Gerald Gruber

On Semidefinite Programming and Applications in Combinatorial Optimization

A dissertation submitted to the
Department of Mathematics
University of Technology, Graz
in candidacy for the degree of
Doctor of Technical Sciences

Supervisor: Professor Dr. Franz Rendl

Abstract

In the $1990's$ striking advances have been achieved in Semidefinite Programming. It is an extension of Linear Programming and belongs to the class of conic convex programming, which means optimizing a linear function subject to the constraint set given by the intersection of an affine space with a closed cone. This dissertation covers Semidefinite Programming in conjunction with Combinatorial Optimization.

This presentation is subdivided into three parts. Part One consisting of two chapters serves as the introductory part. Chapter 1 deals with the semidefinite programming problem in standard form and provides a review of the essential aspects concerning duality and interior-point methods. In chapter 2 we survey recent convergence results and propose a global convergence proof based on the concept of closed mappings.

Part Two presents algorithmic aspects. In chapter 3 we consider linear-quadratic functionals in Semidefinite Programming. We propose a primal-dual path-following algorithm, show that it generates iterates which remain in a compact set and establish the global convergence of the algorithm. In chapter 4 we consider classes of ill-posed semidefinite problems. This chapter provides a numerically stable interior-point approach on how to handle semidefinite problems with zero duality gap for which Slater's condition fails for at least one of the primal and dual problem.

Most of the motivation in Semidefinite Programming comes from the fact that it has found many applications in various fields, including control theory and Combinatorial Optimization. Part Three is devoted to two classical problems in Combinatorial Optimization. In chapter 5 we consider the maximum stable set problem. We investigate various improvements of the well known $\vartheta-$function and present computational results. The last chapter deals with the maximum cut problem. We present an approach to approximate semidefinite relaxations through purely Linear Programming based techniques. First computational results on chordal graphs are very promising and make this an attractive alternative for large scale problems.

Acknowledgments

I am extremely grateful to my supervisor Prof. Franz Rendl, who has introduced me to Semidefinite Programming and its numerous applications in various fields. He led me into active research in this area. His suggestions and innovating ideas have established the basis of this dissertation. I thank him for discussing with me anytime and for pointing out many errors in earlier versions of this manuscript. I had excellent working conditions at his institute all the time. I started my dissertation at the Institute of Mathematics, headed by Prof. Rainer Burkard, at the Technical University Graz. I thank him for providing a stimulating working environment during that time. My thanks also go to the colleagues who encouraged me with their constructive criticism.

I gratefully acknowledge partial financial support from the Austrian Science Foundation (Fonds zur Förderung der wissenschaftlichen Forschung FWF) under Project P12660-MAT.

And above all, my gratitude goes to my friend Christa and to my parents, who gave so much during the years of my education. This dissertation is theirs too. I dedicate this work to them.

April, 2000 Gerald Gruber

Contents

Notations

Below a list of principal symbols used throughout this work is presented for easy reference. Standard notations in mathematics are not listed explicitely. Additional ones have to be used in a minor context. They are all defined in the text as they occur.

The symbols are listed roughly in the order of occurence in chapter sequence. The first digit or a single digit refers to the number of the chapter and the second (if there is one) to the section within the chapter in which the following definitions occur the first time. (An A refers to appendix A.)

LP	Linear Programming: 1
SDP	Semidefinite Programming: 1
$S(n)$	the space of $n \times n$ symmetric matrices: 1.1
$\mathbb{R}^n$	the space of real $n-$dimensional vectors: 1.1
(PSDP)	primal (SDP)-problem in standard form: 1.1
(DSDP)	Lagrangian dual problem of **(PSDP)**: 1.1
min	minimum, minimize: 1.1
max	maximum, maximize: 1.1
$x \mapsto T(x)$	mapping: 1.1
$\boldsymbol{\mathcal{A}}(\cdot)$	linear operator, ($\boldsymbol{\mathcal{A}} : S(n) \mapsto \mathbb{R}^m$): 1.1
$\boldsymbol{\mathcal{A}^t}(\cdot)$	adjoint operator of $\boldsymbol{\mathcal{A}}(\cdot)$, ($\boldsymbol{\mathcal{A}^t} : \mathbb{R}^m \mapsto S(n)$): 1.1
$\succeq$	Löwner partial order: 1.1
$A \succeq B$	$A - B$ is positive semidefinite: 1.1
$A \succ B$	$A - B$ is positive definite: 1.1
$A_{ij} = (a_{ij})$	designates the $ij-$th entry of the matrix A, (both notations are used in this work): 1.1
trace(A)	trace of a square matrix A: 1.1
$\langle A, B\rangle$	Frobenius inner product, ($\langle A, B\rangle :=$ trace$(A^T B)$): 1.1
$\|A\|$	the Frobenius norm of $A \in M(n)$, ($\|A\| := \langle A^T, A\rangle^{1/2}$): 1.1

$\sum_{j=1}^{n} x_j$	sum of a series: 1.1
$\sum_i x_i, \sum_{i\in I} x_i$	sum of a summable family: 1.1
$X \times Y$	product of two sets: 1.1
inf	infimum: 1.1
sup	supremum: 1.1
C^*	dual cone of a cone C, $C^* := \{d : d^T c \geq 0\ \forall c \in C\}$: 1.1
$S(n)^+$	$S(n)^+ := \Big\{X : X \in S(n), X \succeq 0\Big\}$: 1.1
$S(n)^{++}$	$S(n)^{++} := \Big\{X : X \in S(n), X \succ 0\Big\}$: 1.1
A^T	transpose of a matrix A: 1.1
$\boldsymbol{0}^m$	vector of all zeros in $\mathbb{R}^m$: 1.1
(IPM(s))	Interior-Point Method(s): 1.2
(OPT)	optimality conditions: 1.2
$\mathbb{N}$	set of natural numbers: 1.3
$(x^k)_{k\in\mathbb{N}}$	sequence: 1.3
$\mathcal{F}(\mathcal{P})$	primal feasible set: 1.3
$\mathcal{F}(\mathcal{D})$	dual feasible set: 1.3
$\mathcal{AF}(\mathcal{P})$	primal asymptotically feasible set: 1.3
$\mathcal{AF}(\mathcal{D})$	dual asymptotically feasible set: 1.3
$\liminf_{k\to\infty}(x^k)$	limit inferior: 1.3
$\limsup_{k\to\infty}(x^k)$	limit superior: 1.3
p^*	$p^* = \inf\langle C, X\rangle$ over $X \in \mathcal{F}(\mathcal{P})$: 1.3
d^*	$d^* = \inf \liminf_{k\to\infty}\langle C, X^k\rangle$ over $(X^k)_{k\in\mathbb{N}} \in \mathcal{AF}(\mathcal{P})$: 1.3
$p^*_{\mathcal{A}}$	$p^*_{\mathcal{A}} = \sup\langle b, y\rangle$ over $(y, Z) \in \mathcal{F}(\mathcal{D})$: 1.3
$d^*_{\mathcal{A}}$	$d^*_{\mathcal{A}} = \sup \limsup_{k\to\infty}\langle b, y^k\rangle$ over $(y^k)_{k\in\mathbb{N}} \in \mathcal{AF}(\mathcal{D})$: 1.3
$\emptyset$	the empty set: 1.1
F	feasible: 1.3
IF	infeasible: 1.3
AF	asymptotically feasible: 1.3
SIF	strongly infeasible: 1.3
PAF	properly AF: 1.3
IAF	improperly AF: 1.3
BD	bounded: 1.3
UBD	unbounded: 1.3
ABD	asymptotically BD: 1.3
AUBD	asymptotically UBD: 1.3
$f(X) = -\log\det(X)$	barrier function on $S(n)^+$: 1.4
(KKT)	Karush-Kuhn-Tucker conditions: 1.4
(X^*, y^*, Z^*)	optimal solution to **(PSDP)** and **(DSDP)**: 1.4
$\mathcal{L}_\mu(\cdot)$	Lagrangian: 1.4
$\frac{\partial}{\partial x_i} f(x_1, \ldots, x_n)$	partial derivatives: 1.4

$\left\{(X_\mu, y_\mu, Z_\mu), \mu > 0\right\}$	primal-dual central path: 1.4
I	identity matrix: 1.4
$f(X, \mu)$	primal barrier function: 1.4
$g(y, Z, \mu)$	dual barrier function: 1.4
$M(n) = \mathbb{R}^{n\times n}$	the space of $n \times n$ matrices: 1.4
$(\Delta X, \Delta y, \Delta Z)$	search direction: 1.4
∇	nabla operator, gradient: 1.4
$X^{1/2}, \sqrt{X}$	unique positive definite (semidefinite) square root of $X \in S(n)^{++}$ $(X \in S(n)^{+})$: 1.4
$X^{-1/2}, \sqrt{X}^{-1}$	inverse of $X^{1/2}, \sqrt{X}$: 1.4
$H_p(\cdot)$	symmetrization operator, $(H_p(M) : \mathbb{R}^{n\times n} \mapsto S(n))$: 1.4
X^n	product of n sets equal to X: 1.4
$\mathrm{int}(F)$	interior of F: 1.4
$(X^\circ, y^\circ, Z^\circ)$	initial point for **(PSDP)** and **(DSDP)**: 1.4
$x \leftarrow y$	y is allocated to x: 1.4
$\lambda_{\min}(X)$	minimum eigenvalue of $X \in M(n)$: 1.5
$\lambda_{\max}(X)$	maximum eigenvalue of $X \in M(n)$: 1.5
$\lVert x \rVert$	the Euclidean norm: 1.5
$(x^k)_{k\in\mathbb{N}}, (x^k) \searrow 0$	a series of positive numbers converging to zero: 1.5
(HSDP)	homogeneous self dual model for (SDP): 1.5
(EHSDP)	extended homogeneous self dual formulation: 1.5
$\mathbb{R}$	set of real numbers: 1.1
$\mathcal{F}(\mathcal{H})$	set of feasible solutions of **(EHSDP)**: 1.5
$\mathbb{R}^{+}$	the set of real numbers ≥ 0: 1.5
$\succsim$	$\succsim\ :\Leftrightarrow (=, \succeq, \geq, =)^T$: 1.5
$\precsim$	$\precsim\ :\Leftrightarrow (=, \preceq, \leq, =)^T$: 1.5
$\lim\limits_{k\to\infty} (x^k)$	limit of a sequence: 1.5
$f_\mu(\cdot)$	merit function: 1.5
$(\xi^k) \to \xi^*$	a sequence $(\xi^k)_{k\in\mathbb{N}}$ that converges to ξ^* as $k \to \infty$: 1.5
$\prod\limits_{j=1}^{n} x_j$	product of a series: 2.2
$o(\cdot)$	Landau's notation: 2.2
(SDLCP)	the linear complementarity problem in (SDP): 2.2
$\exp$	the exponential function: 1.5
$\mathbf{A}$	point-to-set mapping: 2.2
$\mathcal{N}(\gamma)$	horn neighbourhood of the central path: 2.3
$\mathrm{Diag}(x)$	the diagonal matrix with diagonal x: 2.3
$\mathrm{diag}(X)$	the vector formed from the diagonal elements of $X \in M(n)$: 2.3
$O(\cdot)$	Landau's notation: 2.3
$(X(\varepsilon), y(\varepsilon), Z(\varepsilon))$	an approximate solution for **(PSDP)** and **(DSDP)**: 2.4

$\|A\|_2$	the spectral norm, $(\|A\|_2 := (\lambda_{\max}(A^T A))^{1/2})$: 2.4
(LQ-PSDP)	positive semidefinite linear-quadratic problem: 3.1
(LQ-DSDP)	Lagrangian dual problem of **(LQ-PSDP)**: 3.1
$\boldsymbol{\mathcal{Q}}(\cdot)$	symmetric positive semidefinite linear operator, $(\boldsymbol{\mathcal{Q}} : S(n) \mapsto S(n))$: 3.1
$A \otimes B$	the Kronecker product of the matrices A and B: 3.1
$A_{i.}$	the i-th row of the matrix A: 3.1
$A_{.j}$	the j-th column of the matrix A: 3.1
$\text{vec}(\cdot)$	vec operator, $(\text{vec} : M(n) \mapsto R^{n^2})$: 3.1
$\text{Mat}(\cdot)$	inverse operator of $\text{vec}(\cdot)$, $(\text{Mat} : \mathbb{R}^{n^2} \mapsto M(n))$: 3.1
$\mathcal{F}(\mathcal{LQ} - \mathcal{P})$	set of feasible solutions of **(LQ-PSDP)**: 3.1
$\mathcal{F}(\mathcal{LQ} - \mathcal{D})$	set of feasible solutions of **(LQ-DSDP)**: 3.1
(ECP)	the Equicut Problem: 4.1
(QAP)	the Quadratic Assignment Problem: 4.1
$\mathbf{(GP)}_{\boldsymbol{\alpha}}$	(SDP)-problem: 4.1
$\mathbf{(DGP)}_{\boldsymbol{\alpha}}$	Lagrangian dual problem of $\mathbf{(GP)}_{\boldsymbol{\alpha}}$: 4.3
e	the vector of all ones: 4.1
J	the matrix of all ones: 4.1
(OPT')	perturbed optimality conditions: 4.2
$\vartheta(G)$	Lovász number,(ϑ-function): 5
$G = (V(G), E(G))$	graph with node set $V(G)$ and edge set $E(G)$: 5.1
$e = (ij)$	an edge in a graph joining nodes i and j: 5.1
$\alpha(G)$	stability number of G: 5.1
$\chi(G)$	chromatic number of G: 5.1
$\bar{G}$	complement graph: 5.1
$STAB(G)$	the stable set polytope of G: 5.1
χ^S	characteristic vector of a set S: 5.1
$\text{conv}(F)$	convex hull of F: 5.1
(STAB)	the maximum stable set problem: 5.1
K_n	complete graph on n nodes: 5.1
e_j	j-th canonical unit vector: 5.1
E_{ij}	the matrix $E_{ij} := e_i e_j^t + e_j e_i^t$: 5.1
$\lfloor\gamma\rfloor, \gamma \in \mathbb{R}$	the largest integer not larger than γ: 5.1
$QSTAB(G)$	clique constrained stable set polytope of G: 5.2
$CSTAB(G)$	circuit constrained stable set polytope of G: 5.2
$TH(G)$	semidefinite relaxation of the stable set polytope of G: 5.2
$M(P)$	polytope defined matrix inequalities: 5.3
$N(P)$	projection of $M(P)$ on the subspace $\mathbb{R}^n$: 5.3
$N^k(P)$	repeated $N-$ operator: 5.3
$FRAC(G)$	fractional stable set polytope of G: 5.3
MET	the metric polytope: 5.3
P_{MET}	projection of MET on the subspace $\mathbb{R}^n$: 5.3

$S(G)$	$\Big\{X = (x_{ij}) \in S(n) \mid x_{ij} = 0 \text{ for } (ij) \in E(G)\Big\}$: 5.3
$MET_1(G)$	$MET \cap S(G)$: 5.3
$P_{MET_1(G)}$	projection of $MET_1(G)$ on the subspace $\mathbb{R}^n$: 5.3
$P_{MET_2(G)}$	linear relaxation of the stable set polytope of G: 5.3
$M_+(P)$	$M_+(P) := \Big\{Y \in M(P) \mid Y \in M(P)$ satisfying $\begin{pmatrix} Y & \operatorname{diag}(Y) \\ \operatorname{diag}(Y)^T & 1 \end{pmatrix} \succeq 0\Big\}$: 5.3
$N_+(P)$	projection of $M_+(P)$ on the subspace $\mathbb{R}^n$: 5.3
G^k	the strong product of k graphs equal to G: 5.4
$\Theta(G)$	the Shannon capacity, $\left(\Theta(G) := \sup\limits_{k\to\infty} \sqrt[k]{\alpha(G^k)}\right)$: 5.4
(PSDP$\boldsymbol{\vartheta_S}$)	(SDP)-problem with optimal objective value $\vartheta(G)$: 5.4
(DSDP$\boldsymbol{\vartheta_S}$)	Lagrangian dual problem of **(PSDP$\boldsymbol{\vartheta_S}$)**: 5.4
(PSDP$\boldsymbol{\vartheta_L}$)	(SDP)-problem with optimal objective value $\vartheta(G)$: 5.4
(DSDP$\boldsymbol{\vartheta_L}$)	Lagrangian dual problem of **(PSDP$\boldsymbol{\vartheta_L}$)**: 5.4
$G[U]$	subgraph of G induced by $U \subseteq V$: 6.1
$\delta(S)$	set of $S - V(G)$§ edges (= cut in a graph): 6.1
$w(\delta(S))$	$w(\delta(S)) := \sum_{e\in\delta(S)} w_e$, weight of a cut: 6.1
w_e	the weight on the edge e: 6.1
(MC)	The Maximum Cut Problem: 6.2
$P_c(G)$	the cut polytope: 6.2
(MC)	Max–Cut programming model in matrix variables: 6.2
w^*	the optimal objective value of **(MC)**: 6.2
$\mathcal{W}$	weighted adjacency matrix of G: 6.2
L	Laplacian matrix associated with G, $(L := \operatorname{Diag}(\mathcal{W}e) - \mathcal{W})$: 6.2
(SDPMC)	semidefinite relaxation for Max–Cut: 6.2
(LPMC)	linear relaxation for Max Cut: 6.2
$MET(G)$	metric polytope of a graph G: 6.3
$\varepsilon(G)$	elliptope of G: 6.5
$\mathcal{M}_{m,n}(\mathbf{C})$	vector space of $m \times n$ complex matrices: A
A^*	Hermitian adjoint of $A \in \mathcal{M}_{m,n}(\mathbf{C})$: A

Part One –

Theory

Chapter 1

Semidefinite Programming

Semidefinite Programming (denoted SDP) is an extension of Linear Programming (denoted LP). Because of its numerous important applications in various fields (including control and system theorie [18] and combinatorial optimization [34, 77]) SDP has been the subject of intensive research during the past decade. A comprehensive list of applications is given in [127]. SDP belongs to the class of conic convex programming , where the objective is linear and the constraint set is given by the intersection of an affine space with a closed cone. In particular, the problem is optimizing a linear function of matrix variables subject to finitely many linear constraints and the positive semidefiniteness condition on the matrix variables. For recent surveys on SDP we refer the reader e.g. to [112, 127] and the references quoted therein. In the following section we describe the problem in standard form.

1.1 The Semidefinite Programming Problem

The semidefinite programming problem in standard form is a problem of the following type. For given matrices $C, A_1, \ldots, A_m$ in the space of $n \times n$ symmetric matrices, denoted by $S(n)$ and $b \in \mathbb{R}^m$, solve

$$\textbf{(PSDP)} \qquad \begin{aligned} &\min\ \langle C, X\rangle \\ &\text{s.t. } \boldsymbol{\mathcal{A}}(X) = b \\ &\qquad X \in S(n),\ X \succeq 0 \end{aligned} \tag{1.1}$$

where $\boldsymbol{\mathcal{A}} : S(n) \mapsto \mathbb{R}^m$ denotes a linear operator written in full as

$$\boldsymbol{\mathcal{A}}(X) = \begin{pmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{pmatrix}.$$

The associated adjoint operator $\boldsymbol{\mathcal{A}^t} : \mathbb{R}^m \mapsto S(n)$, defined by the adjoint relation

$$\langle \boldsymbol{\mathcal{A}}(X), y\rangle = \langle X, \boldsymbol{\mathcal{A}^t}(y)\rangle, \text{ for all } X \in S(n),\ y \in \mathbb{R}^m \tag{1.2}$$

is given by $\boldsymbol{\mathcal{A}^t}(y) = \sum\limits_{i=1}^{m} y_i A_i$. We write trace$(X)$ to denote the trace of a square matrix X, i.e. trace$(X) = \sum_i X_{ii}$. The inner product which we use in $S(n)$ is the Frobenius inner product defined by

$$\langle A, B\rangle := \text{trace}(AB) \text{ for given } A \in S(n) \text{ and } B \in S(n).$$

The induced norm is the Frobenius norm which we denote by $\|A\|$, i.e., $\|A\| = (\text{trace}(AA))^{\frac{1}{2}}$ for $A \in S(n)$. (The dimension of $S(n) = n(n+1)/2$.)

The symbol $\succeq$ denotes the Löwner partial order, i.e., $A \succeq B$ if $A - B$ is positive semidefinite. $A \succ 0$ means A is positive definite. Throughout this work we assume that the set $\{A_i\}, (i = 1, \ldots, m)$ is linearly independent.

If $C, A_1, \ldots, A_m$ are diagonal matrices then this type of problem is a linear program. So SDP is a direct generalization of LP. **(PSDP)** denotes the primal problem. A standard way to derive the corresponding dual **(DSDP)** is using Lagrangian duality.

Lemma 1.1 [113, Lemma 36.1] *(minimax inequality) Suppose the product set $X \times Y$ is non-empty. Let $f : X \times Y \mapsto [-\infty, +\infty]$. Then*

$$\inf_{x\in X} \sup_{y\in Y} f(x,y) \geq \sup_{y\in Y} \inf_{x\in X} f(x,y).$$

For the next result we need additional definitions. A set C is said to be a cone if $\alpha c \in C$ for all $\alpha \geq 0$ and $c \in C$. Given a cone C the dual cone of C, denoted C^*, is defined by

$$C^* := \left\{ d : d^T c \geq 0, \text{ for all } c \in C \right\}.$$

(A synonym for "dual cone" used by many authors is "polar cone".)

Proposition 1.2 [113, Theorem 14.1] *Let $C \subseteq \mathbb{R}^m$ be a closed convex cone. Then $(C^*)^* = C$.*

Let $S(n)^+$ denote the cone of symmetric positive semidefinite matrices, i.e., $S(n)^+ := \left\{ X : X \in S(n), X \succeq 0 \right\}$. We will write $S(n)^{++}$ to denote the interior of $S(n)^+$, i.e., $S(n)^{++} := \left\{ X : X \in S(n), X \succ 0 \right\}$.

Lemma 1.3 *$S(n)^+$ is a closed convex cone and is selfdual, i.e., $(S(n)^+)^* = S(n)^+$.*

Proof. It is obvious that $S(n)^+$ is a closed convex cone. The selfduality of $S(n)^+$ is a direct consequence of Fejer's theorem, see [57, Corollary 7.5.4]. Consider $X, Y \in S(n)^+$. Let $\sqrt{X}$ and $\sqrt{Y}$ denote the unique positive semidefinite square root of $X \in S(n)^+$ and $Y \in S(n)^+$, respectively. Then $X = \sqrt{X}\sqrt{X}$, $Y = \sqrt{Y}\sqrt{Y}$ and $\mathrm{trace}(XY) = \mathrm{trace}(\sqrt{X}\sqrt{Y})(\sqrt{X}\sqrt{Y})^T \geq 0$. Hence $S(n)^+ \subseteq (S(n)^+)^*$. Conversely, assume that $Y \in S(n)$ and $Y \notin S(n)^+$. Thus there exists a $y \in \mathbb{R}^n$ such that $y^TYy < 0$. Define $X := yy^T$. Clearly, $X \in S(n)^+$, but $\mathrm{trace}(XY) = y^TXy < 0$, which means that $Y \notin (S(n)^+)^*$. This completes the proof.

$\square$

Let $y \in \mathbb{R}^m$ denote the Lagrangian multiplier for (1.1), then

$$\begin{aligned}
\min\langle C, X\rangle \text{ s.t. } \boldsymbol{\mathcal{A}}(X) = b, X \in S(n)^+ &= \min_{X \in S(n)^+} \max_{y \in \mathbb{R}^m} \langle C, X\rangle - \langle \boldsymbol{\mathcal{A}}(X) - b, y\rangle \\
&\geq \max_{y \in \mathbb{R}^m} \min_{X \in S(n)^+} \langle b, y\rangle + \langle C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y), X\rangle \\
&= \max\langle b, y\rangle \text{ s.t. } C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) \in S(n)^+, y \in \mathbb{R}^m.
\end{aligned} \tag{1.3}$$

The first equation is a simple consequence of Proposition 1.2. If $\boldsymbol{\mathcal{A}}(X) \neq b$, or equivalently $\boldsymbol{\mathcal{A}}(X) - b \notin \boldsymbol{0}^m$, ($\boldsymbol{0}^m$ denotes the vector of all zeros in $\mathbb{R}^m$) $\Rightarrow \boldsymbol{\mathcal{A}}(X) - b \notin (\boldsymbol{0}^m)^{**} \Rightarrow \exists y \in \mathbb{R}^m : \langle \boldsymbol{\mathcal{A}}(X) - b, y\rangle < 0$ and the inner maximization over $y \in \mathbb{R}^m$ is unbounded, i.e., $\max_{y \in \mathbb{R}^m} \langle C, X\rangle - \langle \boldsymbol{\mathcal{A}}(X) - b, y\rangle \to \infty$. On the other hand, if $\boldsymbol{\mathcal{A}}(X) = b$, then $\langle \boldsymbol{\mathcal{A}}(X) - b, y\rangle = 0$ for all $y \in \mathbb{R}^m$ and hence $\max_{y \in \mathbb{R}^m} \langle C, X\rangle - \langle \boldsymbol{\mathcal{A}}(X) - b, y\rangle = \langle C, X\rangle$. Similar ideas work for the last equation.

We refer to problem (1.3) as the dual problem, hence

$$\textbf{(DSDP)} \qquad \begin{aligned}
&\max\ \langle b, y\rangle \\
&\text{s.t. } Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = C \\
&\quad\ y \in \mathbb{R}^m,\ Z \in S(n),\ Z \succeq 0.
\end{aligned} \tag{1.4}$$

The difference between primal objective value and dual objective value is called duality gap.

1.2 On Solving Semidefinite Programs – An Overview

Semidefinite Program(s) (denoted SDP(s)) can be solved very efficiently using interior-point method(s) (denoted IPM(s)), which are generalizations of IPMs in LP. (Note, in what follows the meaning of SDP is twofold. We use it as abbreviation for Semidefinite

Programming and Semidefinite Program as well.) IPMs for convex programming has been studied already in the 1960s, see e.g. [31]. For a note describing the recent history of IPMs for LP we refer the reader to the work of den Hertog [56]. In 1984 Karmarkar published a polynomial time IPM for LP, see [64]. It was Karmarkar's work which set in motion the interior-point machinery. Nesterov and Nemirovskii [97] showed how to extend several classes of interior-point algorithms for LP to nonlinear optimization problems.

IPMs for SDP were proposed by many authors, see e.g. Alizadeh [1], Alizadeh, Haeberly, Overton [2, 3], Monteiro [95], Helmberg, Rendl, Vanderbei, Wolkowicz [54], Nesterov, Todd [99, 98], Kojima, Shindoh, Hara [73], Kojima, Shida, Shindoh [71, 72], Jarre [60], Vandenberghe, Boyd [20], Tseng [124] or Zhang [135].

The most efficient variants solve the primal and dual problem simultaneously. Alizadeh [1] pointed out first that many theoretical results can be transfered in an easy way from LP to SDP. But in spite of all the theoretical as well as algorithmic similarities between SDP and LP, there are also significant differences.

One of the main sources of complication comes from the more complicated duality relations between **(PSDP)** and **(DSDP)**. A discussion of the possible duality states can be found e.g. in [15].

For LP the situation is relatively straightforward. If one of the primal or dual in LP is feasible and has a bounded objective function, then there is strong duality : both optima are attained, and the resulting optimal values are equal. Otherwise, if optimal solutions do not exist, there are two possibilities: either both feasible sets are empty, or else one is empty and the other problem is unbounded. For SDP the situation is significantly more subtle. First, feasibility and boundedness do not imply that the optimum is attained. Moreover, there may be a nonzero, but finite duality gap between the primal and the dual, or the primal may be solveable and the associated dual infeasible, and so on. A sufficient condition to insure that the duality gap is zero is the Slater condition. Duality results for semidefinite programs will be discussed in more detail in section 1.3.

Standard algorithms, based on the interior-point idea, assume Slater's constraint qualification (also known as Slater's regularity condition) to hold for **(PSDP)** and **(DSDP)**, thus insuring that both the primal and the dual are attained, with zero duality gap (see e. g. [29, 97]). By definition, **(PSDP)** satisfies the Slater condition if there exists a feasible matrix X with $X \in S(n)^{++}$. We also call such a matrix X strictly feasible. Similarly, **(DSDP)** satisfies the Slater condition if there exists a feasible pair (y, Z) with $Z \in S(n)^{++}$. If strictly feasible solutions for both problems exist, the following set of constraints is necessary and sufficient to characterize optimality of a primal-dual solution. Suppose **(PSDP)** and **(DSDP)** satisfy Slater's condition. Then (X, y, Z) is optimal if and only if

$$
\textbf{(OPT)} \qquad \begin{array}{l} \boldsymbol{\mathcal{A}}(X) - b = 0 \\ Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = C \\ ZX = 0 \\ X \in S(n)^+, \; Z \in S(n)^+, \; y \in I\!\!R^m. \end{array}
$$

The problem is called primal and dual feasible if the first two conditions together with the fourth condition are satisfied by some (X, y, Z). The third condition is henceforth referred to as the complementary slackness condition. By definition, any pair X and Z satisfying **(OPT)** is called a complementary pair. (X and Z are complementary.)

1.3 Duality in Semidefinite Programming

We do not give an extensive survey in duality theory, but we overview earlier work and results on this topic and include some of it in detail. As we have already mentioned duality theory in SDP is different from duality theory in LP. Generalizations of linear programs to optimization problems over convex cones have been studied extensively. Duality theory for cone programming is essentially due to Duffin [29]. We refer also to the works of Pullan [109], Ben-Israel, Charnes and Kortanek [15] and Wolkowicz [130].

Bellman and Fan [14] showed that many results of duality in LP extends quite naturally to SDP. Note, that they already observed that additional regularity assumptions are required to guarantee for example a zero duality gap. The generalization of duality in LP to SDP has been considered recently by Alizadeh [1] and by Nesterov and Nemirovskii [96].

What are the differences between duality in LP and SDP? As we remarked earlier one of it is that there are instances in SDP for which there is a positive and finite duality gap at an optimal solution. Ramana, [110], proposes an exact dual for which strong duality holds without any additional regularity assumptions. Based on this work Ramana, Tuncel and Wolkowicz [111] proposed alternative dual problems that guarantee a zero duality gap for problems in SDP.

A detailed discussion of the possible duality states for LP over closed convex cones is given in [15]. This work combines results from Duffin [29] together with asymptotic refinements. These results in the setting of SDP are given here. Note, within this extended framework it might happen that the minimum in **(PSDP)** cannot be attained or even does not exist. So it is convenient to exchange "min" with "inf" and analogous "max" with "sup". Hence the semidefinite programming problem in standard form is given by the following primal-dual pair. Since there is no reason for confusion we do not change the denotation of the problems.

$$
\begin{array}{llll}
\textbf{(PSDP)} & \inf\ \langle C, X\rangle & \textbf{(DSDP)} & \sup\ \langle b, y\rangle \\
& \text{s.t. } \boldsymbol{\mathcal{A}}(X) = b & & \text{s.t. } Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = C \\
& \quad X \in S(n)^+ & & \quad y \in I\!R^m,\ Z \in S(n)^+
\end{array}
$$

The construction of **(DSDP)** (see section 1.1) shows that for X feasible in **(PSDP)** and (y, Z) feasible in **(DSDP)** the duality gap is always a nonnegative quantity. This is the well known weak duality theorem. To get more insight in the more complicated duality theory in SDP and its possible duality states we follow the presentation in [15] and consider also asymptotic refinements for SDP.

Definition 1.4 *(asymptotically feasible solutions)*

(1) *A sequence $(X^k)_{k\in I\!N}$ such that $X^k \in S(n)^+$, $\boldsymbol{\mathcal{A}}(X^k) \to b$ as $k \to \infty$ is called an asymptotically feasible solution for* **(PSDP)**.

(2) *A sequence $(y^k)_{k\in I\!N}$, $y^k \in I\!R^m$ such that there exists a sequence $(Z^k)_{k\in I\!N}$, $Z^k \in S(n)^+$, $Z^k + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^k) \to C$ is called an asymptotically feasible solution for* **(DSDP)**.

