 4.5, viz.*

$$p^* = \inf\left\{ 2x_{45} \;\middle|\; \begin{bmatrix} x_{11} & -x_{55}/2 & x_{13} & & \\ & x_{22} & x_{11} & & \\ & & 0 & & \\ & & & 1 & x_{45} \\ & & & & x_{55} \end{bmatrix} \succeq 0 \right\} = 0,$$

for which the dual is weakly infeasible,

$$d^* = \inf\left\{ z_{44} \;\middle|\; \begin{bmatrix} z_{11} & z_{55} & 0 & & \\ & 0 & -z_{11}/2 & & \\ & & z_{33} & & \\ & & & z_{44} & 1 \\ & & & & z_{55} \end{bmatrix} \succeq 0 \right\} = \infty.$$

Notice however, that $d^- = 0$. We construct a weakly centered sequence for small ϵ by

$$X(\epsilon) = \epsilon^{12} I + \begin{bmatrix} \epsilon^6 & -\epsilon^3/2 & 0 & & \\ & 3 - O(\epsilon^3) & \epsilon^6 & & \\ & & \epsilon^{12} & & \\ & & & \epsilon^7 & -\epsilon^5 \\ & & & & \epsilon^3 \end{bmatrix},$$

$$Z(\epsilon) = \epsilon^{12} I + \begin{bmatrix} \epsilon^6 & \epsilon^9 & 0 & & \\ & \epsilon^{12} & -\epsilon^6/2 & & \\ & & 3 - O(\epsilon^3) & & \\ & & & \epsilon^5 & \epsilon^7 \\ & & & & \epsilon^9 \end{bmatrix},$$

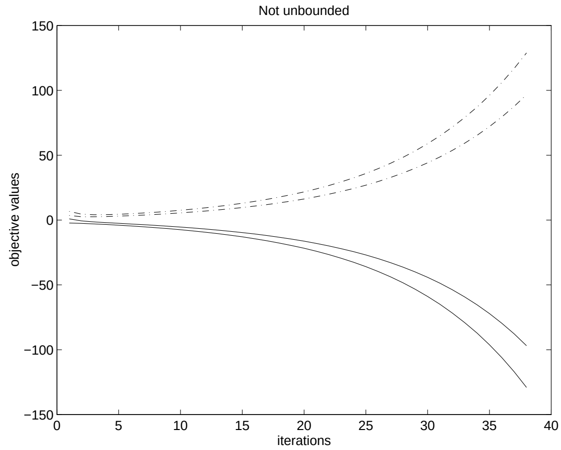

Figure 4.7: Example 4.7

$$y_0(\epsilon) = \epsilon^{12}, \quad x_0(\epsilon) = \epsilon^7 + \epsilon^{12}, \quad z_0(\epsilon) = \epsilon^5 + \epsilon^{12}.$$

We have

$$\lim_{\epsilon \downarrow 0} \frac{c^T x(\epsilon) + z_0(\epsilon)}{x_0(\epsilon)} = \frac{-2\epsilon^5 + \epsilon^5 + \epsilon^{12}}{\epsilon^7 + \epsilon^{12}} = -\infty,$$

but (P) is not unbounded. See Corollary 4.1 and Remark 4.8.

4.5 Existence of Weakly Centered Sequences

In this section, we will prove the existence of a weakly centered sequence for (E), if we use the primal–dual model $\mathrm{CP}(b_{\mathrm{SD}}, c_{\mathrm{SD}}, \mathcal{A}_{\mathrm{SD}}, \mathcal{K}_{\mathrm{SD}})$ of Section 4.4. In fact, we will give a constructive proof using the theory of logarithmically homogeneous barriers, which has been developed by Nesterov and Nemirovsky [110].

Definition 4.3 *Let $\mathcal{K}$ be a closed, solid and pointed convex cone. Then $F : \text{ int } \mathcal{K} \to \Re$ is a ν-*logarithmically homogeneous barrier *for $\mathcal{K}$ if F is a twice continuously differentiable convex function on $\text{ int } \mathcal{K}$ such that $F(x^{(i)}) \to \infty$ for any sequence $x^{(i)}$ in $\text{ int } \mathcal{K}$, $i = 1, 2, \ldots$, that converges to the boundary of $\mathcal{K}$, and*

$$F(tx) = F(x) - \nu \log t \text{ for all } x \in \text{ int } \mathcal{K},\ t > 0, \tag{4.21}$$

where $\nu \geq 1$ is a fixed parameter.

It is known that any closed, pointed and solid convex cone $\mathcal{K}$ is endowed with a logarithmically homogeneous barrier, see Theorem 2.5.1 in [110]. Moreover, Proposition 2.3.5 in [110] states that logarithmically homogeneous barriers are strictly convex functions. Important special cases are the n-logarithmically homogeneous barrier $F(x) = -\sum_{i=1}^{n} \log x_i$

for the cone $\mathcal{K} = \Re^n_+$, and the $\bar{n}$-logarithmically homogeneous barrier $F(X) = -\log\det X$ for the cone $\mathcal{K} = \mathcal{H}^{(\bar{n})}_+$.

Let $F(x)$ be a ν-logarithmically homogeneous barrier for $\mathcal{K}$, and define its conjugate (or Legendre–Young–Fenchel transform) by

$$F^*(z) := \sup_{x \in \text{ int } \mathcal{K}} \{(-z)^{\mathrm{T}}x - F(x)\}. \tag{4.22}$$

Notice that if $z \notin \mathcal{K}^*$, then $(-z)^{\mathrm{T}}x > 0$ for some $x \in \text{ int}\,\mathcal{K}$, which together with (4.21) implies $F^*(z) = \infty$. In fact, Nesterov and Nemirovsky [110] showed that $F^*(z)$ is a ν-logarithmically homogeneous barrier for $\mathcal{K}^*$ (see Theorem 2.4.4 therein). Hence, $F^*(z) < \infty$ if and only if $z \in \text{ int}\,\mathcal{K}^*$. Moreover, the biconjugate F^{**} of F is again F, i.e. $F(x) = F^{**}(x)$, see e.g. Rockafellar [132].

Let $x \in \text{ int}\,\mathcal{K}$. Using definition (4.22) and the first order optimality conditions for concave maximization, it follows that

$$F^*(z) = (-z)^{\mathrm{T}}x - F(x) \text{ if } z = -\nabla F(x),$$

from which we obtain

$$-\nabla F(x) \in \text{ int}\,\mathcal{K}^*, \tag{4.23}$$

and

$$\nabla F^*(z) = -x \text{ if } z = -\nabla F(x). \tag{4.24}$$

Using (4.21), it is straightforward to show that

$$\nabla F(x/t) = t\nabla F(x) \text{ for all } t > 0, \tag{4.25}$$

and

$$\nabla F(x)^{\mathrm{T}}x = -\nu. \tag{4.26}$$

The above properties of F are also listed by Nesterov and Todd [113] and Nesterov, Todd and Ye [114]. Based on the barrier F, one can define a *barrier path* with parameter $\mu > 0$, see Theorem 4.9 below. In the case of semidefinite programming, the central path (see Chapter 3) is the barrier path for $F(x) = -\log(\det X)$.

Theorem 4.9 (barrier path) *Let $\mathcal{K}$ be a convex cone that is closed, pointed and solid. Suppose that $CP(b, c, \mathcal{A}, \mathcal{K})$ is a conic convex program that is primal and dual strongly feasible. Let $F :\ \text{int}\,\mathcal{K} \to \Re$ be a ν-logarithmically homogeneous barrier for $\mathcal{K}$, and define*

$$\phi_\mu(x) := (-c)^T x - \mu F(x),$$

with $\mu > 0$. Then there exists a unique vector $x(\mu) \in (b + \mathcal{A}) \cap \text{ int}\,\mathcal{K}$ such that

$$\phi_\mu(x(\mu)) = \max\{\phi_\mu(x) \mid x \in (b + \mathcal{A}) \cap \text{ int}\,\mathcal{K}\}.$$

Moreover, letting $z(\mu) := -\mu\nabla F(x(\mu))$, *there holds*

$$x(\mu)^T z(\mu) = \nu\mu, \quad x(\mu) = -\mu\nabla F^*(z(\mu)),$$

and

$$z(\mu) = \arg\max\{(-b)^T z - \mu F^*(z) \mid z \in (c + \mathcal{A}^\perp) \cap \text{ int } \mathcal{K}^*\}.$$

Proof. First, notice that $\phi_\mu(\cdot)$ is a concave function. Moreover, for all $x \in (b+\mathcal{A}) \cap \text{ int } \mathcal{K}$ and $z \in (c + \mathcal{A}^\perp) \cap \text{ int } \mathcal{K}^*$, there holds

$$\phi_\mu(x) = b^{\mathrm{T}} z - z^{\mathrm{T}} x - \mu F(x) \leq b^{\mathrm{T}} z + \mu F^*(z/\mu).$$

Hence, $\phi_\mu(x)$ is bounded from above on $(b + \mathcal{A}) \cap \text{ int } \mathcal{K}$, and since $\phi_\mu(\cdot)$ is a strictly concave function, it follows that $\phi_\mu(\cdot)$ achieves a maximum, and the maximizer $x(\mu)$ is unique. From the first–order optimality conditions, we know that $x(\mu)$ satisfies

$$\nabla\phi_\mu(x(\mu)) = -c - \mu\nabla F(x(\mu)) \in \mathcal{A}^\perp.$$

Letting $z(\mu) := -\mu\nabla F(x(\mu))$, it follows from (4.23) that $z(\mu) \in (c + \mathcal{A}^\perp) \cap \text{ int } \mathcal{K}^*$, and using (4.24)–(4.25), we have

$$x(\mu) = -\nabla F^*(z(\mu)/\mu) = -\mu\nabla F^*(z(\mu)),$$

so that

$$-b - \mu\nabla F^*(z(\mu)) \in \mathcal{A}.$$

The above relation shows that $z(\mu)$ satisfies the optimality conditions for

$$\max\{(-b)^{\mathrm{T}} z - \mu F^*(z) \mid z \in (c + \mathcal{A}^\perp) \cap \text{ int } \mathcal{K}^*\}.$$

Finally, it follows from (4.26) that

$$z(\mu)^{\mathrm{T}} x(\mu) = -\mu\nabla F(x(\mu))^{\mathrm{T}} x(\mu) = \nu\mu.$$

Q.E.D.

We now propose the following barrier for $\mathcal{K} \times \mathcal{K}^*$,

$$F_{\mathrm{SD}}(x, z) := F(x) + F^*(z).$$

Using (4.21), we see that $F_{\mathrm{SD}}(x, z)$ is a 2ν–logarithmically homogeneous barrier, and from definition (4.22) and the fact that $F(x) = F^{**}(x)$, we obtain that

$$F^*_{\mathrm{SD}}(z, x) = F_{\mathrm{SD}}(x, z).$$

This leads to the following definition:

Definition 4.4 *Let $\mathcal{K}$ be a solid convex cone such that $\mathcal{K}^* = \Pi\mathcal{K}$ for some symmetric permutation matrix Π. A ν–logarithmically homogeneous barrier $F:$ int $\mathcal{K} \to \Re$ for $\mathcal{K}$ is* Π self–conjugate *if and only if*

$$F^*(\Pi x) = F(x) \text{ for all } x \in \text{ int } \mathcal{K}.$$

For the extended self–dual model (E), we define

$$F_{\mathrm{E}}(x_{\mathrm{SD}}, x_0, z_0) := F_{\mathrm{SD}}(x_{\mathrm{SD}}) - \log x_0 - \log z_0,$$

where F_{SD} is a (2ν)–logarithmically homogeneous Π_{SD} self–conjugate barrier for $\mathcal{K}_{\mathrm{SD}}$. It is easy to verify that F_{E} is then a $(2\nu{+}2)$–logarithmically homogeneous Π_{H} self–conjugate barrier for $\mathcal{K}_{\mathrm{H}}$. Since (E) is strongly feasible (see Theorem 4.3), we can apply Theorem 4.9 to arrive at the following result.

Theorem 4.10 *Define*

$$\phi_\mu(x_E) := (-c_E)^T x - \mu F_E(x_E),$$

and let

$$x_E(\mu) := \arg\max\{\phi_\mu(x_E) \mid x_E \in (b_E + \mathcal{A}_E) \cap \text{ int } \mathcal{K}_E\}.$$

Then

$$x_0(\mu) z_0(\mu) = \mu = \frac{1}{\nu+1} c_E^T x_E(\mu)$$

and $\lim_{\mu\downarrow 0} c_E^T x_E(\mu) = 0$.

Proof. From Theorem 4.9, we know that $x_{\mathrm{E}}(\mu)$ is well defined, and

$$x_{\mathrm{E}}(\mu)^{\mathrm{T}} z_{\mathrm{E}}(\mu) = 2(\nu+1)\mu, \quad x_{\mathrm{E}}(\mu) = -\mu \nabla F_{\mathrm{E}}^*(z_{\mathrm{E}}(\mu)),$$

where $z_{\mathrm{E}}(\mu) := -\mu\nabla F_{\mathrm{E}}(x_{\mathrm{E}}(\mu))$. In addition, Theorem 4.9 tells us that $z_{\mathrm{E}}(\mu)$ is the maximizer of the function $-b_{\mathrm{E}}^{\mathrm{T}} z_{\mathrm{E}} - \mu F_{\mathrm{E}}^*(z_{\mathrm{E}}) = \phi_\mu(\Pi_{\mathrm{H}} z_{\mathrm{E}})$ over all dual interior solutions z_{E}, and therefore

$$x_{\mathrm{E}}(\mu) = \Pi_{\mathrm{H}} z_{\mathrm{E}}(\mu) = -\mu \Pi_{\mathrm{H}} \nabla F_{\mathrm{E}}(x_{\mathrm{E}}(\mu)). \tag{4.27}$$

Noticing that

$$\nabla F_{\mathrm{E}}(x_{\mathrm{E}}) = \left[\ \nabla F_{\mathrm{SD}}(x_{\mathrm{SD}})^{\mathrm{T}}, \quad -1/x_0, \quad -1/z_0\ \right]^{\mathrm{T}},$$

we obtain from (4.27) that

$$x_0(\mu) = \mu / z_0(\mu).$$

Hence,

$$x_0(\mu) z_0(\mu) = \mu = \frac{1}{2(\nu+1)} x_{\mathrm{E}}(\mu)^{\mathrm{T}} z_{\mathrm{E}}(\mu) = \frac{1}{\nu+1} c_{\mathrm{E}}^{\mathrm{T}} x_{\mathrm{E}}(\mu),$$

where we used Theorem 4.1. **Q.E.D.**

Theorem 4.10 shows that $\{x_{\mathrm{E}}(\mu) \mid \mu > 0\}$ is a weakly centered sequence for (E).

We have used the theory of logarithmically homogeneous barriers, to establish the existence of weakly centered sequences for conic convex programming. In the special case of semidefinite programming however, there is no need for this barrier argument. Namely, consider the primal–dual path–following methods that were discussed in Chapter 3. We may initialize these methods with a primal–dual pair $(x_{\mathrm{E}}^{(0)}, z_{\mathrm{E}}^{(0)})$ that satisfies $z_{\mathrm{E}}^{(0)} = \Pi_{\mathrm{H}} x_{\mathrm{E}}^{(0)}$, i.e. the initial primal and dual solutions are essentially the same. In particular, we may start with the identity solution, as explained in Section 4.4.1. It is then easily checked that all subsequent iterates $(x_{\mathrm{E}}^{(k)}, z_{\mathrm{E}}^{(k)})$ also satisfy such a property. Namely, since primal–dual path–following algorithms generate iterates in an $\mathcal{N}_{\infty}^{-}(\beta)$–neighborhood of the central path (see (3.18)), we have

$$x_0^{(k)} z_0^{(k)} \geq (1-\beta) \frac{(x_{\mathrm{E}}^{(k)})^{\mathrm{T}} z_{\mathrm{E}}^{(k)}}{2\bar{n}+2}.$$

This shows that iterative sequences in the $\mathcal{N}_{\infty}^{-}(\beta)$–neighborhood are weakly centered with $\omega = (1-\beta)/(\bar{n}+1)$.

4.6 Discussion

Extensions of the self–dual embedding technique to semidefinite programming were proposed independently by Potra and Sheng [123], De Klerk, Roos and Terlaky [70, 71], Luo, Sturm and Zhang [85] and Nesterov, Todd and Ye [114]. The proposed extensions are based on the self–dual formulations of Ye, Todd and Mizuno [163] and Jansen, Roos and Terlaky [59, 58, 134] for linear programming. Potra and Sheng [123] only consider the problem of finding a complementary solution for a semidefinite program. De Klerk, Roos and Terlaky [70] and Nesterov, Todd and Ye [114] also address cases where complementary solutions do not exist, but do not investigate the more nasty cases that involve weak infeasibility. However, Luo, Sturm and Zhang [85] systematically investigate the most delicate issues in self–dual embeddings. In fact the treatment in this chapter is based on [85]. More recently, De Klerk, Roos and Terlaky also addressed such issues in [71]. The treatment in [123, 70, 71] restricts to the semidefinite programming case, whereas [114, 85] consider the more general setting of conic convex programming. Nesterov, Todd and Ye [114] discuss the application of logarithmically homogeneous barrier techniques to self–dual embeddings.

Chapter 5

Properties of the Central Path

In this chapter, we take a careful look at the behavior of the central path, when it approaches the optimal solution set of a semidefinite program. We will demonstrate that the primal-dual central path converges to the analytic center of the optimal solution set. Moreover, the distance to this analytic center from any point on the central path is shown to converge at the same R-rate as the duality gap. This result can be interpreted as an error–bound for solutions on the central path, with respect to the optimal solution set. Underlying the analysis is an assumption that a strictly complementary solution pair for the semidefinite program exists.

5.1 Introduction

Our study concerns again the semidefinite program (P), given as

$$\text{(P)} \quad \inf\{C \bullet X \mid X \in (B + \mathcal{A}) \cap \mathcal{H}_+\},$$

with feasible solution set $\mathcal{F}_P = (B + \mathcal{A}) \cap \mathcal{H}_+$. For the corresponding dual program (D),

$$(D) \quad \inf\{B \bullet Z \mid Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_+\},$$

the feasible solution set is given by $\mathcal{F}_D := (C + \mathcal{A}^\perp) \cap \mathcal{H}_+$.

Throughout this chapter, we assume that (P) and (D) are both strongly feasible, and that a strictly complementary solution pair exists.

Assumption 5.1 *There exist positive definite solutions $X \in \mathcal{F}_P$ and $Z \in \mathcal{F}_D$ for* (P) *and* (D) *respectively.*

Assumption 5.2 *There exists a pair of strictly complementary primal-dual optimal solutions for* (P) *and* (D). *Specifically, there exists* $(X^*, Z^*) \in \mathcal{F}_P \times \mathcal{F}_D$ *such that*

$$\begin{cases} X^* Z^* = 0, \\ X^* + Z^* \succ 0. \end{cases} \tag{5.1}$$

Remark that (5.1) is indeed equivalent with the definition of strict complementarity in Section 2.5, i.e.

$$X^* \in \text{ rel (face } (\mathcal{H}_+, Z^*)), \quad Z^* \in \text{ rel (face } (\mathcal{H}_+, X^*)).$$

Since $X^*Z^* = Z^*X^* = 0$, we can diagonalize X^* and Z^* simultaneously. Therefore, by applying a unitary transformation to the problem data if necessary, we can assume without loss of generality that X^*, Z^* are both diagonal and of the form

$$X^* = \begin{bmatrix} \Lambda_B & 0 \\ 0 & 0 \end{bmatrix}, \quad Z^* = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_N \end{bmatrix}, \tag{5.2}$$

where $\Lambda_B :=$ diag $(\lambda_1, ..., \lambda_K)$, $\Lambda_N :=$ diag $(\lambda_{K+1}, ..., \lambda_{\bar{n}})$ for some integer $0 \le K \le \bar{n}$ and some positive scalars $\lambda_i > 0$, $i = 1, 2, ..., \bar{n}$. Here the subscripts B and N signify the "basic" and "nonbasic" subspaces (following the terminology of linear programming). Throughout this chapter, the decomposition of any $\bar{n} \times \bar{n}$ Hermitian matrix X is always made with respect to the above partition B and N. In fact, we shall adhere to the following notation throughout:

$$X = \begin{bmatrix} X_B & X_U \\ X_U^{\mathrm{H}} & X_N \end{bmatrix},$$

so X_U will always denote the off-diagonal block of X with size $K \times (\bar{n} - K)$, etc.

Notice that $X \in \mathcal{F}_P$ is an optimal solution to (P) if and only if $XZ^* = 0$. Hence, the primal optimal solution set can be written as

$$\mathcal{F}_P^* := \{X \in \mathcal{F}_P \mid X_U = 0 \text{ and } X_N = 0\}.$$

Analogously, the dual optimal solution set is given by

$$\mathcal{F}_D^* := \{Z \in \mathcal{F}_D \mid Z_U = 0 \text{ and } Z_B = 0\}.$$

Given $\mu \in \Re_{++}$, the pair $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ is said to be the *μ-center* $(X(\mu), Z(\mu))$ if and only if

$$XZ = \mu I, \tag{5.3}$$

see Definition 3.1. We refer to [80, 144, 157] for a proof of the existence and uniqueness of μ-centers; see also Theorem 4.9. The *central path* of the problem (P) is the curve

$$\{(X(\mu), Z(\mu)) \mid \mu > 0\}.$$

Consider the linear subspace $\mathcal{A}_B$ in $\mathcal{H}^{(K)}$ defined as

$$\mathcal{A}_B := \left\{ X_B \in \mathcal{H}^{(K)} \;\middle|\; \begin{bmatrix} X_B & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{A} \right\}. \tag{5.4}$$

The primal optimal solution set can now be formulated as

$$\mathcal{F}_P^* = \{X \in \mathcal{H}^{(\bar{n})} \mid X_B \in (\Lambda_B + \mathcal{A}_B) \cap \mathcal{H}_+^{(K)},\ X_U = 0,\ X_N = 0\},$$

and $\operatorname{rel} \mathcal{F}_P^* = \{X \in \mathcal{F}_P \mid X_B \succ 0\}$. The logarithmically homogeneous barrier function $\log\det X_B$ is strictly concave on $\operatorname{rel} \mathcal{F}_P^*$, see [110] and also Section 4.5. Moreover, Assumption 5.1 implies that $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ are bounded, see Theorem 2.5 or Table 2.1. Hence, there exists a unique maximizer X^a of $\log\det X_B$ on $\operatorname{rel} \mathcal{F}_P^*$; this maximizer X^a is called the *analytic center* of $\mathcal{F}_P^*$. It is characterized by the Karush-Kuhn-Tucker system

$$\begin{cases} X_B^a Z_B = I, \\ X_B^a \in (\Lambda_B + \mathcal{A}_B) \cap \mathcal{H}_{++}, \\ Z_B \in \mathcal{A}_B^\perp \cap \mathcal{H}_{++}, \\ X_U = 0,\ X_N = 0. \end{cases} \tag{5.5}$$

In a similar fashion, we define the analytic center of $\mathcal{F}_D^*$ as the maximizer of the logarithmic barrier $\log\det Z_N$ on the relative interior of $\mathcal{F}_D^*$. Letting

$$\mathcal{A}_N := \left\{ X_N \in \mathcal{H}^{(\bar{n}-K)} \;\middle|\; \begin{bmatrix} X_B & X_U \\ X_U^{\mathrm{H}} & X_N \end{bmatrix} \in \mathcal{A} \text{ for some } X_B, X_U \right\}, \tag{5.6}$$

with orthogonal complement

$$\mathcal{A}_N^\perp = \left\{ Z_N \in \mathcal{H}^{(\bar{n}-K)} \;\middle|\; \begin{bmatrix} 0 & 0 \\ 0 & Z_N \end{bmatrix} \in \mathcal{A}^\perp \right\}.$$

We can characterize Z^a by the Karush-Kuhn-Tucker system

$$\begin{cases} X_N Z_N^a = I, \\ Z_N^a \in (\Lambda_N + \mathcal{A}_N^\perp) \cap \mathcal{H}_{++}, \\ X_N \in \mathcal{A}_N \cap \mathcal{H}_{++} \\ Z_B = 0,\ Z_U = 0. \end{cases} \tag{5.7}$$

5.2 Analysis

The notion of central path plays a fundamental role in the development of interior point methods, see Chapter 3. In this section, we shall study the analytic properties of the central path in the context of semidefinite programming. These properties will be used in Chapter 6, where we perform local convergence analysis of a predictor-corrector algorithm for SDP.

For linear programming (i.e., X is diagonal for all $X \in B + \mathcal{A}$), it is known that the central path curve converges: $(X(\mu), Z(\mu)) \to (X^a, Z^a)$, as $\mu \to 0$, with (X^a, Z^a) being the analytic center of the primal and dual optimal solution sets $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ respectively [90]. Another important property of the central path in the context of linear programming is that it never converges *tangentially* to the optimal face [101]. This means that for any point on the central path, the distance to the end of the central path is of the same order as the distance to the optimal face, viz. $O(\mu)$. The aim of this section is to establish a similar property of the central path for semidefinite programming. More specifically, we shall prove that

$$\|X(\mu) - X^a\| + \|Z(\mu) - Z^a\| = O(\mu).$$

We begin with the following lemma which shows that the set

$$\{(X(\mu), Z(\mu)) \mid 0 < \mu < 1\}$$

is bounded.

Lemma 5.1 *For any $\mu > 0$ there holds*

$$\|X(\mu)\| + \|Z(\mu)\| = O(1 + \mu).$$

Proof. Since $X(\mu) - X(1) \in \mathcal{A}$ and $Z(\mu) - Z(1) \in \mathcal{A}^\perp$, it follows that $(X(\mu) - X(1)) \perp (Z(\mu) - Z(1))$. Using this property, we obtain

$$\begin{aligned} \bar{n}\mu + \bar{n} &= X(\mu) \bullet Z(\mu) + X(1) \bullet Z(1) \\ &= X(1) \bullet Z(\mu) + Z(1) \bullet X(\mu). \end{aligned}$$

Since $X(1) \succ 0$ and $Z(1) \succ 0$, we have

$$\|X(\mu)\| + \|Z(\mu)\| = O(X(1) \bullet Z(\mu) + Z(1) \bullet X(\mu)) = O(1 + \mu).$$

Q.E.D.

It follows from Lemma 5.1 that the central path has a limit point. We will now show that any limit point of the central path $\{(X(\mu), Z(\mu))\}$ is a strictly complementary optimal primal-dual pair.

Lemma 5.2 *For any $\mu \in (0,1)$ there holds*

$$X_B(\mu) = \Theta(1), \quad X_N(\mu) = \Theta(\mu),$$

$$Z_B(\mu) = \Theta(\mu), \quad Z_N(\mu) = \Theta(1).$$

Hence, any limit point of $\{(X(\mu), Z(\mu))\}$ as $\mu \to 0$ is a pair of strictly complementary primal-dual optimal solutions of (P) *and* (D).