We will denote the primal and dual feasible sets by $\mathcal{F}(\mathcal{P})$ and $\mathcal{F}(\mathcal{D})$ and the primal and dual asymptotically feasible sets by $\mathcal{AF}(\mathcal{P})$ and $\mathcal{AF}(\mathcal{D})$, respectively. We use the notation p^*, d^*, $p^*_{\mathcal{A}}$ and $d^*_{\mathcal{A}}$ for the respective optimal values, i.e.,

$$
\begin{array}{lcll}
p^* & := & \inf\langle C, X\rangle & \text{over } X \in \mathcal{F}(\mathcal{P}), \\
p^*_{\mathcal{A}} & := & \inf \liminf_{k\to\infty} \langle C, X^k\rangle & \text{over } (X^k)_{k\in I\!N} \in \mathcal{AF}(\mathcal{P}), \\
d^* & := & \sup\langle b, y\rangle & \text{over } (y, Z) \in \mathcal{F}(\mathcal{D}), \\
d^*_{\mathcal{A}} & := & \sup \limsup_{k\to\infty} \langle b, y^k\rangle & \text{over } (y^k)_{k\in I\!N} \in \mathcal{AF}(\mathcal{D}).
\end{array}
$$

Remark 1.5 *Primal (dual) feasible solutions can be embedded in a natural way in the set of asymptotically feasible solutions for the primal (dual) problem, respectively. Hence, $p^*_{\mathcal{A}} \le p^*$ and $d^*_{\mathcal{A}} \ge d^*$.*

Definition 1.6 [15] *(duality states)*

(i) **(PSDP)** (**(DSDP)**) *is*

F *(feasible) if $\mathcal{F}(\mathcal{P}) \neq \emptyset$ $(\mathcal{F}(\mathcal{D}) \neq \emptyset)$,*

IF *(infeasible) if $\mathcal{F}(\mathcal{P}) = \emptyset$ $(\mathcal{F}(\mathcal{D}) = \emptyset)$,*

AF *(asymptotically feasible) if $\mathcal{AF}(\mathcal{P}) \neq \emptyset$ $(\mathcal{AF}(\mathcal{D}) \neq \emptyset)$,*

SIF *(strongly infeasible) if $\mathcal{AF}(\mathcal{P}) = \emptyset$ $(\mathcal{AF}(\mathcal{D}) = \emptyset)$, respectively.*

(ii) *Let* **(PSDP)** (**(DSDP)**) *be* AF*. Then it is*

PAF *(properly* AF*) if* $p^*_{\mathcal{A}} < \infty$ $(d^*_{\mathcal{A}} > -\infty)$*, respectively; otherwise the respective problem is called* IAF *(improperly* AF*).*

(iii) *Let* **(PSDP)** (**(DSDP)**) *be* F*. Then it is*

BD *(bounded) if* $p^* > -\infty$ $(d^* < +\infty)$*, respectively; otherwise the respective problem is called* UBD *(unbounded).*

(iv) *Let* **(PSDP)** (**(DSDP)**) *be* PAF*. Then it is*

ABD *(asymptotically* BD*) if* $p^*_{\mathcal{A}} > -\infty$ $(d^*_{\mathcal{A}} < \infty)$*, respectively; otherwise the respective problem is called* AUBD *(asymptotically* UBD*).*

The authors in [15] show that exactly eleven duality states may occur. We proceed with a series of lemmas which they used to prove this result.

Lemma 1.7 [15] *(Weak duality)*

(1) *If* **(PSDP)** *is* F *and* **(DSDP)** *is* AF*, then* $d^*_{\mathcal{A}} \le p^*$*. If* **(PSDP)** *is also* BD*, then* $d^*_{\mathcal{A}} = p^*$*.*

(2) *If* **(DSDP)** *is* F *and* **(PSDP)** *is* AF*, then* $p^*_{\mathcal{A}} \ge d^*$*. If* **(DSDP)** *is also* BD*, then* $p^*_{\mathcal{A}} = d^*$*.*

Lemma 1.8 [15]

(1) *If* **(PSDP)** *is* PAF *and* ABD*, then* **(DSDP)** *is* F *and* BD*.*

(2) *If* **(DSDP)** *is* PAF *and* ABD*, then* **(PSDP)** *is* F *and* BD*.*

Lemma 1.9 [15]

(PSDP) *is* F *and* ABD *if and only if* **(DSDP)** *is* F *and* ABD*.*

In this case: $d^* = p^*_{\mathcal{A}} \le d^*_{\mathcal{A}} = p^*$*.*

Lemma 1.10 [15]

(1) *If* **(PSDP)** *is* IAF *then* **(DSDP)** *is* F *and* UBD*.*

(2) *If* **(DSDP)** *is* IAF *then* **(PSDP)** *is* F *and* UBD*.*

Remark 1.11 *The strong duality theorem given in* [29] *is contained within this framework.*

Proof. Let **(PSDP)** be F and BD. Suppose **(DSDP)** is IAF. Then, by Lemma 1.10, **(PSDP)** is F and UBD, which is a contradiction. Hence **(DSDP)** must be at least AF. Lemma 1.7 implies that $d^*_{\mathcal{A}} = p^*$. In particular, **(DSDP)** is PAF and ABD. The same argument can be applied to **(DSDP)**.

□

Now we formulate Duffin's duality theorem .

Theorem 1.12 [29] *(Strong duality)*

(1) **(PSDP)** *is* F *and* BD *if and only if* **(DSDP)** *is* PAF *and* ABD. *In this case,* $d^*_{\mathcal{A}} = p^*$.

(2) **(DSDP)** *is* F *and* BD *if and only if* **(PSDP)** *is* PAF *and* ABD. *In this case,* $d^* = p^*_{\mathcal{A}}$.

Next we restrict our consideration on problems for which p^* is attained and $p^*_{\mathcal{A}} = p^*$, or equivalently d^* is attained and $d^* = d^*_{\mathcal{A}}$. This yields $d^* = p^*_{\mathcal{A}} = d^*_{\mathcal{A}} = p^*$. Stronger assumptions are required in SDP than in LP to ensure this line of equality. For completeness we include the definition of the Slater constraint qualification.

Definition 1.13 *(Slater constraint qualification)*

(1) **(PSDP)** *satisfies the Slater condition if there exists a* $X \in S(n)^{++}$ *such that* $\boldsymbol{\mathcal{A}}(X) = b$.

(2) **(DSDP)** *satisfies the Slater condition if there exists a pair* (y, Z) *with* $Z \in S(n)^{++}$ *such that* $Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = C$.

Theorem 1.14 [29] *If* **(PSDP)** *satisfies the Slater condition and* p^* *is finite, then* **(DSDP)** *is* F, *the dual supremum* d^* *is attained, and* $d^* = p^*_{\mathcal{A}} = d^*_{\mathcal{A}} = p^*$, *and vice versa.*

1.4 Interior-Point Methods

IPMs can be categorized in path-following methods, affine scaling methods and potential reduction methods, see [56]. In this work we are exclusively interested in path-following methods. Each of the stated categories of IPMs for SDP is discussed e.g. in [22].

Nesterov and Nemirovskii showed first that optimization problems over cones can be solved in polynomial time using interior-point techniques. The essential element behind their method is the existence of a function which has the property called self concordance. In particular, optimizing a linear objective over a convex set is always possible in polynomial time if such a function for the convex set is known. We do not use the concept of self concordance, readers interested in it should consult for e.g.

the book of Nesterov and Nemirovskii [97] which contains a comprehensive treatment of this subject. Nesterov and Nemirovskii show that a computeable self concordant barrier function on $S(n)^+$, see [97], is given by

$$f(X) = -\log\det(X).$$

Hence there is a standard way of constructing path-following interior-point algorithms for **(PSDP)**. Note, $f(X)$ is finite for $X \in S(n)^{++}$ and approaches infinity when X approaches the boundary of $S(n)^+$. Using the barrier term $-\log\det(X)$ several classes of primal path-following algorithms for solving semidefinite programs were analyzed, see [30, 5]. Newton's method is used to solve the family of problems (1.5), characterized by the barrier parameter $\mu > 0$, iteratively. The parameter μ is reduced at each iteration.

$$\min \langle C, X\rangle - \mu \log\det(X), \text{ s.t. } \boldsymbol{\mathcal{A}}(X) = b \tag{1.5}$$

This way the positive semidefiniteness condition on the matrix variables has been moved into the objective function. (It has been hidden by adding the additional term $-\log\det(X)$ to the objective function.) Note, the objective in (1.5) is convex in X and the constraints in (1.5) are linear. Hence the Karush-Kuhn-Tucker **(KKT)** conditions associated with this problem guarantee optimality. If y denotes the Lagrangian multiplier for $\boldsymbol{\mathcal{A}}(X) - b$ the optimality conditions are given by

$$\begin{aligned} \mu X^{-1} + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) &= C \\ \boldsymbol{\mathcal{A}}(X) &= b. \end{aligned} \tag{1.6}$$

In the following we switch to the more efficient primal-dual path-following methods. We assume that both, **(PSDP)** and **(DSDP)**, satisfy the Slater constraint qualification. Recall, under this assumption the minimum in **(PSDP)** and the maximum in **(DSDP)** are attained. Both problems have finite solutions and their optimal values are equal, i.e., strong duality holds. Any optimal solution (X^*, y^*, Z^*) is characterized by the optimality conditions **(OPT)**.

The complementary slackness condition is a nonlinear condition. This nonlinearity makes the subject of finding an optimal triple nontrivial. One way going round this complication is analyzing the barrier problem (1.7), associated to the dual problem. It is given by

$$\max \langle b, y\rangle + \mu \log\det(Z), \text{ s.t. } Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = C. \tag{1.7}$$

Note that this problem is well defined for any $\mu > 0$, so it may be considered as sequence of problems indexed by the parameter $\mu > 0,\ \mu \to 0$.

The Lagrangian $\mathcal{L}_\mu(X, y, Z)$ for fixed $\mu > 0$ associated to this problem is

$$\mathcal{L}_\mu(X, y, Z) = \langle b, y\rangle + \mu \log\det(Z) - \langle X, Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - C\rangle.$$

For any fixed value $\mu > 0$ the maximizers for (1.7) satisfy the **(KKT)** conditions

$$\begin{aligned}
\frac{\partial}{\partial y}\mathcal{L}_\mu(X,y,Z) &= b - \boldsymbol{\mathcal{A}}(X) = 0,\\
\frac{\partial}{\partial X}\mathcal{L}_\mu(X,y,Z) &= Z + \boldsymbol{\mathcal{A}^t}(y) - C = 0,\\
\frac{\partial}{\partial Z}\mathcal{L}_\mu(X,y,Z) &= \mu Z^{-1} - X = 0,\\
X, Z \in S(n)^{++}.
\end{aligned}$$

For a proof of this result see Fletcher [32] or Mangasarin [83]. Note that $\frac{\partial}{\partial Z}\log\det Z = Z^{-1}$, see e.g. Graham [39]. For a fixed value $\mu > 0$ the existence of a unique solution (X_μ, y_μ, Z_μ) of **(KKT)** is shown in e.g. [97, 127, 126, 19]. Moreover it can be shown that the sequence (X_μ, y_μ, Z_μ) approaches optimality as μ is driven to zero, see [37].

The set of solutions $\left\{(X_\mu, y_\mu, Z_\mu), \mu > 0\right\}$ defines a smooth curve, the commonly called primal-dual central path or central trajectory. At this point the term "path-following" becomes clear. Primal-dual IPMs can be viewed as an iterative process where the iterates follow the central path in the direction of decreasing μ. In other words, each iteration tries to move towards a point on the central path that is closer to optimality than the current solution.

An equivalent variant of $\mu Z^{-1} - X = 0$ is

$$ZX - \mu I = 0. \tag{1.8}$$

This relationship is similar to the complementary slackness condition $ZX = 0$. It is a slightly modified version of it. For $\mu = 0$ (1.8) represents precisely complementarity. For this reason it is called perturbed complementarity condition or centering property. So for a fixed $\mu > 0$ **(KKT)** for (1.7) may be written as

$$\textbf{(KKT)} \qquad \begin{aligned}
&\boldsymbol{\mathcal{A}}(X) - b = 0\\
&Z + \boldsymbol{\mathcal{A}^t}(y) = C\\
&ZX = \mu I\\
&X \in S(n)^{++},\ Z \in S(n)^{++},\ y \in \mathbb{R}^m.
\end{aligned}$$

Suppose X and Z are complementary. Hence by Lemma A.1 $\langle Z, X\rangle = 0$. Moreover, for X feasible in **(PSDP)** and (y, Z) feasible in **(DSDP)** the duality gap is also given by $\langle Z, X\rangle$ as we can see below.

$$\langle C, X\rangle - \langle b, y\rangle \overset{\text{by (1.4), (1.1)}}{=} \langle Z + \boldsymbol{\mathcal{A}^t}(y), X\rangle - \langle \boldsymbol{\mathcal{A}}(X), y\rangle \overset{\text{by (1.2)}}{=} \langle Z, X\rangle.$$

We conclude that any complementary solution has zero duality gap. Therefore primal and dual methods can also be interpreted as methods minimizing the duality gap.

A classical result of Fiacco and McCormick [31] holds also in SDP. It shows that the primal objective decreases along the central path and the dual objective increases along the central path. Let $\Big\{(X_\mu, y_\mu, Z_\mu), \mu > 0\Big\}$ defined by **(KKT)** , i.e., this point lies on the central path.

Lemma 1.15 *(monotonicity) The objective function $\langle C, X_\mu\rangle$ of* **(PSDP)** *monotonically decreases and the objective function $\langle b, y_\mu\rangle$ of* **(DSDP)** *monotonically increases as μ decreases.*

Proof. Let $f(X,\mu)$ denote the primal barrier function and $g(y, Z, \mu)$ denote the dual barrier function, respectively. Suppose $\mu_1 > \mu_2 > 0$. By definition (y_{μ_1}, Z_{μ_1}) maximizes $g(y, Z, \mu_1)$ and (y_{μ_2}, Z_{μ_2}) maximizes $g(y, Z, \mu_2)$. For this reason the following pair of inequalities holds.

$$g(y_{\mu_1}, Z_{\mu_1}, \mu_1) \geq g(y_{\mu_2}, Z_{\mu_2}, \mu_1) \quad \text{and} \quad g(y_{\mu_2}, Z_{\mu_2}, \mu_2) \geq g(y_{\mu_1}, Z_{\mu_1}, \mu_2)$$

or equivalently

$$\begin{aligned} \langle b, y_{\mu_1}\rangle + \mu_1 \log\det(Z_{\mu_1}) &\geq \langle b, y_{\mu_2}\rangle + \mu_1 \log\det(Z_{\mu_2}) \\ \langle b, y_{\mu_2}\rangle + \mu_2 \log\det(Z_{\mu_2}) &\geq \langle b, y_{\mu_1}\rangle + \mu_2 \log\det(Z_{\mu_1}). \end{aligned}$$

We divide the first inequality by μ_1, the second inequality by μ_2 and add up the two inequalities. This way we get

$$\Big(\langle b, y_{\mu_1}\rangle - \langle b, y_{\mu_2}\rangle\Big)\left(\left(\frac{1}{\mu_1}\right) - \left(\frac{1}{\mu_2}\right)\right) \geq 0.$$

The second term $\left(\left(\frac{1}{\mu_1}\right) - \left(\frac{1}{\mu_2}\right)\right) < 0$ by $\mu_1 > \mu_2 > 0$. Hence $\langle b, y_{\mu_2}\rangle \geq \langle b, y_{\mu_1}\rangle$ must hold. This proves the second part of the lemma.

To prove the statement associated with the primal problem it remains to show that the respective X_μ minimizes $f(X,\mu)$. Then the same style of argument we used to prove the second part completes the proof. Recall that the optimality conditions for (1.5) are given by (1.6). It is easily checked that X_μ solves **(KKT)** (simply set $Z := \mu X^{-1}$).

□

Corollary 1.16 *Let $\mu_1 > \mu_2 > 0$. Suppose $(X_{\mu_1}, y_{\mu_1}, Z_{\mu_1})$ and $(X_{\mu_2}, y_{\mu_2}, Z_{\mu_2})$ on the central path are given. Then:*

$$\langle C, X_{\mu_1}\rangle - \langle C, X_{\mu_2}\rangle \leq n(\mu_1 - \mu_2).$$

Proof.

$$\langle C, X_{\mu_1}\rangle - \langle C, X_{\mu_2}\rangle \leq \Big(\langle C, X_{\mu_1}\rangle - \langle b, y_{\mu_1}\rangle\Big) - \Big(\langle C, X_{\mu_2}\rangle - \langle b, y_{\mu_2}\rangle\Big) = n(\mu_1 - \mu_2).$$

□

1.4.1 Search Directions

Let $p = (X, y, Z) \in S(n) \times \mathbb{R}^m \times S(n)$ and let $F_\mu(p) : S(n) \times \mathbb{R}^m \times S(n) \mapsto \mathbb{R}^m \times S(n) \times M(n)$ be defined as the following nonlinear mapping.

$$F_\mu(p) := \begin{pmatrix} \boldsymbol{\mathcal{A}}(X) - b \\ Z + \boldsymbol{\mathcal{A}}^t(y) - C \\ ZX - \mu I \end{pmatrix}$$

$M(n)$ is used to denote the vector space of $n \times n$ matrices.

Obviously, a point p^* satisfies the **(KKT)** conditions if and only if $F_\mu(p^*) = 0$. Hence finding such a p^* is equivalent to solving the **(KKT)** system. Newton's method is used for solving this problem approximately, i.e., the linearized system to solve for the step direction $\Delta p = (\Delta X, \Delta y, \Delta Z)$ is

$$F_\mu(p) + \nabla F_\mu(p)\Delta p = 0.$$

In our case this system is given by

$$\begin{aligned} \boldsymbol{\mathcal{A}}(\Delta X) &= -r_P \\ \Delta Z + \boldsymbol{\mathcal{A}}^t(\Delta y) &= -r_D \\ \Delta Z\, X + Z\, \Delta X &= -r_{pC} \end{aligned} \tag{1.9}$$

where $r_P = \boldsymbol{\mathcal{A}}(X) - b$ and $r_D = Z + \boldsymbol{\mathcal{A}}^t(y) - C$ stand for the primal and dual residuum, respectively and $r_{pC} = ZX - \mu I$ denotes the perturbed complementarity condition. This direction is known as the Helmberg-Kojima-Monteiro direction (denoted HKM-direction), see [54, 73, 95].

Next the perturbed complementarity condition is used to compute the associated value of μ.

$$\mu = \frac{\langle Z, X \rangle}{n} \tag{1.10}$$

After reducing this parameter the complete process is applied repeatedly.

In general this search direction does not yield a symmetric component ΔX. This comes from the fact that $F_\mu(\cdot)$ is a function between spaces of different dimensions. Since Newton's method is applied iteratively one cannot proceed after computing the search direction without modification. The key is that the ΔX component can be symmetrized before iterating. Consequently, as stated by many authors this method is not strictly a Newton method.

However, this search direction always yields a symmetric component ΔZ because the second equation $\Delta Z + \boldsymbol{\mathcal{A}}^t(\Delta y) = -r_D$ in (1.9) forces ΔZ to be symmetric.

There are several equivalent ways to write the perturbed complementarity condition. In fact, some choices of writing the perturbed complementarity condition lead to symmetric directions while other do not. For interior-point methods, it is crucial to maintain symmetry of the variables. This in turn requires symmetric directions at each iteration. One way to ensure symmetric directions is to take the symmetric part of the third equation in (1.9), see [3]. Another way of symmetrizing this equation was introduced by Monteiro, [95]. He suggested to apply a similarity transformation $P(\cdot)P^{-1}$ to both sides of the third equation in (1.9) and then to symmetrize it. Monteiro called this process the scaling and symmetrization scheme. Monteiro's approach covers only the cases $P = Z^{-1/2}$ and $P = X^{1/2}$. Zhang [135] generalized this scheme to general scaling matrices P. He has introduced a family of symmetrization operators $H_p(M) : \mathbb{R}^{n\times n} \mapsto S(n)$, defined by

$$H_p(M) = \frac{1}{2}\Big(PMP^{-1} + (PMP^{-1})^T\Big)$$

for any nonsingular (but fixed) P. Since, as shown by Zhang [135]

$$H_p(ZX) = \mu I \text{ if and only if } ZX = \mu I, \;\; (Z, X) \in (S(n)^{++})^2$$

system (1.9) is equivalent to

$$\begin{aligned} \boldsymbol{\mathcal{A}}(\Delta X) &= -r_P \\ \Delta Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) &= -r_D \\ H_p(Z\Delta X + \Delta Z X) &= \mu I - H_p(ZX). \end{aligned}$$

This search direction is called the Monteiro-Zhang unified direction (or Monteiro-Zhang family) [94, 135]. It yields a class of search directions parametrized by P. The most commonly used search directions depending on the choice of P are listed below.

(1) NT-direction, $P = W^{-1/2}$ satisfying $WZW = X$, see [99, 120].

(2) HKM-direction, $P = Z^{1/2}$, see [54, 73, 95].

(3) AHO-direction, $P = I$, see [3].

In the sequel, we carry out the basic steps of a primal-dual path-following algorithm, using the HKM-direction, see Table 1.1. We follow the presentation given in [54, 51].

Note that this algorithm generates a sequence $(X^k, y^k, Z^k) \in \mathrm{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$. We mention that a condition on α which guarantees the positive semidefiniteness of the updated variables is given in [73]. The following lemma describes this condition. We include also a proof of the lemma which is not given in [73].

Lemma 1.17 *Suppose that $X \in S(n)^{++}$, $\Delta X \in S(n)$ and $\alpha \geq 0$. Let $\xi_{\min}$ be the minimum eigenvalue of the matrix $X^{-1}\Delta X$ and let*

$$\bar{\alpha} = \sup\Big\{\alpha : 1 + \alpha\,\xi_{\min} \geq 0\Big\} = \begin{cases} -1/\xi_{\min} & if\ \xi_{\min} < 0 \\ +\infty & otherwise. \end{cases}$$

Then $X + \alpha\Delta X \in S(n)^{++}$ if and only if $\alpha < \bar{\alpha}$.

(**algorithm.**)

---

input: initial triple $p^\circ = (X^\circ, y^\circ, Z^\circ)$ with $X^\circ \in S(n)^{++}$ and $Z^\circ \in S(n)^{++}$.

---

begin
  set iteration counter $k = 0$.
  while (p^k not close to optimality) **repeat**
    $p \leftarrow p^k$
    $r_P = \boldsymbol{\mathcal{A}}(X) - b,\ r_D = Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - C,\ \mu = \frac{\langle Z,X\rangle}{n}$
    choose $\sigma \in (0,1)$ and set $r_{pC} = ZX - \sigma\mu I$
    compute the search direction $\Delta p = (\Delta X, \Delta y, \Delta Z)$ by solving (1.9)
    take the symmetric part of ΔX: $\Delta X \leftarrow \frac{\Delta X^T + \Delta X}{2}$
    perform standard line search, i.e., compute step lengths α_P and $\alpha_D \in (0,1)$,
     such that $X + \alpha_P \Delta X \in S(n)^{++}$ and $Z + \alpha_D \Delta Z \in S(n)^{++}$
    update: $\alpha := \min\{\alpha_P, \alpha_D\}$, $X \leftarrow X + \alpha\Delta X$, $y \leftarrow y + \alpha\Delta y$, $Z \leftarrow Z + \alpha\Delta Z$
    $k \leftarrow k+1$ and $p^k \leftarrow (X, y, Z)$
  end
end

Table 1.1: *Path-following algorithm for Semidefinite Programming, HKM-direction*

Proof. Since $X \in S(n)^{++}$ there exists a nonsingular matrix, say $\sqrt{X}$, such that $X + \alpha\Delta X = \sqrt{X}(I + \alpha\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1})\sqrt{X}$. Lemma A.6 shows that $X + \alpha\Delta X \in S(n)^{++}$ if and only if $I + \alpha\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1} \in S(n)^{++}$. In fact, $X^{-1}\Delta X$ is similar to $\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1}$. Hence $\xi_{\min}$ is the minimum eigenvalue of $\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1}$. Let $\lambda_{\min}$ denote the minimum eigenvalue of $I + \alpha\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1} \in S(n)^{++}$. Obviously, $\lambda_{\min} = 1 + \alpha\,\xi_{\min}$. The rest of the proof is quite straightforward.

□

Hence computing the minimum of the eigenvalues of $Z^{-1}\Delta Z$ and $X^{-1}\Delta X$ in the updating step of our path-following algorithm gives a tight upper bound for a step size $\alpha \geq 0$.

1.5 Self-Dual Embedding Technique

Numerous interior-point methods for SDP require that the starting point $(X^\circ, y^\circ, Z^\circ)$ is strictly feasible and therefore are called feasible interior-point methods. For many problems, however, a strictly feasible starting point is not easy to find, even if such a point exists. More recently, infeasible interior-point methods were proposed for SDP, see [73, 135, 108, 70]. The starting point must be an interior point, but it need not be feasible.

Almost all of the feasible and infeasible interior-point algorithms for SDP have made the assumption that the Slater constraint qualification is satisfied. Under this assumption strong duality holds and all of these algorithms find an optimal primal-dual solution.

However, in general it is not known in advance whether or not a semidefinite program does have a complementary solution. Unfortunately, in SDP Slater's condition may fail to hold. The subject of this chapter is to give information in more detail about the solution status of the given problem.

In LP one way to handle the problem of initialization is embedding the given problem in a larger problem, called homogeneous skew symmetric self-dual problem.

The changed problem does always have a solution even in the case that the original does not. Moreover, this solution shows whether the original problem does have a solution and if so there is a natural way to compute the solution to the original problem. Otherwise, if optimal solutions do not exist the solution of the extended problem shows: either both the primal and the dual are infeasible or else one of the primal and the dual is infeasible and the other problem is unbounded.

A homogeneous self-dual linear programming algorithm and corresponding theory was introduced by Ye, Todd and Mizuno [133]. They investigated a homogeneous skew symmetric self-dual model for which a strictly feasible point is always known. Their approach does not require any regularity assumptions. Any feasible IPM is a candidate to solve the embedding problem. In 1996, Xu, Hung and Ye [132] proposed a simplified version. They called it the simplified homogeneous self-dual formulation. All the properties to identify the solution status of the original problem remain preserved. However, since this formulation does not have any strictly feasible point, this variant must be solved using infeasible IPMs.

Potra and Sheng [107] were the first who generalized the embedding strategy to SDP. They proposed a homogeneous primal-dual feasibility model for SDP, which is based on Goldman and Tucker's self-dual feasibility model for LP, see [38, 125]. They restrict their study on characterizing complementary solutions. Other possibilities concerning the solution status of a semidefinite problem are not considered.

Extended variants of the homogeneous formulation for SDP were proposed by De Klerk, Roos and Terlaky [24, 23, 22] (they consider SDP) and independently by Luo, Sturm and Zhang [80] (they consider the more general case of conic convex programming).

This section is organized as follows. Section 1.5.1 provides background material. The homogeneous self-dual formulation for SDP is introduced in section 1.5.2. A generalized model, based on LP (see [133]), is presented in section 1.5.3. In this section we point out the main features of the embedding. In section 1.5.4 we present a formal description of our path-following algorithm for solving the self-dual problem. Implementation details are given in section 1.5.5. In section 1.5.6 we show that the iterates produced by our algorithm remain within a compact set. Finally we consider asymptotic refinements for SDP.

Throughout this section, we use the following notation. For $X \in M(n)$ we will write $\|X\|$ to denote the Frobenius matrix norm. If v denotes a vector in $\mathbb{R}^n$ we shall write $\|v\|$ for the Euclidean norm of v. The eigenvalues of a matrix $X \in M(n)$ are denoted by $\lambda_j(X), (j = 1, \ldots, n)$. We use $\lambda_{\min}(Z^k)$ $(\lambda_{\max}(Z^k))$ to denote the minimum (maximum) eigenvalue of Z^k for every $k = 0, 1, 2, \ldots$, respectively.

1.5.1 Maximal Complementarity

In this section we review some definitions and results related to SDP which will be used in the sequel.

Definition 1.18 *A problem is called self-dual if its own dual is the same as the primal.*

Definition 1.19 [23] *An optimal solution (X^*, Z^*) is called a maximally complementary solution if X^* and Z^* have maximal rank.*

De Klerk et al. [24, 23] show that maximally complementary solutions exist. In particular, they prove that the central path converges to a maximally complementary solution, called analytic center, if μ decreases to zero. An analogous result holds in LP, see [48].

Recall, the central path is formed by the set of solutions $\Big\{(X_\mu, y_\mu, Z_\mu), \mu > 0\Big\}$ of **(KKT)**.

Theorem 1.20 [24, 23] *(maximal complementarity) Consider a fixed sequence $(\mu^k)_{k\in\mathbb{N}} \searrow 0$ with $\mu^k > 0$ and the corresponding solutions $\{(X_\mu, y_\mu, Z_\mu)\}$ of* **(KKT)***. Then there exists a subsequence of (X_{μ^k}, Z_{μ^k}) which converges to a maximally complementary solution.*

1.5.2 The Homogeneous Self-Dual Formulation

In this section we generalize the presentation in LP given by Xu, Hung and Ye [132]. The homogeneous self-dual model **(HSDP)** for SDP is given by

$$
\textbf{(HSDP)} \qquad \min 0
$$

$$
\begin{array}{rrrrcl}
 & & \boldsymbol{\mathcal{A}}(X) & -b\tau & = & 0 \\
\text{s.t.} & -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) & & +C\tau & \succeq & 0 \\
 & \langle b, y\rangle & -\langle C, X\rangle & & \geq & 0
\end{array}
$$

$$
y \in \mathbb{R}^m,\ \tau \geq 0,\ X \in S(n)^+.
$$

Note that problem **(HSDP)** has no strictly feasible point. Such a point $(\bar{X}, \bar{y}, \bar{\tau})$ would satisfy $\boldsymbol{\mathcal{A}}(\bar{X}) = \bar{\tau} b, \ -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\bar{y}) + \bar{\tau} C \in S(n)^{++}\ \langle b, \bar{y}\rangle - \langle C, \bar{X}\rangle > 0$. This yields a strictly feasible solution $(X, y, Z) = (\bar{X}/\bar{\tau}, \bar{y}/\bar{\tau}, C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\bar{y}/\bar{\tau}))$ to **(PSDP)** and **(DSDP)**.

Hence by weak duality $0 < \langle Z, X\rangle = \langle C, X\rangle - \langle b, y\rangle = \frac{1}{\bar{\tau}}(\langle C, \bar{X}\rangle - \langle b, \bar{y}\rangle) < 0$, which is a contradiction. For this reason a strictly feasible point cannot exist for **(HSDP)**.

If **(HSDP)** has a feasible solution (X, y, Z) with $\tau > 0$ (note, the all zero solution is also feasible), an optimal solution to the original problem is given by $(X/\tau, y/\tau, C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y)/\tau)$. The next theorem is a simple generalization from LP, see e.g. Wright [131].

Theorem 1.21 [107] *The primal-dual pair* **(PSDP)** *and* **(DSDP)** *has an optimal solution if and only if a maximally complementary solution of* **(HSDP)** *has* $\tau > 0$ *and* $\langle b, y\rangle = \langle C, X\rangle$.

Potra and Sheng [107] presented two infeasible interior-point algorithms to solve the homogeneous formulation **(HSDP)**.

1.5.3 The Extended Homogeneous Self-Dual Formulation

In this section we generalize the homogeneous and self-dual linear programming algorithm, introduced by Ye, Todd and Mizuno [133]. We denote the generalized model by **(EHSDP)**. As we will see later in this section it is quite easy to give a strictly feasible initial point for this extended model. That means that we can use any feasible interior-point algorithm to solve the problem. This advantage is also stated in [131] for LP and in [24, 22] for SDP. Note, that the extended model proposed by de Klerk et al. [24, 23, 22] has dimension $n+2$, while the dimension of **(EHSDP)** is $n+1$.

Given any $X^\circ \in S(n)^{++}$, $Z^\circ \in S(n)^{++}$ and $y^\circ \in I\!R^m$, **(EHSDP)** is given by

$$
\textbf{(EHSDP)} \qquad \min\ (\langle X^\circ, Z^\circ\rangle + 1)\vartheta
$$

$$
\text{s.t.}\quad \begin{array}{ccccccc}
 & \boldsymbol{\mathcal{A}}(X) & -b\tau & +\bar{b}\vartheta & = & 0 \\
-\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) & & +C\tau & -\bar{C}\vartheta & \succeq & 0 \\
\langle b, y\rangle & -\langle C, X\rangle & & +\bar{z}\vartheta & \geq & 0 \\
-\langle \bar{b}, y\rangle & +\langle \bar{C}, X\rangle & -\bar{z}\tau & & = & -\langle X^\circ, Z^\circ\rangle - 1
\end{array} \tag{1.11}
$$

$$
X \in S(n)^+,\ y \in I\!R^m,\ \tau \geq 0,\ \vartheta \in I\!R,
$$

where

$$
\bar{b} = b - \boldsymbol{\mathcal{A}}(X^\circ), \quad \bar{C} = C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^\circ) - Z^\circ, \quad \bar{z} = \langle C, X^\circ\rangle + 1 - \langle b, y^\circ\rangle. \tag{1.12}
$$

The terms $\bar{b}$ and $\bar{C}$ are error terms to primal and dual infeasibility of $(X^\circ, y^\circ, Z^\circ)$. For $\tau = 1$ and $\vartheta = 0$ the leading two constraints in (1.11) reduce exactly to primal and dual feasibility (1.1) and (1.4). The third constraint in (1.11) shows that the duality gap for the original pair **(PSDP)** and **(DSDP)** is nonpositive. However, since X° is feasible for **(PSDP)** and Z° is feasible for **(DSDP)** weak duality holds, i.e., the duality gap is always nonnegative. Hence the setting $\tau = 1$ and $\vartheta = 0$ yields an optimal solution to the original problem.

The set of feasible solutions (strictly feasible solutions) of (1.11) will be denoted by $\mathcal{F}(\mathcal{H})$ $(\mathrm{int}(\mathcal{F}(\mathcal{H})))$, respectively.

Remark 1.22 *For X° feasible in* **(PSDP)** *and (y°, Z°) feasible in* **(DSDP)** *the terms $\bar{b}$ and $\bar{C}$ vanishes and model* **(EHSDP)** *represents exactly model* **(HSDP)**.