Proof. Let $0 < \mu < 1$. For notational convenience, we will use X and Z to denote the matrices $X(\mu)$ and $Z(\mu)$. Let (X^*, Z^*) be the pair of strictly complementary primal-dual optimal solutions postulated by Assumption 5.2. Since $(X - X^*) \perp (Z - Z^*)$, we have

$$\begin{aligned} 0 &= (X - X^*) \bullet (Z - Z^*) \\ &= X \bullet Z - X^* \bullet Z - X \bullet Z^* \\ &= \operatorname{tr}(\mu I - X^* Z - X Z^*) \\ &= \bar{n}\mu - \sum_{i=1}^{K} \lambda_i Z_{ii} - \sum_{i=K+1}^{\bar{n}} \lambda_i X_{ii}, \end{aligned}$$

where the last step follows from (5.2). Since $\lambda_i > 0$ for all i and $X_{ii} \geq 0$ and $Z_{ii} \geq 0$ (by the positive semidefiniteness of X and Z), we obtain

$$\begin{cases} Z_{ii} = O(\mu), & i = 1, ..., K, \\ X_{ii} = O(\mu), & i = K+1, ..., \bar{n}. \end{cases}$$

Since $X \succeq 0$, $Z \succeq 0$, it follows that

$$X_N = O(\mu), \quad Z_B = O(\mu). \tag{5.8}$$

From $X \succ 0$ and $Z \succ 0$ we obtain

$$X_N - X_U^{\mathrm{H}} X_B^{-1} X_U \succ 0, \quad Z_B - Z_U Z_N^{-1} Z_U^{\mathrm{H}} \succ 0,$$

see Section A.4. Now consider the identities

$$\log\det X = \log\det X_B + \log\det(X_N - X_U^{\mathrm{H}} X_B^{-1} X_U),$$

$$\log\det Z = \log\det Z_N + \log\det(Z_B - Z_U Z_N^{-1} Z_U^{\mathrm{H}}).$$

Since $\det X \det Z = \det(\mu I) = \mu^{\bar{n}}$, it follows that $\log\det X + \log\det Z = \bar{n}\log\mu$ and

$$\begin{aligned} 0 \;=\; & \log\det X_B + \log\det\left(\frac{1}{\mu}(X_N - X_U^{\mathrm{H}} X_B^{-1} X_U)\right) \\ & + \log\det Z_N + \log\det\left(\frac{1}{\mu}(Z_B - Z_U Z_N^{-1} Z_U^{\mathrm{H}})\right). \end{aligned}$$

By the estimates (5.8) and using Lemma 5.1, we see that

$$X_B = O(1), \quad \frac{1}{\mu}(X_N - X_U^{\mathrm{H}} X_B^{-1} X_U) = O(1), \quad Z_N = O(1), \quad \frac{1}{\mu}(Z_B - Z_U Z_N^{-1} Z_U^{\mathrm{H}}) = O(1).$$

Therefore each of the four logarithm terms in the preceding equation are bounded from above as $\mu \to 0$. Since these four terms sum to zero, we must have

$$\begin{aligned} X_B &= \Theta(1), \qquad \frac{1}{\mu}(X_N - X_U^{\mathrm{H}} X_B^{-1} X_U) = \Theta(1), \\ Z_N &= \Theta(1), \qquad \frac{1}{\mu}(Z_B - Z_U Z_N^{-1} Z_U^{\mathrm{H}}) = \Theta(1). \end{aligned}$$

Together with (5.8), this implies

$$X_N = \Theta(\mu), \quad Z_B = \Theta(\mu).$$

This completes the proof of the lemma. **Q.E.D.**

Lemma 5.2 gives a precise result on the order of the eigenvalues of $X_B(\mu), X_N(\mu), Z_B(\mu)$ and $Z_N(\mu)$. We will now prove a preliminary result on the order of the off-diagonal blocks $X_U(\mu)$ and $Z_U(\mu)$.

Lemma 5.3 *For $\mu \in (0,1)$, there holds*

$$\begin{aligned} &\|X_U(\mu)\| = \Theta(1)\,\|Z_U(\mu)\|, \\ &-X_U(\mu) \bullet Z_U(\mu) = \Theta(1)\,\|X_U(\mu)\|^2, \\ &\|X_U(\mu)\| = o(\sqrt{\mu}), \quad \|Z_U(\mu)\| = o(\sqrt{\mu}), \qquad \textit{as } \mu \to 0. \end{aligned} \tag{5.9}$$

Proof. By the central path definition, we have

$$\mu I = \begin{bmatrix} X_B(\mu) & X_U(\mu) \\ X_U(\mu)^{\mathrm{H}} & X_N(\mu) \end{bmatrix} \begin{bmatrix} Z_B(\mu) & Z_U(\mu) \\ Z_U(\mu)^{\mathrm{H}} & Z_N(\mu) \end{bmatrix}.$$

Expanding the right-hand side and comparing the upper-right corner of the above identity, we have

$$0 = X_B(\mu) Z_U(\mu) + X_U(\mu) Z_N(\mu), \tag{5.10}$$

or equivalently,

$$Z_U(\mu) = -X_B(\mu)^{-1} X_U(\mu) Z_N(\mu). \tag{5.11}$$

Using $X_B(\mu) = \Theta(1)$ and $Z_N(\mu) = \Theta(1)$ (see Lemma 5.2), this implies that

$$\|Z_U(\mu)\| = \Theta(1)\,\|X_U(\mu)\|.$$

This proves the first part of the lemma.

We now prove (5.9). Pre-multiplying both sides of (5.11) by the matrix $X_U(\mu)^{\mathrm{H}}$ yields

$$X_U(\mu)^{\mathrm{H}} Z_U(\mu) = -X_U(\mu)^{\mathrm{H}} X_B(\mu)^{-1} X_U(\mu) Z_N(\mu).$$

Now taking the trace of the above matrices, we obtain

$$\begin{aligned} X_U(\mu) \bullet Z_U(\mu) &= -\operatorname{tr} X_U(\mu)^{\mathrm{H}} X_B(\mu)^{-1} X_U(\mu) Z_N(\mu) \\ &= -\operatorname{tr} Z_N(\mu)^{1/2} X_U(\mu)^{\mathrm{H}} X_B(\mu)^{-1} X_U(\mu) Z_N(\mu)^{1/2} \\ &= -\Theta(1) \, \|X_U(\mu)\|^2, \end{aligned}$$

where we used the fact that $X_B(\mu) = \Theta(1)$ and $Z_N(\mu) = \Theta(1)$, as proved in Lemma 5.2. This establishes (5.9).

It remains to prove the last part of the lemma. We consider an arbitrary convergent sequence $\{(X(\mu_k), Z(\mu_k)) \mid k = 1, 2, ...\}$ on the central path with $\mu_k \to 0$; its limit is denoted by X^*, Z^*, so that

$$X(\mu_k) - X^* = o(1), \quad Z(\mu_k) - Z^* = o(1). \tag{5.12}$$

By Lemma 5.2, we have $X_N^* = 0$, $X_U^* = Z_U^* = 0$ and $Z_B^* = 0$. Since $(X(\mu_k) - X^*) \perp (Z(\mu_k) - Z^*)$, we have

$$\begin{aligned} 0 &= (X_B(\mu_k) - X_B^*) \bullet Z_B(\mu_k) + 2X_U(\mu_k) \bullet Z_U(\mu_k) \\ &\quad + X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^*). \end{aligned} \tag{5.13}$$

Using (5.9) and (5.13), we obtain

$$\begin{aligned} \|X_U(\mu_k)\|^2 &= -\Theta(1) \, (X_U(\mu_k) \bullet Z_U(\mu_k)) \\ &= \Theta(1) \, ((X_B(\mu_k) - X_B^*) \bullet Z_B(\mu_k) + X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^*)) \\ &= o(\mu_k), \end{aligned} \tag{5.14}$$

where in the last step we used (5.12) and of Lemma 5.2. This implies that $\|X_U(\mu)\|^2 = o(\mu)$ holds true on the entire central path curve, for otherwise there would exist a convergent subsequence $\{(X(\mu_k), Z(\mu_k)) \mid k = 1, 2, ...\}$ for which

$$\liminf_{k \to \infty} \frac{\|X_U(\mu_k)\|^2}{\mu_k} > 0,$$

contradicting (5.14). The proof is complete. **Q.E.D.**

We now use Lemma 5.2 and Lemma 5.3 to prove that the central path $\{(X(\mu), Z(\mu)) \mid \mu > 0\}$ converges to (X^a, Z^a), and to estimate the rate at which it converges to this limit.

Lemma 5.4 *The primal-dual central path $\{(X(\mu), Z(\mu)) \mid \mu > 0\}$ converges to the analytic centers (X^a, Z^a) of $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ respectively. Moreover, if we let*

$$\epsilon(\mu) := \frac{\|X_U(\mu)\|}{\sqrt{\mu}},$$

then

$$\|X_B(\mu) - X_B^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2), \quad \|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

Proof. Suppose $0 < \mu < 1$. By expanding $X(\mu)Z(\mu) = \mu I$ and comparing the upper-left block, we obtain

$$\mu I_B = X_B(\mu)Z_B(\mu) + X_U(\mu)Z_U(\mu)^{\mathrm{H}}.$$

Pre-multiplying both sides with $(\mu X_B(\mu))^{-1}$ yields

$$X_B(\mu)^{-1} = \frac{1}{\mu}Z_B(\mu) + \frac{1}{\mu}X_B(\mu)^{-1}X_U(\mu)Z_U(\mu)^{\mathrm{H}}. \tag{5.15}$$

Let $\{A_B^{(i)} \mid i = 1, 2, \ldots, \dim \mathcal{A}_B^\perp\}$ be a set of matrices that spans the linear subspace

$$\mathcal{A}_B^\perp = \left\{ Z_B \in \mathcal{H}^{(K)} \;\middle|\; \begin{bmatrix} Z_B & Z_U \\ Z_U^{\mathrm{H}} & Z_N \end{bmatrix} \in \mathcal{A}^\perp \right\},$$

see (5.4). Since $Z_B^* = 0$, it follows from dual feasibility and (5.15) that

$$\begin{aligned} \frac{1}{\mu}Z_B(\mu) &= \sum_{i=1}^{\dim \mathcal{A}_B^\perp} \nu_i(\mu)A_B^{(i)}, \qquad \text{for some scalars } \nu_i(\mu) \\ &= X_B(\mu)^{-1} - \frac{1}{\mu}X_B(\mu)^{-1}X_U(\mu)Z_U(\mu)^{\mathrm{H}}. \end{aligned} \tag{5.16}$$

From Lemma 5.2, we know that $Z_B(\mu)/\mu = \Theta(1)$. Due to the linearindependence of the matrices $A_B^{(i)}$, $i = 1, \ldots, \dim \mathcal{A}_B^\perp$, this implies for all i that the sequences $\{\nu_i(\mu) \mid \mu \in (0, 1)\}$ are bounded. Moreover, we have from Lemma 5.3 that

$$\lim_{\mu \to 0} \frac{1}{\mu}X_B(\mu)^{-1}X_U(\mu)Z_U(\mu)^{\mathrm{H}} = 0.$$

Hence, any limit X^*, ν_i^* ($i \in \{1, \ldots, \dim \mathcal{A}_B^\perp\}$) for $\mu \to 0$ satisfies the following nonlinear system of equations:

$$\left\{ \begin{array}{ll} X_B^{-1} - \displaystyle\sum_{i=1}^{\dim \mathcal{A}_B^\perp} \nu_i A_B^{(i)} = 0, & \\ A_B^{(i)} \bullet X_B = b_i, \qquad i = 1, \ldots, \dim \mathcal{A}_B^\perp. & \end{array} \right. \tag{5.17}$$

Moreover, since $Z_B(\mu)/\mu = \Theta(1)$ and $X_B(\mu) = \Theta(1)$ for $\mu \in (0,1)$, we have

$$\sum_{i=1}^{\dim \mathcal{A}_B^{\perp}} \nu_i^* A_B^{(i)} \succ 0, \quad X_B^* \succ 0.$$

By (5.5), this means that $X^* = X^a$, the analytic center of $\mathcal{F}_P^*$ and hence

$$X(\mu) - X^a = o(1), \quad \text{as } \mu \to 0.$$

Using the linear independence of the matrices $A_B^{(i)}$, and using the fact that X_B^a is positive definite, it can be checked that the Jacobian (with respect to the variables X_B and ν_i of the nonlinear system (5.17)) is nonsingular at the solution X_B^a, ν_i^*, $i \in \{1, \ldots, \dim \mathcal{A}_B^{\perp}\}$. Hence we can apply the classical inverse function theorem to the above nonlinear system at the point: $X_B = X_B^a$, $\nu_i = \nu_i^*$, $i \in \{1, \ldots, \dim \mathcal{A}_B^{\perp}\}$, to obtain

$$\|X_B(\mu) - X_B^a\| = O\left(\|X_B(\mu)^{-1} - \sum_{i=1}^{\dim \mathcal{A}_B^{\perp}} \nu_i(\mu) A_B^{(i)}\| + \sum_{i=1}^{\dim \mathcal{A}_B^{\perp}} |A_B^{(i)} \bullet X_B(\mu) - b_i| \right). \quad (5.18)$$

By (5.16) and the definition of $\epsilon(\mu)$, we obtain from Lemma 5.3

$$\left\| X_B(\mu)^{-1} - \sum_{i=1}^{\dim \mathcal{A}_B^{\perp}} \nu_i(\mu) A_B^{(i)} \right\| = \left\| \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^{\mathrm{H}} \right\| = O(\epsilon(\mu)^2).$$

Also we have from $X(\mu) \in \mathcal{F}_P$

$$\begin{aligned} \left| A_B^{(i)} \bullet X_B(\mu) - b_i \right| &= \left| 2A_U^{(i)} \bullet X_U(\mu) + A_N^{(i)} \bullet X_N(\mu) \right| \\ &= O\left(\epsilon(\mu)\sqrt{\mu} + \mu \right), \qquad \text{for } i = 1, \ldots, \dim \mathcal{A}_B^{\perp}. \end{aligned}$$

Substituting the above two bounds into (5.18) yields

$$\|X_B(\mu) - X_B^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

It can be shown by an analogous argument that

$$\|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

The proof is complete. **Q.E.D.**

Lemma 5.4 only provides a rough sketch of the convergence behavior of the central path as $\mu \to 0$. Our goal is to characterize this convergence behavior more precisely.

Theorem 5.1 *Let $\mu \in (0,1)$. There holds*

$$X_B(\mu) = \Theta(1), \quad Z_N(\mu) = \Theta(1), \quad X_N(\mu) = \Theta(\mu), \quad Z_B(\mu) = \Theta(\mu), \quad (5.19)$$

and

$$\|X(\mu) - X^a\| = O(\mu), \quad \|Z(\mu) - Z^a\| = O(\mu). \quad (5.20)$$

Proof. The estimate (5.19) is already known from Lemma 5.2, so we only need to prove (5.20). By Lemma 5.3 and Lemma 5.4, it is sufficient to show that

$$\|X_U(\mu)\| = O(\mu).$$

Suppose to the contrary that there exists a sequence

$$\{(X(\mu_k), Z(\mu_k)) \mid k = 1, 2, \ldots\} \tag{5.21}$$

with $\|X_U(\mu_k)\| > 0$ for all k and

$$\lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|} = 0. \tag{5.22}$$

With the notation of Lemma 5.4, the condition (5.22) implies

$$\epsilon(\mu_k) + \sqrt{\mu_k} \quad \sim \quad \epsilon(\mu_k) = \frac{\|X_U(\mu_k)\|}{\sqrt{\mu_k}}. \tag{5.23}$$

By virtue of Lemma 5.2, we can choose the subsequence (5.21) such that

$$\lim_{k\to\infty} \frac{Z_B(\mu_k)}{\mu_k}$$

exists, and using Lemma 5.4 and relation (5.23), we can also assume the existence of

$$\Delta_B^x(\infty) := \lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|^2}(X_B(\mu_k) - X_B^a).$$

From the existence of the above limits, we obtain

$$\lim_{k\to\infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} = \lim_{k\to\infty} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k}. \tag{5.24}$$

Notice that the hypothesis (5.22) implies that

$$\lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|^2}(X(\mu_k) - X^a) = \begin{bmatrix} \Delta_B^x(\infty) & 0 \\ 0 & 0 \end{bmatrix}.$$

Using also $Z_B^a = 0$, we thus obtain for any $k = 1, 2, \ldots$ that

$$\begin{aligned} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k} &= \frac{\Delta_B^x(\infty) \bullet (Z_B(\mu_k) - Z_B^a)}{\mu_k} \\ &= \lim_{j\to\infty} \frac{\mu_j}{\mu_k\|X_U(\mu_j)\|^2}((X(\mu_j) - X^a) \bullet (Z(\mu_k) - Z^a)) \\ &= 0, \end{aligned}$$

where the last step is due to the orthogonality condition $(X(\mu_j) - X^a) \perp (Z(\mu_k) - Z^a)$ for all j and k. Therefore,

$$0 = \lim_{k\to\infty} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k} = \lim_{k\to\infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2}, \tag{5.25}$$

where we used (5.24). Analogously, it can be shown that

$$\lim_{k\to\infty} \frac{X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^a)}{\|X_U(\mu_k)\|^2} = 0. \tag{5.26}$$

Since $(X(\mu_k) - X^a) \perp (Z(\mu_k) - Z^a)$, we have from (5.25) and (5.26) that

$$\begin{aligned} 0 &= \lim_{k\to\infty} \frac{(X(\mu_k) - X^a) \bullet (Z(\mu_k) - Z^a)}{\|X_U(\mu_k)\|^2} \\ &= \lim_{k\to\infty} 2\frac{X_U(\mu_k) \bullet Z_U(\mu_k)}{\|X_U(\mu_k)\|^2}, \end{aligned}$$

which clearly contradicts (5.9). The proof is complete. **Q.E.D.**

5.3 Discussion

Theorem 5.1 characterizes completely the limiting behavior of the primal-dual central path as $\mu \to 0$. We point out that this limiting behavior was well understood in the context of linear programming and the monotone horizontal linear complementarity problem, see Güler [51] and Monteiro and Tsuchiya [101] respectively. Notice that under a *Nondegeneracy Assumption*, the estimates (5.20) follow immediately from the application of the classical inverse function theorem. Namely, nondegeneracy requires that the Jacobian of the nonlinear system

$$\begin{cases} P_{\mathcal{H}}(XZ) = \mu I \\ X \in B + \mathcal{A}, \quad Z \in C + \mathcal{A}^\perp \end{cases}$$

is nonsingular at $X = X^a$, $Z = Z^a$ and $\mu = 0$, see Alizadeh, Haeberly and Overton [3, 4] and Kojima, Shida and Shindoh [76, 78]. Thus, the real contribution of Theorem 5.1 lies in establishing these estimates in the absence of the nondegeneracy assumption.

It is known that in the case of linear programming the proof of quadratic convergence of predictor-corrector interior point algorithms required an error bound result of Hoffman. This error bound states that the distance from any vector $x \in \Re^n$ to a polyhedral set $\mathcal{P} := \{x \mid Ax \le a\}$ can be bounded in terms of the "amount of constraint violation" at x, namely $\|[Ax - a]_+\|$, where $[\cdot]_+$ denotes the positive part of a vector. More precisely, Hoffman's error bound [55] states that there exists some constant $\tau > 0$ such that

$$\mathrm{dist}(x, \mathcal{P}) \le \tau \|[Ax - a]_+\|, \qquad \forall x \in \Re^n.$$

Unfortunately, this error bound no longer holds for linear systems over the cone of positive semidefinite matrices (see the example below). In fact, much of the difficulty in the local analysis of interior point algorithms for SDP can be attributed to this lack of an analog of Hoffman's error bound result (see the analysis of [76, 123]). Specifically, without such an error bound result, it is difficult to estimate the distance from the current iterates to the optimal solution set. In essence, what we have established in Theorem 5.1 is an error bound result along the central path. In other words, although a Hoffman type error bound cannot hold over the entire feasible set of (P), it nevertheless still holds true on the restricted region "near the central path". One consequence of this restriction to the central path is that in Chapter 6, we will need to require the iterates of a path–following method to stay "sufficiently close" to the central path, in order to establish superlinear convergence.

Example 5.1 *Consider the semidefinite program $CP(B, C, \mathcal{A}, \mathcal{S}_+)$ with*

$$B = \begin{bmatrix} 0 & 0 \\ & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ & 0 \end{bmatrix}, \quad \mathcal{A} = \{X \in \mathcal{S}^{(2)} \mid X_{22} = 0\}.$$

This semidefinite program is primal and dual strongly feasible, and has optimal value $p^* = 0$. We are interested in the optimal solution set

$$\mathcal{F}_P^* = \{X \succeq 0 \mid X_{11} = 0, \, X_{22} = 1\}.$$

Clearly, there is exactly one solution X^* to the above linear matrix inequality, namely

$$X^* := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

For each $\epsilon > 0$, consider the matrix

$$X(\epsilon) := \begin{bmatrix} \epsilon^2 & \epsilon \\ \epsilon & 1 \end{bmatrix}.$$

Clearly, $X(\epsilon) \succeq 0$. The amount of constraint violation is equal to ϵ^2. However, the distance $\|X(\epsilon) - X^*\|_F = \Theta(\epsilon)$. Thus, there cannot exist any fixed $\tau > 0$ such that $\|X(\epsilon) - X^*\| \leq \tau\epsilon^2$, for all $\epsilon > 0$.

The assumption of strict complementarity plays a crucial role in our analysis of the central path. However, some properties still hold without strictly complementary solutions. Specifically, De Klerk, Roos and Terlaky [70] showed that any limit point of the central path is a maximal complementary solution (cf. Lemma 5.2), and Goldfarb and Scheinberg [41] showed that the central path converges to the analytic center of the optimal solution set (cf. Lemma 5.4).

Chapter 6

Superlinear Convergence

The goal of this chapter is to establish the superlinear convergence of a path–following algorithm for semidefinite programming, without nondegeneracy assumptions. Specifically, we propose a predictor–corrector type algorithm with $(r+1)$–step superlinear convergence of order $2/(1+2^{-r})$, where any positive integer can be assigned to the parameter r. The parameter r is used in the algorithm as an upper bound on the number of successive corrector steps that are allowed between two predictor steps. The proof of superlinear convergence is based on the properties of the central path that were derived in Chapter 5.

6.1 A Predictor–Corrector Algorithm

In Chapter 3, we have discussed the basic form of the predictor–corrector path–following algorithm for semidefinite programming. Below, we propose some modifications with respect to the step length choice in this basic scheme, and we will also consider the option of multiple successive corrector steps. However, we will use the same search directions as in Chapter 3. Namely, we consider a semidefinite program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{H}_+^{(\bar{n})})$ that is both primal and dual strongly feasible. Given an interior solution pair (X, Z),

$$X \in (B + \mathcal{A}) \cap \mathcal{H}_{++}^{(\bar{n})}, \quad Z \in (C + \mathcal{A}^\perp) \cap \mathcal{H}_{++}^{(\bar{n})},$$

the direction for obtaining a $(\gamma\mu)$-center with $\gamma \in [0, 1]$, is defined as the solution $(\Delta X, \Delta Z)$ of the following system of linear equations

$$\begin{cases} \Delta X + D(X, Z)\Delta Z D(X, Z) = \gamma\mu Z^{-1} - X \\ \Delta X \in \mathcal{A}, \quad \Delta Z \in \mathcal{A}^\perp, \end{cases} \tag{6.1}$$

see (3.15). For $\gamma = 0$, we denote the solution of (6.1) by $(\Delta X^p, \Delta Z^p)$, the predictor direction. For $\gamma = 1$, the solution is denoted by $(\Delta X^c, \Delta Z^c)$, the corrector direction.

As in Chapter 3, we treat primal–dual interior point methods in the V-space framework. With each interior solution pair (X, Z), we associate a V–space solution, which is a positive diagonal matrix V satisfying

$$V = L_d^{-1} X L_d^{-\mathrm{H}} = L_d^{\mathrm{H}} Z L_d,$$

for some invertible matrix L_d. On the diagonal of V^2 are then the eigenvalues of the matrix XZ. Recall from definition (3.7) that $D_X := L_d^{-1} \Delta X L_d^{-\mathrm{H}}$ and $D_Z := L_d^{\mathrm{H}} \Delta Z L_d$. Using this notation, (6.1) can be reformulated as

$$\begin{cases} D_X + D_Z = \gamma\mu V^{-1} - V \\ D_X \in \mathcal{A}(L_d), \quad D_Z \in \mathcal{A}^{\perp}(L_d), \end{cases}$$

see (3.11). It follows from orthogonality that

$$\|D_X^p\|_F^2 + \|D_Z^p\|_F^2 = \|D_X^p + D_Z^p\|_F^2 = \|V\|_F^2 = \bar{n}\mu. \tag{6.2}$$

The corrector direction does not change the duality gap,

$$(X + \Delta X^c) \bullet (Z + \Delta Z^c) = X \bullet Z, \tag{6.3}$$

whereas

$$(X + t\Delta X^p) \bullet (Z + t\Delta Z^p) = (1 - t) X \bullet Z, \tag{6.4}$$

for any $t \in \Re$, see (3.14).

We let $\bar{X}(t)$ and $\bar{Z}(t)$ denote the primal and dual solutions in the transformed space, after taking a step of length t in the predictor direction, i.e.

$$\bar{X}(t) = V + tD_X^p, \quad \bar{Z}(t) = V + tD_Z^p.$$

Just as in Section 3.2, we define $G(t) := D(\bar{X}(t), \bar{Z}(t))^{1/2}$ and $V(t) := G(t)\bar{Z}(t)G(t)$, so that

$$V(t) = G(t)\bar{Z}(t)G(t) = G(t)^{-1}\bar{X}(t)G(t)^{-1}.$$

Algorithm 6.1

Given $(X^{(0)}, Z^{(0)}) \in \mathcal{F}_P \times \mathcal{F}_D$ *with* $\delta(V^{(0)}) \leq 1/4$.
Parameter ϵ, $0 < \epsilon \leq (X^{(0)} \bullet Z^{(0)})/\bar{n}$ *and positive integer* r.
Let $k = 0$.
`REPEAT` *(main iteration)*
  Let $X = X^{(k)}, Z = Z^{(k)}$ *and* $\mu_k = X \bullet Z/\bar{n}$.
  Predictor*: compute* $(\Delta X^p, \Delta Z^p)$ *from (6.1) with* $\gamma = 0$.
  Compute the largest step t_k *such that for all* $0 \leq t \leq t_k$ *there holds*

$\delta(V(t)) \leq \min(1/2, ((1-t)\mu_k/\epsilon)^{2^{-r}})$.

Let $X' := X + t_k \Delta X^p, Z' := Z + t_k \Delta Z^p$ *and* $\beta_k = \min(1/4, (1-t_k)\mu_k/\epsilon)$.

Corrector*:*

`FOR` $i = 1$ `to` r `DO`

Let $X = X', Z = Z'$ *and compute* V.

`IF` $\delta(V) \leq \beta_k$ `THEN` *exit loop.*

Compute $(\Delta X^c, \Delta Z^c)$ *from (6.1) with* $\gamma = 1$.

Set $X' = X + \Delta X^c, Z' = Z + \Delta Z^c$.

`END FOR`

$X^{(k+1)} = X', Z^{(k+1)} = Z'$

Set $k = k + 1$.

`UNTIL` *convergence.*

Interestingly, each corrector step reduces $\delta(\cdot)$ at a quadratic rate as stated by Lemma 3.8. This implies that for any $k \geq 1$, we have

$$\delta(V^{(k)}) \leq \beta_{k-1}, \tag{6.5}$$

where $V^{(k)}$ is the V–space solution corresponding to $(X^{(k)}, Z^{(k)})$, i.e.

$$V^{(k)} = (L_d^{(k)})^{-1} X^{(k)} (L_d^{(k)})^{-\mathrm{H}} = (L_d^{(k)})^{\mathrm{H}} Z^{(k)} L_d^{(k)}.$$

Also, it follows from (6.3) and (6.4) that for any $k > 1$

$$\begin{aligned} \delta(V^{(k)}) \leq \beta_{k-1} &\leq (1-t_{k-1})\mu_{k-1}/\epsilon \\ &= \mu_k/\epsilon = O(\mu_k). \end{aligned} \tag{6.6}$$

Furthermore, if $\beta_k = 1/4$, then only one (instead of r) corrector step is needed to recenter the iterate (see Chapter 3). In other words, the iterations of Algorithm SDP(ϵ) are identical to those of the basic primal-dual predictor-corrector algorithm of Section 3.5.2, for all k with

$$\frac{\mu_k}{\epsilon} \geq \frac{1}{4}.$$

We can therefore conclude from Theorem 3.2 that the algorithm yields $\mu_k \leq \epsilon/4$ for all $k \geq \Gamma\sqrt{\bar{n}}\log(\mu_0/\epsilon)$, where Γ is a universal constant, independent of the problem data. Thus, we have the following polynomial complexity result.