Let Z and ν denote the surplus variables associated with the second and the third relationship in (1.11). A simplified version of the fourth constraint in (1.11) is then given by

$$\langle Z^\circ, X\rangle + \langle Z, X^\circ\rangle + \tau + \nu - \vartheta(\langle Z^\circ, X^\circ\rangle + 1) = \langle Z^\circ, X^\circ\rangle + 1. \tag{1.13}$$

This is true since

$$\begin{aligned}
&\langle X^\circ, Z^\circ\rangle + 1 \\
&\quad = \langle \bar{b}, y\rangle - \langle \bar{C}, X\rangle + \bar{z}\tau \\
&\quad = \Big\langle b - \boldsymbol{\mathcal{A}}(X^\circ), y\Big\rangle - \Big\langle C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^\circ) - Z^\circ, X\Big\rangle + \Big(\langle C, X^\circ\rangle + 1 - \langle b, y^\circ\rangle\Big)\ \tau \\
&\quad = \Big\langle \boldsymbol{\mathcal{A}}(X) - b\tau, y^\circ\Big\rangle + \Big\langle -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) + C\tau, X^\circ\Big\rangle + \langle b, y\rangle - \langle C, X\rangle + \langle Z^\circ, X\rangle + \tau \\
&\quad = \langle -\bar{b}\vartheta, y^\circ\rangle + \langle Z + \bar{C}\vartheta, X^\circ\rangle + \nu - \bar{z}\vartheta + \langle Z^\circ, X\rangle + \tau \quad \text{(by (1.11))}.
\end{aligned}$$

Here the second equality follows from (1.12). It is quite straightforward that replacing $\bar{b}$, $\bar{C}$ and $\bar{z}$ in the last equality with its defining terms given in (1.12) yields the equivalence between the fourth constraint in (1.11) and (1.13).

Definition 1.23 *Suppose $(X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ is any optimal solution to* **(EHSDP)**. *It is called a maximally complementary solution if X^* and Z^* have maximal rank, and if the number of strictly positive variables of τ^* and ν^* is maximum.*

The following result is a generalization from LP, see [133].

Theorem 1.24 *(features of* **(EHSDP)***)*

(i) **(EHSDP)** *is self-dual.*

(ii) **(EHSDP)** *has a strictly feasible point.*

(iii) **(EHSDP)** *has an optimal solution.*

(iv) *There exists a maximally complementary solution $(X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ for which*

$$Z^*X^* + \nu^*\tau^* = 0.$$

(v) $\vartheta = 0$ *at any optimal solution* $(X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ *and for any* $(X, y, Z, \tau, \vartheta, \nu) \in \mathcal{F}(\mathcal{H})$

$$((\langle X^\circ, Z^\circ \rangle + 1)\vartheta = \langle Z, X \rangle + \nu\tau.$$

Proof. The homogeneous model **(EHSDP)** has a skew-symmetric system matrix $\mathcal{R}$. In a slight abuse of notation we write it in operator form.

$$\mathcal{R} = \begin{pmatrix} 0 & \boldsymbol{\mathcal{A}} & -b & \bar{b} \\ -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}} & 0 & C & -\bar{C} \\ b^T & -C & 0 & \bar{z} \\ -\bar{b}^T & \bar{C} & -\bar{z} & 0 \end{pmatrix}$$

To get the dual problem **(D)** to **(EHSDP)** we apply duality directly. The rules for obtaining the dual are a direct extension of these rules for duality in LP and summarized in [1]. Denote by $y' \in \mathbb{R}^m, X' \in S(n)^+, \tau' \geq 0$ and $\vartheta' \in \mathbb{R}$ the dual variables. Let $\tilde{c}, \tilde{b}, w$ and v in $(\mathbb{R}^m \times S(n)^+ \times \mathbb{R}^+ \times \mathbb{R})$ be defined as follows. ($\mathbb{R}^+$ denotes the set of real numbers ≥ 0). We write $\tilde{c}^T$ for $(0, 0, 0, \langle X^\circ, Z^\circ \rangle + 1)$, $\tilde{b}^T$ for $(0, 0, 0, -\langle X^\circ, Z^\circ \rangle - 1)$, w^T for (y, X, τ, ϑ) and v^T for $(y', X', \tau', \vartheta')$. For convenience we define

$$\succsim \; :\Leftrightarrow \begin{pmatrix} = \\ \succeq \\ \geq \\ = \end{pmatrix} \qquad \text{and} \qquad \precsim \; :\Leftrightarrow \begin{pmatrix} = \\ \preceq \\ \leq \\ = \end{pmatrix}.$$

With this setting the primal-dual pair **(EHSDP)** and **(D)** can be written as follows.

$$\begin{array}{lll} \textbf{(EHSDP)} & \min & \tilde{c}^T w \\ & \text{s.t.} & \mathcal{R}(w) \succsim \tilde{b} \\ & & w \in (\mathbb{R}^m \times S(n)^+ \times \mathbb{R}^+ \times \mathbb{R}) \end{array}$$

$$\begin{array}{lll} \textbf{(D)} & \max & \tilde{b}^T v \\ & \text{s.t.} & \mathcal{R}^T(v) \precsim \tilde{c} \\ & & v \in (\mathbb{R}^m \times S(n)^+ \times \mathbb{R}^+ \times \mathbb{R}) \end{array}$$

Since $\mathcal{R} = -\mathcal{R}^T$ and $\tilde{b} = -\tilde{c}$ we can rewrite the constraints in **(D)** and get $\mathcal{R}(v) \succsim -\tilde{c}$ which is equivalent to $\mathcal{R}(v) \succsim \tilde{b}$. Likewise we rewrite the objektive function. This way we get $\max\ \tilde{b}^T v = -\min -\tilde{b}^T v = -\min\ \tilde{c}^T w$. This proofs (i).

(ii) A strictly feasible point for **(EHSDP)** is given by $X = X^\circ, y = y^\circ, Z = Z^\circ, \tau = 1, \vartheta = 1, \nu = 1$.
(iii) **(EHSDP)** is self-dual. Thus by (ii) **(D)** has also a strictly feasible point, i.e., the Slater constraint qualification holds for both **(EHSDP)** and **(D)** which ensures that optimal solutions exists.
(iv) is a consequence of Theorem 1.20.
(v) Since $(X, y, Z, \tau, \vartheta, \nu) \in \mathcal{F}(\mathcal{H})$, multiplying the fourth constraint in (1.11) by ϑ yields

$$\begin{aligned}
(\langle X^\circ, Z^\circ\rangle + 1)\vartheta &= \langle \bar{b}\vartheta, y\rangle - \langle \bar{C}\vartheta, X\rangle + \bar{z}\vartheta\tau \\
&= \langle b\tau - \boldsymbol{\mathcal{A}}(X), y\rangle - \langle \bar{C}\vartheta, X\rangle + \bar{z}\vartheta\tau \\
&= \Big\langle -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) + C\tau - \bar{C}\vartheta, X\Big\rangle + \Big(\langle b, y\rangle - \langle C, X\rangle + \bar{z}\vartheta\Big)\tau \\
&= \langle Z, X\rangle + \nu\tau \\
&\geq 0 \qquad (\text{since } (X, y, Z, \tau, \vartheta, \nu) \in \mathcal{F}(\mathcal{H})).
\end{aligned}$$

Let $(X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ denote any optimal solution, hence $Z^*X^* + \nu^*\tau^* = 0$. Since $(\langle X^\circ, Z^\circ\rangle + 1) > 0$, $\vartheta^* = 0$ must hold. This proves (v).

□

We are not restricted in the choice of X°, y° and Z°. Henceforth we set $X^\circ = I$, $Z^\circ = I$ and $y^\circ = 0$, where I denotes the identity matrix of appropriate size. With this setting **(EHSDP)** can be written as

$$\begin{array}{lll}
(\textbf{EHSDP}) & \min & (n+1)\vartheta \\
& \text{s.t.} & \mathcal{R}(w) \succsim \tilde{b} \\
& & w \in (\mathbb{R}^m \times S(n)^+ \times \mathbb{R}^+ \times \mathbb{R}),
\end{array} \tag{1.14}$$

where $\tilde{b}^T = (0, 0, 0, -(n+1))$. This specific choice of the initial point changes the set of relationships (1.12) to

$$\bar{b} = b - \boldsymbol{\mathcal{A}}(I), \quad \bar{C} = C - I, \quad \bar{z} = \langle C, I\rangle + 1 \tag{1.15}$$

and (1.13) to

$$\langle I, X\rangle + \langle Z, I\rangle + \tau + \nu = (1 + \vartheta)(n+1). \tag{1.16}$$

Theorem 1.25 *Suppose $p^* = (X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ denotes any optimal solution to* **(EHSDP)**. *Then*

(i) *If $\tau^* > 0$ then an optimal solution for* **(PSDP)** *and* **(DSDP)** *is given by $(X^*/\tau^*, y^*/\tau^*, Z^*/\tau^*)$*

(ii) *If $\nu^* > 0$ then at least one of* **(PSDP)** *and* **(DSDP)** *is infeasible.*

Proof. By Theorem 1.24 item (v) we have $\vartheta^* = 0$ and $\langle Z^*, X^*\rangle + \nu^*\tau^* = 0$ at any optimal solution. We know that both terms vanishes.
(i) Since $\tau^* > 0$, ν^* must be zero due to optimality. We substitute the decision variables in (1.11) by $p^* = (X^*, y^*, Z^*, \tau^*, \vartheta^* = 0, \nu^* = 0)$ to obtain the following set of equations.

$$\boldsymbol{\mathcal{A}}(X^*) - b\tau^* = 0, \quad Z^* + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*) = C\tau^*, \quad \langle b, y^*\rangle - \langle C, X^*\rangle = 0$$

The setting $(\tilde{X}^*, \tilde{y}^*, \tilde{Z}^*) := (X^*/\tau^*, y^*/\tau^*, Z^*/\tau^*)$ yields a feasible solution for the original primal-dual pair **(PSDP)** and **(DSDP)** with primal cost = dual cost. In fact, we have

$$\begin{aligned} &\boldsymbol{\mathcal{A}}(\tilde{X}^*) - b = 0, \quad \tilde{Z}^* + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\tilde{y}^*) = C, \quad \langle b, \tilde{y}^*\rangle - \langle C, \tilde{X}^*\rangle = 0 \\ &\tilde{X}^* \in S(n)^+, \ \tilde{Z}^* \in S(n)^+, \ \tilde{y}^* \in \mathbb{R}^m. \end{aligned}$$

This shows optimality.
(ii) In this case we must have $\tau^* = 0$, again due to optimality. Since ν^* is the slack variable assigned to the third constraint in (1.11) we have $\langle C, X^*\rangle - \langle b, y^*\rangle = -\nu^* < 0$. That means that at least one of $\langle C, X^*\rangle$ and $-\langle b, y^*\rangle$ must be strictly less than zero.

- Case 1: $\langle C, X^*\rangle < 0$

  The first constraint in (1.11) reduce to $\boldsymbol{\mathcal{A}}(X^*) = 0$. Assume **(DSDP)** is feasible, i.e., there exists a solution $\tilde{y} \in \mathbb{R}^m, \tilde{Z} \in S(n)^+$ for which $\tilde{Z} + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\tilde{y}) = C$. Scalar multiplication by X^* supplies $\langle \tilde{Z}, X^*\rangle + \langle \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\tilde{y}), X^*\rangle = \langle C, X^*\rangle$. From $\tilde{Z} \in S(n)^+$ and $X^* \in S(n)^+$ it follows that $\langle \tilde{Z}, X^*\rangle \geq 0$. Furthermore we have $\langle \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\tilde{y}), X^*\rangle = \langle \tilde{y}, \boldsymbol{\mathcal{A}}(X^*)\rangle = 0$.

  Thus we have $0 \leq \langle \tilde{Z}, X^*\rangle = \langle C, X^*\rangle$. On the other hand by assumption we have $\langle C, X^*\rangle < 0$. This is a contradiction, hence **(DSDP)** must be infeasible.

- Case 2: $-\langle b, y^*\rangle < 0$

  The second constraint in (1.11) reduce to $Z^* + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*) = 0$. Assume **(PSDP)** is feasible, i.e., there exists a solution $\tilde{X} \in S(n)^+$ such that $\boldsymbol{\mathcal{A}}(\tilde{X}) = b$. Multiplying this equation scalar by y^*, we get $\langle \boldsymbol{\mathcal{A}}(\tilde{X}), y^*\rangle = \langle b, y^*\rangle$. This equation is equivalent to $\langle \tilde{X}, \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*)\rangle = \langle b, y^*\rangle$ with $\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*) = -Z^*$. Since $\tilde{X} \in S(n)^+$ and $Z^* \in S(n)^+$ it follows that $\langle \tilde{X}, Z^*\rangle \geq 0$ and therefore $\langle \tilde{X}, -Z^*\rangle = -\langle \tilde{X}, Z^*\rangle \leq 0$.

  Summing up we have $0 \geq \langle \tilde{X}, -Z^*\rangle = \langle b, y^*\rangle > 0$ which is a contradiction. Thus **(PSDP)** must be infeasible.

- Case 3: $\langle C, \tilde{X}^*\rangle < 0$ and $-\langle b, \tilde{y}^*\rangle < 0$

  This is only a combination of the first two cases. **(PSDP)** and **(DSDP)** are both infeasible.

□

Remark 1.26 *Supppose* $p^* = (X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ *denotes a maximally complementary solution to* **(EHSDP)**. *Since* $\vartheta^* = 0$ *in any optimal solution (see Theorem 1.24) Theorem 1.21 remains also valid for* **(EHSDP)**. *Hence, if* $\tau^* = 0$ *and* $\nu^* = 0$, *then an optimal solution for the original primal-dual pair does not exist. The case* $\tau^* = 0$ *and* $\nu^* = 0$ *does not directly permit a conclusion regarding the duality status of the original pair of problems.*

1.5.4 Algorithmic Aspects

Recall **(EHSDP)** is self-dual and satisfies the Slater constraint qualification. Hence $p = (X, y, Z, \tau, \vartheta, \nu)$ is optimal if and only if

$$\begin{array}{ccccrcl} & \boldsymbol{\mathcal{A}}(X) & -b\tau & +\bar{b}\vartheta & = & 0 \\ -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) & & +C\tau & -\bar{C}\vartheta & = & Z \\ \langle b, y\rangle & -\langle C, X\rangle & & +\bar{z}\vartheta & = & \nu \\ -\langle \bar{b}, y\rangle & +\langle \bar{C}, X\rangle & -\bar{z}\tau & & = & -(n+1) \\ & & & ZX & = & 0 \\ & & & \nu\tau & = & 0 \end{array} \tag{1.17}$$

$$X \in S(n)^+,\ y \in \mathbb{R}^m,\ Z \in S(n)^+,\ \tau \geq 0,\ \vartheta \in \mathbb{R},\ \nu \geq 0.$$

We proceed in the same way proposed in section 1.4.1. Let $F_\mu(p)$, $p = (X, y, Z, \tau, \vartheta, \nu)$ be defined as the following nonlinear mapping.

$$F_\mu(p) := \begin{pmatrix} & & \boldsymbol{\mathcal{A}}(X) & -b\tau & +\bar{b}\vartheta \\ -Z & -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) & & +C\tau & -\bar{C}\vartheta \\ -\nu & +\langle b, y\rangle & -\langle C, X\rangle & & +\bar{z}\vartheta \\ (n+1) & -\langle \bar{b}, y\rangle & +\langle \bar{C}, X\rangle & -\bar{z}\tau & \\ & & ZX - \mu I & & \\ & & \nu\tau - \mu & & \end{pmatrix} =: \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \end{pmatrix}$$

The problem is finding a point p^* for which $F_\mu(p^*) = 0$. Newton's method is used for solving this problem approximately. The step direction is obtained as solution of the linearized system

$$F_\mu(p) + \nabla F_\mu(p)\Delta p = 0.$$

which equivalently can be written as

$$
\begin{array}{cccccl}
 & & \boldsymbol{\mathcal{A}}(\Delta X) & -b\Delta\tau & +\bar{b}\Delta\vartheta & = -r_1 \\
-\Delta Z & -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) & & +C\Delta\tau & -\bar{C}\Delta\vartheta & = -r_2 \\
-\Delta\nu & +\langle b, \Delta y\rangle & -\langle C, \Delta X\rangle & & +\bar{z}\Delta\vartheta & = -r_3 \\
 & -\langle \bar{b}, \Delta y\rangle & +\langle \bar{C}, \Delta X\rangle & -\bar{z}\Delta\tau & & = -r_4 \\
 & & \Delta Z X + Z\Delta X & & & = -r_5 \\
 & & \Delta\nu\tau - \nu\Delta\tau & & & = -r_6.
\end{array} \tag{1.18}
$$

A slightly modified version of the path-following algorithm, proposed in section 1.4.1 is used for solving this problem. Recall, $(X^\circ, y^\circ, Z^\circ, \tau^\circ, \vartheta^\circ, \nu^\circ) = (I, 0, I, 1, 1, 1)$ is a strictly feasible point for **(EHSDP)**. If we take this point as initial point for our IPM then it follows from the linearity of the residuals r_1 up to r_4 that he updated point is again a feasible point. That way the method generates a sequence $(X^k, y^k, Z^k, \tau^k, \vartheta^k, \nu^k)_{k\in\mathbb{N}} \in \operatorname{int}(\mathcal{F}(\mathcal{H}))$. For the sake of completeness we briefly carry out the basic steps, see Table 1.2.

1.5.5 A Note on Efficient Implementation

Solving the system (1.18) directly can be very expensive. It turns out that the cost per iteration grows rapidly with the size n of the original problem. From a computational point of view the primary goal must be to increase the performance of our algorithm. In the rest of this section we show that computing the search direction $\Delta p = (\Delta X, \Delta y, \Delta Z, \Delta\tau, \Delta\vartheta, \Delta\nu)$ satisfying (1.18) can be done in an efficient way.

We first eliminate ΔZ from the second equation in (1.18). We get

$$\Delta Z = r_2 - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) + C\Delta\tau - \bar{C}\Delta\vartheta. \tag{1.19}$$

We then eliminate ΔX from the fifth equation in (1.18) and replace ΔZ by (1.19).

$$
\begin{aligned}
\Delta X &= -Z^{-1}r_5 - Z^{-1}\Delta Z X \\
&= -Z^{-1}r_5 - Z^{-1}r_2 X + Z^{-1}\Big(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - C\Delta\tau + \bar{C}\Delta\vartheta\Big)X
\end{aligned}
$$

We also eliminate $\Delta\nu$ from the last equation in (1.18) to get

$$\Delta\nu = -\frac{r_6}{\tau} - \frac{\nu}{\tau}\Delta\tau.$$

This relations substituted into the remaining equations of system (1.18) yields the following linear system for $\Delta y, \Delta\tau$ and $\Delta\vartheta$.

(**algorithm.**)

---

input: initial triple $p^\circ = (X^\circ, y^\circ, Z^\circ, \tau^\circ, \vartheta^\circ, \nu^\circ)$
with $X^\circ \in S(n)^{++},\ Z^\circ \in S(n)^{++},\ \tau^\circ = 1$, $\vartheta^\circ = 1$ and $\nu^\circ = 1$.

---

begin
  set iteration counter $k = 0$.
  while (p^k not close to optimality) **repeat**
    $p \leftarrow p^k$
    $r_1 = 0,\ r_2 = 0,\ r_3 = 0,\ r_4 = 0$, compute $\mu = \frac{\langle Z, X\rangle + \nu\tau}{n+1}$
    choose $\sigma \in (0,1)$ and set $r_5 = ZX - \sigma\mu I$ and $r_6 = \nu\tau - \sigma\mu$
    compute the search direction $\Delta p = (\Delta X, \Delta y, \Delta Z, \Delta\tau, \Delta\vartheta, \Delta\nu)$
     by solving (1.18)
    take the symmetric part of ΔX: $\Delta X \leftarrow \frac{\Delta X^T + \Delta X}{2}$
    perform standard line search, i.e., compute step length $\alpha \in (0,1)$, such that
     $X + \alpha\Delta X \in S(n)^{++}$, $Z + \alpha\Delta Z \in S(n)^{++}$, $\tau + \alpha\Delta\tau > 0$ and $\nu + \alpha\Delta\nu > 0$
    update: $p \leftarrow p + \alpha\Delta p$,
    $k \leftarrow k+1$ and $p^k \leftarrow (X, y, Z, \tau, \vartheta, \nu)$
  end
end

Table 1.2: *Path-following algorithm for the extended homogeneous skew symmetric self-dual model* **(EHSDP)**, *HKM-direction*

$$
\begin{aligned}
&\boldsymbol{\mathcal{A}}\left(Z^{-1}(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - C\Delta\tau + \bar{C}\Delta\vartheta)X\right) - b\Delta\tau + \bar{b}\Delta\vartheta && = \bar{r}_1 \\
&-\left\langle C, Z^{-1}(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - C\Delta\tau + \bar{C}\Delta\vartheta)X\right\rangle + \frac{\nu}{\tau}\Delta\tau + b^T\Delta y + \bar{z}\Delta\vartheta && = \bar{r}_2 \\
&\left\langle \bar{C}, Z^{-1}\left(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - C\Delta\tau + \bar{C}\Delta\vartheta\right)X\right\rangle - \bar{z}\Delta\tau - \bar{b}^T\Delta y && = \bar{r}_3
\end{aligned}
\tag{1.20}
$$

where

$$
\begin{aligned}
\bar{r}_1 &= -r_1 + \boldsymbol{\mathcal{A}}\left(Z^{-1}r_5 + Z^{-1}r_2X\right) \\
\bar{r}_2 &= -r_3 - \frac{r_6}{\tau} - \left\langle C, Z^{-1}r_5 + Z^{-1}r_2X\right\rangle \quad \text{and} \\
\bar{r}_3 &= -r_4 + \left\langle \bar{C}, Z^{-1}r_5 + Z^{-1}r_2X\right\rangle.
\end{aligned}
$$

Next we define a new operator $\boldsymbol{\mathcal{O}}$ by

$$\boldsymbol{\mathcal{O}}(X) = \begin{pmatrix} \boldsymbol{\mathcal{A}}(X) \\ \langle -C, X\rangle \\ \langle \bar{C}, X\rangle \end{pmatrix}.$$

The adjoint operator is given by

$$\boldsymbol{\mathcal{O}}^{\boldsymbol{t}} \begin{pmatrix} y \\ \tau \\ \vartheta \end{pmatrix} = \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - C\tau + \bar{C}\vartheta.$$

Using this setting, system (1.20) is equivalent to the following formulation.

$$\boldsymbol{\mathcal{O}}\left(Z^{-1}\boldsymbol{\mathcal{O}}^{\boldsymbol{t}} \begin{pmatrix} \Delta y \\ \Delta\tau \\ \Delta\vartheta \end{pmatrix} X \right) + \bar{M} \begin{pmatrix} \Delta y \\ \Delta\tau \\ \Delta\vartheta \end{pmatrix} = \boldsymbol{\mathcal{O}}\left(Z^{-1}r_5 + Z^{-1}r_2 X\right) - \begin{pmatrix} r_1 \\ r_3 + \frac{r_6}{\tau} \\ r_4 \end{pmatrix}.$$

Here

$$\bar{M} = \begin{pmatrix} 0 & -b & \bar{b} \\ b^T & \frac{\nu}{\tau} & \bar{z} \\ -\bar{b}^T & -\bar{z} & 0 \end{pmatrix}.$$

Obviously, $\boldsymbol{\mathcal{O}}$ is a linear operator. Therefore $\boldsymbol{\mathcal{O}}(Z^{-1}\boldsymbol{\mathcal{O}}^{\boldsymbol{t}}(\cdot)X)$ can be written as $M(\cdot)$ where M is symmetric and can be computed as follows. Let $M = (m_{ij})$, then

$$m_{ij} = \begin{cases} \operatorname{trace}(A_i Z^{-1} A_j X), & 1 \le i, j \le m \\ -\operatorname{trace}(A_i Z^{-1} C X), & 1 \le i \le m,\ j = m+1 \\ \operatorname{trace}(C Z^{-1} C X), & i = m+1,\ j = m+1 \\ \operatorname{trace}(A_i Z^{-1} \bar{C} X), & 1 \le i \le m+1,\ j = m+2 \\ \operatorname{trace}(\bar{C} Z^{-1} \bar{C} X), & i = m+2,\ j = m+2. \end{cases}$$

Thus system (1.20) is equivalent to

$$(M + \bar{M}) \begin{pmatrix} \Delta y \\ \Delta\tau \\ \Delta\vartheta \end{pmatrix} = \boldsymbol{\mathcal{O}}(Z^{-1}r_5 + Z^{-1}r_2 X) - \begin{pmatrix} r_1 \\ r_3 + \frac{r_6}{t} \\ r_4 \end{pmatrix}.$$

Note, we eliminated the matrix variables $\Delta Z, \Delta X$ and $\Delta\nu$ to get a linear system of size $m+2$ for the remaining variables $\Delta y, \Delta\tau$ and $\Delta\vartheta$. Our numerical testing indicated that approximate solutions can be found without incurring a high computational cost.

1.5.6 Refining the Duality States

The next result is based on one hand on strictly feasible iterates generated by our path-following algorithm and on the other hand on the minimal centrality condition of Todd and Ye [121], defined as $\tau^k \nu^k \geq (1-\omega)\vartheta^k = (1-\omega)\mu^k > 0,\ k \in \mathbb{N}$ for some fixed $0 < \omega < 1$, see also [107].

Theorem 1.27 *Let $(X^k, y^k, Z^k, \tau^k, \vartheta^k, \nu^k)_{k\in\mathbb{N}} \in \operatorname{int}(\mathcal{F}(\mathcal{H}))$ denote a sequence of iteration points generated by our path-following algorithm. Suppose that $\lim_{k\to\infty} ||X^k|| < \infty$, $\lim_{k\to\infty} ||Z^k|| < \infty$ and that the original primal-dual pair has an optimal solution (X^*, y^*, Z^*). Assume there exists a $\omega,\ 0 < \omega < 1$ such that $\tau^k\nu^k \geq (1-\omega)\vartheta^k > 0$ for all $k \in \mathbb{N}$. Then τ^k stays bounded away from zero for all $k \in \mathbb{N}$.*

Proof. Obviously, $(X^*, y^*, Z^*, \tau^* = 1, \vartheta^* = 0, \nu^* = 0)$ is a solution of (1.11). Consider the set of constraints obtained from (1.14) for $\vartheta = 0$, in fact

$$\begin{array}{rrrl} & \boldsymbol{\mathcal{A}}(X^k) & -b\tau^k & = 0 \\ -\boldsymbol{\mathcal{A}^t}(y^k) & & +C\tau^k & = Z^k \\ \langle b, y^k\rangle & -\langle C, X^k\rangle & & = \nu^k \end{array} \tag{1.21}$$

$$y^k \in \mathbb{R}^m,\ (\tau^k, \nu^k) \in (\mathbb{R}^+)^2,\ (X^k, Z^k) \in (S(n)^+)^2.$$

The next thing we show is that $(X^k - \vartheta^k X^\circ, y^k - \vartheta^k y^\circ, Z^k - \vartheta^k Z^\circ, \tau^k - \vartheta^k\tau^\circ, \vartheta^k, \nu^k - \vartheta^k\nu^\circ)$ is a solution of (1.21). (Here $(X^\circ, y^\circ, Z^\circ, \tau^\circ, \vartheta^\circ, \nu^\circ) = (X^\circ, y^\circ, Z^\circ, 1, 1, 1)$.) We substitute (1.12) into (1.11) and carry out some elementary calculations.

$$\begin{aligned} 0 &= \boldsymbol{\mathcal{A}}(X^k) - b\tau^k + (b - \boldsymbol{\mathcal{A}}(X^\circ))\vartheta^k \\ &= \boldsymbol{\mathcal{A}}(X^k - \vartheta^k X^\circ) - b(\tau^k - \vartheta^k) \\ &= \boldsymbol{\mathcal{A}}(X^k - \vartheta^k X^\circ) - b(\tau^k - \vartheta^k\tau^\circ) \quad (\text{since } \tau^\circ = 1). \end{aligned}$$

Similarly we get

$$\begin{aligned} 0 &= Z^k + \boldsymbol{\mathcal{A}^t}(y^k) - C\tau^k + \Big(C - \boldsymbol{\mathcal{A}^t}(y^\circ) - Z^\circ\Big)\vartheta^k \\ &= (Z^k - \vartheta^k Z^\circ) + \boldsymbol{\mathcal{A}^t}(y^k - \vartheta^k y^\circ) - C(\tau^k - \vartheta^k) \\ &= (Z^k - \vartheta^k Z^\circ) + \boldsymbol{\mathcal{A}^t}(y^k - \vartheta^k y^\circ) - C(\tau^k - \vartheta^k\tau^\circ) \quad (\text{since } \tau^\circ = 1) \end{aligned}$$

and

$$\begin{aligned} 0 &= \langle b, y^k\rangle - \langle C, X^k\rangle + \Big(\langle C, X^\circ\rangle + 1 - \langle b, y^\circ\rangle\Big)\vartheta^k - \nu^k \\ &= \langle b, y^k - \vartheta^k y^\circ\rangle - \langle C, X^k - \vartheta^k X^\circ\rangle - (\nu^k - \vartheta^k) \\ &= \langle b, y^k - \vartheta^k y^\circ\rangle - \langle C, X^k - \vartheta^k X^\circ\rangle - (\nu^k - \vartheta^k\nu^\circ) \quad (\text{since } \nu^\circ = 1). \end{aligned}$$

Hence the claimed statement is proved. Scalar multiplication of the first three constraints in (1.21) by y^k, X^k and τ^k, respectively yields

$$\begin{aligned} \langle \boldsymbol{\mathcal{A}}(X^k), y^k\rangle - \langle b\tau^k, y^k\rangle &= 0 \\ \langle C\tau^k, X^k\rangle - \langle Z^k, X^k\rangle - \langle \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^k), X^k\rangle &= 0 \\ \langle b\tau^k, y^k\rangle - \langle C\tau^k, X^k\rangle - \nu^k\tau^k &= 0. \end{aligned}$$

We add up these equations and obtain

$$\langle Z^k, X^k\rangle + \nu^k\tau^k = 0. \tag{1.22}$$

We conclude that all feasible points of system (1.21) satisfy (1.22). Consequently, the point $(X^k - \vartheta^k X^\circ + X^*, y^k - \vartheta^k y^\circ + y^*, Z^k - \vartheta^k Z^\circ + Z^*, \tau^k - \vartheta^k\tau^\circ + \tau^*, \vartheta^k + \vartheta^*, \nu^k - \vartheta^k\nu^\circ + \nu^*)$ solves system (1.21). Thus we have

$$\begin{aligned} 0 &= \langle X^k - \vartheta^k X^\circ + X^*, Z^k - \vartheta^k Z^\circ + Z^*\rangle + (\nu^k - \vartheta^k\nu^\circ + \nu^*)(\tau^k - \vartheta^k\tau^\circ + \tau^*) \\ &= \underbrace{\langle X^k - \vartheta^k X^\circ, Z^k - \vartheta^k Z^\circ\rangle + (\nu^k - \vartheta^k\nu^\circ)(\tau^k - \vartheta^k\tau^\circ)}_{=0 \text{ by } (1.22)} + \underbrace{\langle X^*, Z^*\rangle + \nu^*\tau^*}_{=0 \text{ due to optimality}} \\ &\quad + \langle X^k - \vartheta^k X^\circ, Z^*\rangle + \langle X^*, Z^k - \vartheta^k Z^\circ\rangle + (\nu^k - \vartheta^k\nu^\circ)\tau^* + \nu^*(\tau^k - \vartheta^k\tau^\circ) \\ &= \langle X^k, Z^*\rangle + \langle X^*, Z^k\rangle + \nu^k\tau^* + \nu^*\tau^k - \vartheta^k\left(\langle X^\circ, Z^*\rangle + \langle X^*, Z^\circ\rangle + \nu^\circ\tau^* + \nu^*\tau^\circ\right). \end{aligned}$$

Since $X^k \in S(n)^{++},\ X^* \in S(n)^+,\ Z^k \in S(n)^{++},\ Z^* \in S(n)^+$ we have $\langle X^k, Z^*\rangle + \langle X^*, Z^k\rangle \geq 0$. With our setting $\tau^* = 1,\ \nu^* = 0$ and $(X^\circ, Z^\circ) = (I, I)$ the above equation yields the estimate

$$\begin{aligned} \nu^k &\leq \vartheta^k\left(\langle X^\circ, Z^*\rangle + \langle X^*, Z^\circ\rangle + \nu^\circ\right) \\ &= \vartheta^k\left(\langle I, Z^*\rangle + \langle I, X^*\rangle + 1\right). \end{aligned}$$

Combining this estimate for ν^k with $\tau^k\nu^k \geq (1-\omega)\vartheta^k$ the following chain of inequalities holds true.

$$(1-\omega)\vartheta^k \leq \tau^k\nu^k \leq \tau^k\vartheta^k\left(\langle I, Z^*\rangle + \langle I, X^*\rangle + 1\right)$$

Because X^k and Z^k are bounded for all k there exists a $\xi \in \mathbb{R}^+$ such that $\operatorname{trace}(X^* + Z^*) \leq \xi$ or equivalently $\langle I, X^* + Z^*\rangle \leq \xi$. So we have

$$(1-\omega)\vartheta^k \leq \tau^k\vartheta^k(\xi + 1).$$

We rewrite this inequality and get the desired result $\tau^k \geq \dfrac{1-\omega}{\xi+1} > 0$.