Theorem 6.1 *For each* $0 < \epsilon < (X^{(0)} \bullet Z^{(0)})/\bar{n}$, *Algorithm SDP($\epsilon$) generates an iterate* $(X^{(k)}, Z^{(k)}) \in \mathcal{F}_P \times \mathcal{F}_D$ *with* $(X^{(k)} \bullet Z^{(k)})/\bar{n} \leq \epsilon/4$ *in at most* $\mathbf{O}(\sqrt{\bar{n}}\log(\mu_0/\epsilon))$ *predictor-corrector steps.*

In addition to having polynomial complexity, Algorithm SDP(ϵ) also possesses a superlinear rate of convergence. We prove this in the next section.

6.2 Convergence Analysis

We begin by establishing the global convergence of Algorithm SDP(ϵ). Notice that Algorithm SDP(ϵ) chooses the predictor step length t_k to be the largest step such that for all $0 \le t \le t_k$ there holds

$$\delta(V(t)) \le \min\left(\frac{1}{2}, ((1-t)\mu/\epsilon)^{2^{-r}}\right). \tag{6.7}$$

Recall from (3.23) that

$$(1-t)\delta(V(t)) \le (1-t)\delta(V) + t^2 \left\|D_X^p D_Z^p\right\|_F / \mu. \tag{6.8}$$

Using (6.2), we thus obtain for $0 \le t < 1$ that

$$\delta(V(t)) \le \delta(V) + \frac{\bar{n}t^2}{2(1-t)}. \tag{6.9}$$

Combining (6.7) and (6.9), we can easily establish the global convergence of Algorithm SDP(ϵ).

Theorem 6.2 *There holds*

$$\lim_{k\to\infty} \mu_k = 0,$$

i.e. Algorithm SDP(ϵ) is globally convergent.

Proof. Due to (6.4), the sequence $\mu_0,\ \mu_1, \ldots$ is a monotonically decreasing sequence. Hence, $\lim_{k\to\infty} \mu_k$ exists. Suppose contrary to the statement of the lemma that

$$\mu_\infty = \lim_{k\to\infty} \mu_k, \quad \mu_\infty > 0. \tag{6.10}$$

Consider $k \ge \Gamma\sqrt{\bar{n}}\log(\mu_0/\epsilon)$. From Theorem 6.1, we have

$$\mu_k \le \epsilon/4 \tag{6.11}$$

and hence, using (6.5),

$$\delta(V^{(k)}) \le \beta_{k-1} = \min(\frac{1}{4}, \frac{\mu_k}{\epsilon}) = \frac{\mu_k}{\epsilon}. \tag{6.12}$$

Now consider a step length $0 \le t \le 0.5\sqrt{\mu_k/(\bar{n}\epsilon)}$ and note from (6.11) that

$$1 - t \ge \frac{3}{4}.$$

We obtain from (6.9) and (6.12) that

$$\begin{aligned}
\delta(V^{(k)}(t)) &\leq \delta(V^{(k)}) + \frac{\bar{n}t^2}{2(1-t)} \\
&\leq \frac{\mu_k}{\epsilon} + \frac{\mu_k}{6\epsilon} \\
&\leq \frac{7}{6} \cdot \frac{4}{3} \cdot \frac{(1-t)\mu_k}{\epsilon} \\
&< \left(\frac{(1-t)\mu_k}{\epsilon}\right)^{2^{-r}},
\end{aligned}$$

where we used (6.11) and the fact that $r \geq 1$. By definition of t_k, this implies that

$$t_k \geq \frac{1}{2}\sqrt{\frac{\mu_k}{\bar{n}\epsilon}} \geq \frac{1}{2}\sqrt{\frac{\mu_\infty}{\bar{n}\epsilon}} = \Theta(1).$$

This, together with (6.4), implies that $1 - (\mu_{k+1}/\mu_k) = \Theta(1)$, which contradicts (6.10). **Q.E.D.**

Next we proceed to establish the superlinear convergence of Algorithm SDP(ϵ). In light of (6.4), we only need to show that the predictor step length t_k approaches 1. Hence we are led to bound t_k from below. For this purpose, we note from (6.8) that, for $t \in (0,1)$,

$$\delta(V(t)) \leq \delta(V) + \frac{1}{1-t} \left\| D_X^p D_Z^p \right\|_F / \mu. \tag{6.13}$$

Thus, if we can properly bound $\|D_X^p D_Z^p\|_F$, then we will obtain a lower bound on the predictor step length t_k.

To begin, let us define

$$L_\mu := L_d D(L_d^{-1} X(\mu) L_d^{-\mathrm{H}}, L_d^{\mathrm{H}} Z(\mu) L_d)^{1/2}. \tag{6.14}$$

Remark that

$$\sqrt{\mu} I = L_\mu^{-1} X(\mu) L_\mu^{-\mathrm{H}} = L_\mu^{\mathrm{H}} Z(\mu) L_\mu.$$

Now define the predictor direction starting from the solution $(X(\mu), Z(\mu))$ on the central path as follows:

$$\begin{cases} \hat{D}_X^p(\mu) + \hat{D}_Z^p(\mu) = -\sqrt{\mu} I, \\ \hat{D}_X^p(\mu) \in \mathcal{A}(L_\mu), \quad \hat{D}_Z^p(\mu) \in \mathcal{A}^\perp(L_\mu). \end{cases}$$

Let $(\hat{X}^a, \hat{Z}^a)$ be the analytic center of the optimal solution set in the L_μ-transformed space,

$$\hat{X}^a := L_\mu^{-1} X^a L_\mu^{-\mathrm{H}}, \quad \hat{Z}^a := L_\mu^{\mathrm{H}} Z^a L_\mu.$$

See (5.5) and (5.7) for a definition of the analytic center (X^a, Z^a). We will show in Lemma 6.1 below that $\hat{D}_X^p(\mu)$ is close to the optimal step $\hat{X}^a - \sqrt{\mu} I$ for small μ. We will bound the difference between $\hat{D}_X^p(\mu)$ and D_X^p afterwards.

Lemma 6.1 *There holds*

$$\left\|\sqrt{\mu}I + \hat{D}_X^p(\mu) - \hat{X}^a\right\| + \left\|\sqrt{\mu}I + \hat{D}_Z^p(\mu) - \hat{Z}^a\right\| = O(\mu^{3/2}).$$

Proof. Since

$$\hat{X}^a\hat{Z}^a = L_\mu^{-1}X^aZ^aL_\mu = 0,$$

it follows that

$$(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a) = (\sqrt{\mu}I - \hat{Z}^a)(\sqrt{\mu}I - \hat{X}^a).$$

Therefore, the matrix $(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a)$, or equivalently, the matrix

$$L_\mu^{-1}(X(\mu) - X^a)(Z(\mu) - Z^a)L_\mu,$$

is Hermitian. By definition of the Frobenius norm, we obtain

$$\begin{aligned}\left\|(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a)\right\|_F &= \left\|L_\mu^{-1}(X(\mu) - X^a)(Z(\mu) - Z^a)L_\mu\right\|_F \\ &= \sqrt{\operatorname{tr}\,(X(\mu) - X^a)(Z(\mu) - Z^a)(X(\mu) - X^a)(Z(\mu) - Z^a)} \\ &= O(\mu^2), \qquad (6.15)\end{aligned}$$

where the last step follows from Theorem 5.1. Now since $\hat{X}^a\hat{Z}^a = 0$ and $\hat{D}_X^p(\mu)+\hat{D}_Z^p(\mu) = -\sqrt{\mu}I$, we have

$$\begin{aligned}(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a) &= \mu I - \sqrt{\mu}(\hat{X}^a + \hat{Z}^a) \\ &= \sqrt{\mu}(\sqrt{\mu}I + \hat{D}_X^p(\mu) - \hat{X}^a) \\ &\quad + \sqrt{\mu}(\sqrt{\mu}I + \hat{D}_Z^p(\mu) - \hat{Z}^a).\end{aligned}$$

As

$$\sqrt{\mu}I + \hat{D}_X^p(\mu) - \hat{X}^a \in \mathcal{A}^\perp(L_\mu), \quad \sqrt{\mu}I + \hat{D}_Z^p(\mu) - \hat{Z}^a \in \mathcal{A}(L_\mu),$$

it follows that

$$\begin{aligned}&\left\|\sqrt{\mu}I + \hat{D}_X^p(\mu) - \hat{X}^a\right\|_F^2 + \left\|\sqrt{\mu}I + \hat{D}_Z^p(\mu) - \hat{Z}^a\right\|_F^2 = \\ &\quad \frac{1}{\mu}\left\|(\hat{X}^a - \sqrt{\mu}I)(\hat{Z}^a - \sqrt{\mu}I)\right\|_F^2 \\ &= O(\mu^3),\end{aligned}$$

where the last step is due to (6.15). This proves the lemma. **Q.E.D.**

Lemma 6.1 applies only to $(\hat{D}_X^p(\mu), \hat{D}_Z^p(\mu))$, namely the predictor directions for the points located exactly on the central path. What we need is a similar bound for (D_X^p, D_Z^p) (obtained at points close to the central path). This leads us to bound the difference $\hat{D}_X^p(\mu) - D_X^p$. Indeed, our next goal is to show (Lemma 6.5) that

$$\left\|\hat{D}_X^p(\mu) - D_X^p\right\|_F = O(\sqrt{\mu}\delta(V)).$$

We prove this bound by a sequence of lemmas.

Lemma 6.2 *Suppose $\delta(V) \leq 1/2$. There holds*

$$\left\| L_d^{-1}(X(\mu) - X)L_d^{-H} \right\| + \left\| L_d^H(Z(\mu) - Z)L_d \right\| = O(\sqrt{\mu}\delta(V)).$$

Proof. Let

$$\Delta_x(\mu) := L_d^{-1}(X(\mu) - X)L_d^{-\mathrm{H}}, \quad \Delta_z(\mu) := L_d^{\mathrm{H}}(Z(\mu) - Z)L_d.$$

Clearly, $\Delta_x(\mu)$ and $\Delta_z(\mu)$ are Hermitian and $\Delta_x(\mu) \perp \Delta_z(\mu)$. Since $X \bullet Z = X(\mu) \bullet Z(\mu) = \bar{n}\mu$, we have

$$\begin{aligned} \mathrm{tr}\,(Z(X(\mu) - X) + X(Z(\mu) - Z)) &= \mathrm{tr}\,((X(\mu) - X)Z + X(Z(\mu) - Z)) \\ &= -\,\mathrm{tr}\,((X(\mu) - X)(Z(\mu) - Z)) - \mathrm{tr}\,XZ \\ &\quad + \mathrm{tr}\,X(\mu)Z(\mu) \\ &= 0, \end{aligned}$$

where the last step follows from $(X(\mu) - X) \perp (Z(\mu) - Z)$. Recall that $V = L_d^{-1}XL_d^{-\mathrm{H}} = L_d^{\mathrm{H}}ZL_d$ and hence

$$\begin{aligned} \mathrm{tr}\,(V(\Delta_x(\mu) + \Delta_z(\mu))) &= \mathrm{tr}\,(L_d^{\mathrm{H}}Z(X(\mu) - X)L_d^{-\mathrm{H}} + L_d^{-1}X(Z(\mu) - Z)L_d) \\ &= \mathrm{tr}\,(Z(X(\mu) - X) + X(Z(\mu) - Z)) \\ &= 0. \end{aligned}$$

Using the spectral decomposition $\Delta_x(\mu) + \Delta_z(\mu) = Q^{\mathrm{H}}\Lambda Q$ with Q unitary and Λ real diagonal (see Appendix A), we further obtain

$$0 = \mathrm{tr}\,(V(\Delta_x(\mu) + \Delta_z(\mu))) = \mathrm{tr}\,(VQ^{\mathrm{H}}\Lambda Q) = \mathrm{tr}\,((QVQ^{\mathrm{H}})\Lambda).$$

Notice that QVQ^{H} is a positive definite matrix with eigenvalues V_{jj}, $j = 1, 2, \ldots, \bar{n}$, which are all $\Theta(\sqrt{\mu})$. This implies that the diagonal entries of QVQ^{H} are also $\Theta(\sqrt{\mu})$. Therefore, the preceding equation implies that the diagonal matrix Λ must have a nonpositive eigenvalue and that its diagonal entries are mutually of the same order of magnitude. In other words,

$$\|\Delta_x(\mu) + \Delta_z(\mu)\|_2 = O(|\lambda_{\min}(\Delta_x(\mu) + \Delta_z(\mu))|). \tag{6.16}$$

By the definition of the central path, we have

$$\begin{aligned} \mu I &= (V + \Delta_x(\mu))(V + \Delta_z(\mu)) \\ &= \left(V + \frac{\Delta_x(\mu) + \Delta_z(\mu)}{2} + \frac{\Delta_x(\mu) - \Delta_z(\mu)}{2}\right) \times \\ &\quad \left(V + \frac{\Delta_x(\mu) + \Delta_z(\mu)}{2} - \frac{\Delta_x(\mu) - \Delta_z(\mu)}{2}\right). \end{aligned}$$

Since the left hand side matrix (μI) is Hermitian, the skew–Hermitian cross term must cancel when we expand the matrix product in the right hand side. It follows that

$$\mu I = \left(V + \frac{\Delta_x(\mu) + \Delta_z(\mu)}{2} \right)^2 - \frac{1}{4} \left(\Delta_x(\mu) - \Delta_z(\mu) \right)^2$$

and therefore,

$$V + \frac{\Delta_x(\mu) + \Delta_z(\mu)}{2} \succeq \sqrt{\mu} I.$$

Using (3.17), we obtain

$$|\lambda_{\min}(\Delta_x(\mu) + \Delta_z(\mu))| = O(\sqrt{\mu}\delta(V)).$$

Combining this with (6.16) and using the fact that $\Delta_x(\mu) \perp \Delta_z(\mu)$, we have

$$\|\Delta_x(\mu)\| + \|\Delta_z(\mu)\| = O(|\lambda_{\min}(\Delta_x(\mu) + \Delta_z(\mu))|) = O(\sqrt{\mu}\delta(V)).$$

Q.E.D.

Lemma 6.3 *Suppose* $\delta(V) \leq 1/2$ *and let* $\bar{D}_\mu := D(L_d^{-1} X(\mu) L_d^{-H}, L_d^H Z(\mu) L_d)$. *There holds*

$$\left\| \bar{D}_\mu - I \right\| = O(\delta(V)).$$

Proof. Notice that

$$L_\mu^{-1} X L_\mu^{-\mathrm{H}} = \sqrt{\mu} I + L_\mu^{-1}(X - X(\mu)) L_\mu^{-\mathrm{H}} \tag{6.17}$$

and

$$L_\mu^{\mathrm{H}} Z L_\mu = \sqrt{\mu} I + L_\mu^{\mathrm{H}}(Z - Z(\mu)) L_\mu. \tag{6.18}$$

Recall from definition (6.14) that $L_\mu = L_d \bar{D}_\mu^{1/2}$. Therefore, by pre- and post-multiplying (6.17) by $\bar{D}_\mu^{1/2}$ and (6.18) by $\bar{D}_\mu^{-1/2}$ and rearranging terms,

$$\sqrt{\mu}(\bar{D}_\mu - \bar{D}_\mu^{-1}) = L_d^{-1}(X(\mu) - X) L_d^{-\mathrm{H}} + L_d^{\mathrm{H}}(Z - Z(\mu)) L_d.$$

Together with Lemma 6.2, this implies $\bar{D}_\mu = \Theta(1)$ and

$$\left\| \bar{D}_\mu - I \right\| = O(\delta(V)).$$

The lemma is proved. **Q.E.D.**

Now, let

$$\hat{D}_X^p := L_\mu^{-1}(L_d D_X^p L_d^{\mathrm{H}}) L_\mu^{-\mathrm{H}}, \quad \hat{D}_Z^p := L_\mu^{\mathrm{H}}(L_d^{-\mathrm{H}} D_Z^p L_d^{-1}) L_\mu.$$

Notice that $(\hat{D}_X^p, \hat{D}_Z^p) \in \mathcal{A}(L_\mu) \times \mathcal{A}^\perp(L_\mu)$.

Lemma 6.4 *Suppose $\delta(V) \leq 1/2$. We have*

$$\left\|\hat{D}_X^p - D_X^p\right\| + \left\|\hat{D}_Z^p - D_Z^p\right\| = O(\sqrt{\mu}\delta(V)).$$

Proof. Notice that $L_d^{-1} L_\mu = \bar{D}_\mu^{1/2}$, where $\bar{D}_\mu$ is as in Lemma 6.3. We have

$$\begin{aligned}\hat{D}_Z^p &= \bar{D}_\mu^{1/2} Z_X^p \bar{D}_\mu^{1/2} \\ &= D_Z^p + (\bar{D}_\mu^{1/2} - I) D_Z^p \bar{D}_\mu^{1/2} + D_Z^p (\bar{D}_\mu^{1/2} - I).\end{aligned}$$

Now using Lemma 6.3 and (6.2), we see that

$$\left\|\hat{D}_Z^p - D_Z^p\right\| = O(\sqrt{\mu}\delta(V)).$$

It can be shown in an analogous way that

$$\left\|\hat{D}_X^p - D_X^p\right\| = O(\sqrt{\mu}\delta(V)).$$

Q.E.D.

Now we are ready to bound the difference between $\hat{D}_X^p(\mu)$ and D_X^p.

Lemma 6.5 *Suppose $\delta(V) \leq 1/2$. We have*

$$\left\|\hat{D}_X^p(\mu) - D_X^p\right\| + \left\|\hat{D}_Z^p(\mu) - D_Z^p\right\| = O(\sqrt{\mu}\delta(V)).$$

Proof. By definition of the predictor directions, we have

$$\hat{D}_X^p(\mu) + \hat{D}_Z^p(\mu) = -\sqrt{\mu} I$$

and

$$D_X^p + D_Z^p = -V.$$

Combining these two relations yields

$$\hat{D}_X^p(\mu) - \hat{D}_X^p + \hat{D}_Z^p(\mu) - \hat{D}_Z^p = V - \sqrt{\mu} I + D_X^p - \hat{D}_X^p + D_Z^p - \hat{D}_Z^p.$$

Now using Lemma 6.4 and using the fact that

$$\|V - \sqrt{\mu} I\|_F = \|(V + \sqrt{\mu} I)^{-1}(V^2 - \mu I)\|_F \leq \sqrt{\mu}\delta(V),$$

we obtain

$$\left\|\hat{D}_X^p(\mu) - \hat{D}_X^p + \hat{D}_Z^p(\mu) - \hat{D}_Z^p\right\| = O(\sqrt{\mu}\delta(V)).$$

Since $(\hat{D}_X^p(\mu) - \hat{D}_X^p) \perp (\hat{D}_Z^p(\mu) - \hat{D}_Z^p)$, the lemma follows from the above relation, after applying Lemma 6.4 once more. **Q.E.D.**

Combining (6.15), Lemma 6.1 and Lemma 6.5 we can now estimate the order of $\|D_X^p D_Z^p\|$, and hence, using (6.13), we can estimate the predictor step length t_k.

Lemma 6.6 *We have*

$$\|D_X^p D_Z^p\| = O(\mu(\mu + \delta(V))).$$

Proof. Combining Lemma 6.5 with Lemma 6.1, we have

$$\left\|\sqrt{\mu}I + D_X^p - \hat{X}^a\right\| + \left\|\sqrt{\mu}I + D_Z^p - \hat{Z}^a\right\| = O(\sqrt{\mu}(\mu + \delta(V))), \tag{6.19}$$

so that, using (6.2),

$$\left\|\sqrt{\mu}I - \hat{X}^a\right\| + \left\|\sqrt{\mu}I - \hat{Z}^a\right\| = O(\sqrt{\mu}). \tag{6.20}$$

Moreover,

$$\begin{aligned} D_X^p D_Z^p &= (\hat{X}^a - \sqrt{\mu}I)(\hat{Z}^a - \sqrt{\mu}I) + (\hat{X}^a - \sqrt{\mu}I)(\sqrt{\mu}I + D_Z^p - \hat{Z}^a) \\ &\quad + (\sqrt{\mu}I + D_X^p - \hat{X}^a)D_Z^p. \end{aligned}$$

Applying (6.15), (6.19), (6.20) and (6.2) to the above relation yields

$$\|D_X^p D_Z^p\| = O(\mu(\mu + \delta(V))).$$

Q.E.D.

Theorem 6.3 *The iterates $(X^{(k)}, Z^{(k)})$ generated by Algorithm SDP(ϵ) converge to the analytic center (X^a, Z^a) superlinearly with order $2/(1+2^{-r})$. The duality gap $\mu^{(k)}$ converges to zero at the same rate.*

Proof. From (6.13) we see that for any $t \geq 0$ satisfying

$$\beta_{k-1} + \|D_X^p D_Z^p\|_F / \mu_k \leq (1-t)((1-t)\mu_k/\epsilon)^{2^{-r}},$$

there holds

$$\delta(V(t)) \leq ((1-t)\mu/\epsilon)^{2^{-r}}.$$

This implies using (6.6) and Lemma 6.6 that

$$\begin{aligned} (1-t_k)^{1+2^{-r}} &\leq (\beta_{k-1} + \|D_X^p D_Z^p\|_F / \mu_k)(\mu_k/\epsilon)^{-2^{-r}} \\ &= O(\mu_k^{1-2^{-r}}), \end{aligned}$$

so that

$$\mu_{k+1} = (1-t_k)\mu_k = O(\mu_k^{2/(1+2^{-r})}).$$

This shows that the duality gap converges to zero superlinearly with order $2/(1+2^{-r})$. It remains to prove that the iterates converge to the analytic center with the same order. Notice that

$$\begin{aligned}
\|X^{(k)} - X(\mu_k)\|_F^2 &= \operatorname{tr}\, (X^{(k)} - X(\mu_k)) L_d^{-\mathrm{H}} (L_d^{\mathrm{H}} L_d) L_d^{-1} (X^{(k)} - X(\mu_k)) \\
&\leq \|L_d^{\mathrm{H}} L_d\|_2 \cdot \operatorname{tr}\, (X^{(k)} - X(\mu_k)) L_d^{-\mathrm{H}} L_d^{-1} (X^{(k)} - X(\mu_k)) \\
&= \|L_d^{\mathrm{H}} L_d\|_2 \cdot \operatorname{tr}\, L_d^{-1} (X^{(k)} - X(\mu_k))(X^{(k)} - X(\mu_k)) L_d^{-\mathrm{H}} \\
&\leq \|L_d^{\mathrm{H}} L_d\|_2^2 \cdot \|L_d^{-1} (X^{(k)} - X(\mu_k)) L_d^{-\mathrm{H}}\|_F^2. \qquad (6.21)
\end{aligned}$$

However, using definition (6.14) and applying Lemma 6.3,

$$\|L_d^{\mathrm{H}} L_d\|_F = \|\bar{D}_{\mu_k}^{-1/2} L_{\mu_k}^{\mathrm{H}} L_{\mu_k} \bar{D}_{\mu_k}^{-1/2}\|_F = O(\|L_{\mu_k}^{\mathrm{H}} L_{\mu_k}\|_F).$$

Since $L_{\mu_k} L_{\mu_k}^{\mathrm{H}} = X(\mu_k)/\sqrt{\mu_k}$, it follows that

$$\|L_{\mu_k}^{\mathrm{H}} L_{\mu_k}\|_F^2 = \frac{\operatorname{tr}\, X(\mu_k)^2}{\mu_k} = \frac{1}{\mu_k} \|X(\mu_k)\|_F^2.$$

Using Lemma 5.1, we thus obtain

$$\|L_d^{\mathrm{H}} L_d\|_F = O(\frac{1}{\sqrt{\mu_k}}). \qquad (6.22)$$

Combining (6.21) and (6.22) with Lemma 6.2, we have

$$\|X^{(k)} - X(\mu_k)\|_F = O(\frac{1}{\sqrt{\mu_k}} \|L_d^{-1}(X^{(k)} - X(\mu_k)) L_d^{-\mathrm{H}}\|_F) = O(\delta(V^{(k)})) = O(\mu_k).$$

Hence, we obtain from Theorem 5.1 that

$$\|X^{(k)} - X^a\|_F = O(\mu_k).$$

Similarly, it can be shown that

$$\|Z^{(k)} - Z^a\|_F = O(\mu_k).$$

This shows that the iterates converge to the analytic center R-superlinearly, with the same order as μ_k converges to zero. **Q.E.D.**

6.3 Discussion

We have shown the global and superlinear convergence of the predictor-corrector algorithm SDP(ϵ), assuming only the existence of a strictly complementary solution pair. In particular, our results hold true for degenerate semidefinite programs, with possibly multiple optimal solutions. Such results were previously known only in the context of linear programming and linear complementarity problems, see Ye and Anstreicher [162]. Linear programs are always endowed with a strictly complementary solution pair, but this is not the case for linear complementarity problems. In fact, the Mizuno–Todd–Ye predictor–corrector method cannot be superlinearly convergent for linear complementarity problems if a strictly complementary solution does not exist, see Monteiro and Wright [103] and El-Bakry, Tapia and Zhang [33]. However, recent algorithms of Mizuno [94] and Sturm [140] also achieve superlinear convergence for monotone complementarity problems without strictly complementary solutions.

The local convergence analysis in this chapter is based on Theorem 5.1, which states that $\|X(\mu)-X^a\|+\|Z(\mu)-Z^a\| = O(\mu)$. By enforcing $\delta(V^{(k)}) \to 0$, the iterates "inherit" this property of the central path. For the generalization of the Mizuno-Todd-Ye predictor-corrector algorithm as discussed in Section 3.5.2, we do not enforce $\delta(V^{(k)}) \to 0$, and hence we cannot conclude superlinear convergence for it yet. In this respect, it will be interesting to study the asymptotic behavior of the corrector steps.

Superlinear convergence in the context of semidefinite programming was first studied under nondegeneracy assumptions by Kojima, Shida and Shindoh [76], and Potra and Sheng [120]. Then Luo, Sturm and Zhang [86] extended the superlinearity results to degenerate semidefinite programs. In this thesis, we have followed the approach of [86]. Potra and Sheng [122] adapted some techniques of Luo, Sturm and Zhang [86] to prove superlinear convergence with Helmberg–Rendl–Vanderbei–Wolkowicz directions, instead of Nesterov–Todd directions. Moreover, they show that instead of enforcing $\delta(V^k) \to 0$, it is sufficient to let $\|X^k Z^k\| = o(\sqrt{\mu_k})$. For nondegenerate semidefinite programs, the $XZ + ZX$ method of Alizadeh, Haeberly and Overton is asymptotically very effective in the centering step. This makes it possible to prove superlinear convergence without shrinking the parameter of the $\mathcal{N}_2$–neighborhood, see Kojima, Shida and Shindoh [78] and Potra and Sheng [121].

Chapter 7

Central Region Method

In this chapter, we discuss a modification of the standard path–following scheme that tends to speed up the global convergence. This modification, the central region method, generates iterates that do not really trace the central path, or at least not closely. In this way, it has a relatively large freedom of movement, and consequently the ability to take long steps. This makes it interesting to consider more sophisticated search directions. We propose a search direction that is built up in three phases, viz.

1. Initial centering,
2. Predictor,
3. Second order centrality corrector.

The first two components were already used in the largest step path–following method, with one important difference: the initial centering will now direct towards the central region instead of the central path. The third component combines a second order correction with a partial centering step. The second order correction is based on a second order Taylor expansion of the V-space movement. Unlike the initial centering, the partial centering step is directed towards the central path (instead of the central region). The central region algorithm that uses this sophisticated search direction will be called the centering–predictor–corrector algorithm.

As discussed in Chapter 3, path–following algorithms always use (Newton) directions towards solutions on the central path. The central region method of Sturm and Zhang [146] is an extension of the path–following method, where targets outside the central path are also used. Such targets are called *weighted centers*. Existence of these weighted centers for semidefinite programming was proved by Sturm and Zhang [144]. A different type of weighted centers was defined and analyzed by Monteiro and Pang [100], and De Klerk, Roos and Terlaky [69] proposed a non-convex weighted potential function for semidefinite programming. For linear programming, weighted centers were used in [58, 61, 93, 134, 146].