□

The next result shows that the iterates generated by our algorithm stay within a compact set. The search direction $\Delta p = (\Delta X, \Delta y, \Delta Z, \Delta\tau, \Delta\vartheta, \Delta\nu)$ forms a descent direction with respect to the merit function f_μ, which is defined as

$$f_\mu(p) := \langle Z, X\rangle + \nu\tau - \mu \log\det(ZX) - \mu\log(\nu\tau) + \frac{1}{2}\sum_{i=1}^{4} ||r_i||^2.$$

Note, the global minimum $(n+1)\mu(1-\log\mu)$ of f_μ is attained at p if and only if $F_\mu(p) = 0$. Details will be given in section 2.2 in which we consider f_μ in the context with the semidefinite programming problem in standard form.

Lemma 1.28 *Let $p^\circ = (X^\circ, y^\circ, Z^\circ, \tau^\circ, \vartheta^\circ, \nu^\circ) \in \operatorname{int}(\mathcal{F}(\mathcal{H}))$ and let $\zeta > 0,\ \zeta \in I\!R$ such that $f_\mu(p^\circ) \le \zeta$. Then: The level set*

$$\Gamma = \left\{p = (X, y, Z, \tau, \vartheta, \nu) \in \operatorname{int}(\mathcal{F}(\mathcal{H})) : f_\mu(p) \le \zeta\right\}$$

is contained in a bounded and closed subset of $S(n) \times I\!R^m \times S(n) \times I\!R \times I\!R \times I\!R$.

Since the proof is about the same than that of Lemma 2.3 we do not give it here. We only note that, if X, Z, τ and ν are bounded, then ϑ is bounded by (1.13).

Corollary 1.29 *Let $p^\circ = (X^\circ, y^\circ, Z^\circ, \tau^\circ, \vartheta^\circ, \nu^\circ) \in \operatorname{int}(\mathcal{F}(\mathcal{H}))$ and let $\zeta > 0,\ \zeta \in I\!R$ such that $f_\mu(p^\circ) \le \zeta$. Then: The iterates generated by our path-following algorithm (performing inaccurate line search with respect to the merit function), stay within a compact set.*

The following lemma shows the existence of a certain expression. We will use this result in Theorem 1.32.

Lemma 1.30 *Let $(X^k, y^k, Z^k, \tau^k, \vartheta^k, \nu^k)_{k\in I\!N} \in \operatorname{int}(\mathcal{F}(\mathcal{H}))$ be a sequence generated by our path-following algorithm. Suppose this sequence converges to a maximally complementary solution $(X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ to* **(EHSDP)** *and suppose there exists a $\omega,\ 0 < \omega < 1$ such that $\tau^k\nu^k \ge (1-\omega)\vartheta^k > 0$ for all $k \in I\!N$. If $(\tau^k)_{k\in I\!N}$ and $(\nu^k)_{k\in I\!N}$ converges to zero then the fraction*

$$\left(\frac{\nu^k}{\tau^k - \vartheta^k}\right)$$

is well defined for all $k \in I\!N$.

Proof. By Theorem 1.24 we know that $(\vartheta^k)_{k\in I\!N}$ converges to zero. By assumption there exists a $\omega > 0$ such that $\tau^k\nu^k \ge (1-\omega)\vartheta^k > 0$ for all $k \in I\!N$. Hence for $(\tau^k)_{k\in I\!N} \to 0$ and $(\nu^k)_{k\in I\!N} \to 0$ there exists a $\bar{k}$ such that $(\tau^k + \nu^k) < 1 - \omega$ for all $k > \bar{k}$. From that it follows that $\frac{1-\omega}{\nu^k}\vartheta^k > \vartheta^k$ for all $k > \bar{k}$ and therefore $\tau^k - \vartheta^k > 0$ for all $k > \bar{k}$. That is the proof.

□

Definition 1.31 *Let* $(X^k)_{k\in\mathbb{N}}$ *be an asymptotically feasible solutions for* **(PSDP)** *and let* $(y^k)_{k\in\mathbb{N}}$ *be an asymptotically feasible solutions for* **(DSDP)**, *respectively. Let* $\varepsilon > 0$ *and suppose there exists a* $\bar{k}$ *such that*

$$\|\boldsymbol{\mathcal{A}}(X^k) - b\| \le \varepsilon,\ \|Z^k + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^k) - C\| < \varepsilon \text{ and } \langle C, X^k\rangle - \langle b, y^k\rangle < \varepsilon \quad \text{for all} \quad k \ge \bar{k}.$$

Then any (X^k, y^k, Z^k), $k \ge \bar{k}$ *is called an approximate solution of* **(PSDP)** *and* **(DSDP)**.

The next result is somewhat stronger than that given in Theorem 1.27. To facilitate the notation we will write $(\xi^k) \to \xi^*$ for a sequence $(\xi^k)_{k\in\mathbb{N}}$ that converges to ξ^* instead of $(\xi^k) \to \xi^*$ as $k \to \infty$.

Theorem 1.32 *Let* $(X^k, y^k, Z^k, \tau^k, \vartheta^k, \nu^k)_{k\in\mathbb{N}} \in \operatorname{int}(\mathcal{F}(\mathcal{H}))$ *denote a sequence of iteration points generated by our path-following algorithm. If there exists a* $\omega > 0$ *such that* $\tau^k\nu^k \ge (1-\omega)\vartheta^k > 0$ *for all* $k \in \mathbb{N}$, *then:*

(1) $(\tau^k)_{k\in\mathbb{N}}$ *stays bounded away from zero if and only if the original primal-dual pair* **(PSDP)** *and* **(DSDP)** *has an optimal solution.*

(2) $(\nu^k)_{k\in\mathbb{N}}$ *stays bounded away from zero if and only if at least one of* **(PSDP)** *and* **(DSDP)** *is* SIF. *Moreover, if*

$$(\mathbf{PSDP}) \textit{ is } \mathrm{SIF} \textit{ and} \begin{cases} (\mathbf{DSDP}) \textit{ is } \mathrm{F} & \Rightarrow \quad (\mathbf{DSDP}) \textit{ is } \mathrm{UBD} \\ (\mathbf{DSDP}) \textit{ is } \mathrm{AF} & \Rightarrow \quad (\mathbf{DSDP}) \textit{ is } \mathrm{AUBD}. \end{cases}$$

and similarly for interchanged **(PSDP)** *and* **(DSDP)**.

(3) *If both,* $(\tau^k)_{k\in\mathbb{N}}$ *and* $(\nu^k)_{k\in\mathbb{N}}$ *converges to zero three cases occur.*

(i) *If* $\lim\limits_{k\to\infty}\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) = \infty$ *then one of* **(PSDP)** *and* **(DSDP)** *is* IAF *and the other is* F *and* UBD.

(ii) *If* $\lim\limits_{k\to\infty}\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) = 0$ *then approximate solutions to* **(PSDP)** *and* **(DSDP)** *are detected. We cannot classify the duality status of* **(PSDP)** *and* **(DSDP)**.

(iii) *If* $\lim\limits_{k\to\infty}\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) = \zeta, \quad 0 < \zeta < \infty$ *then asymptotically feasible solutions are detected. We cannot classify the duality status of* **(PSDP)** *and* **(DSDP)**.

Proof. Let $(X^*, y^*, Z^*, \tau^*, \vartheta^*, \nu^*)$ be the maximally complementary solution to **(EHSDP)**, generated by our algorithm, sketched in Table 1.2. (Recall, the central path converges to a maximally complementary solution, see page 18.) Note, any complementary solution is an optimal solution.

(1) The only if part was shown in Theorem 1.27.

Conversely, assume there exists no optimal solution to the original pair of problems **(PSDP)** and **(DSDP)**. Note that the iterates stay within a compact set. In particular, the sequence $(\tau^k)_{k\in\mathbb{N}} \ge 0$ stays bounded. It follows from the classical result that every bounded sequence contains a convergent subsequence (or equivalent every bounded sequence contains at least one cluster value) that $(\tau^{k_j}) \subseteq (\tau^k)_{k\in\mathbb{N}}$, where $j \mapsto k_j$ is a strictly increasing infinite sequence of integers, converges to a finite value $\tau^* > 0$ or (τ^{k_j}) converges to zero. This limit point is also a cluster value of $(\tau^k)_{k\in\mathbb{N}}$. Suppose $\lim_{k\to\infty}(\tau^k) = \tau^* > 0$. Then, by Theorem 1.25 (i) an optimal solution for **(PSDP)** and **(DSDP)** is given by $(X^*/\tau^*, y^*/\tau^*, Z^*/\tau^*)$. This is a contradiction to the assumption that no optimal solution exists. Otherwise, if $\lim_{k\to\infty}(\tau^k) = 0$ this is a contradiction to $(\tau^k)_{k\in\mathbb{N}}$ stays bounded away from zero. Hence the original pair of problems has an optimal solution.

(2) Since $(\nu^k)_{k\in\mathbb{N}}$ stays bounded away from zero there exists a $\bar{\nu} > 0$ such that $\nu^k > \bar{\nu}$ for all k, or equivalently $\nu^* > 0$. Thus τ^* must be zero due to maximal complementarity. From (1.17) we obtain

$$\begin{aligned}
&\boldsymbol{\mathcal{A}}(X^k) \to 0\\
&Z^k + \boldsymbol{\mathcal{A}^t}(y^k) \to 0\\
&\langle C, X^k\rangle - \langle b, y^k\rangle \to -\nu^k.
\end{aligned}$$

The way of argument given in Theorem 1.25 (ii) gives that at least one of **(PSDP)** and **(DSDP)** is SIF.

Conversely, suppose **(PSDP)** is SIF and $\lim_{k\to\infty}(\nu^k) = \nu^* = 0$. Then $\lim_{k\to\infty}(\tau^k) = \tau^* > 0$ cannot occur, because then the pair **(PSDP)** and **(DSDP)** would have an optimal solution by (1). This is a contradiction to the assumption **(PSDP)** is SIF. Now consider the case $\lim_{k\to\infty}(\tau^k) = \tau^* = 0$ and $\lim_{k\to\infty}(\nu^k) = \nu^* = 0$. As we will see in (3) this case occurs if **(PSDP)** is AF, which is again a contradiction to the assumption **(PSDP)** is SIF.

Applying Theorem 1.12 completes the proof of (2).

(3) (i) Note the minimal centrality condition implies $\left(\frac{\vartheta^k}{\nu^k}\right) \to 0$ $\left(\left(\frac{\vartheta^k}{\tau^k}\right) \to 0\right)$ due to $(\tau^k) \to 0$ $((\nu^k) \to 0)$, respectively. From that it follows that $\left(\frac{\tau^k}{\nu^k}\right) \to 0$ or equivalently $\left(\frac{\nu^k}{\tau^k}\right) \to \infty$. From (1.17) we obtain

$$\begin{aligned}
&\boldsymbol{\mathcal{A}}\left(\frac{X^k}{\tau^k-\vartheta^k}\right) - b\left(\frac{\tau^k}{\tau^k-\vartheta^k}\right) + \bar{b}\left(\frac{\vartheta^k}{\tau^k-\vartheta^k}\right) \to 0\\
&\left(\frac{Z^k}{\tau^k-\vartheta^k}\right) + \boldsymbol{\mathcal{A}^t}\left(\frac{y^k}{\tau^k-\vartheta^k}\right) - C\left(\frac{\tau^k}{\tau^k-\vartheta^k}\right) - \bar{C}\left(\frac{\vartheta^k}{\tau^k-\vartheta^k}\right) \to 0\\
&\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) - \left\langle b, \frac{y^k}{\tau^k-\vartheta^k}\right\rangle + \left\langle C, \frac{X^k}{\tau^k-\vartheta^k}\right\rangle - \bar{z}\left(\frac{\vartheta^k}{\tau^k-\vartheta^k}\right) \to 0.
\end{aligned}$$

Since $\left(\frac{\tau^k}{\tau^k-\vartheta^k}\right) \to 1$, $\left(\frac{\vartheta^k}{\tau^k-\vartheta^k}\right) \to 0$ and $\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) \to \infty$ we have

$$\boldsymbol{\mathcal{A}}\left(\frac{X^k}{\tau^k-\vartheta^k}\right) - b \to 0$$

$$\left(\frac{Z^k}{\tau^k-\vartheta^k}\right) + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}\left(\frac{y^k}{\tau^k-\vartheta^k}\right) - C \to 0$$

$$\left\langle C, \frac{X^k}{\tau^k-\vartheta^k}\right\rangle - \left\langle b, \frac{y^k}{\tau^k-\vartheta^k}\right\rangle \to -\infty.$$

Consequently $\left(\frac{X^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ is an asymptotically feasible solution for **(PSDP)** and $\left(\frac{y^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ is an asymptotically feasible solution for **(DSDP)** and at least one of $\left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle$ and $-\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle$ must be strictly less than zero. We first consider the case $\left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle \to \xi < 0$ and $-\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle \to \eta > 0$. Clearly, **(PSDP)** is PAF and if **(DSDP)** is PAF then it is ABD.

Suppose **(DSDP)** is PAF and ABD. Then **(PSDP)** is F and BD by Lemma 1.8. Hence strong duality holds by Theorem 1.12 which is a contradiction. Therefore the assumption **(DSDP)** is PAF and ABD cannot hold, but from considerations above we know that **(DSDP)** is AF. Hence **(DSDP)** must be IAF.

Analogously the case $\left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle \to \xi > 0$ and $-\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle \to \eta < 0$ can be investigated.

Now consider the case $\left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle \to \xi < 0$ and $-\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle \to \eta < 0$. We see that both, **(PSDP)** and **(DSDP)** are PAF. From Theorem 1.12 it follows that none of the primal and the dual can be PAF and ABD. This would imply strong duality which is a contradiction. Moreover, by Lemma 1.7 none of **(PSDP)** and **(DSDP)** can be F. In particular, if one of the primal-dual pair is F, then the other problem is ABD. However from above we know that this is not possible.

Recapitulating, we have that

$$\begin{aligned}
\left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle < 0,\ -\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle > 0 \ &\Rightarrow\ (\textbf{DSDP}) \text{ is IAF} \\
&\Rightarrow\ (\textbf{PSDP}) \text{ is F and UBD (see Lemma 1.10)}\,. \\
-\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle < 0,\ \left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle > 0 \ &\Rightarrow\ (\textbf{PSDP}) \text{ is IAF} \\
&\Rightarrow\ (\textbf{DSDP}) \text{ is F and UBD (see Lemma 1.10)}\,. \\
\left\langle C, \frac{X^k}{\nu^k-\vartheta^k}\right\rangle < 0,\ -\left\langle b, \frac{y^k}{\nu^k-\vartheta^k}\right\rangle < 0 \ &\Rightarrow\ (\textbf{PSDP}) \text{ and } (\textbf{DSDP}) \text{ are IF, PAF and} \\
&\quad\ \text{AUBD.}
\end{aligned}$$

(ii) Since $\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) \to 0$, $\left(\frac{\tau^k}{\tau^k-\vartheta^k}\right) \to 1$ and $\left(\frac{\vartheta^k}{\tau^k-\vartheta^k}\right) \to 0$ we get analogous as under item (i)

$$\mathcal{A}\left(\frac{X^k}{\tau^k-\vartheta^k}\right) - b \to 0$$

$$\left(\frac{Z^k}{\tau^k-\vartheta^k}\right) + \mathcal{A}^t\left(\frac{y^k}{\tau^k-\vartheta^k}\right) - C \to 0$$

$$\left\langle C, \frac{X^k}{\tau^k-\vartheta^k}\right\rangle - \left\langle b, \frac{y^k}{\tau^k-\vartheta^k}\right\rangle \to 0.$$

So $\left(\frac{X^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ is an asymptotically feasible solution for **(PSDP)** and $\left(\frac{y^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ is an asymptotically feasible solution for **(DSDP)**.

If these sequences converge, then the triple $\left(\frac{X^k}{\tau^k-\vartheta^k}, \frac{y^k}{\tau^k-\vartheta^k}, \frac{Z^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ converges to an optimal solution of **(PSDP)** and **(DSDP)**. Then we see from (1) that $(\tau^k)_{k\in\mathbb{N}}$ stays bounded away from zero which is a contradiction to $(\tau^k) \to 0$ and also to maximal complementarity. The only conclusion we can draw here is that the asymptotically feasible solutions for **(PSDP)** and **(DSDP)** gives approximate solutions to the problem. We cannot classify the original pair **(PSDP)** and **(DSDP)** concerning its duality status.

We mention, that Sturm [118] shows that with the additional assumption **(DSDP)** is F and $\mathrm{int}(\mathcal{F}(\mathcal{D})) \neq \emptyset$ it follows: **(PSDP)** F, $\mathrm{int}(\mathcal{F}(\mathcal{P})) = \emptyset$ and the dual supremum is not attained if and only if $\lim_{k\to\infty}(\tau^k) = \tau^* = 0$ and $\lim_{k\to\infty}\left(\frac{\nu^k}{\tau^k-\vartheta^k}\right) = 0$. Hence with this stronger assumption this class of problems can be categorized concerning its duality status.

(iii) Similarly as under (2) we get

$$\mathcal{A}\left(\frac{X^k}{\tau^k-\vartheta^k}\right) - b \to 0$$

$$\left(\frac{Z^k}{\tau^k-\vartheta^k}\right) + \mathcal{A}^t\left(\frac{y^k}{\tau^k-\vartheta^k}\right) - C \to 0$$

$$\left\langle C, \frac{X^k}{\tau^k-\vartheta^k}\right\rangle - \left\langle b, \frac{y^k}{\tau^k-\vartheta^k}\right\rangle \to \zeta, \quad 0 < \zeta < \infty.$$

Again as above, $\left(\frac{X^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ is an asymptotically feasible solution for **(PSDP)** and $\left(\frac{y^k}{\tau^k-\vartheta^k}\right)_{k\in\mathbb{N}}$ is an asymptotically feasible solution for **(DSDP)**. Since ζ is finite we conclude that both problems **(PSDP)** and **(DSDP)** are PAF. (**(PSDP)** and/or **(DSDP)** IAF is not possible due to the detected asymptotically feasible solution to the respective problem). However, we cannot identify clearly the duality status for the class of such problems.

$\square$

Remark 1.33 *Our computational tests show that case (iii) (3) in Theorem 1.32 holds for*

(a) **(PSDP)** *is* F *and* BD*,* **(DSDP)** *is* IF, PAF *and* ABD.

(b) **(PSDP)** *is* IF*,* PAF *and* ABD, **(DSDP)** *is* F *and* BD.

(c) **(PSDP)** *is* F *and* BD*,* **(DSDP)** *is* F *and* BD. *(there is a finite duality gap).*

(d) **(PSDP)** *and* **(DSDP)** *are* IF, PAF *and* AUBD. *(there is an infinite duality gap).*

Remark 1.34 *For the class of problems where both problems* **(PSDP)** *and* **(DSDP)** *are* F *and* BD *may happen that strong duality holds, the primal or dual optimum is finite but not attained or there exists a finite duality gap. Hence fourteen different examples concerning the duality status may be formulated in SDP, however, unfortenately only eight of them can be classified using Theorem 1.32.*

Chapter 2

Convergence Analysis

As we have already mentioned generalizations of IPMs for LP to IPMs for SDP have been proposed by many authors. Not surprisingly, many convergence results can be found in their works and likewise in many other publications. We will mention several recent results. This thesis studies primal-dual path-following algorithms for SDP. For this reason we firstly propose a global convergence proof based on the concept of closed mappings and secondly include one of the convergence proofs establishing the local polynomial convergence for this category of IPMs. Moreover we derive a condition which enables us to decide if a given problem does have an optimal solution in a given domain.

2.1 On Recent Convergence Results in Semidefinite Programming

In this section we do not give a comprehensive survey in convergence analysis, but we will review several recent results in this area.

Convergence of primal-dual path-following methods was analyzed first by Kojima, Shindoh and Hara [73] and Nesterov and Todd [99, 98]. The paper [73] introduces the Kojima, Shindoh and Hara family (denoted KSH-family) of search directions. Besides a feasible path-following IPM, (a generalization of the algorithm given in [68, 89] for LP), Kojima et al. presented also a feasible (a generalization of the algorithm given in [69] for LP) and an infeasible (a generalization of an algorithm given in [86] for LP) potential reduction interior-point algorithm, see [73]. We mention that in an independent paper from [73], Helmberg, Rendl, Vanderbei, and Wolkowicz [54] analyse a very similar path-following interior-point algorithm. In particular, the authors in [54] could show global convergence, while Kojima et al. [73] established the polynomial convergence for the IPMs proposed in their paper. Note, in the particular case of the feasible path-following IPM they established polynomiality based on the HKM search direction, which is contained in the KSH-family of search directions. In [93], Monteiro and Tsuchiya show that this result holds true for any direction of the KSH-family.

Local convergence analysis of interior-point algorithms in SDP was given first by Kojima, Shida and Shindoh [70] and independently by Potra and Sheng [108]. In both papers the authors show superlinear convergence of an algorithm which is an extension of the Mizuno-Todd-Ye predictor corrector algorithm given in [87] for LP based on the HKM-direction. Kojima et al. established the superlinearity under the assumption of three conditions (see [71]). In contrast to this conditions Potra and Sheng suggested a sufficient condition for superlinearity which is satisfied under the three conditions given in [71].

Another article dealing with local convergence of IPMs in SDP was published by Kojima, Shida and Shindoh [71]. In fact, the unpublished paper [70] was the original version of [71]. The authors of [70] observed that they analyse an algorithm like that given in [108]. Moreover, Kojima et al. [70] could not show such profound global convergence results as given by Potra and Sheng [108]. For this reason they published the paper [71] dealing only with local convergence.

The Monteiro Zhang family of search directions has been analyzed in [94]. In this paper Monteiro and Zhang present a unified analysis for a class of long step primal-dual path-following algorithms for SDP whose search directions are based on the scaling and symmetrization scheme introduced by Zhang [135], (see section 1.4.1). The scaling matrix P is chosen in such a way that $PZXP^{-1}$ is symmetric. Considering this restriction the class of search directions obtained this way includes the HKM-direction and the NT-direction but not the AHO-direction.

The AHO-direction was recently analyzed in [72, 88]. In [72], Kojima et al. proposes a primal-dual predictor corrector algorithm for which they show global but not polynomial convergence. They show also quadratic local convergence under the strict complementarity condition. Monteiro [88] establishes the polynomial convergence of the class of primal-dual feasible interior-point algorithms for SDP based on the Monteiro Zhang family of search directions. In contrast to [94] here no condition is given on the scaling matrix P. In particular, since the AHO-direction is contained in the Monteiro Zhang family, Monteiro establishes also the polynomial convergence of algorithms based on this search direction. In this paper, Monteiro shows that polynomial time computational complexities of the short step path-following algorithm of Kojima, Mizuno and Yoshise [68] and Monteiro and Adler [89, 90] and the predictor corrector algorithm of Mizuno, Todd and Ye [87] extends from LP to SDP. Tseng [123] also established polynomial iteration complexity for a predictor corrector algorithm for SDP using the AHO-direction.

In a more recent paper, Monteiro and Tsuchiya [92] establishes the polynomial convergence of a new class of primal-dual path-following feasible algorithm for SDP. They introduce a family of search directions (denoted MT family) which are based on the symmetric central path equation

$$(PZP^T)^{1/2}(P^{-T}XP^{-1})(PZP^T)^{1/2} - \mu I = 0,$$

where P denotes a nonsingular matrix.

In the next section we establish the global convergence of the class of primal-dual IPMs for SDP based on the HKM-direction.

2.2 Global Convergence of a Primal-Dual Algorithm in SDP

In this section the concept of closed maps is used to show global convergence of our algorithm for the semidefinite programming problem in standard form. A unified analysis of the convergence of nonlinear programming algorithms was given first by Zangwill [134]. We use these techniques to show global convergence.

Under the assumption that the set $\{A_i\}, (i = 1, \ldots, m)$ is linearly independent and that both the primal and the dual problem satisfy the Slater constraint qualification it is known that the central path exists for the primal-dual pair **(PSDP)** and **(DSDP)** and that the sequence (X_μ, y_μ, Z_μ) converges to an optimal solution as μ is driven to zero, see section 1.4. Recall, for any fixed $\mu > 0$, (X_μ, y_μ, Z_μ) is characterized as unique solution of the **(KKT)** conditions.

We will write p to denote a triple (X, y, Z) and Δp to denote the search direction $(\Delta X, \Delta y, \Delta Z)$.

In [129], Vial suggests that a merit function can be used to measure the progress towards an optimal solution. This type of function is used in [54, 51] to prove convergence of an interior-point algorithm in SDP. Actually, this proof follows a proof given in Anstreicher and Vial [6] for a convex programming problem. We establish the global convergence in a completely different way but we also take advantage of the features of a merit function. Moreover, we use also techniques given in [73]. We use the merit function defined as

$$f_\mu(p) := \langle Z, X\rangle - \mu \log\det(ZX) + \frac{1}{2}||r_P||^2 + \frac{1}{2}||r_D||^2.$$

The merit function enjoys the following properties. For feasible interior points f_μ is strictly convex. f_μ is continuously differentiable if $(Z, X) \in (S(n)^{++})^2$. Moreover, f_μ is bounded below and tends to infinity on the boundary of the set of feasible points.

Let $p = (X, y, Z) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ and let $\lambda_j(ZX), (j = 1, \ldots, n)$ be the eigenvalues of the matrix ZX. We now consider the terms depending on Z and X.

$$\langle Z, X\rangle - \mu \log\det(ZX) = \operatorname{trace}(ZX) - \mu \log\left(\prod_{j=1}^{n} \lambda_j(ZX)\right)$$

$$= \sum_{j=1}^{n} \Big(\lambda_j(ZX) - \mu \log \lambda_j(ZX) \Big)$$

It follows from Lemma A.5 that $\lambda_j(ZX) > 0$ for every $j = 1, \ldots, n$. Therefore it is easy to show that each term $\lambda_j(ZX) - \mu \log \lambda_j(ZX)$ attains the minimum if and only if $\lambda_j(ZX) = \mu, (j = 1, \ldots, n)$. Hence f_μ is bounded below. In particular

$$f_\mu(p) \geq n\mu\,(1 - \log \mu)$$

and

$$F_\mu(p) = 0 \;\text{ if and only if }\; f_\mu(p) = n\mu\,(1 - \log \mu).$$

The if part follows because $p \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$, from considerations in the last preceding paragraph and the equivalence of $ZX = \mu I$ and $\lambda_j(ZX) = \mu, (j = 1, \ldots, n)$. The only if part follows from the fact that, provided that $\operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ is non-empty, there is a unique global minimizer of f_μ over $\operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$. This follows from Corollary 2.2, Lemma 2.3 and the continuity of f_μ.

The following lemma uses that the matrix function $\log \det X$ is infinite times differentiable on $\Big\{ Y : Y \in M(n), \det Y > 0 \Big\}$ and the fact that the $r-$th differential, see [82], is given by

$$d^r \log \det X = (-1)^{r-1}(r-1)!\operatorname{trace}(X^{-1}\Delta X)^r \quad (r = 1, 2, \ldots).$$

A similar formulation of the following lemma can be found in [73].

Lemma 2.1 *Suppose that* $\mu > 0$ *and* $(Z, X) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$. *Then*

$$\begin{aligned} f_\mu(Z + \Delta Z, X + \Delta X) - f_\mu(Z, X) &= \Phi_1(\Delta Z, \Delta X) + \Phi_2(\Delta Z, \Delta X) \\ &\quad + o(\|\Delta Z\|^2) + o(\|\Delta X\|^2), \end{aligned}$$

where

$$\Phi_1(\Delta Z, \Delta X) = \langle X - \mu Z^{-1}, \Delta Z \rangle + \langle Z - \mu X^{-1}, \Delta X \rangle$$

and

$$\Phi_2(\Delta Z, \Delta X) = \frac{\mu}{2}\|Z^{-1}\Delta Z\|^2 + \frac{\mu}{2}\|X^{-1}\Delta X\|^2$$

for all $(\Delta Z, \Delta X) \in S(n)^2$ *where* $\Delta Z = Z - Z'$ *and* $\Delta X = X - X'$ *for some* $(Z, X) \in \mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})$ *and* $(Z', X') \in \mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})$, *respectively.*

Proof. see [73], but here we give a simplified version. Let $(\Delta Z, \Delta X) \in S(n)^2$, $\xi_1, \ldots, \xi_n$ be the eigenvalues of the symmetric matrix $Z^{-1}\Delta Z$, and $\eta_1, \ldots, \eta_n$ be the eigenvalues of the symmetric matrix $X^{-1}\Delta X$. It suffices to prove the statement under the assumption that $(\Delta Z, \Delta X) \in S(n)^2$ is so small that the absolute values of the eigenvalues $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n$ are less than one. Since $Z \in S(n)^{++}$ there exists a nonsingular matrix, say $\sqrt{Z}$, such that $Z + \Delta Z = \sqrt{Z}(I + \sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1})\sqrt{Z}$.

The assumption that the absolute values of all eigenvalues of $\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}$ are less than one ensures that $I + \sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1} \in S(n)^{++}$. This follows from the fact that $Z^{-1}\Delta Z$ is similar to $\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}$. So they share the same eigenvalues $\xi_1, \ldots, \xi_n$. Hence by Lemma A.6 we see that the matrix $Z + \Delta Z \in S(n)^{++}$ and similarly $X + \Delta X \in S(n)^{++}$. We then have

$$\begin{aligned}
&f_\mu(Z + \Delta Z, X + \Delta X)\\
&\quad= \langle Z, X\rangle + \langle \Delta Z, X\rangle + \langle Z, \Delta X\rangle + \langle \Delta Z, \Delta X\rangle - \mu \log\det(Z + \Delta Z)\\
&\qquad - \mu \log\det(X + \Delta X)\\
&\quad= \langle Z, X\rangle + \langle \Delta Z, X\rangle + \langle Z, \Delta X\rangle + \langle \Delta Z, \Delta X\rangle\\
&\qquad - \mu \left(\log\det Z + \langle Z^{-1}, \Delta Z\rangle - \frac{1}{2}\langle Z^{-1}\Delta Z Z^{-1}, \Delta Z\rangle + o(\|\Delta Z\|^2)\right)\\
&\qquad - \mu \left(\log\det X + \langle X^{-1}, \Delta X\rangle - \frac{1}{2}\langle X^{-1}\Delta X X^{-1}, \Delta X\rangle + o(\|\Delta X\|^2)\right)\\
&\quad= f_\mu(Z, X) + \langle X - \mu Z^{-1}, \Delta Z\rangle + \langle Z - \mu X^{-1}, \Delta X\rangle\\
&\qquad + \frac{\mu}{2}\|Z^{-1}\Delta Z\|^2 + \frac{\mu}{2}\|X^{-1}\Delta X\|^2 + o(\|\Delta Z\|^2) + o(\|\Delta X\|^2).
\end{aligned}$$

Here the last equation matches because for feasible points $p' = (X', y', Z')$ and $p = (X, y, Z)$ the so called orthogonality property holds true. That is,

$$\langle Z - Z', X - X'\rangle \overset{\text{by (1.4)}}{=} \langle \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y - y'), X - X'\rangle \overset{\text{by (1.2)}}{=} \langle y - y', \boldsymbol{\mathcal{A}}(X - X')\rangle \overset{\text{by (1.1)}}{=} 0. \tag{2.1}$$

That is the proof.

□

Corollary 2.2 *Suppose that $\mu > 0$. Then f_μ is strictly convex on* $\operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$.

Proof. Note that the quadratic form $\Phi_2(\Delta Z, \Delta X)$ is positive for any nonzero $(\Delta Z, \Delta X) \in S(n)^2$. Moreover $\Phi_2(\Delta Z, \Delta X)$ dominates $o(\|\Delta Z\|^2) + o(\|\Delta X\|^2)$. Hence

$$f_\mu(Z + \Delta Z, X + \Delta X) - f_\mu(Z, X) > \Phi_1(\Delta Z, \Delta X).$$

Note that $\Phi_1(\Delta Z, \Delta X)$ is the directional derivative in the direction $(\Delta Z, \Delta X)$. In other words, given any nonzero direction $(\Delta Z, \Delta X) \in S(n)^2$ the directional derivative gives a strictly lower bound on the change of the value of f_μ in that direction, i.e. the tangent plane of f_μ in $(Z, X) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ lies strictly below of f_μ. This ensures that f_μ is strictly convex on $\operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$.

□

Kojima et al. [73] used the following result to prove the existence of the central trajectory for the monotone **(SDLCP)**, defined in section 2.3.

Lemma 2.3 *Let* $p^\circ = (X^\circ, y^\circ, Z^\circ) \in \mathrm{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ *and let* $\zeta > 0,\ \zeta \in I\!R$ *such that* $f_\mu(p^\circ) \leq \zeta$. *Then: The level set*

$$\Gamma = \Big\{ p = (X, y, Z) \in \mathrm{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})) : f_\mu(p) \leq \zeta \Big\}$$

is a bounded and closed subset of $S(n) \times I\!R^m \times S(n)$.