7.1 Weighted Centers

As in Chapter 3, we treat primal–dual interior point methods in the V-space framework. Namely, we consider a semidefinite program $\mathrm{CP}(b, c, \mathcal{A}, \mathcal{H}_+^{(\bar{n})})$ that is both primal and dual strongly feasible. With each interior solution pair (X, Z),

$$X \in (B + \mathcal{A}) \cap \mathcal{H}_{++}^{(\bar{n})}, \quad Z \in (C + \mathcal{A}^{\perp}) \cap \mathcal{H}_{++}^{(\bar{n})},$$

we associate a V–space solution, which is a positive diagonal matrix V satisfying

$$V = L_d^{-1} X L_d^{-\mathrm{H}} = L_d^{\mathrm{H}} Z L_d,$$

for some invertible matrix L_d. On the diagonal of V^2 are then the eigenvalues of the matrix XZ. We consider a trajectory of solution pairs $(\bar{X}(t), \bar{Z}(t))$ in the scaled space, i.e.

$$\bar{X}(t) \in (V + \mathcal{A}(L_d)) \cap \mathcal{H}_{++}, \quad \bar{Z}(t) \in (V + \mathcal{A}^{\perp}(L_d)) \cap \mathcal{H}_{++},$$

for t in a neighborhood of zero. We associate with this trajectory the symmetric primal–dual transformation

$$G(t) := D(\bar{X}(t), \bar{Z}(t))^{1/2}, \tag{7.1}$$

and we let $V(t) := G(t)\bar{Z}(t)G(t)$. By definition of $D(\cdot, \cdot)$, we have

$$V(t) = G(t)^{-1}\bar{X}(t)G(t)^{-1} = G(t)\bar{Z}(t)G(t), \tag{7.2}$$

see Lemma 3.1. We know from (3.8) that if $V(0) = V$ and $\bar{X}(t)$ and $\bar{Z}(t)$ are differentiable in $t = 0$, then

$$V^{[1]}(0) = \frac{1}{2}(\bar{Z}^{[1]}(0) + \bar{X}^{[1]}(0)),$$

where the superscript $^{[1]}$ is used to denote the first order derivative with respect to t. This implies that

$$\left\| V(t) - \frac{1}{2}(\bar{X}(t) + \bar{Z}(t)) \right\| = o(t). \tag{7.3}$$

The parameter t above has been used primarily to facilitate the study of derivatives. Below, we drop the argument t, and consider a fixed interior solution pair $(\bar{X}, \bar{Z})$, with a corresponding V-space solution $\bar{V}$, i.e.

$$G = D(\bar{X}, \bar{Z})^{1/2}, \quad \bar{V} = G\bar{Z}G = G^{-1}\bar{X}G^{-1}. \tag{7.4}$$

Based on (7.3), a natural estimate for $\bar{V}$ is V^{E}, defined as

$$V^{\mathrm{E}} := \frac{1}{2}(\bar{X} + \bar{Z}). \tag{7.5}$$

Notice that there is a one-to-one correspondence between V^{E} and $(\bar{X}, \bar{Z})$, namely

$$\bar{X} - V = 2P_{\mathcal{A}(L_d)}(V^{\mathrm{E}} - V), \quad \bar{Z} - V = 2P_{\mathcal{A}^{\perp}(L_d)}(V^{\mathrm{E}} - V). \tag{7.6}$$

This also shows that if $\bar{X} = \bar{Z}$, then $V^{\mathrm{E}} = V = \bar{V}$. With this in mind, we define a residual matrix R as

$$R := \frac{1}{2}(\bar{X} - \bar{Z}).$$

Below, we derive a bound for the error $\|\bar{V} - V^{\mathrm{E}}\|_F$ in terms of the residual $\|R\|_F$.

First, we notice the identities

$$\bar{X} = V^{\mathrm{E}} + R, \quad \bar{Z} = V^{\mathrm{E}} - R. \tag{7.7}$$

Since $\bar{X} \succ 0$ and $\bar{Z} \succ 0$, it follows that $\|(V^{\mathrm{E}})^{-1/2} R (V^{\mathrm{E}})^{-1/2}\|_2 < 1$. The quantity

$$\rho := \|(V^{\mathrm{E}})^{-1/2} R (V^{\mathrm{E}})^{-1/2}\|_2 \tag{7.8}$$

will play a crucial role in estimating $\|\bar{V} - V^{\mathrm{E}}\|_F$. The relation

$$(1+\rho)V^{\mathrm{E}} \succeq \bar{X} \succeq (1-\rho)V^{\mathrm{E}}, \quad (1+\rho)V^{\mathrm{E}} \succeq \bar{Z} \succeq (1-\rho)V^{\mathrm{E}}. \tag{7.9}$$

is an immediate consequence of (7.7). It is known from (7.6) that $(\bar{X} - V) \perp (\bar{Z} - V)$, which implies

$$\|R\|_F = \frac{1}{2}\|(\bar{X} - V) - (\bar{Z} - V)\|_F = \frac{1}{2}\|(\bar{X} - V) + (\bar{Z} - V)\|_F = \left\|V^{\mathrm{E}} - V\right\|_F. \tag{7.10}$$

We will now examine the difference between the V-space solution $\bar{V}$ and its approximation V^{E}.

Lemma 7.1 *There holds*

$$\left(1 - \frac{1}{2}\|I - G\|_2 \|I + G\|_2\right) \|V^E - \bar{V}\|_F \le \frac{1}{2}\|I - G\|_2 \|I + G\|_2 \|R\|_F.$$

Proof. First remark using (7.4) that

$$\bar{Z} - \bar{V} = \bar{Z} - G\bar{Z}G = P_{\mathcal{H}}\left((I - G)\bar{Z}(I + G)\right)$$

and

$$\bar{X} - \bar{V} = G\bar{V}G - \bar{V} = -P_{\mathcal{H}}\left((I - G)\bar{V}(I + G)\right).$$

Adding the above two relations yields

$$\frac{\bar{X}+\bar{Z}}{2}-\bar{V}=\frac{1}{2}P_{\mathcal{H}}\left((I-G)(\bar{Z}-\bar{V})(I+G)\right).$$

By substitution of (7.5) and (7.7) in the above identity, we have

$$V^{\mathrm{E}}-\bar{V}=\frac{1}{2}P_{\mathcal{H}}\left((I-G)(V^{\mathrm{E}}-R-\bar{V})(I+G)\right),$$

and therefore

$$\begin{aligned}\|V^{\mathrm{E}}-\bar{V}\|_F &= \frac{1}{2}\|P_{\mathcal{H}}\left((I-G)(V^{\mathrm{E}}-R-\bar{V})(I+G)\right)\|_F \\ &\leq \frac{1}{2}\|I-G\|_2(\|V^{\mathrm{E}}-\bar{V}\|_F+\|R\|_F)\|I+G\|_2.\end{aligned}$$

Rearranging terms, we get

$$\left(1-\frac{1}{2}\|I-G\|_2\|I+G\|_2\right)\|V^{\mathrm{E}}-\bar{V}\|_F\leq\frac{1}{2}\|I-G\|_2\|I+G\|_2\|R\|_F.$$

Q.E.D.

Based on the above lemma, it is natural to further bound $\left\|V^{\mathrm{E}}-\bar{V}\right\|_F$ by deriving a bound on $\|I-G\|_2$ in terms of ρ. Such a bound is given in the following lemma.

Lemma 7.2 *There holds*

$$\|I-G\|_2\leq\left(\frac{1+\rho}{1-\rho}\right)^{1/4}-1.$$

Proof. Remark from (7.9) and (7.4) that

$$(1-\rho)G^2V^{\mathrm{E}}G^2\preceq G^2\bar{Z}G^2=\bar{X}\preceq(1+\rho)V^{\mathrm{E}}.$$

Together with Lemma A.1, this implies that

$$\|G^2\|_2^2\leq\frac{1+\rho}{1-\rho}.$$

Using the primal-dual symmetry, it follows that also

$$\|G^{-2}\|_2^2\leq\frac{1+\rho}{1-\rho}.$$

Since G^2 is positive definite, we know that $\|G^2\|_2$ is the largest eigenvalue of G^2, and $1/\|G^{-2}\|_2$ is the smallest eigenvalue of G^2. Therefore, the above two estimations imply that

$$\frac{1-\rho}{1+\rho}I\preceq G^4\preceq\frac{1+\rho}{1-\rho}I,$$

and therefore,

$$\|G - I\|_2 \leq \left(\frac{1+\rho}{1-\rho}\right)^{1/4} - 1.$$

Q.E.D.

Based on Lemma 7.2, we immediately obtain a bound on $\|I - G\|_2 \|I + G\|_2$. In the following lemma, we work the resulting bound up into a neat expression.

Lemma 7.3 *There holds*

$$\|I - G\|_2 \|I + G\|_2 \leq \left(\sqrt{\frac{1+\rho}{1-\rho}} - 1\right) = \frac{2\rho}{(1-\rho) + \sqrt{1-\rho^2}}.$$

Proof. Using the triangle inequality, we have

$$\|I - G\|_2 \|I + G\|_2 \leq \|I - G\|_2 (2\|I\|_2 + \|I - G\|_2) = \|I - G\|_2 (2 + \|I - G\|_2).$$

This implies, using Lemma 7.2, that

$$\begin{aligned} \|I - G\|_2 \|I + G\|_2 &\leq \left(\frac{(1+\rho)^{1/4}}{(1-\rho)^{1/4}} - 1\right)\left(\frac{(1+\rho)^{1/4}}{(1-\rho)^{1/4}} + 1\right) \\ &= \frac{\sqrt{1+\rho}}{\sqrt{1-\rho}} - 1 \\ &= \frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{\sqrt{1-\rho}}. \end{aligned}$$

Multiplying numerator and denominator both by $\sqrt{1+\rho} + \sqrt{1-\rho}$, we obtain

$$\|I - G\|_2 \|I + G\|_2 \leq \frac{2\rho}{\sqrt{1-\rho^2} + 1 - \rho}.$$

Q.E.D.

Applying Lemma 7.3 to the bound in Lemma 7.1, we obtain the following estimate for the error term $\|V^{\mathrm{E}} - \bar{V}\|_F$.

Lemma 7.4 *If $\rho < 4/5$, then*

$$\|V^E - \bar{V}\|_F \leq \frac{\rho \|R\|_F}{(1 - 2\rho) + \sqrt{1-\rho^2}},$$

where $\rho = \|(V^E)^{-1/2} R (V^E)^{-1/2}\|_2$.

Proof. Combining Lemma 7.1 and Lemma 7.3, we get

$$\begin{aligned}\frac{\rho}{(1-\rho)+\sqrt{1-\rho^2}}\|R\|_F &\geq \left(1-\frac{\rho}{(1-\rho)+\sqrt{1-\rho^2}}\right)\|V^{\mathrm{E}}-\bar{V}\|_F \\ &= \frac{(1-2\rho)+\sqrt{1-\rho^2}}{(1-\rho)+\sqrt{1-\rho^2}}\|V^{\mathrm{E}}-\bar{V}\|_F.\end{aligned}$$

Since $\rho < 4/5$, we know that $(1-2\rho)+\sqrt{1-\rho^2} > 0$, and we obtain that

$$\|V^{\mathrm{E}}-\bar{V}\|_F \leq \frac{\rho\|R\|_F}{(1-2\rho)+\sqrt{1-\rho^2}}.$$

Q.E.D.

Recall from (7.10) that $\|R\|_F = \|V^{\mathrm{E}} - V\|_F$. Moreover,

$$\rho = \|(V^{\mathrm{E}})^{-1/2}R(V^{\mathrm{E}})^{-1/2}\|_2 \leq \|(V^{\mathrm{E}})^{-1/2}\|_2^2\|R\|_2 = \frac{\|R\|_2}{\lambda_{\min}(V^{\mathrm{E}})}. \tag{7.11}$$

Hence, it follows from Lemma 7.4 that if we take a unit Newton step from V to steer towards a target V^{E} in the V-space, we obtain a V-space solution $\bar{V}$ with $\|V^{\mathrm{E}}-\bar{V}\| = O(\|V^{\mathrm{E}}-V\|^2)$, provided that V^{E} is close enough to V. The following corollary makes this observation more precise.

Corollary 7.1 *If* $\|R\|_2 < \lambda_{\min}(V^E)/\sqrt{5}$, *then*

$$\|V^E-\bar{V}\|_F \leq \frac{\|R\|_2\|R\|_F}{\lambda_{\min}(V^E)} \leq \frac{\|V^E-V\|_F^2}{\lambda_{\min}(V^E)}.$$

It is interesting to compare Corollary 7.1 with Lemma 3.8, which states for the direction $2(D_X+D_Z) = \mu V^{-1} - V$ that $\delta(\bar{V}) \leq \delta(V)^2$ if $\delta(V) \leq 1/2$, where

$$\delta(V) = \left\|I - \frac{1}{\mu}V^2\right\|_F \geq \frac{1+\lambda_{\min}(V)/\sqrt{\mu}}{\sqrt{\mu}}\|\sqrt{\mu}I - V\|_F.$$

Recalling that $V^2 \geq (1-\delta(V))\mu I$, we obtain from the above inequality that

$$\|\sqrt{\mu}I - V\|_F \leq \frac{\delta(V)\sqrt{\mu}}{1+\sqrt{1-\delta(V)}}.$$

Hence, the region of quadratic convergence to the target $V^{\mathrm{E}} = \sqrt{\mu}I$ in Corollary 7.1 is wider than in Lemma 3.8.

7.2 Central Region and its Neighborhood

Although it does not have the best (proved) worst case behavior, the long step path–following method with the $\mathcal{N}_\infty^-$–neighborhood (see Chapter 3) has received much attention, because it is polynomial and it allows for long steps, which is a prerequisite for practical efficiency. For medium sized linear programming problems, the long step path following method works indeed reasonably well, see Lustig, Marsten and Shanno [88]. For semidefinite programming however, the actual performance of the long step path–following method is rather disappointing. This can be explained from the large distance between the iterates and the Newton targets on the central path, which can aversely affect the accuracy of the Newton direction. A good remedy for this problem is the *central region method* of Sturm and Zhang [146].

In [146], Sturm and Zhang proposed a generic v-space method for linear programming, where targets are chosen in a wide central region that includes the central path, and the iterates are restricted to an — even wider — neighborhood of this region. It was demonstrated in [146] that it is efficient to use targets in the central region that are not necessarily on the central path. In this respect, the central region method resembles the target following method of Jansen, Roos, Terlaky and Vial [61]. However, the iterates of the target following method trace a certain target sequence, whereas the iterates of the central region method greedily pursue large reductions in the duality gap, without necessarily tracing their Newton targets.

In Chapter 3, the central path has been defined in terms of V–space solutions, as the half–line of positive multiples of the identity matrix. We have seen that in a neighborhood of the central path, we can efficiently approach the so–called μ–centers, by using Newton directions. This is the basis of the polynomiality results for path–following methods.

However, it is clear from the derivations in Section 7.1 that we do not have to restrict to targets on the central path. In particular, Corollary 7.1 shows that there is a provably large region of quadratic convergence associated with a weighted center V^{E} if $\lambda_{\min}(V^{\mathrm{E}})$ is relatively large. This motivates the definition of a *central region.*

The central region is defined as the closed pointed convex cone $\mathrm{CR}(\theta)$,

$$\mathrm{CR}(\theta) := \{Y \in \mathcal{H}_+^{(\bar{n})} \mid \lambda_{\min}(Y) \succeq \theta \|Y\|_F / \sqrt{\bar{n}}\},$$

where $\theta \in [0,1]$ is a fixed parameter.

Remark that $\mathrm{CR}(\theta)$ is in fact the wide neighborhood $\mathcal{N}_\infty^-(1-\theta^2)$ of the central path. However, we will not use $\mathrm{CR}(\theta)$ as a neighborhood, but as a region from which we choose targets in centering steps. Extremal choices for θ are $\theta = 0$, in which case we obtain the positive semidefinite cone $\mathrm{CR}(0) = \mathcal{H}_+^{(\bar{n})}$, and $\theta = 1$, in which case we obtain the central path $\mathrm{CR}(1) = \{Y \in \mathcal{H} \mid Y = \sqrt{\operatorname{tr} Y^2/\bar{n}}I\}$. We will restrict to $\theta > 0$ in the sequel.

In Corollary 7.1, we have established a region of quadratic convergence for weighted centers in CR(θ). The union of these regions will serve as a neighborhood of CR(θ), in which the iterates of the central region method will be generated. More precisely, for fixed $\theta \in (0,1]$ and $\beta \in (0,1)$ we define a neighborhood of the central region as

$$\mathcal{N}(\theta,\beta) := \left\{ Y \in \mathcal{H}_+^{(\bar{n})} \,\middle|\, \mathrm{dist}_F(Y,\ \mathrm{CR}(\theta)) \leq \frac{\beta\theta \, \|Y\|_F}{\sqrt{\bar{n}+\beta^2\theta^2}} \right\}.$$

Because CR(θ) is a closed convex set, there exists for given Hermitian Y a minimal norm projection Y^θ onto the central region, i.e.

$$Y^\theta := \arg\min_X \{ \|X - Y\|_F \mid X \in \ \mathrm{CR}(\theta)\}.$$

Now, we can write

$$\mathcal{N}(\theta,\beta) = \left\{ Y \in \mathcal{H}_+^{(\bar{n})} \,\middle|\, \left\|Y^\theta - Y\right\|_F \leq \frac{\beta\theta \, \|Y\|_F}{\sqrt{\bar{n}+\beta^2\theta^2}} \right\}. \tag{7.12}$$

As is well known, minimal norm projections on closed convex cones satisfy the normal relation $Y^\theta \perp (Y^\theta - Y)$. Hence, (7.12) states for $Y \in \mathcal{N}(\theta,\beta)$ that the angle ϕ between Y and Y^θ satisfies

$$\sin\phi = \frac{\|Y^\theta - Y\|_F}{\|Y\|_F} \leq \frac{\beta\theta}{\sqrt{\bar{n}+\beta^2\theta^2}},$$

or equivalently,

$$\tan\phi = \frac{\|Y^\theta - Y\|_F}{\|Y^\theta\|_F} = \frac{\sin\phi}{\sqrt{1-(\sin\phi)^2}} \leq \frac{\beta\theta}{\sqrt{\bar{n}}}.$$

This implies that

$$\mathcal{N}(\theta,\beta) = \{ Y \in \mathcal{H}_+^{(\bar{n})} \mid \sqrt{\bar{n}} \left\|Y^\theta - Y\right\|_F \leq \beta\theta \left\|Y^\theta\right\|_F \}. \tag{7.13}$$

We remark that of all Hermitian matrices in CR(θ), the projection Y^θ has the smallest angle with Y, see Lemma 2.1 in [146]. Based on this interpretation, the following result is straightforward.

Lemma 7.5 *Suppose that $Y \in$ CR(θ). Then*

$$\{X \in \mathcal{H} \mid \sqrt{\bar{n}} \, \|X - Y\|_F \leq \beta\theta \, \|Y\|_F\} \subseteq \mathcal{N}(\theta,\beta).$$

Since we want to use the central region concept for V–space solutions, we are interested in the projection V^θ. We will now show that V^θ is a diagonal matrix.

Lemma 7.6 *If V is a nonnegative diagonal matrix, then so is V^θ.*

Proof. Let $\mathcal{D}$ denote the linear space of real diagonal matrices, and let $\mathcal{D}^{\perp}$ denote its orthogonal complement in $\mathcal{H}$, with respect to the standard real valued inner product $X \bullet Y = \operatorname{tr} YX$. Then $\mathcal{D}^{\perp}$ is the space of zero-diagonal Hermitian matrices. Since $V^{\theta} = (P_{\mathcal{D}} V^{\theta}) + P_{\mathcal{D}^{\perp}} V^{\theta}$, we have

$$(P_{\mathcal{D}} V^{\theta})_{ii} = V^{\theta}_{ii} \quad \text{for all } i \in \{1, 2, \ldots, \bar{n}\}.$$

Therefore,

$$\lambda_{\min}(P_{\mathcal{D}} V^{\theta}) = \min_{1 \leq i \leq \bar{n}} V^{\theta}_{ii} \geq \lambda_{\min}(V^{\theta}),$$

where the inequality is easily established from the spectral decomposition of V^{θ}, or immediately from the Rayleigh–Ritz characterization. Since $V^{\theta} \in \operatorname{CR}(\theta)$, we further obtain

$$\lambda_{\min}(P_{\mathcal{D}} V^{\theta}) \geq \frac{\theta \left\| V^{\theta} \right\|_F}{\sqrt{\bar{n}}} \geq \frac{\theta \left\| P_{\mathcal{D}} V^{\theta} \right\|_F}{\sqrt{\bar{n}}}.$$

Hence, $P_{\mathcal{D}} V^{\theta} \in \operatorname{CR}(\theta)$, whereas

$$\left\| (P_{\mathcal{D}} V^{\theta}) - V \right\|_F^2 = \left\| V^{\theta} - V \right\|_F^2 - \left\| P_{\mathcal{D}^{\perp}} V^{\theta} \right\|_F^2,$$

which implies that $V^{\theta} \in \mathcal{D}$. **Q.E.D.**

Since V^{θ} is known to be diagonal, we can compute V^{θ} from V in $\mathbf{O}(\bar{n})$ operations, using the procedure that was described in the appendix of [146]. Remark that $\operatorname{CR}(\theta)$ is invariant under rotations, i.e.

$$(Q \otimes_{\mathrm{H}} Q) \operatorname{CR}(\theta) = \operatorname{CR}(\theta) \quad \text{for unitary matrices } Q.$$

For general positive semidefinite Y, we can thus find Y^{θ} using the spectral decomposition $Y = Q \Lambda_Y Q^{\mathrm{H}}$, viz. $Y^{\theta} = Q \Lambda_Y^{\theta} Q^{\mathrm{H}}$.

7.3 Generic Central Region Method

As stipulated in Section 7.2, each iteration of the central region method starts with a given iterate $V \in \mathcal{N}(\theta, \beta)$. The parameters θ and β must be chosen as universal constants in the following domain:

$$\theta \in (0, 1], \quad \beta \in (0, \frac{1}{\sqrt{5}}].$$

Exactly as in Section 7.1, we consider $(\bar{X}(t), \bar{Z}(t))$ trajectories in the primal–dual transformed space, i.e.

$$\bar{X}(t) \in (V + \mathcal{A}(L_d)) \cap \mathcal{H}_{++}, \quad \bar{Z}(t) \in (V + \mathcal{A}^{\perp}(L_d)) \cap \mathcal{H}_{++},$$

and we define $G(t)$ as in (7.1), so that

$$V(t) = G(t)^{-1}\bar{X}(t)G(t)^{-1} = G(t)\bar{Z}(t)G(t), \tag{7.14}$$

see (7.2). In the central region method, $\bar{X}(t)$ and $\bar{Z}(t)$ are affine functions of the step length t, viz.

$$\bar{X}(t) = \bar{X}(0) + 2tD_X, \quad \bar{Z}(t) = \bar{Z}(0) + 2tD_Z. \tag{7.15}$$

Starting from the V-space solution V, we compute $(\bar{X}(0), \bar{Z}(0))$ by an initial centering step towards V^θ, the projection of V onto $\mathrm{CR}(\theta)$. Specifically, we solve $(\bar{X}(0), \bar{Z}(0))$ from the linear equations

$$\begin{cases} \frac{1}{2}(\bar{X}(0) + \bar{Z}(0)) = V^\theta \\ \bar{X}(0) - V \in \mathcal{A}(L_d), \quad \bar{Z}(0) - V \in \mathcal{A}^\perp(L_d). \end{cases} \tag{7.16}$$

As a first order estimation of $V(t)$, we use

$$V^{\mathrm{E}}(t) = \frac{1}{2}(\bar{X}(t) + \bar{Z}(t)) = V^\theta + t(D_X + D_Z), \tag{7.17}$$

and we define the residual $R(t)$ as

$$R(t) = \frac{1}{2}(\bar{X}(t) - \bar{Z}(t)). \tag{7.18}$$

Since $V^{\mathrm{E}}(0) = V^\theta$, it follows from Corollary 7.1 and the fact that $V \in \mathcal{N}(\theta, \beta)$ with $\beta \leq 1/\sqrt{5}$ that $V(0) \in \mathcal{N}(\theta, \beta^2)$, which implies that the centered iterate $V(0)$ is contained in the interior of $\mathcal{N}(\theta, \beta)$. The step length towards the boundary of the neighborhood $\mathcal{N}(\theta, \beta)$ is denoted by $t(\theta, \beta)$,

$$t(\theta, \beta) := \max\{\bar{t} \mid V(t) \in \mathcal{N}(\theta, \beta) \text{ for } 0 \leq t \leq \bar{t}\}, \tag{7.19}$$

where $\beta \in (0, 1/\sqrt{5}]$.

The generic central region algorithm is defined below. Specific choices for the search direction $D_X + D_Z$ lead to different algorithms in this generic class.

Algorithm 7.1

Parameters: $\epsilon > 0$, $0 < \beta \leq 1/\sqrt{5}$, $0 < \theta \leq 1$ *and* $\theta\beta/\sqrt{n} < \gamma \leq 1$.

Initial solution pair $(X^{(0)}, Z^{(0)})$ *in* $\mathcal{N}(\theta, \beta)$.

Step 0 Initialization. *Set* $k = 0$, *and compute* $(L_d^{(0)}, V^{(0)})$ *such that*

$$V^{(0)} = (L_d^{(0)})^{-1}X^{(0)}(L_d^{(0)})^{-H} = (L_d^{(0)})^H Z^{(0)} L_d^{(0)}.$$

Step 1 Stopping rule. *If* $\left\|V^{(k)}\right\|_F^2 < \epsilon$ *then stop.*

Step 2 Initial centering. *Let* $L_d = L_d^{(k)}$ *and* $V = V^{(k)}$. *Solve* $(\bar{X}(0), \bar{Z}(0))$ *from (7.16).*

Step 3 Search direction. *Compute a nonzero search direction* $D_X + D_Z$ *such that*

$$V^\theta \bullet (D_X + D_Z) \leq -\gamma \left\|V^\theta\right\|_F \|D_X + D_Z\|_F \,, \tag{7.20}$$

and decompose it into $D_X \in \mathcal{A}(L_d)$ *and* $D_Z \in \mathcal{A}^\perp(L_d)$.

Step 4 Line search. *Compute a step length* $t \in [t(\theta,\beta)/2, t(\theta,\beta)]$, *and let*

$$X^{(k+1)} = L_d \bar{X}(t) L_d^H, \quad Z^{(k+1)} = L_d^{-H} \bar{Z}(t) L_d^{-1}.$$

Step 5 Primal–dual transformation. *Compute* $(L_d^{(k+1)}, V^{(k+1)})$ *such that*

$$V^{(k+1)} = (L_d^{(k+1)})^{-1} X^{(k+1)} (L_d^{(k+1)})^{-H} = (L_d^{(k+1)})^H Z^{(k+1)} L_d^{(k+1)}.$$

Set $k = k+1$ *and return to Step 1.*

Remark 7.1 *A possible choice for the search direction in Algorithm 7.1 is* $D_X + D_Z = -V^\theta$. *In the case that* $\theta = 1$, *this choice leads to a largest step path–following algorithm. This largest step algorithm is slightly different from the largest step algorithm of Chapter 3, since the search direction is now based on the linearization of* $V(t)$, *whereas the search directions in Chapter 3 are all based on the linearization of* $V(t)^2$.

Remark 7.2 *Algorithm 7.1 can be initialized with the identity solution* $X^0 = Z^0 = I$, *by using the self–dual embedding technique of Chapter 4.*

In the sequel of this section, we analyze the convergence of Algorithm 7.1.

Lemma 7.7 *In each iteration of Algorithm 7.1, it holds that*

$$t(\theta,\beta) \geq \frac{\beta\theta \left\|V^\theta\right\|_F}{10\sqrt{n}\,\|D_X + D_Z\|_F}.$$

Proof. We will first estimate $\|R(t)\|_F$ and $\lambda_{\min}(V^{\mathrm{E}}(t))$, so that we can apply Lemma 7.4. The result will then follow using Lemma 7.5, with $Y = V^\theta$.