Proof. see [73], here we adapt it to the semidefinite programming problem in standard form and give it in more detail. The minimum of $f(t) = t - \log t,\ t > 0$ is attained at $t = 1$ and therefore $\log t \leq t - 1,\ t > 0$. Then for arbitrary scalars $\alpha > 0$ and $\beta > 0$

$$\log \frac{\beta}{\alpha} \leq \frac{\beta}{\alpha} - 1 \Rightarrow \log \beta = \log(\alpha + \beta - \alpha) \leq \log \alpha + \frac{\beta - \alpha}{\alpha}. \tag{2.2}$$

We will use this estimate in the sequel to derive an upper bound on $\langle Z, X \rangle$. Let $\lambda_j(ZX), (j = 1, \ldots, n)$ denote the eigenvalues of the matrix ZX. By lemma A.5, $\lambda_j(ZX) > 0$ for every $j = 1, \ldots, n$. Assume that $p \in \Gamma$. Then:

$$\begin{aligned}
\zeta &\geq \langle Z, X \rangle - \mu \log \det(ZX) \\
&= \sum_{j=1}^{n} \Big(\lambda_j(ZX) - \mu \log \lambda_j(ZX) \Big) \\
&\geq (n-1)\mu(1 - \log \mu) + \Big(\lambda_j(ZX) - \mu \log \lambda_j(ZX) \Big) \qquad \text{(for some fixed } j) \\
&= (n-1)\mu(1 - \log \mu) + \Big(\lambda_j(ZX) - \mu \log(2\mu + \lambda_j(ZX) - 2\mu) \Big) \\
&\geq (n-1)\mu(1 - \log \mu) + \lambda_j(ZX) - \mu \left(\log 2\mu + \frac{\lambda_j(ZX) - 2\mu}{2\mu} \right) \qquad \text{(by (2.2))} \\
&= (n-1)\mu(1 - \log \mu) - \mu \log 2\mu + \mu + \frac{\lambda_j(ZX)}{2} \\
&= n\mu(1 - \log \mu) - \mu \log 2 + \frac{\lambda_j(ZX)}{2}.
\end{aligned}$$

Hence for every $j = 1, \ldots, n$

$$\lambda_j(ZX) \leq 2\Big(\zeta - n\mu(1 - \log \mu) + \mu \log 2 \Big).$$

We will denote this bound by γ. Since $\mathrm{trace}(ZX) = \sum_{i=1}^{n} \lambda_i(ZX)$ this yields

$$\langle Z, X \rangle \leq n\gamma. \tag{2.3}$$

It follows from (2.1) that

$$\langle Z - Z^\circ, X - X^\circ \rangle = 0.$$

This gives

$$\langle Z, X \rangle + \langle Z^\circ, X^\circ \rangle = \langle Z, X^\circ \rangle + \langle Z^\circ, X \rangle.$$

We use (2.3) to get the estimate

$$n\gamma + \langle Z^\circ, X^\circ \rangle \geq \langle Z, X^\circ \rangle + \langle Z^\circ, X \rangle.$$

We will denote the left hand side terms by γ'. Since $p^\circ \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ and $p \in \Gamma$ we have $\langle Z, X^\circ \rangle \geq 0$ and $\langle Z^\circ, X \rangle \geq 0$. Applying Lemma A.2 and Lemma A.3 shows that X and Z are bounded. In particular,

$$\|X\| \leq \sum_{j=1}^{n} \lambda_j(X) \leq \frac{\gamma'}{\lambda_{\min}(Z^\circ)}$$

and similarly for Z. Here $\lambda_{\min}(Z^\circ)$ denotes the minimum eigenvalue of Z° and $\|X\|$ denotes the Frobenius norm. Consequently, Γ is contained in the bounded set, defined by

$$\Big\{ Z \in S(n)^+ : \langle Z, X^\circ \rangle \geq 0 \Big\} \times \Big\{ X \in S(n)^+ : \langle Z^\circ, X \rangle \geq 0 \Big\}.$$

Since (Z, X) is primal-dual feasible it follows from (1.4) that $\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y)$ is bounded. We will show now that y is bounded. Let A be the matrix whose $j-$th column $(j = 1, \ldots, m)$ is obtained by stacking the columns of A_j one underneath the other. Hence, $A \in \mathbb{R}^{n^2 \times m}$ and Ay is bounded. Recall, the set $\{A_i\}, (i = 1, \ldots, m)$ is linearly independent. Thus A has linearly independent columns and therefore it can be factored into $A = QR$. The columns of Q are orthonormal, and R is upper triangular and invertible. Hence,

$$Ay = QRy \Leftrightarrow R^{-1}Q^T Ay = y \Rightarrow \|y\| \leq \|R^{-1}Q^T\| \|Ay\|.$$

Since $\|R^{-1}Q^T\| \|Ay\|$ is bounded we have that $\|y\|$ is bounded.

It remains to show that Γ is a closed set. For this we follow the usual line of argument and consider a sequence $s^k = (X^k, Z^k)_{k \in \mathbb{N}} \subset \Gamma$. Suppose this sequence converges to some $s^* = (X^*, Z^*) \in S(n)^2$. We have to show that $s^* \in (S(n)^{++})^2$. The boundedness of this sequence implies that there exists a $\delta > 0$ such that

$$\log\det X^k \leq \delta, \quad \text{and } \log\det Z^k \leq \delta \quad \text{for all } k = 1, 2, \ldots$$

We conclude that for all $k = 1, 2, \ldots$

$$\mu \log\det X^k \quad = \quad -f_\mu(p^k) + \langle Z^k, X^k \rangle - \mu \log\det Z^k \geq -\zeta - \mu\delta$$

or equivalently

$$\det X^k = \prod_{j=1}^{n} \lambda_j(X^k) \geq \exp\left(-\frac{c+\mu\delta}{\mu}\right) > 0.$$

Thus

$$\lambda_{\min}(X^k)\lambda_{\max}^{(n-1)}(X^k) \geq \exp\left(-\frac{c+\mu\delta}{\mu}\right) > 0.$$

From this we get

$$\lambda_{\min}(X^k) \geq \exp\left(-\frac{c+\mu\delta}{\mu}\Big/\lambda_{\max}^{(n-1)}(X^k)\right) > 0.$$

Let $\lambda_{\max}$ denote the maximum eigenvalue of all the maximum eigenvalues $\lambda_{\max}(Z^k)$, $(k = 1, 2, \ldots)$. Wlog. we may assume that $\lambda_{\max} \geq 1$ otherwise we set $\lambda_{\max} = 1$.

Hence

$$\lambda_{\min}(X^k) \geq \exp\left(-\frac{c+\mu\delta}{\mu}\Big/\lambda_{\max}^{(n-1)}\right) > 0.$$

Analogous we get

$$\prod_{j=1}^{n} \lambda_j(Z^k) \geq \exp\left(-\frac{c+\mu\delta}{\mu}\right), \qquad \text{for all } k = 1, 2, \ldots.$$

Thus we have shown that $s^* \in (S(n)^{++})^2$. This completes the proof.

□

The next result shows that the search direction Δp, produced by our algorithm (see Table 1.1) is a descent direction for f_μ. For a proof of this result see [54, 51], (see also chapter 3 section 3.3.1, where we show this result in the context with linear quadratic functionals in SDP).

Lemma 2.4 *The directional derivative of f_μ in the direction Δp satisfies*

$$\left\langle \frac{\partial f_\mu(p)}{\partial p}, \Delta p \right\rangle \leq 0.$$

Equality holds if and only if $F_\mu(p) = 0$.

The Goldstein Armijo criterion for the step size α such that $p^{k+1} = p^k + \alpha^k \Delta p^k$ and $0 < \delta < \frac{1}{2}$ is defined by

$$\alpha^k \ \text{ acceptable if and only if } \ \delta \leq \frac{f_\mu(p^{k+1}) - f_\mu(p^k)}{\alpha \left\langle \frac{\partial f_\mu(p)}{\partial p}, \Delta p \right\rangle} \leq 1 - \delta. \tag{2.4}$$

Because f_μ is bounded below and f_μ is continously differentiable on $\text{int}(\mathcal{F}(\mathcal{P})\times\mathcal{F}(\mathcal{D}))$ such an α does always exist. This condition guarantees that the step size α is neither too large nor too small and leads to a closed line search algorithm, see e.g. Luenberger [79]. That is, performing inaccurate line search with respect to the merit function, the Goldstein Armijo criterion and an initial point $p^\circ \in \text{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ ensures that the iterates generated by our algorithm stay within a compact set and move towards an optimal solution. In particular, because the search direction Δp is a descent direction for f_μ and because f_μ tends to infinity on the boundary of the set of feasible points we conclude that all iterates remain strictly feasible interior points. The feasibility remains preserved because the primal and dual residuum r_P and r_D are linear equations.

We think of an algorithm $\mathbf{A}$ as a mapping defined on a space X that assigns to every point $x \in X$ a subset of X. We will denote this subset by $\mathbf{A}(x)$. Hence the algorithm $\mathbf{A}$ is a point-to-set mapping. We next define the notion of closed maps, see e.g. [12, 79].

Definition 2.5 *Let X and Y be nonempty closed sets. Let $\mathbf{A}$: $X \mapsto Y$ be a point-to-set map. Then $\mathbf{A}$ is said to be closed at $x \in X$ if the assumptions*

$$\begin{array}{ll} x_k \in X, & x_k \to x \\ y_k \in \mathbf{A}(x_k), & y_k \to y \end{array}$$

imply that $y \in \mathbf{A}(x)$. The point-to-set map $\mathbf{A}$ is said to be closed on $Z \subseteq X$ if it is closed at each point in Z.

Theorem 2.6 *Suppose that $p \in \text{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ and $\Delta p \in S(n) \times \mathbb{R}^m \times S(n)$. For fixed δ, $0 < \delta < \frac{1}{2}$ the mapping $\mathcal{S} : \text{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))\ \times (S(n) \times \mathbb{R}^m \times S(n)) \mapsto \text{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ defined by*

$$\mathcal{S}(p,\Delta p) = \left\{ \bar{p} : \bar{p} = p + \alpha\Delta p \textit{ for some } \alpha \geq 0,\ \delta \leq \frac{f_\mu(\bar{p}) - f_\mu(p)}{\alpha \left\langle \frac{\partial f_\mu(p)}{\partial p}, \Delta p \right\rangle} \leq 1-\delta \right\}$$

is closed at $(p, \Delta p)$ if $\Delta p \neq 0$.

Proof. Suppose $(p^k)_{k\in\mathbb{N}}$ and $(\Delta p^k)_{k\in\mathbb{N}}$ are sequences such that $(p^k) \to p$ and $(\Delta p^k) \to \Delta p$ and that $(\bar{p}^k) \to \bar{p}$, where $\bar{p}^k \in \mathcal{S}(p^k, \Delta p^k)$. We have to show that $\bar{p} \in \mathcal{S}(p, \Delta p)$. Note for each k, $\bar{p}^k = p^k + \alpha^k \Delta p^k$ for some α^k. Hence

$$\alpha^k = \frac{\|\bar{X}^k - X^k\|}{\Delta X^k} \to \frac{\|\bar{X} - X\|}{\Delta X} \equiv \alpha$$

and similarly for interchanging X by y or Z. Hence α^k converges to some α and $\bar{p} = p + \alpha\Delta p$. Let

$$\phi(p, \Delta p, \alpha) = \frac{f_\mu(p+\alpha\Delta p) - f_\mu(p)}{\alpha \left\langle \frac{\partial f_\mu(p)}{\partial p}, \Delta p \right\rangle}.$$

Note that $\delta \leq \phi(p^k, \Delta p^k, \alpha^k) \leq 1 - \delta$ for all k. Since f_μ is continously differentiable it follows that $\phi(p, \Delta p, \alpha)$ is continous. Hence we have $\phi(p^k, \Delta p^k, \alpha^k) \to \phi(p, \Delta p, \alpha)$ and $\delta \leq \phi(p, \Delta p, \alpha) \leq 1 - \delta$. Thus, $\bar{p} \in \mathcal{S}(p, \Delta p)$, and the proof is complete.

□

If we use the Goldstein Armijo criterion for choosing the step length parameter α then the following theorem shows global convergence of our algorithm.

Theorem 2.7 *Suppose* $\operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})) \neq \emptyset$ *and let the nonempty solution set* $\hat{\Gamma} \subseteq \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ *be defined as*

$$\hat{\Gamma} = \Big\{(X_\mu, y_\mu, Z_\mu), \mu > 0 : (X_\mu, y_\mu, Z_\mu) \textit{ solves } (\mathbf{KKT}) \textit{ approximately}\Big\}.$$

Let $\mathcal{S}(p, \Delta p) : \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})) \times (S(n) \times I\!\!R^m \times S(n)) \mapsto \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ *be the point-to-set mapping defined in Theorem 2.6. Given* $p^\circ \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ *such that* $f_\mu(p^\circ) \leq \zeta$ *for some* $\zeta > 0$ *and an accuracy parameter* $\varepsilon > 0$ *the sequence* $(p^k)_{k \in I\!\!N}$ *is generated iteratively as follows:*

> *If* $p^k \in \hat{\Gamma}$ *and* $\mu^k \leq \varepsilon$ *then stop; otherwise for the current value of* μ *generate* p^{k+1}*, replace* k *by* $k+1$ *and repeat as long as* $\|Z^k X^k - \mu I\| \leq \varepsilon$.
> *If* $p^k \in \hat{\Gamma}$ *reduce* μ *by a constant factor* $\rho \in (0, 1)$.

Then the limit of any convergent subsequence of $(p^k)_{k \in I\!\!N}$ *is a solution.*

Proof. The iterates generated by our algorithm are contained in the compact set Γ, defined above. f_μ is a continous descent function with respect to the search direction Δp and the mapping $\mathcal{S}(p, \Delta p)$ is closed at points outside of $\hat{\Gamma}$. These are the conditions that ensure convergence of algorithmic maps. A detailed proof of this result is given in e.g. [79].

□

2.3 Polynomial Convergence of a Primal-Dual Algorithm in SDP

In [73] Kojima, Shindoh and Hara introduced the semidefinite linear complementarity problem, denoted **(SDLCP)**, in symmetric matrices. In this paper they proposed several interior-point algorithms for solving this problem and established polynomial convergence for these variants.

(SDLCP) is defined as the problem of finding a pair $(Z, X) \in (S(n)^+)^2$ which lies in a given $n(n+1)/2$ dimensional affine subspace $\mathcal{F}$ of $S(n)^2$ and satisfies the complementary slackness condition $\langle Z, X \rangle = 0$, i.e., find a pair $(Z, X) \in S(n)^2$ such that

$$(\mathbf{SDLCP}) \qquad (Z, X) \in \mathcal{F},\ X \in S(n)^+,\ Z \in S(n)^+ \text{ and } \langle Z, X \rangle = 0.$$

Setting

$$\mathcal{F} = \left\{ (Z,X) \in S(n)^2 : \begin{array}{l} \boldsymbol{\mathcal{A}}(X) = b \\ Z + \boldsymbol{\mathcal{A}}^t(y) = C, \text{ for some } y \in I\!R^m \end{array} \right\} \tag{2.5}$$

shows that **(SDLCP)** is a generalization of the semidefinite programming problem in standard form. We review the presentation given in [73]. In this context we will denote by $(Z,X) \in (S(n)^+)^2$ a feasible solution of the **(SDLCP)**. Then it is easy to see that a point $(X,y,Z) \in S(n) \times I\!R^m \times S(n)$ satisfying (2.5) is a feasible solution of the primal-dual pair **(PSDP)** and **(DSDP)** if and only if $(Z,X) \in S(n)^2$ is a feasible solution of the **(SDLCP)**. The reader is referred to [73] for more details.

In this section we are interested in the feasible path-following IPM considered in [73] and its convergence analysis. The theoretical results are adapted to the primal-dual pair **(PSDP)** and **(DSDP)** and to the primal-dual path-following algorithm for SDP given in Table 1.1.

Kojima et al. [73] fix throughout their work that the $n(n+1)/2$ dimensional affine subspace $\mathcal{F}$ associated with the **(SDLCP)** is monotone. In particular, they arrange that

$$\langle Z' - Z, X' - X \rangle \geq 0 \text{ for every } (Z',X') \in \mathcal{F} \text{ and } (Z,X) \in \mathcal{F}.$$

The **(SDLCP)** is then called the monotone **(SDLCP)**.

They present a theoretical framework for the monotone **(SDLCP)** where the iterates move in a Newton direction to approximate a point on the central path which is closer to optimality than the current point. As stated in section 1.4 this is the specific attribute of path-following algorithms.

It is shown that the central trajectory for the monotone **(SDLCP)** exists under the following assumptions.

(1) The interior of the set of feasible points is nonempty.

(2) The affine subspace $\mathcal{F}$ is monotone.

In particular, under this assumptions they show that the set $\mathcal{C} = \Big\{(Z_\mu, X_\mu) : Z_\mu X_\mu = \mu I,\ \mu > 0\Big\}$ forms a smooth trajectory (the central path). It is also shown that (Z_μ, X_μ) is the unique minimizer of the logarthmic barrier function defined as $\langle Z, X\rangle - \mu \log\det(ZX)$ over the interior of the feasible set and that the sequence (Z_μ, X_μ) converges to a solution of the monotone **(SDLCP)**, as μ decreases to zero.

For the special setting of the semidefinite programming problem in standard form the monotonicity condition holds with equality for any feasible triple (X,y,Z). In particular, it is easy to verify that this so called orthogonality property under the imposed conditions is satisfied automatically. Hence it has not to be presupposed as for the monotone **(SDLCP)**. In fact,

$$\langle Z - Z', X - X' \rangle = 0 \text{ for every } (X',y',Z'),\ (X,y,Z) \in \mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}). \tag{2.6}$$

Moreover, the existence of the central trajectory and the convergence of the sequence (X_μ, y_μ, Z_μ) for decreasing μ to an optimal solution is known for the pair **(PSDP)** and **(DSDP)**, (see the previous sections and the references quoted therein).

Kojima, Shindoh and Hara [73] establish the polynomial convergence of the central trajectory following method for the monotone **(SDLCP)** by restricting the iterates to a narrow neighbourhood of the central path. They introduced a horn neighbourhood $\mathcal{N}(\gamma)$ which can be defined for the semidefinite programming problem in standard form as follows. Let $p = (X, y, Z)$, then

$$\mathcal{N}(\gamma) = \left\{ p \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})) : \begin{array}{l} \left(\sum\limits_{j=1}^{n} (\lambda_j - \mu)^2 \right)^{1/2} \leq \gamma\mu, \text{ where } \mu = \frac{\langle Z,X \rangle}{n}, \\ \text{and } \lambda_1, \dots, \lambda_n \text{ are the eigenvalues of } ZX \end{array} \right\}$$

$$= \left\{ p \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})) : \|\sqrt{Z} X \sqrt{Z} - \mu I\| \leq \gamma\mu, \text{ where } \mu = \tfrac{\langle Z,X \rangle}{n} \right\}.$$

The parameter $\gamma > 0$ determines the wide of the neighbourhood. To show polynomial convergence of the algorithm given in Table 1.1 we modify it and start from a feasible interior point $p^\circ = (X^\circ, y^\circ, Z^\circ) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$. A sequence $p^k = (X^k, y^k, Z^k) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ is generated such that $\langle Z^k, X^k \rangle \to 0$ as k tends to ∞. Note, this is guaranteed if we perform inaccurate line search with respect to the merit function, defined in the previous section. (For feasible points $p = (X, y, Z)$ we have $f_\mu(p) = f(X, \mu) - g(y, Z, \mu) \to 0$ if μ decreases to zero; $f(X, \mu)$ denotes the primal barrier function and $g(y, Z, \mu)$ denotes the dual barrier function, respectively.) The orthogonality property yields

$$\langle Z^k, X^k \rangle + \langle Z^\circ, X^\circ \rangle = \langle Z^k, X^\circ \rangle + \langle Z^\circ, X^k \rangle \text{ for every } k = 0, 1, \dots$$

Hence there exists a positive number γ by which the right hand side is bounded above. This implies the following result, see also Lemma 2.3.

Lemma 2.8 *The sequence* $(X^k, y^k, Z^k) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ *is contained in a bounded and closed subset of*

$$\left\{ Z \in S(n)^+ : \langle Z, X^\circ \rangle \geq 0 \right\} \times I\!R^m \times \left\{ X \in S(n)^+ : \langle Z^\circ, X \rangle \geq 0 \right\}.$$

Hence the sequence p^k has at least one accumulation point and every accumulation point solves the primal-dual pair **(PSDP)** and **(DSDP)**.

In the next section we state and proof a series of lemmas which will be used to establish the polynomial convergence of the algorithm presented in Table 1.1.

2.3.1 Technical results

For a positive semidefinite matrix X there exists always a unique symmetric positive semidefinite matrix B such that $X = BB$. Wlog. we will denote such a matrix B

by $\sqrt{X}$. We will often use this notation in this section. Throughout this section we write $(\Delta\hat{X}, \Delta y, \Delta Z)$ if we refer to a solution of the linearized Newton system (1.9). We will write $\operatorname{diag}(X)$ to denote the vector formed from the diagonal elements of a square matrix X and $\operatorname{Diag}(x)$ to denote the diagonal matrix with diagonal x.

We mention that Kojima et al. [73] formulated the following lemma for a general matrix $Q \in S(n)$ for the monotone **(SDLCP)**. For our purpose it is sufficient to proof the lemma for the specific choice $Q = \sigma\mu I - ZX$, $\sigma \in (0, 1)$.

Lemma 2.9 [73] *Suppose that* $(Z, X) \in (S(n)^{++})^2$. *Let* $Q = \sigma\mu I - ZX$ *and let* $(\Delta\hat{X}, \Delta y, \Delta Z)$ *be a solution of the system (1.9) of equations. Then*

$$\|\sqrt{Z}^{-1} Q \sqrt{X}^{-1}\|^2 = \|\sqrt{Z}^{-1} \Delta Z \sqrt{X}\|^2 + \|\sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}\|^2.$$

Proof. Recall, μ is reduced at each iteration by a constant factor $\sigma \in (0, 1)$. Considering that, the third equation in (1.9) is equivalent to

$$\Delta Z\, X + Z\, \Delta\hat{X} = Q. \tag{2.7}$$

Since $(Z, X) \in (S(n)^{++})^2$ we can replace Z with $\sqrt{Z}\sqrt{Z}$ and X with $\sqrt{X}\sqrt{X}$, respectively. $\sqrt{Z}$ and $\sqrt{X}$ are unique nonsingular matrices. Using this and multiplying (2.7) by $\sqrt{Z}^{-1}$ from left and $\sqrt{X}^{-1}$ from right leads to

$$\sqrt{Z}^{-1} \Delta Z \sqrt{X} + \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1} = \sqrt{Z}^{-1} Q \sqrt{X}^{-1}.$$

Hence

$$\begin{aligned}
\|\sqrt{Z}^{-1} Q \sqrt{X}^{-1}\|^2 &= \|\sqrt{Z}^{-1} \Delta Z \sqrt{X} + \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}\|^2 \\
&= \langle \sqrt{Z}^{-1} \Delta Z \sqrt{X} + \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}, \sqrt{Z}^{-1} \Delta Z \sqrt{X} + \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1} \rangle \\
&= \langle \sqrt{Z}^{-1} \Delta Z \sqrt{X}, \sqrt{Z}^{-1} \Delta Z \sqrt{X} \rangle + \langle \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}, \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1} \rangle \\
&\quad + 2\langle \sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}, \sqrt{Z}^{-1} \Delta Z \sqrt{X} \rangle \\
&= \|\sqrt{Z}^{-1} \Delta Z \sqrt{X}\|^2 + \|\sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}\|^2 \\
&\quad + 2 \operatorname{trace}\left((\sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1})^T (\sqrt{Z}^{-1} \Delta Z \sqrt{X}) \right) \\
&= \|\sqrt{Z}^{-1} \Delta Z \sqrt{X}\|^2 + \|\sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}\|^2 \\
&\quad + 2 \operatorname{trace}\left(\sqrt{X}^{-1} \Delta\hat{X} \sqrt{Z} \sqrt{Z}^{-1} \Delta Z \sqrt{X} \right) \\
&= \|\sqrt{Z}^{-1} \Delta Z \sqrt{X}\|^2 + \|\sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}\|^2 \\
&\quad + 2 \operatorname{trace}\left(\Delta\hat{X} \sqrt{Z} \sqrt{Z}^{-1} \Delta Z \sqrt{X} \sqrt{X}^{-1} \right) \\
&= \|\sqrt{Z}^{-1} \Delta Z \sqrt{X}\|^2 + \|\sqrt{Z} \Delta\hat{X} \sqrt{X}^{-1}\|^2 + 2\langle \Delta Z, \Delta\hat{X} \rangle.
\end{aligned}$$

This completes the proof because (2.6) implies that $\langle \Delta Z, \Delta \hat{X} \rangle = 0$.

$\square$

Lemma 2.10 [73] *Suppose that* $(Z, X) \in (S(n)^{++})^2$. *Let* $(\Delta Z, \Delta X) \in M(n)^2$ *and let* $\lambda_{\min}(ZX)$ *and* $\lambda_{\max}(ZX)$ *denote the minimum and the maximum eigenvalues of* ZX, *respectively. Then*

(i) $\|\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}\| \leq \dfrac{\|\sqrt{Z}^{-1}\Delta Z\sqrt{X}\|}{\sqrt{\lambda_{\min}(ZX)}}$,

(ii) $\|\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1}\| \leq \dfrac{\|\sqrt{Z}\Delta X\sqrt{X}^{-1}\|}{\sqrt{\lambda_{\min}(ZX)}}$,

(iii) $\|\sqrt{Z}\Delta X\sqrt{Z}\| \leq \sqrt{\lambda_{\max}(ZX)}\|\sqrt{Z}\Delta X\sqrt{X}^{-1}\|$.

Proof.

$$\begin{aligned}
\|\sqrt{Z}^{-1}\Delta Z\sqrt{X}\|^2 &= \langle \sqrt{Z}^{-1}\Delta Z\sqrt{X}, \sqrt{Z}^{-1}\Delta Z\sqrt{X}\rangle \\
&= \operatorname{trace}(\sqrt{Z}^{-1}\Delta Z\sqrt{X})^T(\sqrt{Z}^{-1}\Delta Z\sqrt{X}) \\
&= \operatorname{trace}(\sqrt{X}\Delta Z^T\sqrt{Z}^{-1})(\sqrt{Z}^{-1}\Delta Z\sqrt{X}) \\
&= \operatorname{trace}(\sqrt{Z}^{-1}\Delta Z\sqrt{X})(\sqrt{X}\Delta Z^T\sqrt{Z}^{-1}) \\
&= \operatorname{trace}(\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}\sqrt{Z}\sqrt{X})(\sqrt{X}\sqrt{Z}\sqrt{Z}^{-1}\Delta Z^T\sqrt{Z}^{-1}) \\
&= \operatorname{trace}(\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1})(\sqrt{Z}X\sqrt{Z})\underbrace{(\sqrt{Z}^{-1}\Delta Z^T\sqrt{Z}^{-1})}_{=(\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1})^T} \\
&= \operatorname{trace}(\sqrt{Z}X\sqrt{Z})(\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1})^T(\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}) \\
&\geq \lambda_{\min}(ZX)\|\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}\|^2 \quad \text{(by Lemma A.6 and A.4).}
\end{aligned}$$

This proves (i).

$$\begin{aligned}
\|\sqrt{Z}\Delta X\sqrt{X}^{-1}\|^2 &= \langle \sqrt{Z}\Delta X\sqrt{X}^{-1}, \sqrt{Z}\Delta X\sqrt{X}^{-1}\rangle \\
&= \operatorname{trace}(\sqrt{Z}\Delta X\sqrt{X}^{-1})^T(\sqrt{Z}\Delta X\sqrt{X}^{-1}) \\
&= \operatorname{trace}(\sqrt{X}^{-1}\Delta X^T\sqrt{Z})(\sqrt{Z}\Delta X\sqrt{X}^{-1}) \\
&= \operatorname{trace}(\sqrt{X}^{-1}\Delta X^T\sqrt{X}^{-1}\sqrt{X}\sqrt{Z})(\sqrt{Z}\sqrt{X}\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1}) \\
&= \underbrace{\operatorname{trace}(\sqrt{X}^{-1}\Delta X^T\sqrt{X}^{-1})}_{=(\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1})^T}(\sqrt{X}Z\sqrt{X})(\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1})
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{trace}(\sqrt{X}Z\sqrt{X})(\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1})(\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1})^T \\
&\geq \lambda_{\min}(ZX)\|\sqrt{X}^{-1}\Delta X\sqrt{X}^{-1}\|^2 \quad \text{(by Lemma A.6 and A.4).}
\end{aligned}
$$

This proves (ii).

$$
\begin{aligned}
\|\sqrt{Z}\Delta X\sqrt{X}^{-1}\|^2 &= \langle \sqrt{Z}\Delta X\sqrt{X}^{-1}, \sqrt{Z}\Delta X\sqrt{X}^{-1}\rangle \\
&= \operatorname{trace}(\sqrt{Z}\Delta X\sqrt{X}^{-1})^T(\sqrt{Z}\Delta X\sqrt{X}^{-1}) \\
&= \operatorname{trace}(\sqrt{X}^{-1}\Delta X^T\sqrt{Z})(\sqrt{Z}\Delta X\sqrt{X}^{-1}) \\
&= \operatorname{trace}(\sqrt{Z}\Delta X\sqrt{X}^{-1})(\sqrt{X}^{-1}\Delta X^T\sqrt{Z}) \\
&= \operatorname{trace}(\sqrt{Z}\Delta X\sqrt{Z}\sqrt{Z}^{-1}\sqrt{X}^{-1})(\sqrt{X}^{-1}\sqrt{Z}^{-1}\sqrt{Z}\Delta X^T\sqrt{Z}) \\
&= \operatorname{trace}(\sqrt{Z}\Delta X\sqrt{Z})(\sqrt{Z}^{-1}(\sqrt{X}\sqrt{X})^{-1}\sqrt{Z}^{-1})(\sqrt{Z}\Delta X^T\sqrt{Z}) \\
&= \operatorname{trace}(\sqrt{Z}\Delta X\sqrt{Z})(\sqrt{Z}X\sqrt{Z})^{-1}\underbrace{(\sqrt{Z}\Delta X^T\sqrt{Z})}_{=(\sqrt{Z}\Delta X\sqrt{Z})^T} \\
&= \operatorname{trace}(\sqrt{Z}X\sqrt{Z})^{-1}(\sqrt{Z}\Delta X\sqrt{Z})^T(\sqrt{Z}\Delta X\sqrt{Z}) \\
&\geq \frac{1}{\lambda_{\max}(ZX)}\|\sqrt{Z}\Delta X\sqrt{Z}\|^2.
\end{aligned}
$$

The last inequality follows from the fact that $1/\lambda_{\max}(ZX)$ is the minimum eigenvalue of the matrix $(\sqrt{Z}X\sqrt{Z})^{-1}$. Applying Lemma A.6 and Lemma A.4 completes the proof.

□

Lemma 2.11 [73] *Suppose that $(Z,X) \in (S(n)^{++})^2$ and let $\lambda_{\min}(ZX)$ and $\lambda_{\max}(ZX)$ denote the minimum and the maximum eigenvalues of ZX, respectively. Let $(\Delta\bar{Z}, \Delta\bar{X}) \in M(n)^2$, $\Delta Z = (\Delta\bar{Z} + \Delta\bar{Z}^T)/2$, $\Delta X = (\Delta\bar{X} + \Delta\bar{X}^T)/2$, $\xi_1, \ldots, \xi_n$ be the eigenvalues of $Z^{-1}\Delta Z$ and $\eta_1, \ldots, \eta_n$ be the eigenvalues of $X^{-1}\Delta X$. Then*

(i) $\displaystyle\sum_{j=1}^n \xi_j^2 \leq \|\sqrt{Z}^{-1}\Delta\bar{Z}\sqrt{Z}^{-1}\|^2 \leq \frac{\|\sqrt{Z}^{-1}\Delta\bar{Z}\sqrt{X}\|^2}{\lambda_{\min}(ZX)},$

(ii) $\displaystyle\sum_{j=1}^n \eta_j^2 \leq \|\sqrt{X}^{-1}\Delta\bar{X}\sqrt{X}^{-1}\|^2 \leq \frac{\|\sqrt{Z}\Delta\bar{X}\sqrt{X}^{-1}\|^2}{\lambda_{\min}(ZX)}.$

Proof. The matrix $\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}$ is similar to the matrix $Z^{-1}\Delta Z$. Hence this two matrices share the same spectrum. In fact,

$$\sum_{j=1}^n \xi_j^2 = \|\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}\|^2$$

$$
\begin{aligned}
&= \left\| \sqrt{Z}^{-1} \left(\frac{\Delta\bar{Z} + \Delta\bar{Z}^T}{2} \right) \sqrt{Z}^{-1} \right\|^2 \quad (\text{since } \Delta Z = (\Delta\bar{Z} + \Delta\bar{Z}^T)/2) \\
&\leq \left(\frac{\|\sqrt{Z}^{-1}\Delta\bar{Z}\sqrt{Z}^{-1}\|}{2} + \frac{\|\sqrt{Z}^{-1}\Delta\bar{Z}^T\sqrt{Z}^{-1}\|}{2} \right)^2 \\
&= \|\sqrt{Z}^{-1}\Delta\bar{Z}\sqrt{Z}^{-1}\|^2 \\
&\leq \frac{\|\sqrt{Z}^{-1}\Delta\bar{Z}\sqrt{X}\|^2}{\lambda_{\min}(ZX)}.
\end{aligned}
$$

The last inequality follows from Lemma 2.10, item (i). Similarly this style of argument works for (ii).

□

Lemma 2.12 [73] *Suppose that* $(Z, X) \in (S(n)^{++})^2$, $\mu = \langle Z, X\rangle/n$, $0 \leq \sigma \leq 1$, $Q = \sigma\mu I - ZX$ *and let* $\lambda_1(ZX), \ldots, \lambda_n(ZX)$ *be the eigenvalues of the matrix* ZX. *Let* $\lambda_{\min}(ZX)$ *and* $\lambda_{\max}(ZX)$ *denote the minimum and the maximum eigenvalues of* ZX, *respectively. Assume that* $(X, y, Z) \in \mathcal{N}(\gamma)$ *for some* $\gamma > 0$. *Then*

(i) $(1-\gamma)\mu \leq \lambda_{\min}(ZX) \leq \lambda_j(ZX) \leq \lambda_{\max}(ZX) \leq (1+\gamma)\mu$ *for every* $j = 1, \ldots, n$,

(ii) $\|\sqrt{Z}^{-1}Q\sqrt{X}^{-1}\| \leq \min\left\{ \dfrac{((1-\sigma)\sqrt{n}+\gamma)\mu}{\sqrt{\lambda_{\min}(ZX)}}, \dfrac{\sqrt{2n}\mu}{\sqrt{\lambda_{\min}(ZX)}} \right\}$.

Proof. Since $(X, y, Z) \in \mathcal{N}(\gamma)$ we have

$$
\left(\sum_{j=1}^{n} (\lambda_j(ZX) - \mu)^2 \right)^{1/2} \leq \gamma\mu. \tag{2.8}
$$

This implies that $|\lambda_j(ZX) - \mu| \leq \gamma\mu$ for every $j = 1, \ldots, n$. Hence

$$
(1-\gamma)\mu \leq \lambda_{\min}(ZX) \leq \lambda_j(ZX) \leq \lambda_{\max}(ZX) \leq (1+\gamma)\mu \text{ for every } j = 1, \ldots, n.
$$

Now we will consider item (ii). We have

$$
\begin{aligned}
&\|\sqrt{Z}^{-1}Q\sqrt{X}^{-1}\|^2 \\
&\quad = \|\sigma\mu\sqrt{Z}^{-1}\sqrt{X}^{-1} - \sqrt{Z}^{-1}\sqrt{Z}\sqrt{Z}\sqrt{X}\sqrt{X}\sqrt{X}^{-1}\|^2 \\
&\quad = \|\sigma\mu\sqrt{Z}^{-1}\sqrt{X}^{-1} - \sqrt{Z}\sqrt{X}\|^2 \\
&\quad = \langle \sigma\mu\sqrt{Z}^{-1}\sqrt{X}^{-1} - \sqrt{Z}\sqrt{X}, \sigma\mu\sqrt{Z}^{-1}\sqrt{X}^{-1} - \sqrt{Z}\sqrt{X} \rangle
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma^2\mu^2\langle \sqrt{Z}^{-1}\sqrt{X}^{-1}, \sqrt{Z}^{-1}\sqrt{X}^{-1}\rangle - 2\sigma\mu\langle \sqrt{Z}^{-1}\sqrt{X}^{-1}, \sqrt{Z}\sqrt{X}\rangle \\
&\quad + \langle \sqrt{Z}\sqrt{X}, \sqrt{Z}\sqrt{X}\rangle \\
&= \sigma^2\mu^2 \operatorname{trace}(\sqrt{X}^{-1}\sqrt{Z}^{-1}\sqrt{Z}^{-1}\sqrt{X}^{-1}) - 2\sigma\mu \operatorname{trace}(\sqrt{X}^{-1}\sqrt{Z}^{-1}\sqrt{Z}\sqrt{X}) \\
&\quad + \operatorname{trace}(\sqrt{X}\sqrt{Z}\sqrt{Z}\sqrt{X}) \\
&= \sigma^2\mu^2 \operatorname{trace}(\sqrt{X}\sqrt{Z}\sqrt{Z}\sqrt{X})^{-1} - 2\sigma\mu \operatorname{trace}((\sqrt{Z}\sqrt{Z}^{-1})^T(\sqrt{X}\sqrt{X}^{-1})) \\
&\quad + \operatorname{trace}(\sqrt{X}Z\sqrt{X}) \\
&= \sigma^2\mu^2 \operatorname{trace}(\sqrt{X}Z\sqrt{X})^{-1} - 2\sigma\mu \operatorname{trace}(I) + \operatorname{trace}(\sqrt{X}Z\sqrt{X}) \\
&= \sum_{j=1}^{n}\left(\frac{\sigma^2\mu^2}{\lambda_j(ZX)} - 2\sigma\mu + \lambda_j(ZX)\right) \\
&= \sum_{j=1}^{n}\left(\frac{\sigma\mu}{\sqrt{\lambda_j(ZX)}} - \sqrt{\lambda_j(ZX)}\right)^2 \\
&= \left\|\operatorname{Diag}\left(\frac{\sigma\mu}{\sqrt{\lambda_1(ZX)}} - \sqrt{\lambda_1(ZX)}, \ldots, \frac{\sigma\mu}{\sqrt{\lambda_n(ZX)}} - \sqrt{\lambda_n(ZX)}\right)\right\|^2.
\end{aligned}
$$

We will denote the diagonal matrix with diagonal formed from the vector of the eigenvalues $\lambda_j(ZX), (j = 1, \ldots, n)$ by Λ, that is $\Lambda = \operatorname{Diag}(\lambda_1(ZX), \lambda_2(ZX), \ldots, \lambda_n(ZX))$. Then we see that

$$\sqrt{\Lambda} = \operatorname{Diag}\left(\sqrt{\lambda_1(ZX)}, \ldots, \sqrt{\lambda_n(ZX)}\right)$$

and

$$\sqrt{\Lambda}^{-1} = \operatorname{Diag}\left(\frac{1}{\sqrt{\lambda_1(ZX)}}, \ldots, \frac{1}{\sqrt{\lambda_n(ZX)}}\right).$$

Hence

$$\operatorname{Diag}\left(\frac{\sigma\mu}{\sqrt{\lambda_1(ZX)}} - \sqrt{\lambda_1(ZX)}, \ldots, \frac{\sigma\mu}{\sqrt{\lambda_n(ZX)}} - \sqrt{\lambda_n(ZX)}\right) = \sigma\mu\sqrt{\Lambda}^{-1} - \sqrt{\Lambda}.$$

Then we have

$$
\begin{aligned}
\|\sigma\mu\sqrt{\Lambda}^{-1} - \sqrt{\Lambda}\| &= \|(\sigma-1)\mu\sqrt{\Lambda}^{-1} + \mu\sqrt{\Lambda}^{-1} - \sqrt{\Lambda}\| \\
&\leq (1-\sigma)\mu\|\sqrt{\Lambda}^{-1}\| + \|\mu\sqrt{\Lambda}^{-1} - \sqrt{\Lambda}\| \\
&= (1-\sigma)\mu\sqrt{\sum_{j=1}^{n}\left(\frac{1}{\lambda_j(ZX)}\right)} + \sqrt{\sum_{j=1}^{n}\left(\frac{\mu}{\sqrt{\lambda_j(ZX)}} - \sqrt{\lambda_j(ZX)}\right)^2} \\
&\leq (1-\sigma)\mu\sqrt{\sum_{j=1}^{n}\left(\frac{1}{\lambda_{\min}(ZX)}\right)} + \sqrt{\sum_{j=1}^{n}\left(\frac{\mu - \lambda_j(ZX)}{\sqrt{\lambda_{\min}(ZX)}}\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
&\le (1-\sigma)\mu\sqrt{\frac{n}{\lambda_{\min}(ZX)}} + \frac{\gamma\mu}{\sqrt{\lambda_{\min}(ZX)}} \quad \text{(by (2.8))} \\
&= \frac{((1-\sigma)\sqrt{n}+\gamma)\mu}{\sqrt{\lambda_{\min}(ZX)}}.
\end{aligned}
$$

We also have

$$
\begin{aligned}
\|\sigma\mu\sqrt{\Lambda}^{-1} - \sqrt{\Lambda}\|^2 &= \sum_{j=1}^{n}\left(\frac{\sigma\mu}{\sqrt{\lambda_j(ZX)}} - \sqrt{\lambda_j(ZX)}\right)^2 \\
&\le \sum_{j=1}^{n}\left(\left(\frac{\sigma\mu}{\sqrt{\lambda_j(ZX)}}\right)^2 + \left(\sqrt{\lambda_j(ZX)}\right)^2\right) \\
&\le \frac{n\mu^2}{\lambda_{\min}(ZX)} + n\mu \quad \left(\text{since } 0 \le \sigma \le 1 \text{ and } \sum_{j=1}^{n}\lambda_j(ZX) = n\mu\right) \\
&\le \frac{n\mu^2}{\lambda_{\min}(ZX)} + n\mu\frac{\mu}{\lambda_{\min}(ZX)} \\
&\quad \left(\text{since } \sum_{j=1}^{n}\lambda_j(ZX) = n\mu \Rightarrow \lambda_{\min}(ZX) \le \mu \Rightarrow 1 \le \mu/\lambda_{\min}(ZX)\right) \\
&= \frac{2n\mu^2}{\lambda_{\min}(ZX)}.
\end{aligned}
$$

This completes the proof.

□

Lemma 2.13 *Suppose that* $(Z,X) \in (S(n)^{++})^2$*,* $Z = UU^T$ *for some matrix* $U \in M(n)$*, and* $\mu = \langle Z, X\rangle/n$*. Then*

$$\min_{\nu\in\mathbb{R}} \|U^TXU - \nu I\| = \|U^TXU - \mu I\|..$$

Proof. Let $f(\nu) = \|U^TXU - \nu I\|^2$. Then

$$
\begin{aligned}
\|U^TXU - \nu I\|^2 &= \langle U^TXU - \nu I, U^TXU - \nu I\rangle \\
&= \langle U^TXU, U^TXU\rangle - 2\nu\langle U^TXU, I\rangle + \nu^2\langle I, I\rangle.
\end{aligned}
$$

Let ν^* denote the minimizer of $f(\nu)$. By the first order necessary condition we have

$$\frac{\partial}{\partial\nu}f(\nu^*) = 0.$$

We see that

$$\frac{\partial}{\partial\nu}f(\nu) = 2n\nu - 2\langle U^TXU, I\rangle$$

Hence

$$\frac{\partial}{\partial \nu} f(\nu) = 0 \Leftrightarrow \nu = \frac{\langle U^T X U, I\rangle}{n} \Leftrightarrow \nu = \frac{\mathrm{trace}(UU^T X)}{n} \Leftrightarrow \nu = \frac{\langle Z, X\rangle}{n} \Leftrightarrow \nu = \mu.$$

$\square$

2.3.2 The Convergence Theorem

In this section we will present the convergence theory given by Kojima et al. [73]. In particular we suit the main convergence result to the special setting of the semidefinite programming problem in standard form. As we will see the polynomial convergence of the algorithm proposed in Table 1.1 is proved for a well suited setting of the parameters α, γ and σ. For simplicity we omit the iteration counter in this section. We will write (X, y, Z) to denote the current iteration point and add the superscript "+" to denote the updated variables.

Theorem 2.14 [73] *Let $\gamma \in (0, 0.1]$. Suppose that $(X, y, Z) \in \mathcal{N}(\gamma)$. Let $\sigma = 1 - \gamma/\sqrt{n}$ and $\alpha = 1$. Let $(\Delta\hat{X}, \Delta y, \Delta Z)$ be a solution of the system (1.9) of equations, $\Delta X = (\Delta\hat{X} + \Delta\hat{X}^T)/2$ and let $\mu^+ = \langle Z^+, X^+\rangle/n$. Then*

$$(X^+, y^+, Z^+) = (X, y, Z) + (\Delta X, \Delta y, \Delta Z) \in \mathcal{N}(\gamma)$$

$$\mu^+ = \left(1 - \frac{\gamma}{\sqrt{n}}\right)\mu. \tag{2.9}$$

Proof. Note that $\langle Z, \Delta X\rangle = \langle Z, \Delta X^T\rangle$. Hence $\langle \Delta Z, X\rangle + \langle Z, \Delta X\rangle = n\sigma\mu - \langle Z, X\rangle$. Then

$$\begin{aligned}
n\mu^+ &= \langle Z^+, X^+\rangle \\
&= \langle Z + \Delta Z, X + \Delta X\rangle \\
&= \langle Z, X\rangle + \langle \Delta Z, X\rangle + \langle Z, \Delta X\rangle + \langle \Delta Z, \Delta X\rangle \\
&= n\mu + n\sigma\mu - \langle Z, X\rangle + \langle \Delta Z, \Delta X\rangle \\
&= n\mu + n\sigma\mu - n\mu + \langle \Delta Z, \Delta X\rangle \\
&= n\sigma\mu + \langle \Delta Z, \Delta X\rangle.
\end{aligned} \tag{2.10}$$

Let $\xi_1, \ldots, \xi_n$ be the eigenvalues of $Z^{-1}\Delta Z$ and $\eta_1, \ldots, \eta_n$ be the eigenvalues of $X^{-1}\Delta X$. Suppose $\lambda_{\min}(ZX)$ and $\lambda_{\max}(ZX)$ denotes the minimum and the maximum eigenvalues of ZX, respectively. Then

$$\sum_{j=1}^{n} \left(\xi_j^2 + \eta_j^2\right) \leq \frac{\|\sqrt{Z}^{-1}\Delta Z\sqrt{X}\|^2}{\lambda_{\min}(ZX)} + \frac{\|\sqrt{Z}\Delta X\sqrt{X}^{-1}\|^2}{\lambda_{\min}(ZX)} \quad \text{(by Lemma 2.11)}$$

$$
\begin{aligned}
&= \frac{\|\sqrt{Z}^{-1}(\sigma\mu I - ZX)\sqrt{X}^{-1}\|^2}{\lambda_{\min}(ZX)} \quad \text{(by Lemma 2.9)}\\
&\leq \frac{1}{\lambda_{\min}(ZX)} \min\left\{\frac{((1-\sigma)\sqrt{n}+\gamma)\mu}{\sqrt{\lambda_{\min}(ZX)}}, \frac{\sqrt{2n}\mu}{\sqrt{\lambda_{\min}(ZX)}}\right\}^2\\
&\qquad \text{(by Lemma 2.12 item (ii))}\\
&= \frac{1}{\lambda_{\min}(ZX)} \min\left\{\frac{4\gamma^2\mu^2}{\lambda_{\min}(ZX)}, \frac{2n\mu^2}{\lambda_{\min}(ZX)}\right\} \quad (\text{since } \sigma = 1-\gamma/\sqrt{n})\\
&\leq \min\left\{\frac{4\gamma^2}{(1-\gamma)^2}, \frac{2n}{(1-\gamma)^2}\right\} \quad \text{(by Lemma 2.12 item (i))}\\
&= \frac{4\gamma^2}{(1-\gamma)^2} \quad (\text{because } 4\gamma^2 \leq 2n)\\
&< 1 \quad (\text{because } \gamma \in (0, 0.1]).
\end{aligned}
$$

This implies that $|\xi_j| < 1$ and $|\eta_j| < 1$ for every $j = 1, \dots, n$. We will denote by $\xi_{\min}$ the minimum of the absolute values of all this eigenvalues. Hence $1 + \xi_{\min} > 0$ and therefore applying Lemma 1.17 ensures that

$$
(X^+, y^+, Z^+) = (X, y, Z) + (\Delta X, \Delta y, \Delta Z) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D})). \tag{2.11}
$$

In particular, the updated point (X^+, y^+, Z^+) is again a feasible point and therefore $\langle \Delta Z, \Delta X\rangle = 0$ by (2.6). Hence from (2.10) $\mu^+ = \sigma\mu = (1-\gamma/\sqrt{n})\mu$ and we have shown (2.9). Because $1-\gamma/\sqrt{n} < 1$ it follows that the sequence of $(\mu^k)_{k\in\mathbb{N}}$ approaches zero. To complete the proof, it remains to show that

$$
\|\sqrt{\bar{Z}}^T X^+ \sqrt{\bar{Z}} - \mu^+ I\| \leq \gamma\mu^+ \tag{2.12}
$$

for some $\sqrt{\bar{Z}}$ such that $Z^+ = \sqrt{\bar{Z}}\sqrt{\bar{Z}}^T$. We see that

$$
\begin{aligned}
\|\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}\| &\leq \frac{\|\sqrt{Z}^{-1}\Delta Z\sqrt{X}\|}{\sqrt{\lambda_{\min}(ZX)}} \quad \text{(by Lemma 2.10 item (i))}\\
&\leq \frac{\|\sqrt{Z}^{-1}Q\sqrt{X}^{-1}\|}{\sqrt{\lambda_{\min}(ZX)}}, \text{ where } Q = \sigma\mu I - ZX \quad \text{(by Lemma 2.9)}\\
&\leq \frac{1}{\sqrt{\lambda_{\min}(ZX)}} \min\left\{\frac{((1-\sigma)\sqrt{n}+\gamma)\mu}{\sqrt{\lambda_{\min}(ZX)}}, \frac{\sqrt{2n}\mu}{\sqrt{\lambda_{\min}(ZX)}}\right\}\\
&\qquad \text{(by Lemma 2.12 item (ii))}\\
&\leq \min\left\{\frac{2\gamma}{1-\gamma}, \frac{\sqrt{2n}}{1-\gamma}\right\}\\
&\qquad (\text{since } \sigma = 1-\gamma/\sqrt{n} \text{ and by Lemma 2.12 item (i)})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2\gamma}{1-\gamma} \quad (\text{because } 2\gamma \le \sqrt{2n}) \\
&< 1 \quad (\text{because } \gamma \in (0, 0.1]). \qquad (2.13)
\end{aligned}
$$

We also see that

$$
\begin{aligned}
\|\sqrt{Z}\Delta\hat{X}\sqrt{Z}\| &\le \sqrt{\lambda_{\max}(ZX)}\|\sqrt{Z}\Delta\hat{X}\sqrt{X}^{-1}\| \quad (\text{by Lemma 2.10 item (iii)}) \\
&\le \sqrt{\lambda_{\max}(ZX)}\|\sqrt{Z}^{-1}Q\sqrt{X}^{-1}\|, \text{ where } Q = \sigma\mu I - ZX \\
&\quad (\text{by Lemma 2.9}) \\
&\le \sqrt{(1+\gamma)\mu}\min\left\{\frac{((1-\sigma)\sqrt{n}+\gamma)\mu}{\sqrt{(1-\gamma)\mu}}, \frac{\sqrt{2n}\mu}{\sqrt{(1-\gamma)\mu}}\right\} \quad (\text{by Lemma 2.12}) \\
&= \sqrt{(1+\gamma)\mu}\min\left\{\frac{2\gamma\sqrt{\mu}}{\sqrt{1-\gamma}}, \frac{\sqrt{2n}\sqrt{\mu}}{\sqrt{1-\gamma}}\right\} \quad (\text{since } \sigma = 1-\gamma/\sqrt{n}) \\
&= \sqrt{(1+\gamma)\mu}\frac{2\gamma\sqrt{\mu}}{\sqrt{1-\gamma}} \quad (\text{because } 2\gamma \le \sqrt{2n}) \\
&= 2\mu\gamma\frac{\sqrt{1+\gamma}}{\sqrt{1-\gamma}}. \qquad (2.14)
\end{aligned}
$$

Since $(X^+, y^+, Z^+) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$, see (2.11), we have

$$
Z^+ = Z + \Delta Z = \sqrt{Z}\sqrt{Z} + \Delta Z = \sqrt{Z}(I + \sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1})\sqrt{Z}.
$$

It follows from (2.13) that the absolute values of all eigenvalues of the symmetric matrix $\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}$ are less than one. This implies that the symmetric matrix $I + \sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1} \in S(n)^{++}$. Hence there exists an orthogonal matrix P and a diagonal matrix $\Pi = \operatorname{Diag}(\pi_1, \ldots, \pi_n)$, where $\pi_1, \ldots, \pi_n$ are the eigenvalues of $I+\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1}$, such that $I+\sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1} = P\Pi P^T$. Thus we get the following chain of inequalities.

$$
\frac{1-3\gamma}{1-\gamma} = 1 - \frac{2\gamma}{1-\gamma} \le \pi_j \le 1 + \frac{2\gamma}{1-\gamma} = \frac{1+\gamma}{1-\gamma} \text{ for every } j = 1, \ldots, n. \qquad (2.15)
$$

Now set $\sqrt{\bar{Z}} = \sqrt{Z}P\sqrt{\Pi}$. Then

$$
Z^+ = \sqrt{Z}(I + \sqrt{Z}^{-1}\Delta Z\sqrt{Z}^{-1})\sqrt{Z} = \sqrt{Z}P\Pi P^T\sqrt{Z} = \sqrt{Z}P\sqrt{\Pi}\sqrt{\Pi}P^T\sqrt{Z} = \sqrt{\bar{Z}}\sqrt{\bar{Z}}^T.
$$

The third equation in the linearized Newton system (1.9) is used to get

$$
\begin{aligned}
Z^+(X + \Delta\hat{X}) - \sigma\mu I &= (Z + \Delta Z)(X + \Delta\hat{X}) - \sigma\mu I \\
&= Z\Delta\hat{X} + \Delta Z X + Z\hat{X} + \Delta Z\Delta\hat{X} - \sigma\mu I \\
&= \Delta Z\Delta\hat{X}.
\end{aligned}
$$

Hence, multiplying from left by $\sqrt{\bar{Z}}^{-1}$ and from right by $\sqrt{\bar{Z}}$ leads to

$$\sqrt{\bar{Z}}^T (X + \Delta\hat{X})\sqrt{\bar{Z}} - \sigma\mu I = \sqrt{\bar{Z}}^{-1} \Delta Z \Delta\hat{X}\sqrt{\bar{Z}}. \tag{2.16}$$

In the sequel we will show the validity of (2.12).

$$\begin{aligned}
&\|\sqrt{\bar{Z}}^T X^+ \sqrt{\bar{Z}} - \mu^+ I\| \\
&\leq \|\sqrt{\bar{Z}}^T X \sqrt{\bar{Z}} - \sigma\mu I\| \quad \text{(by Lemma 2.13)} \\
&= \left\|\sqrt{\bar{Z}}^T \left(X + \frac{\Delta\hat{X} + \Delta\hat{X}^T}{2}\right)\sqrt{\bar{Z}} - \sigma\mu I\right\| \\
&\leq \left\|\sqrt{\bar{Z}}^T \left(\frac{X + \Delta\hat{X}}{2}\right)\sqrt{\bar{Z}} - \frac{\sigma}{2}\mu I\right\| + \left\|\sqrt{\bar{Z}}^T \left(\frac{X + \Delta\hat{X}^T}{2}\right)\sqrt{\bar{Z}} - \frac{\sigma}{2}\mu I\right\| \\
&= \|\sqrt{\bar{Z}}^T (X + \Delta\hat{X})\sqrt{\bar{Z}} - \sigma\mu I\| \\
&= \|\sqrt{\bar{Z}}^{-1} \Delta Z \Delta\hat{X}\sqrt{\bar{Z}}\| \quad \text{(by (2.16))} \\
&= \|\sqrt{\Pi}^{-1} P^T \sqrt{Z}^{-1} \Delta Z \sqrt{Z}^{-1} \sqrt{Z}\Delta\hat{X}\sqrt{Z}P\sqrt{\Pi}\| \quad (\text{since } \sqrt{\bar{Z}} = \sqrt{Z}P\sqrt{\Pi}) \\
&\leq \frac{\sqrt{(1+\gamma)/(1-\gamma)}}{\sqrt{(1-3\gamma)/(1-\gamma)}} \|P^T \sqrt{Z}^{-1} \Delta Z \sqrt{Z}^{-1} \sqrt{Z}\Delta\hat{X}\sqrt{Z}P\| \\
&\qquad \text{(by Lemma A.7 and (2.15))} \\
&\leq \frac{\sqrt{1+\gamma}}{\sqrt{1-3\gamma}} \|\sqrt{Z}^{-1} \Delta Z \sqrt{Z}^{-1}\| \|\sqrt{Z}\Delta\hat{X}\sqrt{Z}\| \quad (\text{note } P \text{ is orthogonal}) \\
&\leq \frac{\sqrt{1+\gamma}}{\sqrt{1-3\gamma}} \frac{2\gamma}{1-\gamma} 2\mu\gamma \frac{\sqrt{1+\gamma}}{\sqrt{1-\gamma}} \quad \text{(by (2.13) and (2.14))} \\
&= \frac{4\gamma^2(1+\gamma)\mu}{(1-\gamma)^{\frac{3}{2}}\sqrt{1-3\gamma}} \\
&= \frac{4\gamma^2(1+\gamma)\mu^+}{(1-\gamma)^{\frac{3}{2}}\sqrt{1-3\gamma}\,\sigma} \quad (\text{since } \mu^+ = \sigma\mu) \\
&\leq \frac{4\gamma^2(1+\gamma)\mu^+}{(1-\gamma)^{\frac{3}{2}}\sqrt{1-3\gamma}(1-\gamma)} \quad (\text{since } \sigma = 1 - \gamma/\sqrt{n} \geq 1-\gamma)
\end{aligned}$$

Hence

$$\|\sqrt{\bar{Z}}^T X^+ \sqrt{\bar{Z}} - \mu^+ I\| \leq \frac{4\gamma(1+\gamma)}{(1-\gamma)^{\frac{5}{2}}\sqrt{1-3\gamma}}\gamma\mu^+ \leq \gamma\mu^+.$$

The last inequality follows from $\gamma \in (0, 0.1]$. This completes the proof.

□

A consequence of Theorem 2.14 is that for given $\varepsilon > 0$ the algorithm terminates in at most $O\left(\sqrt{n}\log\frac{\langle Z^\circ, X^\circ\rangle}{\varepsilon}\right)$ steps. This is subject of the following corollary.

Corollary 2.15 *Let $\varepsilon > 0$ and let $(X^k, y^k, Z^k) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ denote the $k-$th iteration point generated by the IPM given in Table 1.1. Then (X^k, y^k, Z^k) is an approximate solution of the primal-dual pair* **(PSDP)** *and* **(DSDP)** *(in the current context that is a solution $(X^k, y^k, Z^k) \in \operatorname{int}(\mathcal{F}(\mathcal{P}) \times \mathcal{F}(\mathcal{D}))$ such that $\langle Z^k, X^k\rangle \le \varepsilon$) if*

$$k \geq \frac{\sqrt{n}}{\gamma}\log\frac{\langle Z^\circ, X^\circ\rangle}{\varepsilon} = O\left(\sqrt{n}\log\frac{\langle Z^\circ, X^\circ\rangle}{\varepsilon}\right).$$

Proof. Let $\varepsilon > 0$. From Theorem 2.14 we see that the algorithm generates a sequence (X^k, y^k, Z^k) such that

$$(X^k, y^k, Z^k) \in \mathcal{N}(\gamma) \quad \text{and} \quad \langle Z^k, X^k\rangle = \left(1 - \frac{\gamma}{\sqrt{n}}\right)^k \langle Z^\circ, X^\circ\rangle$$

for every $k = 0, 1, \ldots$. Hence (X^k, y^k, Z^k) is an approximate solution of the problem if

$$\left(1 - \frac{\gamma}{\sqrt{n}}\right)^k \langle Z^\circ, X^\circ\rangle \le \varepsilon$$

or equivalently if

$$k \geq \frac{\log\varepsilon - \log\langle Z^\circ, X^\circ\rangle}{\log\left(1 - \frac{\gamma}{\sqrt{n}}\right)}. \quad \left(\text{note } \log\left(1 - \frac{\gamma}{\sqrt{n}}\right) < 0\right)$$

Since $\gamma \in (0, 0.1]$ we have $1 - \gamma/\sqrt{n} > 0$. Hence $\log(1 - \gamma/\sqrt{n}) \le -\gamma/\sqrt{n}$. Therefore the estimate above can be simplified. In particular, (X^k, y^k, Z^k) is an approximate solution of the problem for

$$k \geq \frac{\sqrt{n}}{\gamma}\log\frac{\langle Z^\circ, X^\circ\rangle}{\varepsilon}.$$

That is the proof.

□

2.4 Solutions in a Prespecified Domain

In this section we allow an initial point $(X^\circ, y^\circ, Z^\circ) \in S(n)^{++} \times \mathbb{R}^m \times S(n)^{++}$ which may be infeasible. We propose a condition that indicates the algorithm to terminate if the given primal-dual problem has no solution in a given domain. We mention that Lin and Saigal generalized the infeasible IPM of Mizuno [85] for LP to the more complicated setting of SDP, see [75]. They do not require that the primal-dual pair does have feasible solutions. Their method either shows that there is no optimal solution with duality gap zero in a well defined bounded region or for given $\varepsilon > 0$ finds an approximate solution $(X(\varepsilon), y(\varepsilon), Z(\varepsilon))$ such that $\langle Z(\varepsilon), X(\varepsilon)\rangle \le n\varepsilon$, $\|\boldsymbol{\mathcal{A}}(X(\varepsilon)) - b\| \le \varepsilon$ and $\|Z(\varepsilon) + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y(\varepsilon)) - C\| \le \varepsilon$ within this region. Their IPM is similar to a method proposed by Potra and Sheng [108]. It differs only in the directions used in the predictor and the corrector system.

Let $\Gamma(\gamma)$ denote the set of points $p = (X, y, Z)$ for which the primal and dual residuums are uniformly bounded by some multiple of μ, i.e.,

$$\Gamma(\gamma) = \left\{ p \in S(n)^{++} \times I\!R^m \times S(n)^{++} : \|r_P\| \le \frac{\|r_P^\circ\|}{\mu^\circ}\gamma\mu, \|r_D\| \le \frac{\|r_D^\circ\|}{\mu^\circ}\gamma\mu, \gamma \ge 1 \right\}.$$

We will show that the updated iterates produced by our algorithm stay in $\Gamma(\gamma)$ for a conveniently chosen step size parameter α at each iteration. The prove of this result is based on the following quite natural condition. For $(X, y, Z) \in \Gamma(\gamma)$ and the solution $(\Delta\hat{X}, \Delta y, \Delta Z)$ of (1.9) belonging to it we assume that after taking the symmetric part of $\Delta\hat{X}$

$$-(1-\sigma)\alpha\mu + \frac{\alpha^2\langle \Delta Z, \Delta X\rangle}{n} \le 0. \tag{2.17}$$

Note that this condition is always satisfied for feasible iterates. This follows from (2.1) and $\sigma \in (0, 1)$. We will see later that in case of infeasible iterates this condition guarantees that $\mu^{k+1} \le \mu^k$ for every $k = 0, 1, \ldots$.

For simplicity we somewhere omit the iteration counter. We then write (X, y, Z) to denote the current iteration point and add the superscript "+" to denote the updated variables. We will write $\|X\|_2$ to denote the spectral norm, i.e., $\|X\|_2 = (\lambda_{\max}(A^TA))^{1/2}$. (Note, $\|X\|_2 = \lambda_{\max}(A)$ if $A \in S(n)^+$.) We need the following technical results.

Lemma 2.16 *Assume that* $p = (X, y, Z) \in S(n)^{++} \times I\!R^m \times S(n)^{++}$ *and suppose* $(\Delta\hat{X}, \Delta y, \Delta Z)$ *is a solution of the system (1.9) of equations. Let* $\Delta X = (\Delta\hat{X} + \Delta\hat{X}^T)/2$, $\alpha \in [0, 1]$ *and let* $(X^+, y^+, Z^+) = (X, y, Z) + \alpha(\Delta X, \Delta y, \Delta Z)$. *Then*

$$\begin{aligned} \boldsymbol{\mathcal{A}}(X^+) - b &= (1-\alpha)(\boldsymbol{\mathcal{A}}(X) - b) \\ Z^+ + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^+) - C &= (1-\alpha)(Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - C). \end{aligned}$$

Proof.

$$\begin{aligned} \boldsymbol{\mathcal{A}}(X^+) - b &= \boldsymbol{\mathcal{A}}(X + \alpha\Delta X) - b \\ &= \boldsymbol{\mathcal{A}}(X) - b + \alpha\boldsymbol{\mathcal{A}}(\Delta X) \\ &= (1-\alpha)(\boldsymbol{\mathcal{A}}(X) - b) \quad \text{(by (1.9))}. \end{aligned}$$

$$\begin{aligned} Z^+ + \mathcal{A}^t(y^+) - C &= (Z + \alpha\Delta Z) + \mathcal{A}^t(y + \alpha\Delta y) - C \\ &= Z + \mathcal{A}^t(y) - C + \alpha(\Delta Z + \mathcal{A}^t(\Delta y)) \\ &= (1-\alpha)(Z + \mathcal{A}^t(y) - C) \quad \text{(by (1.9))}. \end{aligned}$$

□

Let $\nu_k = \prod_{j=0}^{k-1}(1-\alpha_j)$, ($\nu_0 = 1$). Then an immediate consequence of Lemma 2.16 is that all iterates (X^k, y^k, Z^k) generated by the central trajectory following IPM satisfy

$$\mathcal{A}(X^k) - b = \nu_k(\mathcal{A}(X^\circ) - b) \quad \text{and} \quad Z^k + \mathcal{A}^t(y^k) - C = \nu_k(Z^\circ + \mathcal{A}^t(y^\circ) - C). \tag{2.18}$$

We next show that the primal-dual IPM given in Table 1.1 generates a sequence $((X^k, y^k, Z^k))_{k\in\mathbb{N}} \in S(n)^{++} \times \mathbb{R}^m \times S(n)^{++}$. Lemma 1.17 is used to specify a valid step size α.