Consider an arbitrary step length $t > 0$. Using (7.10) and (7.17), we have

$$\|R(t)\|_F = \|V^{\mathrm{E}}(t) - V\|_F = \left\|t(D_X + D_Z) + (V^\theta - V)\right\|_F.$$

Because $V \in \mathcal{N}(\theta, \beta)$, this implies using (7.13) that

$$\|R(t)\|_F \leq t\,\|D_X + D_Z\|_F + \frac{\beta\theta\left\|V^\theta\right\|_F}{\sqrt{\bar{n}}}. \tag{7.21}$$

Notice from the identity $V^{\mathrm{E}}(t) = V^\theta + t(D_X + D_Z)$ that

$$\lambda_{\min}(V^{\mathrm{E}}(t)) \geq \lambda_{\min}(V^\theta) - t\,\|D_X + D_Z\|_2 \geq \frac{\theta\left\|V^\theta\right\|_F}{\sqrt{\bar{n}}} - t\,\|D_X + D_Z\|_2\,, \tag{7.22}$$

where we used the fact that $V^\theta \in \mathrm{CR}(\theta)$.

Before we combine the inequalities (7.21)–(7.22) with Lemma 7.4, it is convenient to introduce the quantity

$$\sigma := \frac{\sqrt{\bar{n}}\,\|D_X + D_Z\|_F}{\theta\,\|V^\theta\|_F},$$

so that we can rewrite (7.21)–(7.22) as

$$\|R(t)\|_F \leq (t\sigma + \beta)\frac{\theta\|V^\theta\|_F}{\sqrt{\bar{n}}}, \quad \lambda_{\min}(V^{\mathrm{E}}(t)) \geq (1 - t\sigma)\frac{\theta\|V^\theta\|_F}{\sqrt{\bar{n}}}. \tag{7.23}$$

In the sequel of the proof, we consider a step length $0 < t < (4 - 5\beta)/(9\sigma)$. For such t, we have from (7.23) that $\|R(t)\|_2 < (4(1-\beta)/[5(1+\beta)])\lambda_{\min}(V^{\mathrm{E}}(t))$, and hence $\bar{X}(t) \succ 0$ and $\bar{Z}(t) \succ 0$, see (7.7). The error term $\left\|V(t) - V^{\mathrm{E}}(t)\right\|_F$ can now be estimated by Lemma 7.4 and relation (7.11), yielding

$$\|V(t) - V^{\mathrm{E}}(t)\|_F \leq \frac{\|R(t)\|_F^2}{\lambda_{\min}(V^{\mathrm{E}}(t)) - 2\|R(t)\|_2 + \sqrt{\lambda_{\min}(V^{\mathrm{E}}(t))^2 - \|R(t)\|_2^2}}. \tag{7.24}$$

Combining this with (7.23) yields

$$\sqrt{\bar{n}}\left\|V(t) - V^{\mathrm{E}}(t)\right\|_F \leq \frac{(t\sigma + \beta)^2\theta\|V^\theta\|_F}{(1 - t\sigma) - 2(t\sigma + \beta) + \sqrt{(1 - t\sigma)^2 - (t\sigma + \beta)^2}} \tag{7.25}$$

Applying the triangle inequality, we have

$$\begin{aligned} \left\|V(t) - V^\theta\right\|_F &\leq \left\|V(t) - V^{\mathrm{E}}(t)\right\|_F + \left\|V^{\mathrm{E}}(t) - V^\theta\right\|_F \\ &= \left\|V(t) - V^{\mathrm{E}}(t)\right\|_F + t\,\|D_X + D_Z\|_F\,, \end{aligned} \tag{7.26}$$

where we used the identity (7.17). Together with (7.25) and the definition of σ, this implies that

$$\begin{aligned} \frac{\sqrt{\bar{n}}\left\|V(t) - V^\theta\right\|_F}{\beta\theta\,\|V^\theta\|_F} \leq\ & \frac{(1 + t\sigma/\beta)^2}{(\beta^{-1} - t\sigma/\beta) - 2(1 + t\sigma/\beta) + \sqrt{(\beta^{-1} - t\sigma/\beta)^2 - (1 + t\sigma/\beta)^2}} \\ & + \frac{t\sigma}{\beta}. \end{aligned}$$

Since $\beta \le 1/\sqrt{5}$, we get for all $0 \le t \le \beta/(10\sigma)$ that

$$\frac{\sqrt{\bar{n}}\left\|V(t) - V^\theta\right\|_F}{\beta\theta\left\|V^\theta\right\|_F} < 1.$$

Using Lemma 7.5, the above relation implies $V(t) \in \mathcal{N}(\theta, \beta)$. Therefore, $t(\theta, \beta)$ must be larger than $\beta/(10\sigma)$. **Q.E.D.**

The above lemma establishes a lower bound on the step length per iteration. We will now estimate the reduction in the duality gap, as a function of the step length.

Lemma 7.8 *The duality gap $\|V(t)\|_F^2$ is monotonically decreasing in t. Moreover, there holds*

$$\|V(t)\|_F^2 - \|V\|_F^2 \le 2tV^\theta \bullet (D_X + D_Z) + \frac{t^2}{2}\|D_X + D_Z\|_F^2$$

for all feasible t.

Proof. Using the fact that $(\bar{X}(t) - V) \perp (\bar{Z}(t) - V)$, we have

$$\|V(t)\|_F^2 = \bar{X}(t) \bullet \bar{Z}(t) = \|V\|_F^2 + 2V \bullet (V^{\mathrm{E}}(t) - V), \tag{7.27}$$

so that, using (7.17),

$$\begin{aligned} \|V(t)\|_F^2 - \|V\|_F^2 &= 2V \bullet (V^\theta + t(D_X + D_Z) - V) \\ &= 2tV^\theta \bullet (D_X + D_Z) + 2t(V - V^\theta) \bullet (D_X + D_Z) + 2V \bullet (V^\theta - V). \end{aligned}$$

As $V^\theta \perp (V^\theta - V)$, it holds that $V \bullet (V^\theta - V) = -\left\|V^\theta - V\right\|_F^2$. Therefore,

$$\frac{\|V(t)\|_F^2 - \|V\|_F^2}{2} \le tV^\theta \bullet (D_X + D_Z) + t\left\|V^\theta - V\right\|_F \|D_X + D_Z\|_F - \left\|V^\theta - V\right\|_F^2.$$

Using (7.20) and the fact that $V \in \mathcal{N}(\theta, \beta)$, it is now clear that $\|V(t)\|_F^2$ is a monotone decreasing function of t. Furthermore, we have

$$\frac{\|V(t)\|_F^2 - \|V\|_F^2}{2} \le tV^\theta \bullet (D_X + D_Z) + \frac{t^2}{4}\|D_X + D_Z\|_F^2.$$

Q.E.D.

Combining Lemma 7.7 and 7.8, we conclude that Algorithm 7.1 is polynomially convergent.

Theorem 7.1 *Algorithm 7.1 computes an ϵ–optimal solution in*

$$\mathbf{O}\left(\frac{\sqrt{\bar{n}}}{\beta\theta\gamma}\log\frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}\right)$$

iterations.

Proof. We will prove the lemma by showing that Algorithm 7.1 has a linear reduction rate of $1 - 1/\mathbf{O}(\sqrt{\bar{n}}/(\beta\theta\gamma))$. We know from Lemma 7.8 that the reduction in duality gap per iteration is at least

$$\begin{aligned} \|V\|_F^2 - \|V(t)\|_F^2 &\geq -2tV^\theta \bullet (D_X + D_Z) - \frac{t^2}{2}\|D_X + D_Z\|_F^2 \\ &\geq t(2\gamma - \frac{t\|D_X + D_Z\|_F}{2\|V^\theta\|_F})\left\|V^\theta\right\|_F \|D_X + D_Z\|_F, \end{aligned} \tag{7.28}$$

where we used (7.20). As shown in Lemma 7.7, the step length t is at least

$$t \geq \frac{1}{2}t(\theta,\beta) \geq \frac{\beta\theta\left\|V^\theta\right\|_F}{20\sqrt{\bar{n}}\|D_X + D_Z\|_F}. \tag{7.29}$$

Since the duality gap is monotonically decreasing in t, it follows from (7.28)–(7.29) that the reduction in duality gap is at least $\beta\theta\gamma\left\|V^\theta\right\|_F^2/\mathbf{O}(\sqrt{\bar{n}})$. Moreover, $\left\|V^\theta\right\|_F \geq (1 - \beta\theta/\sqrt{\bar{n}})\|V\|_F$ because $V \in \mathcal{N}(\theta,\beta)$. It thus follows that

$$\frac{\|V(t)\|_F^2}{\|V\|_F^2} \leq 1 - \frac{1}{\mathbf{O}(\sqrt{\bar{n}}/(\beta\theta\gamma))},$$

which completes the proof. **Q.E.D.**

7.4 Second Order Search Direction

In the preceding, we have based our search directions on the first order approximation $V^{\mathrm{E}}(t)$ of $V(t)$. In this section, we proceed by computing the second order Taylor expansion of $V(t)$. We assume the same setting as in Section 3.2, where the first order approximation was derived. In particular, we assume that $\bar{X}(t)$ and $\bar{Z}(t)$ are smooth trajectories, defined for t in a neighborhood zero, and that $\bar{X}(0) = \bar{Z}(0) = V$. We should be aware of a slight inconsistency here with (7.15). This inconsistency is due the fact that in Algorithm 7.1, we dispense with updating primal–dual transformations after initial centering steps.

We let $G(t) = D(\bar{X}(t), \bar{Z}(t))^{1/2}$, so that $G(t)$ is positive definite. Recall from (3.5) that

$$\begin{aligned} V^{[1]}(t) &= P_{\mathcal{H}}([G^{[1]}(t)G(t)^{-1} - G(t)^{-1}G^{[1]}(t)]V(t)) \\ &\quad + \frac{1}{2}G(t)\bar{Z}^{[1]}(t)G(t) + \frac{1}{2}G(t)^{-1}\bar{X}^{[1]}(t)G(t)^{-1}, \end{aligned}$$

and from (3.8),

$$V^{[1]}(0) = \frac{1}{2}(\bar{Z}^{[1]}(0) + \bar{X}^{[1]}(0)), \tag{7.30}$$

where we use the convention that superscripts [1] and [2] denote first and second order derivatives.

The relation $G(0) = I$ implies that

$$P_{\mathcal{H}^\perp}(G^{[1]}(0)G(0)^{-1}) = 0, \quad \frac{\mathrm{d}}{\mathrm{d}\,t}P_{\mathcal{H}^\perp}\left(G^{[1]}(t)G(t)^{-1}\right)|_{t=0} = 0,$$

so that

$$\begin{aligned} V^{[2]}(0) &= \frac{\mathrm{d}}{\mathrm{d}\,t}\left(\frac{1}{2}G(t)\bar{Z}^{[1]}(t)G(t) + \frac{1}{2}G(t)^{-1}\bar{X}^{[1]}(t)G(t)^{-1}\right)|_{t=0} \\ &= P_{\mathcal{H}}\left(G^{[1]}(0)(\bar{Z}^{[1]}(0) - \bar{X}^{[1]}(0))\right) + \frac{1}{2}(\bar{Z}^{[2]}(0) + \bar{X}^{[2]}(0)). \end{aligned} \tag{7.31}$$

The identities (7.30) and (7.31) show how we can steer the trajectory $V(t)$. Namely, we have

$$V(t) = V + tV^{[1]}(0) + \frac{t^2}{2}V^{[2]}(0) + o(t^2).$$

After specifying the desired values for $V^{[1]}(0)$ and $V^{[2]}(0)$, we can solve $(\bar{X}^{[1]}(0), \bar{Z}^{[1]}(0))$ and $(\bar{X}^{[2]}(0), \bar{Z}^{[2]}(0))$ from

$$\left\{ \begin{array}{c} \bar{Z}^{[1]}(0) + \bar{X}^{[1]}(0) = 2V^{[1]}(0) \\ \bar{X}^{[1]}(0) \in \mathcal{A}(L_d), \quad \bar{Z}^{[1]}(0) \in \mathcal{A}^\perp(L_d), \end{array} \right. \tag{7.32}$$

and

$$\left\{ \begin{array}{c} \bar{Z}^{[2]}(0) + \bar{X}^{[2]}(0) = 2V^{[2]}(0) + 2P_{\mathcal{H}}(G^{[1]}(0)(\bar{X}^{[1]}(0) - \bar{Z}^{[1]}(0))) \\ \bar{X}^{[2]}(0) \in \mathcal{A}(L_d), \quad \bar{Z}^{[2]}(0) \in \mathcal{A}^\perp(L_d). \end{array} \right. \tag{7.33}$$

Relation (7.32) shows that the first order step $(\bar{X}^{[1]}(0), \bar{Z}^{[1]}(0))$ is an orthogonal decomposition of the matrix $V^{[1]}(0)$. We will show in the next paragraph that given this first order step, we can easily compute $G^{[1]}(0)$. After specifying a desired matrix $V^{[2]}(0)$, the second order step $(\bar{X}^{[2]}(0), \bar{Z}^{[2]}(0))$ follows from (7.33), again as an orthogonal decomposition.

In order to obtain an expression for $G^{[1]}(t)$, we remark from (7.14) that

$$G(t)\bar{Z}(t)G(t) - G(t)^{-1}\bar{X}(t)G(t)^{-1} = 0.$$

By implicit differentiation we get

$$\begin{aligned} 0 &= 2P_{\mathcal{H}}\left(G^{[1]}(t)\bar{Z}(t)G(t)\right) + G(t)\bar{Z}^{[1]}(t)G(t) \\ &\quad +2P_{\mathcal{H}}\left(G(t)^{-1}G^{[1]}(t)G(t)^{-1}\bar{X}(t)G(t)^{-1}\right) - G(t)^{-1}\bar{X}^{[1]}(t)G(t)^{-1}. \end{aligned}$$

Since $G(0) = I$ and $\bar{X}(0) = \bar{Z}(0) = V$, we have

$$P_{\mathcal{H}}(VG^{[1]}(0)) = \frac{1}{4}(\bar{X}^{[1]}(0) - \bar{Z}^{[1]}(0)). \tag{7.34}$$

Since V is a positive diagonal matrix, we can compute $G^{[1]}(0)$ from the above relation in $\mathbf{O}(\bar{n}^2)$ operations, see Lemma 3.2. Notice the similarities between (3.46) and (7.34).

7.5 The Centering–Predictor–Corrector Algorithm

We will now describe an iteration of the wide region method with second order correction. In this algorithm, the search direction and the step length are both determined adaptively. The adaptiveness makes the algorithm remarkably efficient in practice, but it also makes its theoretical treatment more involved. There are basically three stages in determining a step, viz.

- Initial centering towards the central region,
- Use of a predictor for determining a sensible target on the central path,
- Second order search direction towards this target.

The initial centering step is part of the generic central region algorithm as discussed in Section 7.3. We will now describe the predictor–corrector technique for determining the direction $D_X + D_Z$ in Algorithm 7.1.

7.5.1 The Predictor

In the framework of Algorithm 7.1, the trajectories $\bar{X}(t)$ and $\bar{Z}(t)$ are affine functions of the step length t, as given by (7.15). After the initial centering step, the V–space solution is close to its estimate $V^{\mathrm{E}}(0) = V^{\theta}$. In the predictor step, we greedily try to approximate the optimal solution, using the direction $D_X^P + D_Z^P = -V^{\theta}$. After decomposing $-V^{\theta}$ into the orthogonal components $D_X^P \in \mathcal{A}(L_d)$ and $D_Z^P \in \mathcal{A}^{\perp}(L_d)$, we calculate the maximal step length towards the boundary,

$$t_P^* := \max\{t \mid \bar{X}(0) + 2tD_X^P \succeq 0,\ \bar{Z}(0) + 2tD_Z^P \succeq 0\}.$$

By taking a step of length $t \in (0, t_P^*)$ along the predictor direction, we obtain a primal–dual pair for which the V–space solution approximates $V^{\theta} + t(D_X^P + D_Z^P) = (1-t)V^{\theta}$. Based on this first order approximation, we may expect that the resulting duality gap will be close to

$$\left\|(1-t)V^{\theta}\right\|_F^2 = (1 - 2t + t^2)\left\|V^{\theta}\right\|_F^2.$$

However, we know from (7.27) that the exact value of the duality gap is

$$\begin{aligned}(\bar{X}(0) + 2tD_X^P) \bullet (\bar{Z}(0) + 2tD_Z^P) &= \|V\|_F^2 + 2V \bullet \left(V^{\theta} - V + t(D_X^P + D_Z^P)\right) \\ &= \|V\|_F^2 + 2V \bullet (V^{\theta} - V) - 2tV \bullet V^{\theta} \\ &= (1-2t)\left\|V^{\theta}\right\|_F^2 - \left\|V^{\theta} - V\right\|_F^2,\end{aligned}$$

where we used the fact that $V^\theta \perp (V^\theta - V)$. The above relation shows, among others, that $t_P^* \leq 1/2$. This painfully makes clear that the first order approximation of the V–space solution, viz. $(1-t)V^\theta$, is always inaccurate for large t. For such t, a second order approximation is usually much better. Experience shows that we may expect reasonable quality of the second order approximation for $t \in [0, 2t_P^*]$.

7.5.2 Second Order Correction

Based on the predictor direction (D_X^P, D_Z^P) and the predictor step length t_P^*, we want to compute a second order approximation for the target $(1-2t_P^*)V^\theta$. However, the results of Section 7.4 cannot be applied immediately, since we are reusing the factorizations that predate the initial centering step. In a sense, the initial centering step causes a discontinuity at $t=0$ for $\bar{X}(t)$, $\bar{Z}(t)$ and $V(t)$.

We propose to use V^θ as an approximation for $V(0)$, just as we did in the derivation of the predictor direction. Naturally, we use $2D_X^P$ and $2D_Z^P$ as approximations for $\bar{X}^{[1]}(0)$ and $\bar{Z}^{[1]}(0)$. Using (7.34), this results in an approximation D_G for $G^{[1]}(0)$, where D_G is the solution of

$$P_{\mathcal{H}}(V^\theta D_G) = \frac{1}{2}(D_X^P - D_Z^P) \tag{7.35}$$

see also Lemma 3.2. Based on (7.30)–(7.31), a second order search direction $(D_X + D_Z)$ targeting at $(1-2t_P^*)V^\theta$ can now be solved from

$$V^\theta + (D_X + D_Z) = (1-2t_P^*)V^\theta + (2t_P^*)^2 P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)). \tag{7.36}$$

The relation (7.36) can also be written as

$$D_X + D_Z = -(2t_P^*)V^\theta + (2t_P^*)^2 P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)),$$

in which we clearly recognize a first order and a second order component. However, we propose an even more clever search direction, inspired by the very successful second order corrector of Mehrotra for linear programming [91].

In the preceding, we have rather optimistically proposed the target $(1-2t_P^*)V^\theta$. We argued that the second order approximation usually performs much better than the first order approximation, and that it is reasonable to expect that the solution $(\bar{X}(0) + 2D_X, \bar{Z}(0) + 2D_Z)$ is (almost) feasible. However, we cannot expect that the corresponding V–space solution is close to its target $(1-2t_P^*)V^\theta$. Since V^θ is typically at the boundary of $\mathrm{CR}(\theta)$, we thus run a serious risk of leaving the neighborhood $\mathcal{N}(\theta, \beta)$ soon. In order to avoid this phenomenon, we propose to choose a target in the interior of the central region. At first sight, a natural choice for such a target is the projection of $(1-2t_P^*)V^\theta$ onto the central path, i.e.

$$(1-2t_P^*)\frac{\operatorname{tr} V^\theta}{\bar{n}} I. \tag{7.37}$$

However, this amounts to a very heavy centering component, especially if t_P^* is small. Therefore, we propose to move the target from $(1-2t_P^*)V^\theta$ towards its projection (7.37) only by a proportion $(2t_P^*)^2$. The search direction (D_X, D_Z) is thus defined as

$$\begin{aligned} D_X + D_Z &= -(2t_P^*)V^\theta + (2t_P^*)^2 P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)) \\ &\quad +(2t_P^*)^2(1-2t_P^*)(\frac{\operatorname{tr} V^\theta}{\bar{n}} I - V^\theta). \end{aligned} \tag{7.38}$$

We call Algorithm 7.1 with search direction (7.38) the centering–predictor–corrector method.

7.5.3 Polynomiality

In order to prove the polynomiality of the centering–predictor–corrector method, we only need to show that the search direction D_X+D_Z, as defined in (7.38), satisfies the condition (7.20) of Algorithm 7.1. The following lemma is crucial in estimating $\|D_X + D_Z\|_F$.

Lemma 7.9 *There holds*

$$2t_P^* \|D_G\|_2 \leq 1 + \beta - t_P^*.$$

Proof. Since $V \in \mathcal{N}(\theta, \beta)$ and $V^{\mathrm{E}}(0) = V^\theta$, we have using (7.10),

$$\|R(0)\|_F = \left\|V^{\mathrm{E}}(0) - V\right\|_F \leq \frac{\beta\theta}{\sqrt{\bar{n}}} \left\|V^\theta\right\|_F \leq \beta\lambda_{\min}(V^\theta),$$

so that $-\beta V^\theta \preceq R(0) \preceq \beta V^\theta$. It thus follows for all $0 \leq t \leq t_P^*$ that

$$\begin{aligned} 0 &\preceq \bar{X}(0) + 2tD_X^P \\ &= (1-t)V^\theta + R(0) + t(D_X^P - D_Z^P) \\ &\preceq (1+\beta-t)V^\theta + t(D_X^P - D_Z^P) \\ &= P_{\mathcal{H}}\left(V^\theta((1+\beta-t)I + 2tD_G)\right), \end{aligned}$$

where the last step follows from definition (7.35). Using Lemma A.2, the above relation implies that

$$\lambda_{\min}\left(V^\theta((1+\beta-t)I + 2tD_G)\right) \geq 0. \tag{7.39}$$

Similarly, it follows from the dual feasibility of the step length t_P^* that

$$\lambda_{\min}\left(V^\theta((1+\beta-t)I - 2tD_G)\right) \geq 0. \tag{7.40}$$

Combining (7.39) and (7.40) and using that V^θ is positive definite, we obtain

$$1 + \beta - t \geq 2t\|D_G\|_2.$$

Since the above relation holds for any $t \in [0, t_P^*]$, the lemma follows. **Q.E.D.**

The following lemma shows how the second order component in (7.38) affects the duality gap.

Lemma 7.10 *There holds*

$$V^\theta \bullet P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)) = \frac{1}{2}\left\|V^\theta\right\|_F^2.$$

Proof. By definition of the matrix inner product, we have

$$\begin{aligned}
V^\theta \bullet P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)) &= \operatorname{tr} V^\theta P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)) \\
&= \frac{1}{2}\operatorname{tr} V^\theta D_G(D_X^P - D_Z^P) \\
&\quad + \frac{1}{2}\operatorname{tr} V^\theta (D_X^P - D_Z^P) D_G \\
&= \frac{1}{2}\operatorname{tr}\,(V^\theta D_G + D_G V^\theta)(D_X^P - D_Z^P).
\end{aligned}$$

Together with definition (7.35), this implies that

$$V^\theta \bullet P_{\mathcal{H}}(D_G(D_X^P - D_Z^P)) = \frac{1}{2}\left\|D_X^P - D_Z^P\right\|_F^2 = \frac{1}{2}\left\|D_X^P + D_Z^P\right\|_F^2 = \frac{1}{2}\left\|V^\theta\right\|_F^2.$$

Q.E.D.

We can use Lemma 7.10 to estimate $\|D_X + D_Z\|_F$. Namely,

$$\begin{aligned}
\left\|-V^\theta + tP_{\mathcal{H}}(D_G(D_X^P - D_Z^P))\right\|_F^2 &= (1-t)\left\|V^\theta\right\|_F^2 + t^2\left\|P_{\mathcal{H}}(D_G(D_X^P - D_Z^P))\right\|_F^2 \\
&\le (1 - t + t^2\|D_G\|_2^2)\left\|V^\theta\right\|_F^2.
\end{aligned}$$

Filling in the predicted step length $t = 2t_P^*$ and applying Lemma 7.9, we obtain

$$1 - 2t_P^* + (2t_P^*\|D_G\|_2)^2 \le 1 - 2t_P^* + (1 + \beta - t_P^*)^2 < 2(1 + \beta - t_P^*)^2.$$

Using the triangle inequality, it thus follows from definition (7.38) that

$$\begin{aligned}
\frac{\|D_X + D_Z\|_F}{2t_P^*} &\le \left\|-V^\theta + 2t_P^* P_{\mathcal{H}}(D_G(D_X^P - D_Z^P))\right\|_F \\
&\quad + 2t_P^*(1 - 2t_P^*)\left\|(\operatorname{tr} V^\theta/\bar{n})I - V^\theta\right\|_F \\
&\le (\sqrt{2}(1 + \beta - t_P^*) + 2t_P^*(1 - 2t_P^*)\sin\phi)\left\|V^\theta\right\|_F, \qquad (7.41)
\end{aligned}$$

where ϕ denotes the angle between the identity matrix and V^θ. It also follows from Lemma 7.10 and (7.38) that

$$\frac{V^\theta \bullet (D_X + D_Z)}{2t_P^*} = -(1 - t_P^* + 2t_P^*(1 - 2t_P^*)(\sin\phi)^2)\left\|V^\theta\right\|_F^2, \qquad (7.42)$$

where we used the fact that $I \perp ((\operatorname{tr} V^\theta/\bar{n})I - V^\theta)$ to conclude that

$$\begin{aligned} V^\theta \bullet (\frac{\operatorname{tr} V^\theta}{\bar{n}} I - V^\theta) &= -\left\| (\operatorname{tr} V^\theta/\bar{n})I - V^\theta \right\|_F^2 \\ &= -(\sin\phi)^2 \left\| V^\theta \right\|_F^2 . \end{aligned}$$

Combining (7.41) and (7.42), and using the fact that $\beta \leq 1/\sqrt{5}$, it can be verified that

$$\left\| V^\theta \right\|_F \|D_X + D_Z\|_F \leq -\frac{3\sqrt{2}}{2} V^\theta \bullet (D_X + D_Z),$$

so that $D_X + D_Z$ satisfies (7.20) with $\gamma = \sqrt{2}/3 > 1/\sqrt{5}$. Theorem 7.1 is therefore applicable for the centering–predictor–corrector algorithm, stating an

$$\mathbf{O}(\sqrt{\bar{n}} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iteration complexity, if we choose the parameters β and θ to be independent of $\bar{n}$.

7.6 Discussion

Path–following methods are included in the central region framework by choosing $\theta = 1$. However, the choice $\theta < 1$ is more attractive, since it allows long steps to be taken, without sacrificing the superb worst case behavior. An additional performance gain is achieved by incorporating a Mehrotra–type second order centrality corrector [91]. In the special case of linear programming, Mehrotra–type correctors are used in the most successful interior point codes, including the later versions of OB1 [89]. However, there has been a wide gap between theory and practice for these techniques. In the case of LP and monotone LCP, Zhang and Zhang [165] proved polynomiality for a modification of Mehrotra's algorithm. However, this modification has often been criticized for its step length selection: it requires searching along a quadratic curve in t. From a computational point of view, a standard inexact line search is certainly more appealing, and this is what we use in the centering-predictor-corrector method. Moreover, the established worst case iteration bound for the centering–predictor–corrector algorithm improves the bound of Zhang and Zhang by a factor $\bar{n}$. Therefore, the results in this chapter mean a substantial progress, even for the special case of linear programming.

This chapter provides the first theoretical convergence results in semidefinite programming for interior point methods with second order correctors. In numerical experiments, such correctors have been used by Todd, Toh and Tütüncü [151], and Alizadeh, Haeberly and Overton [4]. Their second order correctors were derived in the matrix target framework. In this chapter, we obtained second order directions based on the derivatives of the function $V(t)$. Using the chain rule, it is easy to construct the first two derivatives of $V(t)^2$ in terms of the first two derivatives of $V(t)$, and this leads to a natural generalization of Mehrotra's approach. However, this technique yields a different second order direction than the Todd–Toh–Tütüncü direction.

The generalization of weighted centers to semidefinite programming is also a controversial subject. Monteiro and Pang [100] define a W-weighted center as an interior feasible primal–dual pair (X, Z) satisfying $P_{\mathcal{H}}(XZ) = W$, where W is a given positive definite matrix. Interestingly, the relation $P_{\mathcal{H}}(XZ) = W$ uniquely defines an interior solution pair (X, Z), for each positive definite matrix W. However, the converse is not true: given an interior feasible solution pair (X, Z), the matrix $P_{\mathcal{H}}(XZ)$ is in general *not* positive definite [138, 144, 151], see also Lemma A.2. In fact, the Hermitian part of XZ can be indefinite for arbitrarily well centered primal–dual pairs (X, Z). For instance, consider

$$V(\epsilon) = \begin{bmatrix} 1+\epsilon & 0 \\ 0 & 1 \end{bmatrix}, \quad L_d(\epsilon) = \begin{bmatrix} 1 & 0 \\ 1/\epsilon^2 & 1 \end{bmatrix}, \quad L_d(\epsilon)^{-1} = \begin{bmatrix} 1 & 0 \\ -1/\epsilon^2 & 1 \end{bmatrix},$$

then

$$X(\epsilon)Z(\epsilon) = L_d(\epsilon)V(\epsilon)^2 L_d(\epsilon)^{-1} = \begin{bmatrix} (1+\epsilon)^2 & 0 \\ 1+\epsilon^{-2}+\epsilon^{-4} & 1 \end{bmatrix},$$

for which the Hermitian part is indefinite for $\epsilon \downarrow 0$. Moreover, the Monteiro–Pang weighted center is *not* invariant with respect to linear transformations of the form $L^{-1} \otimes_{\mathrm{H}} L^{-1}$, as were discussed in Section 3.1. Finally, there are (still) no algorithms for calculating Monteiro–Pang weighted centers.

In this chapter, we have used the weighted centers of Sturm and Zhang [144]. These weighted centers do not have any of the weak points of the Monteiro–Pang weighted centers. However, they lack the appealing property of uniqueness. Recall that an interior solution pair (X, Z) is a W-weighted center in the sense of [144] if the eigenvalues of XZ are the diagonal entries of W^2, where W is a positive diagonal matrix. The difference in the approach to weighted centers of [100] and [144] is similar to the difference between the derivation of V–space search directions with eigenvalue targets, and the Monteiro–Zhang path–following directions with matrix targets, see Section 3.6. It is not clear whether weighted centers can be introduced in the self–scaled barrier framework. However, a nonconvex weighted potential function for semidefinite programming was proposed by De Klerk, Roos and Terlaky [69].

We have seen that eigenvalue based weighted centers can be approached very efficiently, and this forms the basis of the central region method. The algorithms in this chapter are the first algorithms for solving semidefinite programs using weighted centers. Algorithms using weighted centers in linear programming are the target–following method of Jansen, Roos, Terlaky and Vial [58, 61, 134], the α–sequence method of Mizuno [93] and the central region method of Sturm and Zhang [146].

Chapter 8

Directions for Further Research

The study of semidefinite programming involves a number of fascinating issues that are due to its nonlinear structure. Researchers in nonlinear programming have often avoided these issues, by imposing various assumptions, and in particular constraint qualifications. However, if we want to solve semidefinite programs in a reliable way, then all aspects of the problem have to be faced. This task has been taken up seriously in this thesis, by starting from a thorough investigation of duality, and proceeding all the way up to the design of an efficient numerical algorithm: the central region method with second order centrality correction.

As mentioned in Chapter 1, there are many (potential) areas of application for semidefinite programming, and experiments with semidefinite programming have already been carried out in some of these areas. To facilitate the application of semidefinite programming, the development of modeling tools will be very important. The parser of Boyd and Wu [21] is a first step in this direction. Techniques as discussed in Appendix B seem to be most suitable for future modeling tools. This will then facilitate the use of semidefinite programming for solving classes of convex programming problems, such as contained in the CUTE test set [14].

The ability of exploiting problem structure is crucial in large scale optimization. A well known and often very successful technique is to exploit a general sparsity structure. In particular, the performance of interior point methods for linear programming is largely determined by the procedure that is used for sparse Cholesky decomposition. However, the applicability of this technique to semidefinite programming has often been questioned [15, 79], with the argument that any sparsity structure can be destroyed by primal–dual transformations (such as L_d in Chapters 5 and 6). Fortunately, this argument does not apply to semidefinite programs with a large number of matrix variables, or matrix variables with a block diagonal structure. In fact, many constraints for these problems will only involve one or two matrix variables at a time, thus generating a highly favorable sparsity structure in the normal equations. Moreover, by using preprocessing techniques such as column splitting, we may be able to split large sparse matrix variables into a number of smaller matrix variables, thus promoting an amenable structure. Other advantages of having a large number of medium sized matrix variables (instead of a few large ones) relate to parallel computations and low rank updates. Of course, there are more opportunities for reducing the computational costs than sparsity. For instance, Vandenberghe and Boyd [157] successfully used a Sylvester structure that is inherent to some semidefinite programs arising in system theory.

Goldfarb and Scheinberg [42] have studied parametric semidefinite programming. Their results concern the behavior of the optimal solution value and the optimal solution set, with respect to data perturbations. An understanding of this behavior is very important, since problem data is often not completely accurate. However, the nonlinearity of semidefinite programming gives rise to truly challenging difficulties here, and many of them are still unsolved.

In Chapter 6, we obtained superlinear convergence results for a predictor–corrector algorithm with a shrinking $\mathcal{N}_2$–neighborhood [87]. In practice however, we also observe superlinear convergence for the standard predictor–corrector algorithm with a fixed $\mathcal{N}_2$–neighborhood. Despite numerous efforts [4, 76, 78, 87, 120, 122, 121, 137], there is still no theoretical proof for this behavior. In [87, 122] and in Chapter 6, superlinear convergence is established for semidefinite programs that are possibly degenerate. However, strict complementarity is required, since, even for quadratic programming, Mizuno–Todd–Ye type predictor–corrector methods cannot be superlinearly convergent otherwise, see [33, 103, 162]. Superlinearly convergent algorithms for quadratic programming without strict complementarity were developed by Mizuno [94] and Sturm [140]. The study of local convergence for semidefinite programming without strict complementarity is a subject of future research.

More research is needed to understand the extensions of weighted centers to semidefinite programming. In particular, it is not known whether there exist smooth trajectories of weighted centers. Moreover, there are different definitions of weighted centers, namely the weighted centers of Monteiro and Pang [100], the V–space weighted centers of Sturm and Zhang [144], and the non-convex weighted potential function of De Klerk, Roos and Terlaky [69]. We have seen in Chapter 7 that the V–space centers are attractive from an algorithmic point of view.

Several extensions of semidefinite programming are certainly worth further investigation. In particular, Vandenberghe, Boyd and Wu [159] consider a generalization of semidefinite programming, where the objective function is the sum of a linear and a logarithmic function. Such problems arise, among others, in statistics. The generalized eigenvalue problem is another interesting extension of semidefinite programming, with applications in system and control theory. This class of problems, which can also be viewed as an extension of generalized fractional programming, is studied by Boyd, El Ghaoui, Feron and Balakrishnan [20] and Nesterov and Nemirovsky [111].

Appendix A

Hermitian Matrices

In this appendix, we list some of the most fundamental results for Hermitian matrices. For a more thorough treatment of Hermitian matrices, we refer to Horn and Johnson [56].

A.1 Eigenvalue Decomposition

Suppose that X is a Hermitian matrix, i.e. $X = X^{\mathrm{H}}$. Then X has an eigenvalue decomposition, or spectral decomposition,

$$X = Q\Lambda_X Q^{\mathrm{H}},$$

where Q is a unitary matrix, i.e. $QQ^{\mathrm{H}} = Q^{\mathrm{H}}Q = I$, and Λ_X is a real diagonal matrix. The diagonal entries of Λ_X are the eigenvalues of X, and the columns of Q are corresponding eigenvectors. If X is an order $\bar{n}$ Hermitian matrix, i.e. $X \in \mathcal{H}^{(\bar{n})}$, then Q and Λ_X can be computed numerically in $\mathbf{O}(\bar{n}^3)$ floating point operations.

The eigenvalues of X are all nonnegative (positive) if and only if X is positive semidefinite (positive definite). Recall that by definition, $X \in \mathcal{H}^{(\bar{n})}$ is positive definite if and only if

$$y^{\mathrm{H}}Xy > 0 \text{ for all } 0 \neq y \in \mathcal{C}^{\bar{n}}.$$

Similarly, X is positive semidefinite if and only if $y^{\mathrm{H}}Xy \geq 0$ for all y. Suppose that X is positive semidefinite. Letting $\Lambda_X^{1/2}$ denote the positive diagonal matrix satisfying $(\Lambda_X^{1/2})^2 = \Lambda_X$, we define the matrix square root $X^{1/2} := Q\Lambda_X^{1/2}Q^{\mathrm{H}}$. Obviously, $X^{1/2}$ is a positive definite matrix, and $(X^{1/2})^2 = X$.

If X is Hermitian, then $\|X\|_2$ is the spectral radius of X, i.e. $\|X\|_2 = \max_{1\leq j\leq \bar{n}} |\lambda_j(X)|$, where $\lambda_1(X), \ldots, \lambda_{\bar{n}}(X)$ are the eigenvalues of X.

Lemma A.1 *Let X be Hermitian and Y positive definite. Then*

$$XYX \preceq \alpha Y \Longrightarrow \|X\|_2^2 \leq \alpha.$$

Proof. Let λ be an eigenvalue of X with the largest absolute value, and let q be a corresponding eigenvector. Then

$$Xq = \lambda q, \quad \lambda^2 = \|X\|_2^2.$$

From $XYX \preceq \alpha Y$, we have

$$\|X\|_2^2\, q^{\mathrm{H}} Y q = q^{\mathrm{H}}(XYX)q \le \alpha\, q^{\mathrm{H}} Y q.$$

Since Y is positive definite, we know that $q^{\mathrm{H}} Y q > 0$, concluding the proof. **Q.E.D.**

A.2 Cholesky Decomposition

For positive definite matrices, we can compute a Cholesky decomposition

$$X = L_X L_X^{\mathrm{H}},$$

where L_X is a nonsingular lower triangular matrix. The Cholesky factor L_X can be computed numerically in $\mathbf{O}(\bar{n}^3)$ floating point operations. Once the Cholesky factor L_X is known, we can solve a system of linear equations

$$\text{Find } y \text{ satisfying } Xy = r$$

in $\mathbf{O}(\bar{n}^2)$ operations, for given $r \in \mathcal{C}^{\bar{n}}$.

A.3 Similarity

Suppose that $X \in \mathcal{H}^{(\bar{n})}$ and $Y \in \mathcal{C}^{\bar{n}\times\bar{n}}$. If there exists a nonsingular $\bar{n} \times \bar{n}$ matrix R such that $Y = RXR^{-1}$, then Y is said to be similar to X. If we let $X = Q\Lambda_X Q^{\mathrm{H}}$ denote the spectral decomposition of X, then $Y = RXR^{-1}$ implies that

$$Y(RQ) = (RQ)\Lambda_X.$$

It follows that Y has the same eigenvalues as X, and the columns of RQ are the eigenvectors of Y. Hence, if an $\bar{n} \times \bar{n}$ matrix is similar to a Hermitian matrix, it has $\bar{n}$ real eigenvalues, and if it is similar to a positive definite matrix, it has $\bar{n}$ positive eigenvalues.

Lemma A.2 *Let Y be a square matrix, and suppose that Y is similar to a Hermitian matrix, which implies that it has only real eigenvalues. Then*

$$\lambda_{\min}(Y) \ge \lambda_{\min}(P_{\mathcal{H}} Y).$$

Proof. Let q be an eigenvector corresponding to the smallest eigenvalue of Y, i.e. $Yq = \lambda_{\min}(Y)q$, and assume without loss of generality that $q^{\mathrm{H}}q = 1$. Then

$$\lambda_{\min}(Y) = q^{\mathrm{H}}Yq = \frac{1}{2}q^{\mathrm{H}}Yq + \frac{1}{2}(q^{\mathrm{H}}Yq)^{\mathrm{H}} = q^{\mathrm{H}}(P_{\mathcal{H}}Y)q \geq \lambda_{\min}(P_{\mathcal{H}}Y),$$

where we used the fact that

$$P_{\mathcal{H}}Y \succeq \lambda_{\min}(P_{\mathcal{H}}Y)I,$$

which is an immediate consequence of the spectral decomposition of Hermitian matrices. **Q.E.D.**

For the real symmetric case, Lemma A.2 is a slight variation on Lemma 3.3 of Monteiro [97] and Lemma 5.3 of Monteiro and Zhang [105]. Since the largest eigenvalue of Y is $-\lambda_{\min}(-Y)$, it follows from Lemma A.2 that

$$\|Y\|_2 \leq \|P_{\mathcal{H}}Y\|_2 \leq \|P_{\mathcal{H}}Y\|_F \leq \|Y\|_F,$$

for any square matrix Y that is similar to a Hermitian matrix.

A.4 Schur Complement

Consider a partitioned Hermitian matrix X,

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^{\mathrm{H}} & X_{22} \end{bmatrix},$$

and let $X_{22}^{\#}$ denote a Hermitian generalized inverse of X_{22}, i.e.

$$X_{22}^{\#} \in \mathcal{H}, \quad X_{22}^{\#}X_{22}X_{22}^{\#} = X_{22}^{\#}, \quad X_{22}X_{22}^{\#}X_{22} = X_{22}.$$

It is easily verified that

$$X \succeq 0 \Longrightarrow X_{12}(I - X_{22}^{\#}X_{22}) = 0. \tag{A.1}$$

Consider the unit–diagonal lower triangular (and hence invertible) matrix L, defined by

$$L := \begin{bmatrix} I & 0 \\ -X_{22}^{\#}X_{12}^{\mathrm{H}} & I \end{bmatrix}.$$

Since L is invertible, we know that $X \succeq 0$ if and only if $L^{\mathrm{H}}XL \succeq 0$. However,

$$\begin{aligned} L^{\mathrm{H}}XL &= \begin{bmatrix} I & -X_{12}X_{22}^{\#} \\ 0 & I \end{bmatrix} \begin{bmatrix} X_{11} - X_{12}X_{22}^{\#}X_{12}^{\mathrm{H}} & X_{12} \\ X_{12}^{\mathrm{H}} - X_{22}X_{22}^{\#}X_{12}^{\mathrm{H}} & X_{22} \end{bmatrix} \\ &= \begin{bmatrix} X_{11} - X_{12}X_{22}^{\#}X_{12}^{\mathrm{H}} & X_{12}(I - X_{22}^{\#}X_{22}) \\ (I - X_{22}X_{22}^{\#})X_{12}^{\mathrm{H}} & X_{22} \end{bmatrix}. \end{aligned}$$

Combining the above identity with (A.1), it follows that

$$X \succeq 0 \iff \begin{cases} X_{12}(I - X_{22}^{\#}X_{22}) = 0 \\ X_{11} \succeq X_{12}X_{22}^{\#}X_{12}^{\mathrm{H}} \\ X_{22} \succeq 0. \end{cases} \tag{A.2}$$

For positive semidefinite X, we see that $L^{\mathrm{H}}XL$ is block diagonal. Since L is unit diagonal lower triangular, we have $\det L = 1$ and

$$\det X = \det(L^{\mathrm{H}}XL) = \det(X_{11} - X_{12}X_{22}^{\#}X_{12}^{\mathrm{H}}) \det X_{22}.$$

For the case that X_{22} is invertible, the matrix $X_{11} - X_{12}X_{22}^{-1}X_{12}^{\mathrm{H}}$ is known as the Schur complement of X.

If we partition X such that X_{11} is a scalar, then (A.2) takes a special form. Namely, since X_{11} is then a diagonal element of a Hermitian matrix, it must be real. Furthermore, by writing $X_{12} = (\operatorname{Re} X_{12}) + \mathbf{i}(\operatorname{Im} X_{12})$, it follows that (A.2) can be reformulated as

$$X \succeq 0 \iff \begin{cases} X_{12} = X_{12}X_{22}^{\#}X_{22} \\ X_{11} \geq (\operatorname{Re} X_{12})X_{22}^{\#}(\operatorname{Re} X_{12})^{\mathrm{T}} + (\operatorname{Im} X_{12})X_{22}^{\#}(\operatorname{Im} X_{12})^{\mathrm{T}} \\ X_{22} \succeq 0. \end{cases} \tag{A.3}$$

Appendix B

Semidefinite Modeling

Semidefinite programming has a wealth of applications. However, the formulation of semidefinite programming models can be nontrivial, to say the least. To give you some inspiration, we propose a number of modeling techniques in this appendix.

Despite many obvious similarities, semidefinite programming is considerably more general than linear programming. Broad classes of convex constraints can be modeled with semidefinite programming by using *semidefinite envelopes*. With this technique, we try to model the epigraph of a convex function, by using only semidefinite constraints.

A function $g : \Re^l \to \Re$ has a semidefinite envelope representation if it satisfies a relation like

$$g(y) = \min_{y_0, x}\{y_0 \mid (y_0, y, x) \in b + \mathcal{A},\ x \in \mathcal{H}_+\} \quad \forall y \in \operatorname{dom}(g). \tag{B.1}$$

Here, b is a vector, $\mathcal{A}$ is a linear subspace and the domain of $g(\cdot)$ is

$$\operatorname{dom}(g) = \{y \in \Re^l \mid (y_0, y, x) \in b + \mathcal{A} \text{ for some } y_0 \in \Re,\ x \in \mathcal{H}_+\}.$$

The right hand side of the identity (B.1) is called the semidefinite envelope representation of $g(y)$. A constraint of the form

$$g(y) \leq 0$$

can now be modeled in a semidefinite program by including a semidefinite envelope representation of $g(y)$ in the constraint set, and adding the constraint $y_0 \leq 0$. Similarly, we can minimize $g(y)$ using the objective function $\inf y_0$.

Suppose that $y^{(1)}$ and $y^{(2)}$ are two vectors in the domain of $g(\cdot)$. From the semidefinite envelope representation (B.1) of $g(\cdot)$, it follows that

$$g((1-\lambda)y^{(1)} + \lambda y^{(2)}) \leq (1-\lambda)g(y^{(1)}) + \lambda g(y^{(2)}) \text{ for all } 0 \leq \lambda \leq 1.$$

Consequently, semidefinite envelope representations of the form (B.1) can exist only for convex functions. For concave functions, we can obviously define semidefinite envelope representations by replacing 'min' by 'max' in (B.1). Envelope representations are often used in non-smooth optimization, as in Rockafellar [133]. However, semidefinite envelopes also exist for many types of smooth convex functions. As a first example, we will discuss convex quadratic functions.

B.1 Convex Quadratic Functions

The basic observation here is

$$\begin{bmatrix} x_1 & y_1 + \mathbf{i}\, y_2 \\ & 1 \end{bmatrix} \succeq 0 \iff x_1 \geq y_1^2 + y_2^2.$$

(See the characterization (A.3).) Let $g(y)$ be a general convex quadratic function, i.e.

$$g(y) = g(0) + \nabla g(0)^{\mathrm{T}} y + \frac{1}{2} y^{\mathrm{T}} \nabla^2 g(0) y,$$

and let $r =$ rank $\nabla^2 g(0)$. Since the Hessian of a convex function is positive semidefinite, we can factorize $\nabla^2 g(0) = LL^{\mathrm{T}}$, for some $n \times r$ matrix L. We have

$$y_0 \geq g(y) \iff \begin{cases} y_0 \geq g(0) + \nabla g(0)^{\mathrm{T}} y + \frac{1}{2} \sum_{j=1}^{\lceil r/2 \rceil} x_j \\ \begin{bmatrix} x_i & \xi_{2j-1} + \mathbf{i}\, \xi_{2j} \\ & 1 \end{bmatrix} \succeq 0 \quad \text{for } j = 1, 2, \ldots, \lceil r/2 \rceil \\ \xi = L^{\mathrm{T}} y, \quad \xi_{r+1} = 0. \end{cases}$$

Instead of $\lceil r/2 \rceil$ Hermitian semidefinite constraints of order 2, we can also use r real symmetric semidefinite constraints of order 2. However, the above formulation with Hermitian matrices is more efficient.

B.2 Compositions

A powerful technique is nesting of semidefinite envelopes, which yields representations of composite convex functions. The constraints in a semidefinite envelope representation (B.1) basically model the epigraph

$$\{(y_0, y) \mid y_0 \geq g(y)\}.$$

Now suppose that $g_1(\cdot), \ldots, g_k(\cdot)$ are convex functions for which semidefinite envelope representations are known, and further assume that we have a semidefinite envelope representation for a monotone nondecreasing convex function $f : \Re^k \to \Re$. Then

$$f(g_1(y), \ldots, g_k(y)) = \min\{y_0 \mid y_0 \geq f(u_1, \ldots, u_k),\ u_j \geq g_j(y)\ \forall j\}.$$

In this way, we can construct semidefinite envelope representations for certain classes of convex polynomials, by composing convex quadratic functions. Remark also that the convexity of convex composite functions of the above considered type is well known, see e.g. Proposition 2.16 in [6].

B.3 Concave Products

The basic observation here is that

$$\begin{bmatrix} x_1 & y_0 \\ & x_2 \end{bmatrix} \succeq 0 \iff | y_0 | \leq \sqrt{x_1 x_2}, \quad x_1, x_2 \geq 0,$$

which yields a semidefinite representation for the concave function $\sqrt{x_1 x_2}$. By composing concave functions of the form $\sqrt{x_1 x_2}$, we can get a representation for

$$\prod_{i=1}^{2^k} x_i^{2^{-k}},$$

where k is an arbitrary positive integer. After adding suitable linear equality constraints, like $x_i = 1$ or $x_i = x_j$, we can represent any function of the form

$$\prod_{i=1}^{r} x_i^{\alpha_i}, \quad \sum_{i=1}^{r} \alpha_i \leq 1, \tag{B.2}$$

with $x_i \geq 0$, $\alpha_i \geq 0$ and

$$\alpha_i = \sum_{j=0}^{k} \alpha_{ij} 2^{-j}, \quad \alpha_{ij} \in \{0, 1\}.$$

The above relation simply states that α_i is a base 2 floating point number. See e.g. Higham [54] for a discussion of the floating point number system.

It is well known that products of the form (B.2) are concave, see Corollary 5.19 in [6]. Remark that compositions can be made by modeling the variables x_i as concave functions with semidefinite envelope representations. Our treatment of concave products extends the technique of Nesterov and Nemirovsky [110] for representing certain geometric means.

B.4 Convex Reciprocal Products

Simply observe that

$$x > 0 \text{ and } \begin{bmatrix} y_1 & 1 \\ & x \end{bmatrix} \succeq 0 \iff y_1 \geq \frac{1}{x},$$

yielding a semidefinite representation for the convex reciprocal $1/x$. For $k = 1, 2, \ldots$, consider the semidefinite constraints

$$\begin{bmatrix} y_{2k} & y_k \\ & 1 \end{bmatrix} \succeq 0, \quad \begin{bmatrix} y_{2k+1} & y_{k+1} \\ & x \end{bmatrix} \succeq 0.$$

By induction, it is then easily checked that

$$y_k \geq \frac{1}{x^k} \quad \text{for all } k = 1, 2, \ldots,$$

and we have constructed a semidefinite envelope representation for the convex reciprocal $1/x^k$ with k an arbitrary positive integer. Moreover, interesting composite convex functions are created by modeling x as a concave product (see Section B.3). This results in semidefinite envelope representations for functions of the form

$$1/\prod_{i=1}^{r} x_i^{\alpha_i}, \quad \text{with } x_i > 0,\ \alpha_i \geq 0, \tag{B.3}$$

where each α_i is a floating point number. The fact that functions of the form (B.3) are convex is known from Corollary 5.18 in [6].

B.5 Convex Fractions

Remark first that

$$x_2 > 0 \text{ and } \begin{bmatrix} y_2 & x_1 \\ & x_2 \end{bmatrix} \succeq 0 \iff y_2 \geq \frac{x_1^2}{x_2},$$

yielding a semidefinite representation of the convex fraction x_1^2/x_2, with $x_2 > 0$. Notice that there is no sign restriction on x_1 in the above formulation. However, we will restrict to nonnegative x_1 below. For $k = 2, 3, \ldots$, we consider the semidefinite constraints

$$\begin{bmatrix} y_{2k-1} & y_k \\ & x_1 \end{bmatrix} \succeq 0, \begin{bmatrix} y_{2k} & y_k \\ & x_2 \end{bmatrix} \succeq 0.$$

Using an inductive argument, we see that

$$y_{k+1} \geq \frac{x_1^{k+1}}{x_2^k} \quad \text{for all } k = 1, 2, \ldots,$$

and we have constructed a semidefinite envelope representation of the convex fraction x_1^{k+1}/x_2^k with $x_1 \geq 0$, $x_2 > 0$ and k an arbitrary positive integer. Moreover, interesting composite convex functions are created by modeling x_2 as a concave product (see

Section B.3). For $x_1 \geq 0$ and $x_i > 0$, $i = 2, 3, \ldots, r$, this yields semidefinite envelope representations for functions of the form

$$x_1^{k+1} / \prod_{i=2}^{r} x_i^{\alpha_i}, \qquad \text{with } \alpha_i \geq 0, \sum_{i=2}^{r} \alpha_i \leq k, \tag{B.4}$$

where each α_i is a floating point number. The convexity for this kind of fractional function is known from Theorem 5.16 in [6].

B.6 Quasiconvex Fractions

As pointed out in the introduction of this appendix, semidefinite envelope representations can only exist for convex and concave functions. Still, it is possible to model some classes of generalized convex constraints, by using different techniques.

As is well known, quasiconvex linear fractional constraints can simply be represented by linear constraints, viz.

$$x_2 > 0 \text{ and } x_1 \leq \beta x_2 \iff \frac{x_1}{x_2} \leq \beta.$$

In semidefinite programming however, we are not restricted to linear fractions. For instance, we can model x_1 as a (nonnegative) convex fraction of the form (B.4), i.e.

$$x_1 \geq x_1^{k+1} / \prod_{i=2}^{r} x_i^{\alpha_i}, \qquad \text{with } \alpha_i \geq 0, \sum_{i=2}^{r} \alpha_i \leq k,$$

and we obtain the quasiconvex fractional constraint

$$x_1^{k} / \prod_{i=2}^{r} x_i^{\alpha_i} \leq \beta, \tag{B.5}$$

with floating point numbers $\alpha_i \geq 0$, $\sum_{i=2}^{r} \alpha_i \leq k$, and k integer. Quasiconvexity of this type of fractions is known from Theorem 5.15 in [6].

If $x_1 > 0$ and $\beta > 0$ then we can rewrite (B.5) as

$$\prod_{i=2}^{r} x_i^{\alpha_i} / x_1^{k} \geq 1/\beta, \tag{B.6}$$

which is a quasiconcave fraction.

We obtain an interesting special case of (B.6) by letting $x_1 = 1$, namely

$$\prod_{i=2}^{r} x_i^{\alpha_i} \geq 1/\beta, \tag{B.7}$$

with floating point numbers $\alpha_i \geq 0$. The above constraint involves a quasiconcave product (see Corollary 5.18 in [6]). Alternatively, we can transform quasiconcave products into concave products, simply by raising both sides of (B.7) to the power $1/\sum_{i=2}^{r} \alpha_i$, see Section B.3.