Let $(X^k, y^k, Z^k) \in \Gamma(\gamma)$. Let $\xi_{\min}$ be the minimum of all eigenvalues of $Z^{-1}\Delta Z$ and $X^{-1}\Delta X$. Then applying Lemma 1.17

$$\bar{\alpha} = \begin{cases} \min\{-1/\xi_{\min}, 1\} & if\ \xi_{\min} < 0 \\ 1 & otherwise \end{cases}$$

gives the upper bound for a step size $\alpha^k \in [0, 1)$ such that

$$(X^{k+1}, y^{k+1}, Z^{k+1}) = (X, y, Z) + \alpha^k(\Delta X, \Delta y, \Delta Z) \in S(n)^{++} \times \mathbb{R}^m \times S(n)^{++}.$$

We take the largest value $\alpha^k \in [0, \bar{\alpha})$ for which

$$\mu^{k+1} \geq \frac{\mu^\circ}{\gamma}\nu_{k+1}. \tag{2.19}$$

Since $(X^k, y^k, Z^k) \in \Gamma(\gamma)$ wee see, using (2.18), that $\nu_k \leq (\gamma/\mu^\circ)\mu^k$. Hence $\alpha^k \in [0, \bar{\alpha})$ satisfying (2.19) does always exist. Note that

$$\begin{aligned} n\mu^+ &= \langle Z^+, X^+\rangle \\ &= \langle Z + \alpha\Delta Z, X + \alpha\Delta X\rangle \\ &= \langle Z, X\rangle + \alpha\Big(\langle \Delta Z, X\rangle + \langle Z, \Delta X\rangle\Big) + \alpha^2\langle \Delta Z, \Delta X\rangle \\ &= n\mu + \alpha\sigma\mu n - \alpha n\mu + \alpha^2\langle \Delta Z, \Delta X\rangle \quad \text{(by (1.10) and (1.9))}. \end{aligned}$$

Note, in the last line we have implicitely used that μ is reduced at each iteration by a constant factor $\sigma \in (0, 1)$. Hence for every $k = 0, 1, \ldots$

$$\mu^{k+1} = \mu^k - (1-\sigma^k)\alpha^k\mu^k + \frac{(\alpha^k)^2\langle \Delta Z^k, \Delta X^k\rangle}{n} \leq \mu^k \text{ by (2.17)}. \tag{2.20}$$

We also see that

$$\|r_P^{k+1}\| = (1-\alpha^k)\|r_P^k\| \le \|r_P^k\| \le \frac{\|r_P^\circ\|}{\mu^\circ}\gamma\mu^k$$

and similarly for the dual residuum. Here the equality follows from Lemma 2.16, the first inequality from $\alpha^k \in [0,1)$ and the second inequality holds because $(X^k, y^k, Z^k) \in \Gamma(\gamma)$. It also follows that

$$\frac{\|r_P^{k+1}\|\mu^\circ}{\|r_P^\circ\|\gamma} \overset{\text{by (2.18)}}{=} \frac{\mu^\circ}{\gamma}\nu_{k+1} \overset{\text{by (2.19)}}{\le} \mu^{k+1} \overset{\text{by (2.20)}}{\le} \mu^k.$$

Thus we have shown that $(X^{k+1}, y^{k+1}, Z^{k+1}) \in \Gamma(\gamma)$. We select $\sigma^{k+1} \in (0,1)$ such that

$$\frac{\mu^\circ}{\gamma}\nu_{k+1} \le \sigma^{k+1}\mu^{k+1} \le \mu^k.$$

This is always possible except $\alpha^k = 0$ for some $k \in \mathbb{N}$. Under the assumtion that both the primal and the dual problem satisfy the Slater constraint qualification this happens only if $\mu^k = 0$. Our IPM is a method minimizing the duality gap, see section 1.4. In other words for given $\varepsilon > 0$ our algorithm terminates with an approximate solution $(X^{\bar{k}}, y^{\bar{k}}, Z^{\bar{k}}) \in S(n)^{++} \times \mathbb{R}^m \times S(n)^{++}$ for which $\|\boldsymbol{\mathcal{A}}(X^{\bar{k}}) - b\| \le \varepsilon$ and $\|Z^{\bar{k}} + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^{\bar{k}}) - C\| \le \varepsilon$ and $\langle Z^{\bar{k}}, X^{\bar{k}}\rangle \le \varepsilon$. Since μ^k decreases to zero we have $\mu^k > \varepsilon/n$ for every $k < \bar{k}$.

If the iterates are updated in the proposed way we have

$$\nu_k \le \gamma\frac{\mu^k}{\mu^\circ} \text{ for every } k = 0, 1, \ldots. \tag{2.21}$$

Lemma 2.17 *Suppose that $p = (X, y, Z) \in S(n) \times \mathbb{R}^m \times S(n)$ satisfies the conditions $\boldsymbol{\mathcal{A}}(X) = 0$ and $\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) + Z = 0$. Then $\langle Z, X\rangle = 0$.*

Proof.

$$\langle Z, X\rangle = \langle -\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y), X\rangle = \langle y, \boldsymbol{\mathcal{A}}(X)\rangle = 0.$$

□

The following result is a generalization from LP to SDP, see [131].

Lemma 2.18 *Suppose that $(X^\circ, y^\circ, Z^\circ) \in S(n)^{++} \times \mathbb{R}^m \times S(n)^{++}$ and $(X^k, y^k, Z^k) \in \Gamma(\gamma)$ for every $k = 1, 2, \ldots$. Let $\nu_k = \prod_{j=0}^{k-1}(1-\alpha_j)$, $(\nu_0 = 1)$. Then there is a positive constant C such that*

$$\nu_k \max\left\{\|Z^k\|, \|X^k\|\right\} \le C\mu^k$$

Proof. Suppose that (X^*, y^*, Z^*) denotes any optimal solution. That is

$$\boldsymbol{\mathcal{A}}(X^*) = b,\ Z^* + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*) = C,\ Z^*X^* = 0,\ Z^* \in S(n)^+,\ X^* \in S(n)^+,\ y^* \in \mathbb{R}^m.$$

Let (X, y, Z) be defined as

$$(X, y, Z) = \nu_k(X^\circ, y^\circ, Z^\circ) + (1 - \nu_k)(X^*, y^*, Z^*) - (X^k, y^k, Z^k).$$

We see that

$$\begin{aligned} \boldsymbol{\mathcal{A}}(X) &= \nu_k \boldsymbol{\mathcal{A}}(X^\circ) + (1 - \nu_k)\boldsymbol{\mathcal{A}}(X^*) - \boldsymbol{\mathcal{A}}(X^k) \\ &= \nu_k(\boldsymbol{\mathcal{A}}(X^\circ) - b) + (b - \boldsymbol{\mathcal{A}}(X^k)) \qquad (\text{since } \boldsymbol{\mathcal{A}}(X^*) = b) \\ &= 0 \qquad (\text{by } 2.18) \end{aligned}$$

and similarly

$$\begin{aligned} Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) &= \nu_k Z^\circ + (1 - \nu_k)Z^* - Z^k + \nu_k \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^\circ) + (1 - \nu_k)\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*) - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^k) \\ &= \nu_k(Z^\circ + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^\circ) - C) + (-Z^k - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^k) + C) \\ &\quad\ (\text{since } Z^* + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^*) = C) \\ &= 0 \qquad (\text{by } 2.18). \end{aligned}$$

Thus by Lemma 2.17 we get

$$\begin{aligned} 0 &= \langle Z, X \rangle \\ &= \Big\langle \nu_k Z^\circ + (1 - \nu_k)Z^* - Z^k, \nu_k X^\circ + (1 - \nu_k)X^* - X^k \Big\rangle \\ &= \nu_k^2 \langle Z^\circ, X^\circ \rangle + (1 - \nu_k)^2 \langle Z^*, X^* \rangle + \nu_k(1 - \nu_k)\Big(\langle Z^*, X^\circ \rangle + \langle Z^\circ, X^* \rangle\Big) \\ &\quad + \langle Z^k, X^k \rangle - \nu_k\Big(\langle Z^\circ, X^k \rangle + \langle Z^k, X^\circ \rangle\Big) - (1 - \nu_k)\Big(\langle Z^k, X^* \rangle + \langle Z^*, X^k \rangle\Big) \end{aligned}$$

Since $(Z^k, X^k) \in (S(n)^{++})^2$ and $(Z^*, X^*) \in (S(n)^+)^2$ we have $\langle Z^k, X^* \rangle + \langle Z^*, X^k \rangle \geq 0$. By definition of (X^*, y^*, Z^*) we have $\langle Z^*, X^* \rangle = 0$. Hence we get

$$\begin{aligned} &\nu_k\Big(\langle Z^\circ, X^k \rangle + \langle Z^k, X^\circ \rangle\Big) \\ &\quad \leq \nu_k^2 \langle Z^\circ, X^\circ \rangle + \langle Z^k, X^k \rangle + \nu_k(1 - \nu_k)\Big(\langle Z^*, X^\circ \rangle + \langle Z^\circ, X^* \rangle\Big) \end{aligned} \tag{2.22}$$

Denote the minimum eigenvalue of the initial matrices Z° and X° by $\lambda_{\min}(Z^\circ)$ and $\lambda_{\min}(X^\circ)$, respectively. Let $\xi = \min\Big\{\lambda_{\min}(Z^\circ), \lambda_{\min}(X^\circ)\Big\}$. Note that $\xi > 0$. Now we see that

$$\begin{aligned} \xi \max\Big\{\|Z^k\|, \|X^k\|\Big\} &\leq \xi\Big(\|X^k\| + \|Z^k\|\Big) \qquad\qquad (2.23) \\ &\leq \lambda_{\min}(Z^\circ)\|X^k\| + \lambda_{\min}(X^\circ)\|Z^k\| \\ &\leq \lambda_{\min}(Z^\circ)\mathrm{trace}(X^k) + \lambda_{\min}(X^\circ)\mathrm{trace}(Z^k) \quad (\text{by Lemma A.2}) \\ &\leq \langle Z^\circ, X^k \rangle + \langle Z^k, X^\circ \rangle \quad (\text{by Lemma A.4}) \end{aligned}$$

We also get that

$$\begin{aligned}\langle Z^\circ, X^*\rangle &\leq n\lambda_{\max}(Z^\circ)\lambda_{\max}(X^*) \quad \text{(by applying Lemma A.4 twice times)}\\ &= n\lambda_{\max}(Z^\circ)\|X^*\|_2\\ &\leq n\lambda_{\max}(Z^\circ)\|X^*\|\end{aligned}$$

Then in view of (2.22) and (1.10) (using the two estimates just derived) we obtain

$$\begin{aligned}&\xi\nu_k \max\left\{\|Z^k\|, \|X^k\|\right\}\\ &\leq \nu_k^2 n\mu^\circ + n\mu^k + \nu_k(1-\nu_k)n\Big(\lambda_{\max}(X^\circ)\|Z^*\| + \lambda_{\max}(Z^\circ)\|X^*\|\Big)\\ &\leq \nu_k n\mu^\circ + n\mu^k + \nu_k n\Big(\lambda_{\max}(X^\circ)\|Z^*\| + \lambda_{\max}(Z^\circ)\|X^*\|\Big) \quad (\text{since } \nu_k \in (0,1))\\ &\leq \nu_k n\mu^\circ + n\mu^k + \nu_k n \max\left\{\lambda_{\max}(Z^\circ), \lambda_{\max}(X^\circ)\right\}\Big(\|X^*\| + \|Z^*\|\Big)\\ &\leq \nu_k n\mu^\circ + n\mu^k + 2\nu_k n \max\left\{\lambda_{\max}(Z^\circ), \lambda_{\max}(X^\circ)\right\} \max\left\{\|X^*\|, \|Z^*\|\right\}\\ &\leq \left(\gamma + 1 + \frac{2\gamma}{\mu^\circ}\max\left\{\lambda_{\max}(Z^\circ), \lambda_{\max}(X^\circ)\right\}\max\left\{\|X^*\|, \|Z^*\|\right\}\right)n\mu^k\\ &\quad \text{(by (2.21))}\end{aligned}$$

Hence defining C as

$$C = \frac{n}{\xi}\left(\gamma + 1 + \frac{2\gamma}{\mu^\circ}\max\left\{\lambda_{\max}(Z^\circ), \lambda_{\max}(X^\circ)\right\}\max\left\{\|X^*\|, \|Z^*\|\right\}\right)$$

completes the proof.

□

The following corollary is an immediate consequence of Lemma 2.18. It is based on a special chosen initial point for the algorithm. We take an initial point $(X^\circ, y^\circ, Z^\circ) = (\zeta I, 0, \zeta I), \zeta > 0$ and assume that the iterates produced by our algorithm converges to an approximate solution (X^*, y^*, Z^*) such that

$$\|Z^*\| \leq \zeta < \infty \quad \text{and} \quad \|X^*\| \leq \zeta < \infty \tag{2.24}$$

Then we can formulate

Corollary 2.19 *Suppose that* $(X^\circ, y^\circ, Z^\circ) = (\zeta I, 0, \zeta I), \zeta > 0$ *and* $(X^k, y^k, Z^k) \in \Gamma(\gamma)$ *for every* $k = 1, 2, \ldots$. *Let* $\nu_k = \prod_{j=0}^{k-1}(1-\alpha_j)$, $(\nu_0 = 1)$. *Then*

$$\zeta\nu_k \max\left\{\|Z^k\|, \|X^k\|\right\} \leq 4n\gamma\mu^k.$$

Proof. In the proof of Lemma 2.18 we defined $\xi = \min\left\{\lambda_{\min}(Z^\circ), \lambda_{\min}(X^\circ)\right\}$. Hence $\zeta = \xi$ and $\mu^\circ = \zeta^2$. In this special situation the constant C from Lemma 2.18 is

$$C = \frac{n}{\zeta}\left(\gamma + 1 + \frac{2\gamma}{\zeta}\max\left\{\|X^*\|, \|Z^*\|\right\}\right) \leq \frac{n}{\zeta}(\gamma + 1 + 2\gamma) \text{ by } (2.24)$$

Since $\gamma \geq 1$ the proof is complete.

□

The following lemma yields a condition which can be used as certificate if a given problem does have an approximate solution in a given bounded set.

Lemma 2.20 *Suppose that* $(X^\circ, y^\circ, Z^\circ) = (\zeta I, 0, \zeta I), \zeta > 0$ *and* $(X^k, y^k, Z^k) \in \Gamma(\gamma)$ *for every* $k = 1, 2, \ldots$. *Let* $\nu_k = \prod_{j=0}^{k-1}(1 - \alpha_j)$, $(\nu_0 = 1)$. *Let* $j \in I\!N$ *such that*

(i) $\|Z^k\| + \|X^k\| > \frac{4n\gamma\mu^k}{\zeta\nu_k}$, $\quad k = j$ *and* (ii) $\|Z^k\| + \|X^k\| \leq \frac{4n\gamma\mu^k}{\zeta\nu_k}$, $\quad k < j$.

Then there exists no approximate solution in the set Π *defined by*

$$\Pi = \left\{(X, y, Z) \in S(n) \times I\!R^m \times S(n) : \frac{\zeta}{n}I \succeq X \succeq 0, \frac{\zeta}{n}I \succeq Z \succeq 0\right\}$$

Proof. Suppose $((Z^k, y^k, X^k))_{k\in I\!N} \in \Gamma(\gamma)$ converges to an approximate solution $(Z^*, y^*, X^*) \in \Pi$. Hence $\frac{\zeta}{n}I - X^* \in S(n)^+$. By definition of positive semidefiniteness we have $\zeta/n \geq X^*_{ii}$ for every $i = 1, \ldots, n$ and therefore

$$\zeta \geq \operatorname{trace}(X^*) \geq \|X^*\|.$$

Here the second inequality follows from Lemma A.2. Analogous considerations lead to $\|Z^*\| \leq \zeta$. Hence in view of (2.23) Corollary 2.19 implies that

$$\|X^k\| + \|Z^k\| \leq \frac{4n\tau\mu^k}{\zeta\nu_k}$$

for every $k = 0, 1, \ldots$, which is a contradiction.

□

Part Two –

Algorithms

Chapter 3

Linear-Quadratic Functionals in Semidefinite Programming

In this chapter we consider the problem of minimizing a linear-quadratic functional subject to finitely many linear equality constraints over the cone of positive semidefinite matrices $S(n)^+$. This problem is also considered in [30]. In this paper two versions of primal path-following algorithms for solving this type of problems were analyzed. In particular, Faybusovich extends a version of a path-following algorithm for a standard convex quadratic programming problem given in [55] to SDP. Moreover he established complexity estimates similar to the best known for the case of the standard convex quadratic programming problem. We mention that he imposed that the set of feasible points is compact and has nonempty interior. We propose a primal-dual path-following algorithm, show that it generates iterates which remain in a compact set and establish the global convergence of the algorithm. The reader is referred to the work of Faybusovich and the references quoted therein to see the numerous applications of the linear-quadratic semidefinite programming problem.

3.1 The Problem

In this section we introduce the semidefinite problem considered in this chapter. We deal with the positive semidefinite linear-quadratic problem of the following form. For given matrices $C, A_1, \ldots, A_m \in S(n)$ and $b \in \mathbb{R}^m$, solve

$$
\begin{aligned}
(\mathbf{LQ-PSDP}) \qquad & \min\ \langle C, X\rangle + \frac{1}{2}\langle \mathcal{Q}(X), X\rangle \\
& \text{s.t. } \mathcal{A}(X) = b \\
& \quad X \in S(n)^+ .
\end{aligned}
\tag{3.1}
$$

Here **(LQ-PSDP)** denotes the primal problem and $\mathcal{A} : S(n) \mapsto \mathbb{R}^m$ denotes a linear operator written in full as

$$
\mathcal{A}(X) = \begin{pmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{pmatrix}.
$$

We will denote the associated adjoint operator to $\mathcal{A}$ by $\mathcal{A}^t : \mathbb{R}^m \mapsto S(n)$. This operator defined by the adjoint relation

$$
\langle \mathcal{A}(X), y\rangle = \langle X, \mathcal{A}^t(y)\rangle, \text{ for all } X \in S(n),\ y \in \mathbb{R}^m \tag{3.2}
$$

is given by $\mathcal{A}^t(y) = \sum_{i=1}^{m} y_i A_i$. Throughout this chapter we will assume that the set $\{A_i\}, (i = 1, \ldots, m)$ is linearly independent over the field of $\mathbb{R}$.

$\mathcal{Q} : S(n) \mapsto S(n)$ denotes a symmetric positive semidefinite linear operator. In particular,

$$
\langle \mathcal{Q}(A), B\rangle = \langle A, \mathcal{Q}(B)\rangle,\ \langle \mathcal{Q}(A), A\rangle \geq 0 \text{ for any } A, B \in S(n). \tag{3.3}
$$

We now derive the Lagrangian dual associated to **(LQ-PSDP)**. Let $y \in \mathbb{R}^m$ ($Z \in S(n)^+$) denote the Lagrangian multiplier for $\mathcal{A}(X) = b$ ($X \in S(n)^+$), respectively then

$$
\begin{aligned}
&\min\langle C, X\rangle + \frac{1}{2}\langle \mathcal{Q}(X), X\rangle \text{ s.t. } \mathcal{A}(X) = b, X \in S(n)^+ \\
&= \min_{X\in S(n)} \max_{\substack{y\in\mathbb{R}^m \\ Z\in S(n)^+}} \langle C, X\rangle + \frac{1}{2}\langle \mathcal{Q}(X), X\rangle - \langle \mathcal{A}(X) - b, y\rangle - \langle Z, X\rangle \\
&\geq \max_{\substack{y\in\mathbb{R}^m \\ Z\in S(n)^+}} \min_{X\in S(n)} \langle b, y\rangle + \left\langle C + \frac{1}{2}\mathcal{Q}(X) - \mathcal{A}^t(y) - Z, X\right\rangle \\
&= \max\langle b, y\rangle - \frac{1}{2}\langle \mathcal{Q}(X), X\rangle \text{ s.t. } Z = C + \mathcal{Q}(X) - \mathcal{A}^t(y) \in S(n)^+, \\
&\qquad\qquad\qquad\qquad\qquad\qquad X \in S(n),\ y \in \mathbb{R}^m .
\end{aligned}
$$

The first and the last equation are in need of some explanation. Suppose that $X \notin S(n)^+$. Hence this matrix has at least one negative eigenvalue. We will denote it by λ. Let v denote its associated eigenvector. Then $u^T vv^T u = (v^T u)^2 \geq 0$ for every $u \in \mathbb{R}^n$ or equivalently $vv^T \in S(n)^+$. We also see that $\langle vv^T, X\rangle = v^T X v = v^T \lambda v = \lambda \|v\|^2 < 0$. Hence the inner maximization is bounded only if $X \in S(n)^+$. Since X is unconstrained it follows that $\langle C + \frac{1}{2}\mathcal{Q}(X) - \mathcal{A}^t(y) - Z, X\rangle$ has a minimum if and only if $Z = C + \mathcal{Q}(X) - \mathcal{A}^t(y)$. This explains the last equation.

Therefore the dual **(LQ-DSDP)** is given by

$$
\textbf{(LQ} - \textbf{DSDP)} \qquad \begin{array}{ll} \max & \langle b, y\rangle - \frac{1}{2}\langle \mathcal{Q}(X), X\rangle \\ \text{s.t.} & Z - \mathcal{Q}(X) + \mathcal{A}^t(y) = C \\ & Z \in S(n)^+, \; X \in S(n), \; y \in \mathbb{R}^m. \end{array} \tag{3.4}
$$

3.1.1 Notation

This section provides some basic definitions and notation which will be used throughout. The Kronecker product maps two matrices $A = (a_{ij})$ and $B = (b_{rs})$ into a matrix $C = (a_{ij} b_{rs})$. Let A be a matrix of order $m \times n$ and B be a matrix of order $p \times q$. Then the Kronecker product $A \otimes B$ is defined as

$$
A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}.
$$

Thus $A \otimes B$ is a matrix of order $mp \times nq$. A smart way adapted for specifying a column or row of a matrix is the following notation. If $A \in \mathbb{R}^{m \times n}$, then we denote the n columns and the m rows of A by $A_{.1}, \dots, A_{.n}$ and $A_{1.}, \dots, A_{m.}$, respectively, so that $A_{.j} = (a_{1j}, a_{2j}, \dots, a_{mj})^T$ and $A_{i.} = (a_{i1}, a_{i2}, \dots, a_{in})$.

The vec operator $\mathrm{vec}(\cdot)$, see Neudecker [100], transforms a matrix into a vector by stacking its columns one underneath the other. For $S \in M(n)$ the vector $s = \mathrm{vec}(S) \in \mathbb{R}^{n^2}$, $(s = (S_{.1}^T, S_{.2}^T, \dots, S_{.n}^T)^T)$. Thus $\mathrm{vec} : M(n) \mapsto R^{n^2}$. The inverse of the vec operator is the matrix operator $\mathrm{Mat} : \mathbb{R}^{n^2} \mapsto M(n)$, i.e., $\mathrm{Mat}(\mathrm{vec}(S)) = S$.

The following properties from Kronecker products are used in this thesis, see Graham [39]. If AC and BD exist, then

$$
(A \otimes B)(C \otimes D) = AC \otimes BD. \tag{3.5}
$$

If α is a scalar, then

$$
\begin{array}{l} \alpha \otimes A = \alpha A = A\alpha = A \otimes \alpha \\ A \otimes (\alpha B) = \alpha(A \otimes B). \end{array} \tag{3.6}
$$

The transpose of a Kronecker product is

$$
(A \otimes B)^T = A^T \otimes B^T. \tag{3.7}
$$

Let A, B and C be three matrices such that the matrix product ABC is defined. Then,

$$\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B). \tag{3.8}$$

Let $A \in M(m)$ with eigenvalues $\lambda_i, (i = 1, \dots, m)$ and corresponding eigenvectors $x_i, (i = 1, \dots, m)$ and $B \in M(n)$ with eigenvalues $\mu_j, (j = 1, \dots, n)$ and corresponding eigenvectors $y_j, (j = 1, \dots, n)$, respectively. Then,

$$A \otimes B \text{ has eigenvalues } \lambda_i \mu_j \text{ with corresponding eigenvectors } x_i \otimes y_j. \quad (i = 1, \dots, m; j = 1, \dots, n) \tag{3.9}$$

As a direct consequence of (3.9) we get (3.10).

$$\text{If } (A, B) \in (S(n)^+)^2 \ \left((A, B) \in (S(n)^{++})^2\right) \text{ then } A \otimes B \in S(n^2)^+ \ \left(S(n^2)^{++}\right), \text{ respectively.} \tag{3.10}$$

We will denote the primal and dual feasible sets by $\mathcal{F}(\mathcal{LQ} - \mathcal{P})$ and $\mathcal{F}(\mathcal{LQ} - \mathcal{D})$, respectively. For primal and dual strictly feasible sets we write $\mathrm{int}(\mathcal{F}(\mathcal{LQ} - \mathcal{P}))$ and $\mathrm{int}(\mathcal{F}(\mathcal{LQ} - \mathcal{D}))$, respectively.

3.2 A Central Trajectory Following Method

In this section we give a detailed description of an infeasible primal-dual IPM that solves **(LQ-PSDP)** and **(LQ-DSDP)** simultaneously. We make the assumption that both, **(LQ-PSDP)** and **(LQ-DSDP)**, satisfy the Slater constraint qualification. Hence strong duality holds, see section 1.3 Theorem 1.14. For $\mu > 0$ we consider the barrier problem (3.11), associated to the primal problem.

$$\min \langle C, X \rangle + \frac{1}{2}\langle \boldsymbol{\mathcal{Q}}(X), X \rangle - \mu \log\det(X) \text{ s.t. } \boldsymbol{\mathcal{A}}(X) = b \tag{3.11}$$

We now take the Lagrangian of this problem. Let $y \in \mathbb{R}^m$ be the Lagrange multiplier for $\boldsymbol{\mathcal{A}}(X) = b$. Then the Lagrangian $\mathcal{L}_\mu(X, y)$ is given by

$$\mathcal{L}_\mu(X, y) = \langle C, X \rangle + \frac{1}{2}\langle \boldsymbol{\mathcal{Q}}(X), X \rangle - \mu \log\det(X) + \langle y, b - \boldsymbol{\mathcal{A}}(X) \rangle.$$

For fixed $\mu > 0$ the minimizers for (3.11) satisfy the necessary and sufficient first order optimality conditions. In fact,

$$\begin{aligned}
\frac{\partial}{\partial X}\mathcal{L}_\mu(X, y) &= C + \frac{1}{2}\boldsymbol{\mathcal{Q}}(X) + \frac{1}{2}\boldsymbol{\mathcal{Q}}(X) - \mu X^{-1} - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = 0 \\
\frac{\partial}{\partial y}\mathcal{L}_\mu(X, y) &= b - \boldsymbol{\mathcal{A}}(X) = 0 \\
X \in S(n)^{++}&, \ y \in \mathbb{R}^m.
\end{aligned}$$

We set $Z = \mu X^{-1}$ which is equivalent to $ZX = \mu I$. Then another way of writing the optimality conditions is

$$\textbf{(KKT)} \qquad \begin{array}{l} \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - \boldsymbol{\mathcal{Q}}(X) + Z = C \\ \boldsymbol{\mathcal{A}}(X) = b \\ ZX = \mu I \\ X \in S(n)^{++},\ Z \in S(n)^{++},\ y \in \mathbb{R}^m. \end{array}$$

For any $\mu > 0$ there is a unique minimizer to **(KKT)**, see [30]. The central trajectory is defined precisely as in section 1.4. The μ values corresponding to points on this trajectory are again calculated using (1.10).

Let $p = (X, y, Z) \in S(n) \times \mathbb{R}^m \times S(n)$ and let $F_\mu(p) : S(n) \times \mathbb{R}^m \times S(n) \mapsto S(n) \times \mathbb{R}^m \times M(n)$ be defined as

$$F_\mu(p) := \begin{pmatrix} \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - \boldsymbol{\mathcal{Q}}(X) + Z - C \\ \boldsymbol{\mathcal{A}}(X) - b \\ ZX - \mu I \end{pmatrix}.$$

Again, as in section 1.4.1, Newton's method is used for solving the problem $F_\mu(p) = 0$. Clearly, $ZX - \mu I = 0$ is nonlinear. Therefore we obtain the step direction $\Delta p = (\Delta X, \Delta y, \Delta Z)$ as solution of the linearized system (3.12).

$$\begin{aligned} \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - \boldsymbol{\mathcal{Q}}(\Delta X) + \Delta Z &= -r_D \\ \boldsymbol{\mathcal{A}}(\Delta X) &= -r_P \\ \Delta Z\, X + Z\, \Delta X &= -r_{pC} \end{aligned} \tag{3.12}$$

where $r_P = \boldsymbol{\mathcal{A}}(X) - b$ and $r_D = \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - \boldsymbol{\mathcal{Q}}(X) + Z - C$ stand for the primal and dual residuum, respectively and $r_{pC} = ZX - \mu I$ denotes the perturbed complementarity condition.

From the third equation in (3.12) we get

$$\Delta X = \mu Z^{-1} - X - Z^{-1} \Delta Z X. \tag{3.13}$$

Then eliminating ΔZ from the first equation in (3.12) and substituting it into (3.13) yields

$$\Delta X = \mu Z^{-1} - X - Z^{-1}\Big(\boldsymbol{\mathcal{Q}}(\Delta X) - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - r_D\Big)X \tag{3.14}$$

or equivalently

$$\mathrm{vec}(\Delta X) + (X \otimes Z^{-1})\mathrm{vec}(\boldsymbol{\mathcal{Q}}(\Delta X)) = \mathrm{vec}\left(\mu Z^{-1} - X + Z^{-1}\left(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) + r_D\right)X\right).$$

Here we have used (3.8). Since $\boldsymbol{\mathcal{Q}}$ is a symmetric positive semidefinite finite dimensional linear operator there is a representation of it given by a positive semidefinite matrix $\bar{Q} \in S(n^2)$ (depends on the basis) such that

$$\boldsymbol{\mathcal{Q}}(\Delta X) = \mathrm{Mat}(\bar{Q}\, \mathrm{vec}(\Delta X)).$$

Hence

$$\big(I + (X \otimes Z^{-1})\bar{Q}\big)\,\mathrm{vec}(\Delta X) = \mathrm{vec}\left(\mu Z^{-1} - X + Z^{-1}\left(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) + r_D\right) X\right).$$

Since $(X, Z^{-1}) \in (S(n)^{++})^2$ it follows from (3.10) that $X \otimes Z^{-1} \in S(n)^{++}$. Moreover, because $\bar{Q} \in S(n^2)^+$ we have that $\big(I + (X \otimes Z^{-1})\bar{Q}\big)$ is nonsingular. Consequently, $(I + (X \otimes Z^{-1})\bar{Q})^{-1}$ exists and we may write

$$\mathrm{vec}(\Delta X) = \big(I + (X \otimes Z^{-1})\bar{Q}\big)^{-1}\mathrm{vec}\left(\mu Z^{-1} - X + Z^{-1}\left(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) + r_D\right) X\right). \tag{3.15}$$

The second equation in (3.12) can be written equivalently as $\boldsymbol{\mathcal{A}}(\mathrm{Mat}(\mathrm{vec}(\Delta X))) = -r_P$ and therefore

$$\begin{aligned}&\boldsymbol{\mathcal{A}}\left\{\mathrm{Mat}\left(\big(I + (X \otimes Z^{-1})\bar{Q}\big)^{-1}\mathrm{vec}\left(Z^{-1}\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) X\right)\right)\right\}\\ &\qquad = \boldsymbol{\mathcal{A}}\left\{\mathrm{Mat}\left(\big(I + (X \otimes Z^{-1})\bar{Q}\big)^{-1}\mathrm{vec}\left(X - \mu Z^{-1} - Z^{-1} r_D X\right)\right)\right\} - r_P.\end{aligned}$$

Let $\boldsymbol{\mathcal{O}}$ be the linear operator defined by

$$\boldsymbol{\mathcal{O}}(\cdot) = \boldsymbol{\mathcal{A}}\left\{\mathrm{Mat}\left(\big(I + (X \otimes Z)^{-1}\bar{Q}\big)^{-1}\mathrm{vec}\left(Z^{-1}\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\cdot) X\right)\right)\right\}.$$

Thus we get

$$\boldsymbol{\mathcal{O}}(\Delta y) = \boldsymbol{\mathcal{A}}\left\{\mathrm{Mat}\left(\big(I + (X \otimes Z^{-1})\bar{Q}\big)^{-1}\mathrm{vec}\left(X - \mu Z^{-1} - Z^{-1} r_D X\right)\right)\right\} - r_P. \tag{3.16}$$

This operator has also a unique matrix $S \in S(m)^+$ associated with it, i.e., $\boldsymbol{\mathcal{O}}(\Delta y) = S\Delta y$, see section 3.2.1 for details.