B.7 Quasiconvex Compositions

Recall that the quasiconvex fraction (B.5) was formed as a composition of the quasiconvex linear fraction y_0/x_1, and the convex fraction $y_0 \geq x_1^{k+1}/\prod_{i=2}^r x_i^{\alpha_i}$. This illustrates the concept quasiconvex compositions, which we will now describe in a more general fashion. Suppose that we have semidefinite envelope representations for convex functions $g_1(\cdot), \ldots, g_k(\cdot)$, and further assume that we have a semidefinite envelope representation for a monotone nondecreasing quasiconvex function $f : \Re^k \to \Re$. Then

$$f(g_1(y), \ldots, g_k(y)) \leq 0$$

if and only if

$$f(u_1, \ldots, u_k) \leq 0 \text{ for some } u_1 \geq g_1(y), \ldots u_k \geq g_k(y).$$

This shows how we can embed semidefinite envelope representations of convex functions in quasiconvex constraints. Quasiconvexity for this type of compositions is well known, see Proposition 5.3 in [6].

B.8 Quasiconvex Quadratic Functions

Merely quasiconvex quadratic functions are quadratic functions for which the Hessian is not positive semidefinite, although its lower level sets are convex on the domain. There exists a nice characterization of merely quasiconvex quadratic functions and their domain of quasiconvexity, which we will now summarize from Section 6.1 in Avriel, Diewert, Schaible and Zang [6]. The Hessian of a merely quasiconvex quadratic function has exactly one negative eigenvalue, i.e.

$$\nabla^2 g(0) = LL^{\mathrm{T}} - qq^{\mathrm{T}}, \quad L^{\mathrm{T}} q = 0,$$

where L has $r \leq n-1$ columns, and $q \in \Re^n$. Moreover, the gradient of a merely quasiconvex quadratic function is contained in the image of the Hessian,

$$\nabla g(0) = \nabla^2 g(0)\lambda \quad \text{for some } \lambda \in \Re.$$

Applying the affine transformation

$$\xi_0 = q^{\mathrm{T}}(y+\lambda), \quad \xi = L^{\mathrm{T}}(y+\lambda),$$

we obtain

$$\begin{aligned}
\frac{1}{2}(\xi^{\mathrm{T}}\xi - \xi_0^2) &= \frac{1}{2} y^{\mathrm{T}}(LL^{\mathrm{T}} - qq^{\mathrm{T}})y + \lambda^{\mathrm{T}}(LL^{\mathrm{T}} - qq^{\mathrm{T}})y \\
&\quad + \frac{1}{2}\lambda^{\mathrm{T}}(LL^{\mathrm{T}} - qq^{\mathrm{T}})\lambda \\
&= \frac{1}{2} y^{\mathrm{T}} \nabla^2 g(0) y + \nabla g(0)^{\mathrm{T}} y + \frac{1}{2}\lambda^{\mathrm{T}} \nabla g(0) \\
&= g(y) - \alpha,
\end{aligned}$$

with $\alpha = g(0) - (\lambda^{\mathrm{T}}\nabla g(0)/2)$. The above relation states the well known fact that merely quasiconvex quadratic functions can be transformed to the normal form

$$(\sum_{i=1}^{r} x_i^2) - x_0^2. \tag{B.8}$$

The (maximal) domain of quasiconvexity of this function is the convex cone

$$\{(x_0, x) \mid x_0 \geq 0, \sum_{i=1}^{r} x_i^2 \leq x_0^2\},$$

which is sometimes called a second–order cone, a Lorentz cone, a circular cone or an ice–cream cone, and is a special case of the norm cone, discussed in Section 2.8. Remark also that (B.8) is the difference of two convex functions; such a construction is sometimes called a dc–function.

Suppose that we want to model the constraint

$$(\sum_{i=1}^{r} x_i^2) - x_0^2 + \alpha \leq 0, \quad x_0 \geq 0. \tag{B.9}$$

We only consider $\alpha \geq 0$, since otherwise the constraint does not define a convex region. The model is as follows:

$$\left\{\begin{array}{c} (\sum_{j=1}^{\lceil r/2 \rceil} y_j) - x_0 + y_0 \leq 0 \\ \begin{bmatrix} y_0 & \sqrt{\alpha} \\ & x_0 \end{bmatrix} \succeq 0 \\ \begin{bmatrix} y_j & x_{2j-1} + \mathbf{i}\, x_{2j} \\ & x_0 \end{bmatrix} \succeq 0 \text{ for } j = 1, 2, \ldots, \lceil r/2 \rceil \\ x_{r+1} = 0. \end{array}\right.$$

To see the equivalence with (B.9), notice that $x_0 = 0$ is feasible for the above semidefinite constraints if and only if the constant α is zero and $x_i = 0$, $y_i = 0$ for all i. This corresponds indeed to (B.9). For $x_0 > 0$, the semidefinite constraints are equivalent with

$$y_0 \geq \frac{\alpha}{x_0} \text{ and } y_j \geq \frac{x_{2j-1}^2 + x_{2j}^2}{x_0} \text{ for } j = 1, 2, \ldots, \lceil r/2 \rceil,$$

and the correspondence with (B.9) becomes obvious.

It is interesting to note that a quasiconvex quadratic function is essentially the composition of a linear fraction with convex quadratic functions. Namely, for $x_0 > 0$, we can rewrite (B.9) as

$$\frac{-x_0 + (\alpha/x_0) + \sum_{i=1}^{r}(x_i^2/x_0)}{x_0} \leq 0.$$

This explains the semidefinite envelope representation of merely quasiconvex quadratic functions: y_0 represents the convex reciprocal α/x_0, and y_i models the convex fraction x_i^2/x_0. What remains is a linear fraction, where the numerator is a summation (which is convex and monotone) of the convex functions $-x_0$, y_0 and $y_1,\ldots,y_r$.

Finally, we notice that by chosing $\alpha = 0$ in (B.9), we have a semidefinite envelope representation of Euclidean norms. Therefore, we can use semidefinite programming to minimize a sum of Euclidean norms, a problem that arises in Euclidean facility location. More generally, we can model optimization problems over second–order cones, such as studied in Xue and Ye [161].

B.9 Determinant of a Semidefinite Matrix

Up to now, we only discussed semidefinite constructions with 2×2 blocks. The 2×2 semidefinite block is just a very special case of the semidefinite cone: in the real symmetric case, it is simply a 3-dimensional second–order cone. We will now discuss some interesting (quasi)convex (concave) functions for which the semidefinite envelope representations involve higher order semidefinite cones.

Consider the linear matrix inequality

$$\begin{bmatrix} X & L \\ & \operatorname{diag} L \end{bmatrix} \succeq 0, \quad L \text{ lower triangular.} \tag{B.10}$$

Here, $\operatorname{diag} L$ denotes the diagonal matrix whose diagonal entries correspond with the diagonal entries of the matrix L. Both X and L are decision variables.

If X and L satisfy (B.10) and L is not invertible, then $0 = \det L \leq \det X$. Otherwise, i.e. if L is invertible, then (B.10) is equivalent with

$$X \succeq L(\operatorname{diag} L)^{-1} L^{\mathrm{H}} \succ 0, \quad L \text{ lower triangular,}$$

(see Appendix A), from which we obtain

$$\det X \geq \det(L(\operatorname{diag} L)^{-1} L^{\mathrm{H}}) = \frac{(\det L)^2}{\det(\operatorname{diag} L)} = \det L,$$

where we used that the determinant of a lower triangular matrix is simply the product of its diagonal entries, i.e. $\det L = \det(\operatorname{diag} L)$. We have shown that $\det X \geq \det L$ for all X and L satisfying (B.10).

Our next goal is to show that for any $X \in \mathcal{H}_+$, the determinant of X is the maximum of $\det L$, over all lower triangular L satisfying (B.10). Let L_X denote the Cholesky factor of X, i.e. $X = L_X L_X^{\mathrm{H}}$, and define

$$L_{ii} = (L_X)_{ii}^2 \text{ for all } i,$$

$$L_{ij} = (L_X)_{ij} \text{ for } i > j.$$

Then X and L satisfy (B.10) and $\det X = \det L$. Consequently, $\det X = \max\{\det L \mid L \text{ satisfying (B.10)}\}$.

In Section B.3, we proposed a semidefinite envelope representation for the function

$$(\prod_{i=1}^{n} L_{ii})^{\alpha}, \quad 0 < \alpha \leq 1/n, \tag{B.11}$$

where α is a floating point number and n denotes the order of X. Combining (B.10) and (B.11), we obtain a semidefinite envelope representation for the concave function

$$(\det X)^{\alpha}, \quad 0 < \alpha \leq 1/n.$$

Nesterov and Nemirovsky [110] gave a semidefinite envelope representation of the above function for a specific choice of α, which is at most $1/(2n)$, and X real symmetric.

B.10 Other Eigenvalue Functions

Envelope representations for sums of eigenvalues are classical. Consider an $n \times n$ Hermitian matrix X, and let $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(X)$ denote its eigenvalues, in descending order (recall that a Hermitian matrix of order n has n real eigenvalues). For any $k \in \{1, 2, \ldots, n\}$, it holds that

$$\sum_{i=1}^{k} \lambda_i(X) = \max\{ \operatorname{tr} P^{\mathrm{H}} X P \mid P^{\mathrm{H}} P = I,\ P \in \mathcal{C}^{n \times k}\},$$

see Fan [35] and Corollary 4.3.18 in [56]. More recently, Overton and Womersley [118] proposed a slightly different envelope representation, viz.

$$\sum_{i=1}^{k} \lambda_i(X) = \max\{ \operatorname{tr} UX \mid \operatorname{tr} U = k,\ 0 \preceq U \preceq I\}.$$

However, neither of the above envelope representations is semidefinite, because $\operatorname{tr} P^{\mathrm{H}} X P$ and $\operatorname{tr} UX$ are nonlinear in the variables P, X and U. Semidefinite envelope representations for $\sum_{i=1}^{k} \lambda_i(X)$ were independently proposed for the real symmetric case by Alizadeh [1] and Nesterov and Nemirovsky [110]. The two constructions are identical, but the derivation in [1] is based on the characterization of Overton and Womersley [118], whereas the derivation of [110] is based on Weyl's characterization of eigenvalues (this is also true for Fan's characterization [35]). An instructive derivation of the semidefinite envelope representation, again based on Weyl's theorem, is given by Jarre [63]. Below, we simply derive the semidefinite envelope representation of $\sum_{i=1}^{k} \lambda_i(X)$ from the spectral decomposition of X. Namely, let $q_1(X), q_2(X), \ldots, q_n(X)$ denote the eigenvectors of X such that

$$X = \sum_{i=1}^{n} \lambda_i(X) q_i(X) q_i(X)^{\mathrm{H}}.$$

Observe that for any $\lambda_0 \in \Re$ and $Y \in \mathcal{H}^{(n)}$, it holds

$$Y \succeq 0 \text{ and } Y \succeq X - \lambda_0 I \tag{B.12}$$

if and only if

$$Y \succeq \sum_{i=1}^{n} \left((\lambda_i(X) - \lambda_0)_+ \, q_i(X) q_i(X)^{\mathrm{H}} \right),$$

where $(\cdot)_+$ denotes the nonnegative part. Hence, (B.12) implies

$$\mathrm{tr}\, Y \geq \sum_{i=1}^{n} (\lambda_i(X) - \lambda_0)_+,$$

with equality if $Y = \sum_{i=1}^{n} \left((\lambda_i(X) - \lambda_0)_+ q_i(X) q_i(X)^{\mathrm{H}} \right)$. Furthermore, we have

- If $\lambda_0 > \lambda_k(X)$ then
$$\sum_{i=1}^{n} (\lambda_i(X) - \lambda_0)_+ \geq (\lambda_0 - \lambda_k(X)) + \sum_{i=1}^{k} (\lambda_i(X) - \lambda_0) > \left(\sum_{i=1}^{k} \lambda_i(X) \right) - k\lambda_0.$$
- If $\lambda_0 < \lambda_{k+1}(X)$ then
$$\sum_{i=1}^{n} (\lambda_i(X) - \lambda_0)_+ \geq (\lambda_{k+1}(X) - \lambda_0) + \sum_{i=1}^{k} (\lambda_i(X) - \lambda_0) > \left(\sum_{i=1}^{k} \lambda_i(X) \right) - k\lambda_0.$$
- If $\lambda_k(X) \geq \lambda_0 \geq \lambda_{k+1}(X)$ then
$$\sum_{i=1}^{n} (\lambda_i(X) - \lambda_0)_+ = (\sum_{i=1}^{k} \lambda_i(X)) - k\lambda_0.$$

The above relations show that

$$\mathrm{tr}\, Y + k\lambda_0 \geq \sum_{i=1}^{k} \lambda_i(X),$$

with equality if and only if

$$Y = \sum_{i=1}^{k} \lambda_i(X) q_i q_i^{\mathrm{H}}, \quad \lambda_k(X) \geq \lambda_0 \geq \lambda_{k+1}(X).$$

Hence, combining (B.12) with the linear equality

$$y_0 = \mathrm{tr}\, Y + k\lambda_0, \tag{B.13}$$

yields a semidefinite envelope representation for the (non-smooth) convex function

$$\sum_{i=1}^{k} \lambda_i(X).$$

Again, some interesting compositions can be made, e.g. let

$$X = -X_1,$$

then $-y_0$ models the sum of the k smallest eigenvalues of X_1, a concave function [1]. We can then consider also convex and quasiconvex fractions of the k_1 largest over the k_2 smallest eigenvalues of a positive semidefinite matrix X with $\operatorname{rank} X > n - k_2$, see Section B.5 and Section B.6. The condition number of X, i.e. $\lambda_1(X)/\lambda_n(X)$, is a special case of such a fraction. Another possibility is to let

$$X \succeq X_1, \quad X \succeq -X_1,$$

then y_0 models the sum of the k largest eigenvalues of X_1 in absolute value, and if we let

$$\begin{bmatrix} X & X_1 \\ & I \end{bmatrix} \succeq 0,$$

then y_0 models the sum of squares of the k largest singular values of a (possibly non Hermitian) matrix X_1. Weighted sums of eigenvalues can be created by convex composite functions of $\sum_{i=1}^{k_1} \lambda_i(X)$ and $\sum_{i=1}^{k_2} \lambda_i(X)$ with $k_1 \neq k_2$, see Alizadeh [1], Rockafellar [133] and Jarre [63].

If we restrict to diagonal X, then (B.12) reduces to n linear inequalities, as can even be handled by linear programming. In this way, we obtain a semidefinite envelope representation for the k largest components of a vector $x \in \Re^n$. Using the composition technique, these components x_i can represent other convex functions, such as convex fractions.

B.11 Convex and Quasiconvex Objective Functions

The preceding discussion has focussed on modeling (generalized) convex constraints with semidefinite programming. It is clear that if a convex function $f(y)$ has a semidefinite envelope representation (B.1), i.e.

$$f(y) = \min_{y_0}\{y_0 \mid (y_0, y, x) \in b + \mathcal{A} \text{ for some } x \in \mathcal{H}_+\},$$

then minimizing the objective function $f(y)$ is equivalent to minimizing its envelope representation y_0, over all feasible y. Unfortunately, the situation is not so straightforward if the objective function is merely quasiconvex.

For quasiconvex fractional objective functions, we can use the homogenization technique of Charnes and Cooper, that was originally proposed for linear fractional programming [24].

Suppose that we want to solve the following problem:

$$\inf \left\{ \frac{y_1}{y_2} \middle| a_{i1}y_1 + a_{i2}y_2 + \sum_{j=3}^{n} A_{ij} \bullet Y_j = b_i, \ Y_j \succeq 0 \, \forall i, j \right\}, \tag{B.14}$$

where $y_2 > 0$ for each $(y_1, y_2, Y_3, \ldots, Y_n)$ satisfying the constraints of (B.14). Recall from Section B.6 that the linear fraction y_1/y_2 in (B.14) may actually represent a quasiconvex fraction of the form (B.5), if y_1 comes from the semidefinite envelope representation of a convex fraction. The Charnes–Cooper transformation of (B.14) results in the model

$$\inf \left\{ x_1 \middle| a_{i1}x_1 - b_i x_2 + \sum_{j=3}^{n} A_{ij} \bullet X_j = -a_{i2}, \ x_2 \geq 0, \ X_j \succeq 0 \, \forall i, j \right\}. \tag{B.15}$$

Obviously, if $(y_1, y_2, Y_3, \ldots, Y_n)$ satisfies the constraints of (B.14), then

$$x_1 = \frac{y_1}{y_2}, \ x_2 = \frac{1}{y_2}, \ X_j = \frac{Y_j}{y_2} \tag{B.16}$$

is feasible for (B.15), and the objective values coincide. Moreover, if the infimum in (B.14) is attained for $(y_1, y_2, Y_3, \ldots, Y_n)$, then $x_1 = y_1/y_2$ is the infimum for (B.15), as can easily be shown by contradiction. Namely, if $x_1', x_2', X_3', \ldots, X_n'$ is feasible for (B.15) and $x_1' < x_1$, then

$$y_1' = \frac{x_1 + x_1'}{x_2 + x_2'}, \ y_2' = \frac{2}{x_2 + x_2'}, \ Y_3 = \frac{2X_3}{x_2 + x_2'}, \ldots, Y_n = \frac{2X_n}{x_2 + x_2'}$$

is feasible for (B.14), but

$$\frac{y_1'}{y_2'} = \frac{x_1 + x_1'}{2} < x_1 = \frac{y_1}{y_2},$$

a contradiction. We have shown that if the infimum in (B.14) is attained, then (B.15) has an optimal solution with $x_2 > 0$, from which an optimal solution for (B.14) follows using the inverse transformation of (B.16). It is however possible that (B.15) has multiple optimal solutions, and x_2 may be zero for some of them. Fortunately, path–following methods, such as discussed in this book, always find a solution with $x_2 > 0$, provided that such a solution exists (this property can be shown using the Güler–Ye argument [52], cf. Lemma 4.4).

Bibliography

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization problems. *SIAM Journal on Optimization*, 5:13–51, 1995.

[2] F. Alizadeh, J.A. Haeberly, and M. Overton. A new primal–dual interior point method for semidefinite programming. In J.G. Lewis, editor, *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, pages 113–117. SIAM, Philadelpha, 1994.

[3] F. Alizadeh, J.A. Haeberly, and M. Overton. Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming*, 77(2):111–128, 1997.

[4] F. Alizadeh, J.A. Haeberly, and M. Overton. Primal–dual interior–point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.

[5] E. Anderson and P. Nash. *Linear programming in infinite dimensional spaces.* John Wiley and Sons, 1987.

[6] M. Avriel, W.E. Diewert, S. Schaible, and I. Zang. *Generalized Concavity*, volume 36 of *Mathematical concepts and methods in engineering.* Plenum Press, 1988.

[7] A. Ben-Israel. Linear equations and inequalities on finite dimensional, real or complex, vector spaces: a unified theory. *Journal of Mathematical Analysis and Applications*, 27:367–389, 1969.

[8] A. Ben-Tal and M.P. Bendsoe. A new method for optimal truss topology design. *SIAM Journal on Optimization*, 3:322–358, 1993.

[9] A. Ben-Tal and A. Nemirovsky. Potential reduction polynomial time method for truss topology design. *SIAM Journal on Optimization*, 4:596–612, 1994.

[10] A. Ben-Tal and A. Nemirovsky. Optimal design of engineering structures. *Optima*, 47:4–9, October 1995.

[11] A.B. Berkelaar, J.F. Sturm, and S. Zhang. Polynomial primal–dual cone affine scaling for semidefinite programming. *Applied Numerical Mathematics*, 29:317–333, 1999.

[12] A. Berman. *Cones, matrices and mathematical programming*, volume 73 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1979.

[13] A. Berman and A. Ben-Israel. Linear inequalities, mathematical programming and matrix theory. *Mathematical Programming*, 1:291–300, 1971.

[14] I. Bongartz, A.R. Conn, I.M. Gould, and Ph.L. Toint. CUTE: Constrained and unconstrained testing environment. Technical Report 18860, IBM T.J. Watson Research Center, Yorktown Heights, New York, USA, 1993.

[15] B. Borchers. CSDP, a C library for semidefinite programming. Technical report, New Mexico Tech Mathematics Faculty, USA, 1997.

[16] J.M. Borwein. A note on perfect duality and limiting Lagrangians. *Mathematical Programming*, 18:330–337, 1980.

[17] J.M. Borwein. The limiting Lagrangian as a consequence of Helly's theorem. *Journal of Optimization Theory and Applications*, 33:497–513, 1981.

[18] J.M. Borwein and H. Wolkowicz. Characterizations of optimality for the abstract convex program with finite dimensional range. *Journal of the Australian Mathematical Society*, 30:390–411, 1981.

[19] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *Journal of Mathematical Analysis and Applications*, 83:495–530, 1981.

[20] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[21] S. Boyd and S. Wu. *SDPSOL. A parser/solver for semidefinite programs with matrix structure.* Stanford, USA, 1995.

[22] R.E. Burkard, S.E. Karisch, and F. Rendl. QAPLIB – a quadratic assignment problem library. *European Journal of Operational Research*, 55:115–119, 1991. Available from `http://www.diku.dk/~karisch/qaplib`.

[23] W.B. Carver. Systems of linear inequalities. *Annals of Mathematics*, 23:212–220, 1921.

[24] A. Charnes and W.W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9:181–186, 1962.

[25] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. *LANCELOT: A Fortran package for large–scale nonlinear optimization (release A)*, volume 17 of *Springer Series in Computation Mathematics*. Springer-Verlag, Berlin, 1992.

[26] B.D. Craven. Modified Kuhn-Tucker conditions when a minimum is not attained. *Operations Research Letters*, 3:47–52, 1984.

[27] B.D. Craven and J.J. Koliha. Generalizations of Farkas' theorem. *SIAM Journal on Mathematical Analysis*, 8:983–997, 1977.

[28] J. Cullum, W.E. Donath, and P. Wolfe. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. *Mathematical Programming Study*, 3:35–55, 1975.

[29] J. Czyzyck and S. Wright. Interior point methods online. Located at URL `http://www.mcs.anl.gov/otc/InteriorPoint`.

[30] I.I. Dikin. Iterative solution of problems of linear and quadratic programming. *Soviet Mathematics Doklady*, 8:674–675, 1967.

[31] A.S. Drud. *CONOPT-A GRG code for large–scale nonlinear optimization — Reference manual*. Denmark, 1992.

[32] R.J. Duffin. Infinite programs. In H.W. Kuhn and A.W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 157–170. Princeton University Press, Princeton, NJ, 1956.

[33] A.S. El-Bakry, R.A. Tapia, and Y. Zhang. On the convergence rate of Newton interior point methods in the absence of strict complementarity. *Computational Optimization and Applications*, 6:157–167, 1996.

[34] M. Epelman and R.M. Freund. Condition number complexity of an elementary algorithm for resolving a conic linear system. Technical report, Massachusetts Institute of Technology, Operations Research Center, Cambridge, Massachusetts, USA, 1997.

[35] Ky Fan. On a theorem of Weyl concerning the eigenvalues of linear transformations. In *Proceedings of the National Academy of Sciences of U.S.A.*, volume 35, pages 652–655. National Academy of Sciences, 1949.

[36] Ky Fan. On systems of linear inequalities. In H.W. Kuhn and A.W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 99–156. Princeton University Press, Princeton, NJ, 1956.

[37] J. Farkas. Über die Theorie des einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik*, 124:1–24, 1902.

[38] J.B.G. Frenk, J.F. Sturm, and S. Zhang. An interior-point based subgradient method for nondifferentiable convex optimization. *Optimization Methods and Software*, 10(2):197–215, 1998. Special Issue Celebrating the 65th Birthday of Professor Masao Iri.

[39] K. Glashoff and S. Gustafson. *Linear optimization and approximation*, volume 45 of *Applied Mathematical Sciences*. Springer-Verlag, Basel, 1978.

[40] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfyability problems using semidefinite programming. *Journal ACM*, 42:1115–1145, 1995.

[41] D. Goldfarb and K. Scheinberg. Interior point trajectories in semidefinite programming. Technical report, Columbia University, Department of IEOR, New York, USA, 1996.

[42] D. Goldfarb and K. Scheinberg. On parametric semidefinite programming. Technical report, Columbia University, Department of IEOR, New York, USA, 1997.

[43] A.J. Goldman and A.W. Tucker. Polyhedral convex cones. In H.W. Kuhn and A.W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 19–40. Princeton University Press, Princeton, NJ, 1956.

[44] A.J. Goldman and A.W. Tucker. Theory of linear programming. In H.W. Kuhn and A.W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 53–98. Princeton University Press, Princeton, NJ, 1956.

[45] C.C. Gonzaga. Polynomial affine algorithms for linear programming. *Mathematical Programming*, 49:7–21, 1990.

[46] C.C. Gonzaga. Interior path following algorithms. In J.R. Birge and K.G. Murty, editors, *Mathematical Programming, State of the Art 1994*, pages 93–101, Ann Arbor, 1994. University of Michigan.

[47] C.C. Gonzaga. The largest step path following algorithm for monotone linear complementarity problems. *Mathematical Programming*, 76:309–332, 1997.

[48] C.C. Gonzaga and R.A. Tapia. On the quadratic convergence of the simplified Mizuno–Todd–Ye algorithm for linear programming. *SIAM Journal on Optimization*, 7(1):66–85, 1997.

[49] P. Gordan. Über die Auflösungen linearer Gleichungen mit reelen Coefficienten. *Mathematische Annalen*, 6:23–28, 1873.

[50] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, New York, Berlin, 1988.

[51] O. Güler. Limiting behavior of weighted central paths in linear programming. *Mathematical Programming*, 65:347–363, 1994.

[52] O. Güler and Y. Ye. Convergence behavior of interior-point algorithms. *Mathematical Programming*, 60:215–228, 1993.

[53] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. An interior–point method for semidefinite programming. *SIAM Journal on Optimization*, 6:342–361, 1996.

[54] N.J. Higham. *Accuracy and stability of numerical algorithms*. SIAM, Philadelphia, 1996.

[55] A.J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.

[56] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, New York, 1985. Corrected reprint 1990.

[57] J.E. Ingersoll. *Theory of financial decision making*. Rowman & Littlefield, Totowa, NJ, 1987.

[58] B. Jansen. *Interior point techniques in optimization. Complexity, sensitivity and algorithms*. PhD thesis, Faculty of Mathematics and Computer Science, Technical University Delft, Delft, The Netherlands, December 1995. To be published by Kluwer.

[59] B. Jansen, C. Roos, and T. Terlaky. The theory of linear programming: skew symmetric self–dual problems and the central path. *Optimization*, 29:225–233, 1994.

[60] B. Jansen, C. Roos, and T. Terlaky. A polynomial Dikin–type primal–dual algorithm for linear programming. *Mathematics of Operations Research*, 21:341–353, 1996.

[61] B. Jansen, C. Roos, T. Terlaky, and J.-Ph. Vial. Primal–dual target following algorithms for linear programming. *Annals of Operations Research*, 62:197–231, 1996.

[62] F. Jarre. An interior–point method for minimizing the maximum eigenvalue of a linear combination of matrices. *SIAM Journal on Control and Optimization*, 31:1360–1377, 1993.

[63] F. Jarre. Innere–Punkte–Methoden mit Anwendungen auf kombinatorische Optimierungsprobleme. Lecture notes, Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Am Hubland, D–97074 Hubland, Germany, 1996. In German.

[64] R.G. Jeroslow. A limiting Lagrangian for infinitely convex optimization in $\Re^n$. *Journal of Optimization Theory and Applications*, 33:479–495, 1981.

[65] N.K. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[66] D.F Karney. Duality gaps in semi–infinite linear porgramming — an approximation problem. *Mathematical Programming*, 20:129–143, 1981.

[67] D.F. Karney and T.D. Morley. Limiting Lagrangians: a primal approach. *Journal of Optimization Theory and Applications*, 48:163–174, 1986.

[68] W.K. Klein Haneveld. Enkele stochastische optimaliseringsproblemen en hun dualen. Lecture notes of LNMB course, Department of Econometrics, University of Groningen, 1994. In Dutch.

[69] E. de Klerk, C. Roos, and T. Terlaky. A nonconvex weighted potential function for polynomial target following methods. Technical Report 95–127, Delft University of Technology, Faculty of Technical Mathematics and Informatics, Delft, The Netherlands, 1995.

[70] E. de Klerk, C. Roos, and T. Terlaky. Initialization in semidefinite programming via a self–dual skew–symmetric embedding. *Operations Research Letters*, 20(5):213–221, 1997.

[71] E. de Klerk, C. Roos, and T. Terlaky. Infeasible start semidefinite programming algorithms via self–dual embeddings. In *The Fields Institute for Research in Mathematical Sciences, Communications Series*. 1998.

[72] E. de Klerk, C. Roos, and T. Terlaky. Polynomial primal–dual affine scaling algorithms in semidefinite programming. *Journal of Combinatorial Optimization*, 2(1):51–70, 1998.

[73] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*. Springer-Verlag, Berlin, 1991.

[74] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial-time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[75] M. Kojima, S. Mizuno, and A. Yoshise. A primal–dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in mathematical programming: interior point and related methods*, pages 29–37. Springer-Verlag, New York, 1989.

[76] M. Kojima, M. Shida, and S. Shindoh. Local convergence of predictor-corrector infeasible–interior-point algorithms for SDPs and SDLCPs. Technical Report B-306, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1995.

[77] M. Kojima, M. Shida, and S. Shindoh. A note on the Nesterov–Todd and the Kojima–Shindoh–Hara search directions in semidefinite programming. Technical Report B–313, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1996.

[78] M. Kojima, M. Shida, and S. Shindoh. A predictor–corrector interior–point algorithm for the semidefinite linear complementarity problem using the alizadeh–haeberly–overton search direction. Technical Report B–311, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1996.

[79] M. Kojima, M. Shida, and S. Shindoh. Search directions in the SDP and the monotone SDLCP: Generalization and inexact computation. Technical Report B–327, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1997.

[80] M. Kojima, S. Shindoh, and S. Hara. Interior–point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7(1):86–125, 1997.

[81] S.Y. Kung, C.K. Lo, and R. Foka. A Toeplitz approximation approach to coherent source direction finding. In *International Conference on ASSP*, Tokyo, 1986.

[82] C.J. Lin and R. Saigal. An infeasible start predictor corrector method for semi–definite linear programming. Technical report, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, USA, 1995.

[83] Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[84] Z.-Q. Luo. private communication, 1996.

[85] Z.-Q. Luo, J.F. Sturm, and S. Zhang. Duality and self-duality for conic convex programming. Technical Report 9620/A, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1996.

[86] Z.-Q. Luo, J.F. Sturm, and S. Zhang. Duality results for conic convex programming. Technical Report 9719/A, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1997.

[87] Z.-Q. Luo, J.F. Sturm, and S. Zhang. Superlinear convergence of a symmetric primal–dual path following algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8(1):59–81, 1998.

[88] I.J. Lustig, R.E. Marsten, and D.F. Shanno. Computational experience with a primal–dual interior point method for linear programming. *Linear Algebra and its Applications*, 152:191–222, 1991.

[89] I.J. Lustig, R.E. Marsten, and D.F. Shanno. On implementing Mehrotra's predictor–corrector interior point method for linear programming. *SIAM Journal on Optimization*, 2:435–449, 1992.

[90] N. Megiddo. Pathways to the optimal solution set in linear programming. In N. Megiddo, editor, *Interior–Point and Related Methods*, Progress in Mathematical Programming, pages 131–158. Springer-Verlag, New York, 1989.

[91] S. Mehrotra. On the implementation of a primal–dual interior point method. *SIAM Journal on Optimization*, 2:575–601, 1992.

[92] B. Mirkin. Approximation clustering: a mine of semidefinite programming problems. Presented during the workshop on semidefinite programming and interior–point approaches for combinatorial optimization problems in Toronto, May 15–17 1996.

[93] S. Mizuno. A new polynomial time method for a linear complementarity problem. *Mathematical Programming*, 56(1):31–43, 1992.

[94] S. Mizuno. A superlinearly convergent infeasible-interior-point algorithm for geometical LCPs without a strictly complementarity condition. *Mathematics of Operations Research*, 2:382–400, 1996.

[95] S. Mizuno, M.J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 1993.

[96] R.D.C. Monteiro. Polynomial convergence of primal–dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions. Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1996.

[97] R.D.C. Monteiro. Primal–dual path following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7(3):663–678, 1997.

[98] R.D.C. Monteiro and I. Adler. Interior path following primal-dual algorithms. part 1: linear programming. *Mathematical Programming*, 44:27–41, 1989.

[99] R.D.C Monteiro, I. Adler, and M.G.C. Resende. A polynomial–time primal–dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. *Mathematics of Operations Research*, 15:191–214, 1990.

[100] R.D.C. Monteiro and J.-S. Pang. On two interior point mappings for nonlinear semidefinite complementarity problems. *Mathematics of Operations Research*, 23:39–60, 1998.

[101] R.D.C. Monteiro and T. Tsuchiya. Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem. *Mathematics of Operations Research*, 21(4):793–814, 1996.

[102] R.D.C. Monteiro and T. Tsuchiya. Polynomial convergence of a new family of primal–dual algorithms for semidefinite programming. Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1996.

[103] R.D.C. Monteiro and S. Wright. Local convergence of interior–point algorithms for degenerate monotone LCP. *Computational Optimization and Applications*, 3:131–155, 1993.

[104] R.D.C. Monteiro and P.R. Zanjácomo. A note on the existence of the Alizadeh–Haeberly–Overton direction for semidefinite programming. Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1996. Revised January, 1997.

[105] R.D.C. Monteiro and Y. Zhang. A unified analysis for a class of path–following primal–dual interior point algorithms for semidefinite programming. *Mathematical Programming*, 81:281–299, 1998.

[106] M. Muramatsu. Affine scaling algorithm fails for semidefinite programming. Technical Report 16, Department of Mechanical Engineering, Sophia University, Japan, 1996.

[107] M. Muramatsu and R.J. Vanderbei. Primal–dual affine scaling fails for semidefinite programming. Technical report, Department of Mechanical Engineering, Sophia University, Japan, 1997.

[108] W. Murray. Algorithms for large nonlinear problems. In J.R. Birge and K.G. Murty, editors, *Mathematical Programming, State of the Art 1994*, pages 172–185, Ann Arbor, 1994. University of Michigan.

[109] B.A. Murtagh and M.A. Saunders. MINOS 5.4 user's guide. Technical Report SOL 93–20R, Department of Operations Research, Stanford University, Stanford, USA, 1993.

[110] Y. Nesterov and A. Nemirovsky. *Interior point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, 1994.

[111] Y. Nesterov and A. Nemirovsky. An interior point method for generalized linear–fractional problems. *Mathematical Programming*, 69:117–204, 1995.

[112] Y. Nesterov and M.J. Todd. Self–scaled barriers and interior–point methods for convex programming. *Mathematics of Operations Research*, 22(1):1–42, 1997.

[113] Y. Nesterov and M.J. Todd. Primal–dual interior–point methods for self–scaled cones. *SIAM Journal on Optimization*, 8:324–364, 1998.

[114] Y. Nesterov, M.J. Todd, and Y. Ye. Infeasible–start primal–dual methods and infeasibility detectors for nonlinear programming problems. Technical Report 1156, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1996.

[115] J.A. Olkin. Using semidefinite programming for controller design in active noise control. *SIAG/OPT Views-and-News*, 8:1–5, Fall 1996.

[116] J.A. Olkin and P.J. Titterton. Using semi–definite programming for H^2 controller design with multiple simultaneous H^2 constraints. *Journal of VLSI Signal Processing*, 14, 1996.

[117] M.L. Overton. Large scale optimization of eigenvalues. *SIAM Journal on Optimization*, 2:88–120, 1992.

[118] M.L. Overton and R.S. Womersley. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Mathematical Programming*, 62:321–357, 1993.

[119] P. Pardalos. On the passage from local to global in optimization. In J.R. Birge and K.G. Murty, editors, *Mathematical Programming, State of the Art 1994*, pages 220–247, Ann Arbor, 1994. University of Michigan.

[120] F.A. Potra and R. Sheng. A superlinearly convergent primal–dual infeasible–interior–point algorithm for semidefinite programming. Technical Report 78, Department of Mathematics, The University of Iowa, Iowa City, USA, 1995.

[121] F.A. Potra and R. Sheng. Superlinear convergence of a predictor–corrector method for semidefinite programming without shrinking central path neighborhood. Technical Report 91, Department of Mathematics, The University of Iowa, Iowa City, USA, 1996.

[122] F.A. Potra and R. Sheng. Superlinear convergence of interior point algorithms for semidefinite programming. Technical Report 86, Department of Mathematics, The University of Iowa, Iowa City, USA, 1996.

[123] F.A. Potra and R. Sheng. On homogeneous interior–point algorithms for semidefinite programming. *Optimization Methods and Software*, 9:161–184, 1998. Special Issue Celebrating the 60th Birthday of Professor Naum Shor.

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

[125] M.V. Ramana and R. Freund. One the ELSD duality theory for SDP. Technical report, Center of Applied Optimization, University of Florida, Gainesville, Florida, USA, 1996.

[126] M.V. Ramana and P.M. Pardalos. Semidefinite programming. In T. Terlaky, editor, *Interior point methods of mathematical programming*, volume 5 of *Applied Optimization*, chapter 9, pages 369–398. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996. Series editors P.M. Pardalos and D. Hearn.

[127] M.V. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. Technical Report 85-12, University of Waterloo, Waterloo, Canada, 1995. To appear in *SIAM Journal on Optimization* 7 (1997) 3.

[128] F. Rendl and H. Wolkowicz. A projection technique for partitioning the nodes of a graph. Technical Report 90–20, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, 1990.

[129] J. Renegar. A polynomial time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40:59–93, 1988.

[130] J. Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65(1):73–91, 1994.

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

[132] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[133] R.T. Rockafellar. Nonsmooth optimization. In J.R. Birge and K.G. Murty, editors, *Mathematical Programming, State of the Art 1994*, pages 248–258, Ann Arbor, 1994. University of Michigan.

[134] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and algorithms for linear optimization. An interior point approach.* Series in discrete mathematics and optimization. John Wiley & Sons, New York, 1997.

[135] C. Scherer, P. Gahinet, and M. Chilali. Multi–objective output–feedback control via LMI optimization. Technical report, Mechanical Engineering Systems and Control Group, Delft University of Technology, Delft, The Netherlands, 1995.

[136] A. Shapiro and J.D. Botha. Dual algorithms for orthogonal Procrustes rotations. *SIAM Journal on Matrix Analysis and Applications*, 9:378–383, 1988.

[137] R. Sheng, F.A. Potra, and J. Ji. On a general class of interior–point algorithms for semidefinite programming with polynomial complexity and superlinear convergence. Technical Report 89, Department of Mathematics, The University of Iowa, Iowa City, USA, 1996.

[138] M. Shida, S. Shindoh, and M. Kojima. Existence and uniqueness of search directions in interior–point algorithms for the SDP and the monotone SDLCP. *SIAM Journal on Optimization*, 8(2):387–396, 1998.

[139] E. Stiemke. Über positiven Lösungen homogener linearer Gleichungen. *Mathematische Annalen*, 76:340–342, 1915.

[140] J.F. Sturm. Superlinear convergence of an algorithm for monotone linear complementarity problems, when no strictly complementary solution exists. *Mathematics of Operations Research*, 24(1):72–94, 1999.

[141] J.F. Sturm and S. Zhang. A dual and interior point approach to solve convex min–max problems. In D.-Z. Du and P.M. Pardalos, editors, *Minimax and Applications*, pages 69–78. Kluwer Academic Publishers, The Netherlands, 1995.

[142] J.F. Sturm and S. Zhang. Potential reduction method for harmonically convex programming. *Journal of Optimization Theory and Applications*, 84:181–205, 1995.

[143] J.F. Sturm and S. Zhang. New complexity results for the Iri-Imai method. *Annals of Operations Research*, 62:539–564, 1996.

[144] J.F. Sturm and S. Zhang. On weighted centers for semidefinite programming. Technical Report 9636/A, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1996.

[145] J.F. Sturm and S. Zhang. An $O(\sqrt{n}L)$ iteration bound primal–dual cone affine scaling algorithm for linear programming. *Mathematical Programming*, 72:177–194, 1996.

[146] J.F. Sturm and S. Zhang. On a wide region of centers and primal–dual interior point algorithms for linear programming. *Mathematics of Operations Research*, 22(2):408–431, 1997.

[147] J.F. Sturm and S. Zhang. An interior point method, based on rank-one updates, for linear programming. *Mathematical Programming*, 81:77–87, 1998.

[148] J.F. Sturm and S. Zhang. On the long-step path-following method for semidefinite programming. *Operations Research Letters*, 22:145–150, 1998.

[149] J.F. Sturm and S. Zhang. Symmetric primal–dual path following algorithms for semidefinite programming. *Applied Numerical Mathematics*, 29:301–315, 1999.

[150] T. Terlaky. On ℓ_p programming. *European Journal of Operational Research*, 22:70–10, 1985.

[151] M.J. Todd, K.C. Toh, and R.H. Tütüncü. On the Nesterov–Todd direction in semidefinite programming. *SIAM Journal on Optimization*, 8(3):769–796, 1998.

[152] M.J. Todd and Y. Ye. A centered projective algorithm for linear programming. *Mathematics of Operations Research*, 15:508–529, 1990.

[153] M.J. Todd and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible–point algorithms for linear programming. *Mathematical Programming*, 81:1–21, 1998.

[154] P. Tseng. Search directions and convergence analysis of some semidefinite path–following methods for the monotone semi–definite LCP. Technical report, Department of Mathematics, University of Washington, Seattle, Washington, USA, 1996.

[155] A.W. Tucker. Dual systems of homogeneous linear relations. In H.W. Kuhn and A.W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 3–18. Princeton University Press, Princeton, NJ, 1956.

[156] M. Ujvàri. Simple proofs of duality theorems in semidefinite and cone linear programming. Technical report, Department of Operations Research, Eötvös Lorànd University, Budapest, Hungary, 1996.

[157] L. Vandenberghe and S. Boyd. A primal–dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69:205–236, 1995.

[158] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.

[159] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. Technical report, Information Systems Laboratory, Electrical Engineering Department, Stanford University, Stanford, USA, April 1996.

[160] H. Wolkowicz. Some applications of optimization in matrix theory. *Linear Algebra and its Applications*, 40:101–118, 1981.

[161] G. Xue and Y. Ye. An efficient algorithm for minimizing a sum of Euclidean norms with applications. *SIAM Journal on Optimization*, 7(4):1017–1036, 1997.

[162] Y. Ye and K. Anstreicher. On quadratic and O($\sqrt{n}L$) convergence of a predictor-corrector algorithm for LCP. *Mathematical Programming*, 62:537–551, 1993.

[163] Y. Ye, M.J. Todd, and S. Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self–dual linear programming algorithm. *Mathematics of Operations Research*, 19:53–67, 1994.

[164] Y. Zhang. On extending some primal–dual interior–point algorithms from linear programming to semidefinite programming. *SIAM Journal on Optimization*, 8(2):365–386, 1998.

[165] Y. Zhang and D. Zhang. On polynomiality of the Mehrotra–type predictor–corrector interior–point algorithms. *Mathematical Programming*, 68:303–318, 1995.

[166] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. Technical Report 95-27, University of Waterloo, Waterloo, Canada, 1996.

Index

Samenvatting

Semidefiniet programmeren vormt een belangrijke klasse van optimalisatie problemen onder convexe nevenvoorwaarden. Kenmerkend voor semidefiniet programmeren is dat deze nevenvoorwaarden worden gespecificeerd door middel van lineaire matrix ongelijkheden. De vele toepassingsmogelijkheden enerzijds en de fraaie wiskundige eigenschappen anderzijds maken semidefiniet programmeren tot een aantrekkelijk onderzoeksgebied. De al wat langer bekende toepassingen in systeemtheorie en grafentheorie zijn met name een grote stimulans geweest voor de snelle ontwikkeling van semidefiniet programmeren. In eerste instantie werden semidefiniete programmeringsproblemen wel opgelost met de ellipsoide methode, maar gelukkig beschikken we nu over een heel wat efficientere methode, namelijk de inwendige punt methode.

De inwendige punt methode maakt het mogelijk om relatief grote semidefiniete programmeringsproblemen in korte tijd op te lossen. Belangrijk voor het succes van deze oplostechniek zijn de fraaie dualiteitseigenschappen van semidefiniete programmeringsproblemen. In de dualiteitstheorie wordt aan het primale (d.w.z. het originele) semidefiniete programmeringsprobleem een duaal semidefiniet programmeringsprobleem gekoppeld. In termen van primale en duale variabelen kunnen optimaliteitscriteria worden geformuleerd. Bovendien is het mogelijk om op basis van een duale oplossing een afschatting te geven van de mogelijk aanwezige ruimte tot verbetering van de primale oplossing.

Binnen de optimalisering neemt lineair programmeren traditioneel een centrale plaats in. De theorie van lineair programmeren is bijzonder goed ontwikkeld, en er is betrouwbare programmatuur op de markt om lineaire programmeringsproblemen op te lossen. Het is daarom interessant om te weten in hoeverre het mogelijk is om resultaten uit de lineaire programmering te generaliseren naar de semidefiniete programmering. Van lineair naar semidefiniet is echter een grote stap, en hierbij dienen zich veel complicaties aan.

Dit proefschrift draagt bij aan de ontwikkeling van efficiente en betrouwbare technieken om semidefiniete programmeringsproblemen op te lossen. Er is hierbij voor gekozen om de uiterst succesvolle primaal-duaal inwendige punt methode voor lineair programmeren uit te breiden naar semidefiniet programmeren. In deze methode wordt een zodanige rij van primale en duale oplossingen gegenereerd dat ieder limiet punt ervan een optimaal primaal–duaal paar is. De primale component van een optimaal primaal–duaal paar is een optimale oplossing van het originele (primale) semidefiniete programmeringsprobleem, terwijl de duale component een certificaat van optimaliteit is. Als leidraad voor het genereren van een rij primale en duale oplossingen wordt vaak het primaal–duaal centrale pad gebruikt. Het primaal–duaal centrale pad is een gladde curve van primale en duale oplossingen, die eindigt in een optimaal primaal–duaal paar. Inwendige punt methoden die bij benadering dit centrale pad volgen worden wel primaal–duaal padvolgende methoden genoemd.

Er zijn verschillende manieren om het primaal–duaal centrale pad te karakteriseren, en deze verschillende manieren leveren niet allemaal dezelfde zoekrichting op, in tegenstelling tot de situatie in lineair programmeren. In dit proefschrift wordt het centrale pad gedefinieerd in termen van de eigenwaarden van het matrix–produkt van primale en duale matrix–variabelen. In onze analyse richten we ons voornamelijk op de ontwikkeling van deze eigenwaarden. Aangezien deze eigenwaarden worden aangegeven met de letter V, spreken we hier wel van de V–ruimte aanpak. Binnen de V–ruimte aanpak blijkt het niet moeilijk om resultaten van polynomiale convergentie uit de lineaire programmering uit te breiden naar het semidefiniete geval.

Zoals de naam al aangeeft, wordt in de primaal–duaal inwendige punt methode gewerkt met inwendige oplossingen voor het primale en duale probleem. In het algemeen hoeven zulke oplossingen echter niet te bestaan. De techniek van de zelf–duale herformulering stelt ons in staat om toch in alle gevallen gebruik te kunnen maken van de inwendige punt methode. Bij toepassing van deze techniek moet in semidefiniet programmeren rekening worden gehouden met enige lastige situaties, die zich niet in lineair programmeren voor kunnen doen. Deze lastige gevallen krijgen dan ook bijzondere aandacht.

Anders dan in de lineaire programmering is het bij semidefiniet programmeren in het algemeen niet mogelijk om in een eindig aantal stappen tot een exacte oplossing te komen. Het is echter wel redelijk om te streven naar oplossingen met een hoge nauwkeurigheid. Als een hoge mate van nauwkeurigheid wordt vereist, dan is het wenselijk om gebruik te maken van een superlineair convergent algoritme. In dit proefschrift wordt superlineaire convergentie bewezen voor een padvolgend algoritme, onder de voorwaarde van strikte complementariteit.

Tenslotte worden in dit proefschrift verschillende technieken besproken om de globale lineaire convergentie op te krikken. Zo wordt de stap gemaakt van padvolgende algoritmen naar algoritmen waarin een centraal gebied wordt gevolgd. In de methode van het centrale gebied kunnen vaak grotere stappen worden genomen, wat resulteert in een globaal snellere convergentie. Verder besteden we aandacht aan tweede–orde centraliteitscorrecties. Toepassing van dergelijke correcties op de zoekrichting blijkt ook een zeer effectieve manier te zijn om het aantal benodigde iteraties te reduceren.

About the Author

Jos F. Sturm was born on August 13, 1971 in Rotterdam. He studied Econometrics at the University of Groningen from 1989 until 1993. At the same university, he was appointed as a student–assistant in statistics and econometrics from 1991 until 1993. After his graduation in 1993, he became a research assistant at the Tinbergen Institute of Erasmus University Rotterdam. Under supervision of Shuzhong Zhang, he specialized in interior point methods, for solving various types of mathematical programming problems.

In 1997, the Netherlands Organization for Scientific Research (NWO) awarded him a TALENT stipendium, to be spent as a postdoctoral fellow at McMaster University, Canada.

The Tinbergen Institute is the Netherlands Research Institute and Graduate School for Economics and Business, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University in Rotterdam, the University of Amsterdam and the Free University in Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. If available, trade editions of the books which are published in the Tinbergen Institute Research Series can be ordered through Thesis Publishers, P.O. Box 14791, 1001 LG Amsterdam, the Netherlands, phone: +3120 6255429; fax: +3120 6203395. The following books recently appeared in this series:

89. A.I. MARTINS BOTTO DE BARROS, *Discrete and fractional programming techniques for location models.*
90. J.A. DOS SANTOS GROMICHO, *Quasiconvex optimization and location theory.*
91. R.B. KOOL, *Aspects of enlarging the market effects in the Dutch health care.*
92. B. LEEFTINK, *The desirability of currency unification theory and some evidence.*
93. A. VAN VLIET, *Heuristic and optimal methods for the three-dimensional bin packing problem.*
94. C.C.J.M. HEYNEN, *Models for option evaluation in alternative price-movements.*
95. Y. VAN EVERDINGEN, *Adoption and diffusion of the European currency unit. An empirical study among European companies.*
96. R. DE VILDER, *Endogenous business cycles.*
97. J.J. STIBORA and A. DE VAAL, *Services and services trade: a theoretical inquiry.*
98. R. VAN DER BIE, *"Een doorlopende groote roes". De economische ontwikkeling van Nederland, 1913-1921.*
99. W.J. JANSEN, *International capital mobility and asset demand. Six empirical studies.*
100. N.E. STROEKER, *Second-hand markets for consumer durables.*
101. A. VAN DEN BERG, *Trade union growth and decline in the Netherlands.*
102. P. VAN DER LAAG, *An analysis of refinement operators in inductive logic programming.*
103. E. BEINAT, *Multiattribute value functions for environmental management.*
104. H.A. VAN KLINK, *Towards the borderless mainport Rotterdam. An analysis of functional, special and administrative dynamics in port systems.*
105. W.H.J. HASSINK, *Worker flows and the employment adjustment of firms. An empirical investigation.*
106. S. LIU, *Contributions to matrix calculus and applications in econometrics.*

107. M. VAN PRAAG, *Determinants of successful entrepreneurship.*

108. E. VERHOEF, *Economic efficiency and social feasibility in the regulation of road transport externalities.*

109. I. BRAMEZZA, *The competitiveness of a European city and the role of urban management in improving the city's performance. The cases of the Central Veneto and Rotterdam regions.*

110. G. BIESSEN, *East European foreign trade and system changes.*

111. A. LUCAS, *Outlier robust unit root analysis.*

112. T. FOKKEMA, *Residential moving behaviour of the elderly: an explanatory analysis for the Netherlands.*

113. T. EROL, *Exchange rate policy in Turkey. External competitiveness and internal stability studied through a macromodel.*

114. T. DE RUYTER VAN STEVENINCK, *The impact of capital imports on the Argentine economy, 1970-1989*

115. D. DANNENBURG, *Actuarial credibility models: evaluations and extensions.*

116. C. FOLKERTSMA, *On equivalence scales.*

117. A. PLAAT, *Research re: search & re-search.*

118. A.W. VAN DER KROGT, *Service development in less developed countries. The case of telecom services in Chile and Argentina.*

119. O.W. STEENBEEK, *Financial regulation in Japan. Systemic risks and the Nikkei futures market.*

120. T.R.P.J. KROES, *Financial intermediation and monetary transmission.*

121. T.H.F. CHEUK, *Exotic options.*

122. K. TEPLA, *Accounting for the quality of work-life-putting price tags on sources of job (dis)satisfaction.*

123. R.H.M. EMMERINK, *Information and pricing in road transport: theoretical and applied models.*

124. T.J.S. OFFERMAN, *Beliefs and decision rules in public good games – theory and experiments.*

125. F. MERCURIO, *Claim pricing and hedging under market imperfections.*

126. A.W.M. ODÉ, *Migrant workers in the Dutch labour market today.*

127. R.E. WILDEMAN, *The art of grouping maintenance.*

128. M.T. ROCHA DE MAGALHÃES MELO, *Stochastic lot-sizing in production planning: strategies for make-to-order and make-to-stock.*

129. G. RUSSO, *Firms' recruitment behaviour: an empirical investigation of the use of search channels, the rate of arrival of applicants, and the spatial radius of search.*

130. J.N. VAN OMMEREN, *Commuting and relocation of jobs and residences: a search perspective.*

131. F.N. GOUWELEEUW, *A general approach for computing loss probabilities in finite-buffer queues.*

132. M.R.H. MANDJES, *Rare event analysis of communication networks.*

133. E.J. VISSER, *Local sources of competativeness: spatial clustering and organisation dynamics in small-scale clothing in Lima, Peru.*

134. M.H.H. HUYSMAN, *Dynamics of organizational learning.*

135. J.K. GOEREE, *Applications of discrete choice models.*

136. N.J. VAN DER LIJN, *Consumer markets and welfare under central planning.*

137. E.A.A. DE LAAT, *Essays on patent policy: the multi-dimensionality of patents and asymmetric information.*

138. C.M. VAN DER TAK, *International commodity-related environmental agreements: an analysis of international environmental agreements for domestic environmental externalities.*

139. M.P.E. MARTENS, *Interaction between financial markets.*

140. L.W.N. VAN DER TORRE, *Reasoning about obligations. Defeasibility in preference-based deontic logic.*

141. W. BOLT, *Theory of negotiation: a game-theoretic approach.*

142. M.A.J. SMITH, *On the availability of failure prone systems.*

143. R. SLOOF, *Game-theoretic models of the political influence of interest groups.*

144. J.N.M. WIT, *Dynamics and information aggregation in elections and markets.*

145. N.T. BUDINA, *Essays on consistency of fiscal and monetary policy in Eastern Europe.*

146. F. LIPPI, *Central Bank independence and credibility: essays on delegation arrangements for monetary policy.*

147. H. HOEK, *Variable trends: a Bayesian perspective.*

148. E.J.S. PLUG, *Leyden welfare and beyond.*

149. C.S.J. JIE-A-JOEN, *Strategic trade policy, multinational firms, and vertically related industries.*

150. R.G.J. DEN HERTOG, *Empirical essays on growth, uncertainty and prediction.*

151. M. GELDERMAN, *Success of management support systems: a review of the literature and an empirical investigation.*

152. M.A. VAN DER ENDE, *On the values of jobs and specific training.*

153. M. KLAPWIJK, *Rural industry clusters in central Java, Indonesia; an empirical assessment of their role in rural industralization.*

154. R.J.G. VENNIKER, *Transition processes from constrained to Walrasian equilibrium.*

155. J.M. MORALEDA, *On the pricing of interest-rate options.*