We now summarize the essential steps of the path-following algorithm for solving (3.12). We first solve (3.16) for Δy. Then substituting Δy into (3.15) gives $\Delta\tilde{X}$. The "$\tilde{\ }$" indicates that $\Delta\tilde{X}$ is not symmetric in general. We take the symmetric part $\Delta X = (\Delta\tilde{X} + \Delta\tilde{X}^T)/2$ and substitute ΔX and Δy into the first equation of (3.12). This enables us to compute ΔZ. In fact,

$$\Delta Z = \boldsymbol{\mathcal{Q}}(\Delta X) - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) - r_D.$$

Then we determine primal and dual step size parameters α_P and α_D using standard line search (backtracking starting from a full Newton step) such that $X + \alpha_P\Delta X$ and $Z + \alpha_D\Delta Z$ are positive definite. We then take the minimum $\alpha = \min\{\alpha_P, \alpha_D\}$. The update is as follows

$$\begin{aligned} X &\leftarrow X + \alpha\Delta X\\ y &\leftarrow y + \alpha\Delta y\\ Z &\leftarrow Z + \alpha\Delta Z\end{aligned}$$

Finally we use (1.10) to compute the corresponding value of μ, choose $\sigma \in (0,1)$ and set $r_{pC} = ZX - \sigma\mu I$. We repeat these steps until (X, y, Z) satisfies primal and dual feasibility and the duality gap is reduced to a value below a prescribed threshold.

We now show that there is an efficient way to compute the search direction $(\Delta X, \Delta y, \Delta Z)$.

3.2.1 Practical Aspects

In this section we study in some detail the implementation of the primal-dual IPM proposed in the section beforehand. In particular, we now emphasize some implementation details which have proven to be very helpful in practice (concerning computational effort) for solving (3.16) to get Δy and (3.14) to get ΔX. It is easy to see that $\boldsymbol{\mathcal{O}}(\Delta y)$ can equivalently be expressed as follows.

$$\boldsymbol{\mathcal{O}}(\Delta y) = \begin{pmatrix} (\mathrm{vec}(A_1))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y)X\right)\right) \\ (\mathrm{vec}(A_2))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y)X\right)\right) \\ \vdots \\ (\mathrm{vec}(A_m))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y)X\right)\right) \end{pmatrix}$$

Now we consider the $i-$th row of this representation. We get

$$\begin{aligned} &(\mathrm{vec}(A_i))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y)X\right)\right) \\ &\quad = (\mathrm{vec}(A_i))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}\sum_{k=1}^{m}(\Delta y_k A_k)X\right)\right) \\ &\quad = \sum_{k=1}^{m}(\mathrm{vec}(A_i))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}(\Delta y_k A_k)X\right)\right) \\ &\quad = \sum_{k=1}^{m}(\mathrm{vec}(A_i))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}A_k X\right)\right)\Delta y_k. \end{aligned}$$

The $j-$th term of this expression is given by

$$(\mathrm{vec}(A_i))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}A_j X\right)\right)\Delta y_j. \tag{3.17}$$

Evidently, the operator $\boldsymbol{\mathcal{O}}$ satisfies $\boldsymbol{\mathcal{O}}(\Delta y) = S\Delta y$ where $S = (s_{ij})$. In fact,

$$s_{ij} = (\mathrm{vec}(A_i))^T \left(\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} \mathrm{vec}\left(Z^{-1}A_j X\right)\right), \text{ by (3.17)}.$$

We shall now simplify the calculation of the entries of the matrix S in the case where $(X \otimes Z^{-1})\bar{Q}$ is symmetric and hence $(X \otimes Z^{-1})\bar{Q} = V\Pi V^T$ with the orthonormal eigenvectors $v_j, (j = 1, \ldots, n^2)$ in V and the eigenvalues $\pi_j, (j = 1, \ldots, n^2)$ in Π. Then

$$(X \otimes Z^{-1})\bar{Q} = \sum_{j=1}^{n^2} \pi_j v_j v_j^T.$$

Now using Lemma A.8, we get

$$\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1} = \sum_{j=1}^{n^2} \frac{1}{1+\pi_j} v_j v_j^T.$$

Hence the $kl-$th element of the matrix S describing the operator $\boldsymbol{\mathcal{O}}(\Delta y)$ is given by

$$s_{kl} = (\text{vec}(A_k))^T \left(\sum_{j=1}^{n^2} \frac{1}{1+\pi_j} v_j v_j^T \text{vec}\left(Z^{-1} A_l X\right) \right).$$

The formula for $\left(I + (X \otimes Z^{-1})\bar{Q}\right)^{-1}$ can be used analogously to simplify the right hand side of (3.16).

The calculation can be further simplified if $\boldsymbol{\mathcal{Q}}$ is the identity operator, i.e., $\boldsymbol{\mathcal{Q}}(\cdot) = I(\cdot)I$. In what follows we consider this special operator. Let $X = P\Lambda P^T$ with the orthonormal eigenvectors $p_i, (i = 1, \ldots, n)$ in P and the eigenvalues $\lambda_i, (i = 1, \ldots, n)$ in Λ and let $Z^{-1} = Q\Sigma Q^T$ with the orthonormal eigenvectors $q_j, (j = 1, \ldots, n)$ in Q and the eigenvalues $\sigma_j, (j = 1, \ldots, n)$ in Σ, respectively. Then from (3.9) we get that

$$(X \otimes Z^{-1}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \sigma_j (p_i \otimes q_j)(p_i \otimes q_j)^T.$$

By Lemma A.8, we get

$$\left(I + (X \otimes Z^{-1})\right)^{-1} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{1+\lambda_i \sigma_j} (p_i \otimes q_j)(p_i \otimes q_j)^T.$$

Hence the $kl-$th element of the matrix S describing the operator $\boldsymbol{\mathcal{O}}(\Delta y)$ is given by

$$s_{kl} = (\text{vec}(A_k))^T \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{1+\lambda_i \sigma_j} (p_i \otimes q_j)(p_i \otimes q_j)^T \text{vec}\left(Z^{-1} A_l X\right) \right).$$

We see that

$$\begin{aligned}
(p_i \otimes q_j)^T \text{vec}\left(Z^{-1} A_l X\right) &= \left((p_i^T \otimes q_j^T)(X \otimes Z^{-1})\right) \text{vec}(A_l) \quad \text{(by (3.7) and (3.8))} \\
&= (p_i^T X \otimes q_j^T Z^{-1}) \text{vec}(A_l) \quad \text{(by (3.5))} \\
&= (\lambda_i p_i^T \otimes \sigma_j q_j^T) \text{vec}(A_l) \\
&= \lambda_i \sigma_j (q_j^T A_l p_i) \quad \text{(by (3.6) and (3.8))}.
\end{aligned}$$

From this we get

$$\begin{aligned} s_{kl} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\lambda_i\sigma_j}{1+\lambda_i\sigma_j}(q_j^T A_l p_i)(\mathrm{vec}(A_k))^T(p_i\otimes q_j)\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\lambda_i\sigma_j}{1+\lambda_i\sigma_j}(q_j^T A_l p_i)(p_i^T\otimes q_j^T)(\mathrm{vec}(A_k))\right) \quad \text{(by (3.7))} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\lambda_i\sigma_j}{1+\lambda_i\sigma_j}(q_j^T A_l p_i)(q_j^T A_k p_i)\right). \quad \text{(by (3.8))} \end{aligned}$$

Let us look now at the right hand side of (3.16), which also can be simplified in this manner using the formula for $\left(I+(X\otimes Z^{-1})\right)^{-1}$, just derived. Here we get for the $k-$th component

$$\begin{aligned} &\boldsymbol{\mathcal{A}}\left\{\mathrm{Mat}\left(\left(I+(X\otimes Z^{-1})\right)^{-1}\mathrm{vec}\left(X-\mu Z^{-1}-Z^{-1}r_D X\right)\right)\right\}_{k.} - (r_P)_{k.} \\ &\quad= (\mathrm{vec}(A_k))^T\left(\left(I+(X\otimes Z^{-1})\right)^{-1}\mathrm{vec}\left(X-\mu Z^{-1}-Z^{-1}r_D X\right)\right) - (r_P)_{k.} \\ &\quad= (\mathrm{vec}(A_k))^T\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{1+\lambda_i\sigma_j}(p_i\otimes q_j)(p_i\otimes q_j)^T\right)\mathrm{vec}\left(X-\mu Z^{-1}-Z^{-1}r_D X\right) \\ &\qquad - (r_P)_{k.} \\ &\quad= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{1+\lambda_i\sigma_j}(p_i^T\otimes q_j^T)(\mathrm{vec}(A_k))(p_i\otimes q_j)^T\right)\mathrm{vec}\left(X-\mu Z^{-1}-Z^{-1}r_D X\right) \\ &\qquad - (r_P)_{k.} \\ &\quad= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{1+\lambda_i\sigma_j}\left(q_j^T(X-\mu Z^{-1}-Z^{-1}r_D X)p_i\right)(q_j^T A_k p_i)\right) - (r_P)_{k.}. \end{aligned}$$

This special operator makes it also simpler to compute ΔX. In particular, multiplication on the left of (3.14) by Z produces

$$Z\Delta X + \Delta X X = \mu I - ZX + \left(\boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) + r_D\right)X. \tag{3.18}$$

This representation is called the Sylvester equation. Denoting the right hand side by R and multiplying (3.18) on the right by P (recall $X = P\Lambda P^T$) yields

$$Z\Delta X P + \Delta X P\Lambda = RP. \tag{3.19}$$

This is an equivalent equation of the same form as the original (3.18), but it is easier to solve this representation. The solution to the original version can be recovered easily from the solution of the transformed equation. Equation (3.19) may be written as

$$((I \otimes Z) + (\Lambda \otimes I))\,\mathrm{vec}(\Delta X P) = \mathrm{vec}(RP).$$

Here we have used (3.8). From this representation it is not difficult to see that we must not solve a linear system of equations with system matrix of dimension $n^2 \times n^2$. It is evident that this matrix has exploitable block structure. It is easy to see that we can solve alternatively $k, (k = 1, \ldots, n)$ linear equation systems of the following form.

$$(Z + \lambda_k I)\,\delta_k = (RP)_{.k} \quad \text{for every} \quad k = 1, \ldots, n.$$

We then see that ΔX is given by

$$\Delta X = (\delta_1, \delta_2, \ldots, \delta_n)P^T.$$

3.3 Convergence Analysis

In this section we will anlyse the convergence of the primal-dual path-following algorithm presented in section 3.2. We follow the style of the convergence proof given in section 2.2.

3.3.1 Descent Direction

In this section we will show that the search direction $\Delta p = (\Delta X, \Delta y, \Delta Z)$ forms a descent direction with respect to a conveniently chosen merit function. Let $p = (X, y, Z)$ and let the merit function f_μ be defined as

$$f_\mu(p) := \langle Z, X\rangle - \mu \log\det(ZX) + \frac{1}{2}||r_P||^2 + \frac{1}{2}||r_D||^2.$$

The merit function is used to decide whether the step Δp makes sufficient progress towards the optimal solution of the problem under consideration. The most interesting property of this function is that the point p where the global minimum $n\mu(1 - \log\mu)$ of it is attained, is the solution of the nonlinear problem $F_\mu(p) = 0$.

Lemma 3.1 *The directional derivative of f_μ in the direction Δp satisfies*

$$\left\langle \frac{\partial f_\mu(p)}{\partial p}, \Delta p \right\rangle \leq 0.$$

Equality holds if and only if $F_\mu(p) = 0$.

Proof. We have to show that

$$\left\langle \frac{\partial f_\mu}{\partial X}, \Delta X \right\rangle + \left\langle \frac{\partial f_\mu}{\partial y}, \Delta y \right\rangle + \left\langle \frac{\partial f_\mu}{\partial Z}, \Delta Z \right\rangle \leq 0.$$

The partial derivatives of f_μ are

$$\begin{aligned}
\frac{\partial f_\mu}{\partial X} &= Z - \mu X^{-1} + \boldsymbol{\mathcal{A}^t}(r_P) - \boldsymbol{\mathcal{Q}}(r_D) \\
\frac{\partial f_\mu}{\partial Z} &= X - \mu Z^{-1} + r_D \\
\frac{\partial f_\mu}{\partial y} &= \boldsymbol{\mathcal{A}}(r_D).
\end{aligned}$$

We see that

$$\begin{aligned}
\left\langle \boldsymbol{\mathcal{A}^t}(r_P), \Delta X \right\rangle &= \left\langle r_P, \boldsymbol{\mathcal{A}}(\Delta X) \right\rangle \\
&= \left\langle r_P, \boldsymbol{\mathcal{A}}(\Delta \tilde{X}) \right\rangle \quad (\text{ since } \boldsymbol{\mathcal{A}}(\Delta X) = \boldsymbol{\mathcal{A}}(\Delta \tilde{X})) \\
&= \left\langle r_P, -r_P \right\rangle \quad (\text{ by } (3.12)) \\
&= -||r_P||^2.
\end{aligned}$$

We also see that

$$\begin{aligned}
\left\langle -\boldsymbol{\mathcal{Q}}(r_D), \Delta X \right\rangle + \left\langle r_D, \Delta Z \right\rangle + \left\langle \boldsymbol{\mathcal{A}}(r_D), \Delta y \right\rangle &= \left\langle r_D, -\boldsymbol{\mathcal{Q}}(\Delta X) \right\rangle + \langle r_D, \Delta Z \rangle \\
&\quad + \left\langle r_D, \boldsymbol{\mathcal{A}^t}(\Delta y) \right\rangle \\
&= \left\langle r_D, \boldsymbol{\mathcal{A}^t}(\Delta y) - \boldsymbol{\mathcal{Q}}(\Delta X) + \Delta Z \right\rangle \\
&= \left\langle r_D, -r_D \right\rangle \quad (\text{ by } (3.12)) \\
&= -||r_D||^2.
\end{aligned}$$

Here the first equation follows from (3.3) and (3.2). Then collecting the remaining terms yields

$$\begin{aligned}
\left\langle Z - \mu X^{-1}, \Delta X \right\rangle + \left\langle X - \mu Z^{-1}, \Delta Z \right\rangle &= \left\langle I - \mu Z^{-1} X^{-1}, Z \Delta \tilde{X} + \Delta Z X \right\rangle \\
&= \left\langle I - \mu Z^{-1} X^{-1}, \mu I - Z X \right\rangle \\
&= -\operatorname{trace}\left(Z X - 2\mu I + \mu^2 (Z X)^{-1} \right) \\
&= -\sum_{i=1}^{n} \left(\lambda_i(Z X) - 2\mu + \mu^2 \lambda_i^{-1}(Z X) \right) \\
&= -\sum_{i=1}^{n} \left(\lambda_i(Z X) \left(1 - \mu \lambda_i^{-1}(Z X) \right)^2 \right).
\end{aligned}$$

Here the first equation is true since $\mathrm{trace}(SX) = \mathrm{trace}(S\tilde{X})$, where $(S, X) \in S(n)^2$ and X is the symmetrized $\tilde{X}$.

From $(Z, X) \in (S(n)^{++})^2$ it follows that the matrix $ZX \in S(n)^{++}$. Moreover, $\left(1 - \mu\lambda_i^{-1}(ZX)\right)^2 \geq 0$ and equality holds if and only if $\mu = \lambda_i(ZX)$. Hence,

$$\sum_{i=1}^{n} \left(\lambda_i(ZX)\left(1 - \mu\lambda_i^{-1}(ZX)\right)^2\right) = 0 \text{ if and only if } ZX = \mu I.$$

Combining all, we get

$$\left\langle \frac{\partial f_\mu(p)}{\partial p}, \Delta p \right\rangle = -||r_P||^2 - ||r_D||^2 - \sum_{i=1}^{n} \left(\lambda_i(ZX)\left(1 - \mu\lambda_i^{-1}(ZX)\right)^2\right) \leq 0.$$

It is now quite straightforward to see that equality holds if and only if $F_\mu(p) = 0$.

□

3.4 The Main Convergence Statement

In this section we first prepare some technical results which are the basis for the main convergence theorem.

Firstly, note that for feasible points $(X', y', Z') \in \mathcal{F}(\mathcal{LQ} - \mathcal{P}) \times \mathcal{F}(\mathcal{LQ} - \mathcal{D})$ and $(X, y, Z) \in \mathcal{F}(\mathcal{LQ} - \mathcal{P}) \times \mathcal{F}(\mathcal{LQ} - \mathcal{D})$ we have

$$\begin{aligned}\langle Z - Z', X - X' \rangle &= \left\langle C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y') + \boldsymbol{\mathcal{Q}}(X') - C + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - \boldsymbol{\mathcal{Q}}(X), X - X' \right\rangle \\ &= \left\langle y - y', \boldsymbol{\mathcal{A}}(X' - X) \right\rangle + \left\langle \boldsymbol{\mathcal{Q}}(X' - X), X - X' \right\rangle \quad \text{(by (3.2)}.\end{aligned}$$

It follows from (3.1) that the first term add up to zero and from (3.3) that the second term is nonnegative.

Hence

$$\langle Z - Z', X - X' \rangle \geq 0. \tag{3.20}$$

Lemma 3.2 *Let* $p^\circ = (X^\circ, y^\circ, Z^\circ) \in \mathrm{int}(\mathcal{F}(\mathcal{LQ} - \mathcal{P}) \times \mathcal{F}(\mathcal{LQ} - \mathcal{D}))$ *and let* $\zeta > 0$, $\zeta \in I\!R$ *such that* $f_\mu(p^\circ) \leq \zeta$. *Then: The set*

$$\Gamma = \left\{ p \in \mathrm{int}(\mathcal{F}(\mathcal{LQ} - \mathcal{P}) \times \mathcal{F}(\mathcal{LQ} - \mathcal{D})) : f_\mu(p) \leq \zeta \right\}$$

is contained in a bounded and closed subset of $S(n) \times I\!R^m \times S(n)$.

Proof. We already have shown in Lemma 2.3 that for $p \in \Gamma$

$$\langle Z, X \rangle \leq n\gamma \quad \text{where} \quad \gamma = 2\Big(\zeta - n(\mu - \mu \log \mu) + \mu \log 2\Big).$$

It follows from (3.20) that

$$\langle Z - Z^\circ, X - X^\circ \rangle \geq 0.$$

This gives

$$\langle Z, X \rangle + \langle Z^\circ, X^\circ \rangle \geq \langle Z, X^\circ \rangle + \langle Z^\circ, X \rangle$$

The rest of the argumentation is analogous to that in Lemma 2.3.

□

Recall, we have shown in section 2.2 that f_μ is bounded below on $\operatorname{int}(\mathcal{F}(\mathcal{LQ}-\mathcal{P}) \times \mathcal{F}(\mathcal{LQ}-\mathcal{D}))$. Note, f_μ is continuously differentiable on $\operatorname{int}(\mathcal{F}(\mathcal{LQ}-\mathcal{P}) \times \mathcal{F}(\mathcal{LQ}-\mathcal{D}))$ and hence there always exists a step size parameter α such that the Goldstein Armijo criterion (2.4) is satisfied. For the same reasons as stated in section 2.2 all the iterates generated by our algorithm are strictly feasible interior points contained in a compact set. If we use the Goldstein Armijo criterion for choosing the step length parameter α then Theorem 2.7 ensures the global convergence of our algorithm.

The identity operator $\boldsymbol{\mathcal{Q}}$ in **(LQ-PSDP)** leads to the following application of linear-quadratic semidefinite programming.

3.5 An Application

For given $\bar{X} \in M(n)$ we study the projection of $\bar{X}$ onto $S(n)^+$. In other words we consider the problem of determining that matrix $X \in S(n)^+$ which has minimum distance to a given matrix $\bar{X} \in M(n)$.

Clearly, given $\bar{X} \in M(n)$ this problem may be stated as

$$\|\bar{X} - X\|^2 = \|X\|^2 + \|\bar{X}\|^2 - 2\langle \bar{X}, X \rangle \to \min, \quad \text{such that} \quad X \in S(n)^+.$$

More general, one can restrict the feasible region through adding linear matrix constraints $\boldsymbol{\mathcal{A}}(X) = b$. Then, in terms of the positive semidefinite linear-quadratic problem, introduced in section (3.1), the problem reads

$$\begin{array}{ll} \min & -\langle \bar{X}, X \rangle + \frac{1}{2}\langle X, X \rangle + \|\bar{X}\|^2 \\ \text{s.t.} & \boldsymbol{\mathcal{A}}(X) = b \\ & X \in S(n)^+. \end{array}$$

Consequently, solving this problem gives the projection of $\bar{X}$ onto $\mathcal{F}(\mathcal{LQ}-\mathcal{D})$.

Remark 3.3 *Observe, setting $\bar{X} = 0$ yields the special case of determining that matrix $X \in \mathcal{F}(\mathcal{LQ}-\mathcal{P})$ which is closest to the origin. As we have remarked earlier there are numerous further applications of the linear-quadratic semidefinite programming problem, see [30].*

Chapter 4

Computational Experience with Ill-Posed Problems in Semidefinite Programming

This chapter presents work of Gruber and Rendl, see [45]. Many theoretical and algorithmic results in Semidefinite Programming are based on the assumption that Slater's constraint qualification is satisfied for the primal and the associated dual problem. In this chapter we consider semidefinite problems with zero duality gap for which Slater's condition fails for at least one of the primal and dual problem. We propose a numerically feasible way of dealing with such semidefinite programs. The new method is based on a standard search direction with damped Newton steps towards primal and dual feasibility.

4.1 Preliminaries

We already have introduced the semidefinite programming problem in standard form in section 1.1. We also mentioned that semidefinite programs can be solved very efficiently using IPMs. Standard algorithms, based on the interior-point idea, assume Slater's constraint qualification to hold for **(PSDP)** and **(DSDP)**, thus insuring that both the primal and the dual are attained, with zero duality gap (see section 1.3, Theorem 1.14).

In this chapter we investigate how standard interior-point algorithms run into difficulties, as this constraint qualification is violated. We propose a computational procedure which permits to recognize the different duality states of a semidefinite program in a computationally reasonable way. We also discuss the advantages and disadvantages of existing procedures to deal with the same problem.

The reader might wonder whether primal-dual pairs of SDP, where one of the problems does not satisfy the Slater condition, are likely to appear in practical applications. We recall here two recently introduced relaxations of combinatorial problems, which are general enough to deserve wide attention, and which have turned out to provide tight approximations to the original combinatorial problem.

The *Equicut Problem* **(ECP)** takes as input an edge-weighted graph G with an even number of vertices. It asks to partition the vertices of G into two sets of equal cardinality, so as to minimize (or maximize) the weight of the edges joining the two partition blocks. If we represent partitions by partition vectors x, where $x_i \in \{-1, 1\}$, then the parity condition amounts to asking that $\sum_i x_i = 0$. The value of a bisection is $x^T Lx/4$, if L denotes the weighted Laplacian of the graph, see [63, 36]. Rewriting the cost function as $\operatorname{trace} L(xx^T)$ and linearizing, i.e. substituting xx^T by X yields a semidefinite program of the form

$$\min \operatorname{trace} LX/4 \text{ such that } \operatorname{diag}(X) = e,\ e^T Xe = 0,\ X \in S(n)^+.$$

Clearly, $X \in S(n)^+, e^T Xe = 0$ imply that X is singular, so the Slater condition will not hold for this problem. On the other hand, it is a simple exercise to verify that the dual satisfies the Slater condition, hence the duality gap is zero.

The *Quadratic Assignment Problem* **(QAP)** can also be approximated through SDP. If A and B are two symmetric matrices of order n, the **QAP** consists in minimizing $\operatorname{trace} AXBX^T$ where X runs through all permutation matrices of order n. In [136] this problem is relaxed to SDP as follows. We let $x = \operatorname{vec}(X)$, so x is the vector formed from the columns of X. In this case $\operatorname{trace} AXBX^T = x^T(B \otimes A)x = \operatorname{trace}(B \otimes A)(xx^T)$. Replacing again xx^T by a new variable $X \succeq 0$, we obtain an SDP relaxation. After adding linear constraints expressing the fact that x is built from a matrix X with row and column sums equal one, it is shown in [136, Proposition 2.1] that the resulting (primal) SDP does not satisfy the Slater condition, but again the dual problem does.

Both these problems can be remodeled in such a way that the modified primal problem satisfies the Slater condition as well. The point however is, that this remodeling needs additional information about the problem, which need not be available in general.

To get started we look at the following harmlessly looking type of problem $\mathbf{(GP)}_{\boldsymbol{\alpha}}$, depending on a parameter α.

$$\begin{array}{lll}
\mathbf{(GP)}_{\boldsymbol{\alpha}} \qquad & \min & \langle C, X\rangle \\
& \text{s.t.} & \operatorname{diag}(X) = e \\
& & \langle J, X\rangle = \alpha \\
& & X \in S(n),\ X \succeq 0
\end{array}$$

Here e denotes the vector of all ones, of appropriate dimension, $J = ee^T$ denotes the matrix of all ones, $\mathrm{diag}(X)$ is the vector of the diagonal entries of X and C is a given matrix in $S(n)$.

For $0 < \alpha < n^2$ the problem is strictly feasible. For $\alpha \in \{0, n^2\}$ strong duality still holds, but the primal does not have strictly feasible solutions. For $\alpha < 0$ or $\alpha > n^2$ the primal is infeasible. For $\alpha = 0$ we get the semidefinite relaxation of the equicut problem **(ECP)**.

In Table 4.1 we summarize some computational results from the publicly available software packages SDPT3 (a MATLAB software package for SDP, see [122]). and SeDuMi (a MATLAB toolbox for solving optimization problems over self-dual homogeneous cones, see [119]). We report results obtained by solving the original semidefinite problem and also its homogeneous formulation. As search directions we use the AHO-direction [3] and the HKM-direction [54]. We write AHO-S if **(PSDP)** and **(DSDP)** are solved directly and AHO-H if the corresponding homogeneous model is solved. An analogous notation is used for the HKM-direction. The line labeled SeDuMi gives the results from solving the homogeneous model using the toolbox SeDuMi. The remaining parameters are default parameters.

We denote by n the size of the problem, m is the number of constraints, and *iter* represents the number of iterations. $r_P = \boldsymbol{\mathcal{A}}(X) - b$ and $r_D = Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) - C$ stand for the primal and dual residuum, respectively. $\|X\|$ denotes the Frobenius matrix norm and $\|v\|$ stands for the Euclidean norm of $v \in \mathbb{R}^n$. The duality gap is denoted by $gap = |\langle C, X\rangle - b^T y|$. The column *Comment* indicates the termination status of the code SDPT3; 0 represents normal termination and 1 indicates that the Schur matrix is too ill-conditioned for further progress.

From this table we can draw the following conclusions. If $\alpha < n^2$ is positive and large enough, e.g. $\alpha = 1$, $\mathbf{(GP)}_{\boldsymbol{\alpha}}$ is well posed, and none of the algorithms has difficulties. As α is decreased to zero, the number of iterations grows, and the codes start to have difficulties. In fact, each model terminates prematurely with an ill-conditioned schur matrix, but SeDuMi seems to be best.

In this chapter we take a closer look at situations similar to the above example, where the resulting problem may have feasible solutions, which are bounded, but there need not be strictly feasible solutions. We restrict our considerations on feasible problems which have zero duality gap.

The problem of recognizing strict infeasibility can be handled using the selfdual embedding technique [24], and is not adressed in this chapter. Recognizing "weakly infeasible" problems is more subtle, and is also a challenge for the present approach. (A problem is weakly infeasible if it is IF but AF, see section 1.3.)

In section 4.2 we describe our approach for dealing with the type of problems introduced above. It is based on the HKM search direction, [54]. The main idea behind our method is to slow down the convergence towards the boundary of the feasible sets. In section 4.3 we present two test examples. The computational results with the new method appear in section 4.4. We conclude this chapter with a discussion in section 4.5.

search direction	α	*iter*	$\|r_D\|$	$\|r_P\|$	*gap*	*Comment*
AHO-S	1	13	1.2427e-14	4.0058e-14	6.0870e-07	0
HKM-S	1	16	1.6458e-14	1.4717e-14	1.4735e-06	0
AHO-H	1	13	9.1232e-08	1.0281e-07	1.3502e-06	0
HKM-H	1	15	9.0869e-09	4.7452e-08	6.2721e-07	0
SeDuMi	1	11	4.5722e-15	3.2122e-11	5.9799e-11	0
AHO-S	1e-7	16	2.8693e-12	1.4451e-06	3.1791e-02	1
HKM-S	1e-7	21	7.5685e-12	2.3774e-07	1.2574e-05	1
AHO-H	1e-7	29	4.7352e-06	1.0303e-04	1.7525e-03	1
HKM-H	1e-7	30	1.3608e-07	4.5060e-04	1.1031e-03	1
SeDuMi	1e-7	32	0.0000e+00	5.0009e-07	1.4155e-05	1
AHO-S	0	16	2.1927e-12	1.5394e-06	3.2734e-02	1
HKM-S	0	23	3.1848e-10	9.6005e-06	3.1045e-04	1
AHO-H	0	29	7.1918e-06	1.5156e-04	3.3823e-03	1
HKM-H	0	33	6.6927e-08	2.9510e-04	3.9075e-03	1
SeDuMi	0	34	0.0000e+00	4.3977e-07	5.9729e-04	1

Table 4.1: *Numerical results obtained by solving* $(\mathbf{GP})_{\boldsymbol{\alpha}}$ *with SDPT3, with* $n = 30$ *and* $m = 31$. *The last column indicates the termination status (0 for normal termination and 1 for numerical instabilities.)*

4.2 An Algorithm with Damped Steps Towards Feasibility

In this section we propose a way of dealing with feasible semidefinite problems for which the constraint qualification fails for at least one of **(PSDP)** and **(DSDP)**. Strong duality is assumed to hold.

The motivation for this approach comes from LP, see [131]. Many linear programs have no strictly feasible points although they are feasible and have finite optimal solutions. In LP standard algorithms were adapted to handle problems without strictly feasible solutions. We follow this idea and do not require feasible iterates.

For SDP we propose a computational way, which is based on the HKM search direction, [54]. (The approach can equally well be applied to other search directions.)

To derive the HKM-direction we replace in **(OPT)**, see section 1.2, the complementary slackness condition $ZX = 0$ by

$$ZX - \mu I = 0,\ \mu \geq 0.$$

This way we get the perturbed system

$$(\mathbf{OPT'}) \qquad \begin{aligned} &\boldsymbol{\mathcal{A}}(X) - b = 0 \\ &Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) = C \\ &ZX = \mu I \\ &X \in S(n)^+,\ Z \in S(n)^+,\ y \in \mathbb{R}^m. \end{aligned}$$

Note that, assuming strict primal and dual feasibility, the solution (X_μ, y_μ, Z_μ) of **(OPT')** is unique for any $\mu > 0$. The set of solution triplets (X_μ, y_μ, Z_μ) parameterized by $\mu > 0$ form the central path . The primal-dual interior-point methods generate iterates close to the central path. As we have already mentioned for a sequence $(\mu^k)_{k\in\mathbb{N}}$, $(\mu^k) \searrow 0$ it can be shown that the corresponding solutions (X_μ, y_μ, Z_μ) converge to an optimal solution triple (X^*, y^*, Z^*). The reader is referred to section 1.4 and the references quoted therein for a comprehensive treatment of this subject.

If Slater's constraint qualification does not hold then **(OPT')** has no feasible solution for (any) $\mu > 0$. Nevertheless, because of the successful handling of such optimization problems in LP, we try to solve a sequence of such systems where μ is driven to zero. Obviously, we can not require the iterates to be feasible. Like in LP, we try to reduce the norm of the residuals of feasibility and complementarity simultaneously. The central point for achieving this is to modify the HKM-direction so as to slow down the convergence towards primal and dual feasibility.

Applying Newton's method the standard search direction is computed by solving the following linear system of equations.

$$\begin{aligned} \boldsymbol{\mathcal{A}}(\Delta X) &= -r_P \\ \Delta Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) &= -r_D \\ \Delta Z\, X + Z\, \Delta X &= -r_{pC} \end{aligned} \tag{4.1}$$

where $r_{pC} = ZX - \sigma\mu I,\ \sigma \in (0, 1)$.

Our idea is to keep the iterates away from the boundary of the set of positive semidefinite matrices. That is why we modify (4.1) as follows. (The superscript denotes the iteration number.) We demand for all $k \in \mathbb{N}$ and for a damping parameter $\delta,\ (0 < \delta \leq 1)$

$$\boldsymbol{\mathcal{A}}(X^{k+1}) = (1 - \delta)\boldsymbol{\mathcal{A}}(X^k) + \delta b\ .$$

Using $X^{k+1} = X^k + \Delta X^k$ we get $\boldsymbol{\mathcal{A}}(\Delta X^k) = \delta(b - \boldsymbol{\mathcal{A}}(X^k))$. Since $r_P = \boldsymbol{\mathcal{A}}(X) - b$ we have

$$\boldsymbol{\mathcal{A}}(\Delta X^k) = -\delta r_P^k\ .$$

Analogous considerations lead to

$$\Delta Z^k + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y^k) = -\delta(Z^k + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y^k) - C) = -\delta r_D^k \ .$$

Summarizing, the damped search direction is computed by solving the following system of equations.

$$\begin{aligned} \boldsymbol{\mathcal{A}}(\Delta X) &= -\delta r_P \\ \Delta Z + \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(\Delta y) &= -\delta r_D \\ \Delta Z\, X + Z\, \Delta X &= -r_{pC} \end{aligned} \tag{4.2}$$

A damping factor $\delta = 1$ corresponds to the standard search direction. δ controls, how quickly $\|r_P\|$ and $\|r_D\|$ approach zero; the closer δ is to zero, the slower we approach feasibility. A choice of $\delta < 1$ prevents X and Z from approaching the boundary of the corresponding feasible sets prematurely. In general we obtain a solution triple $(\Delta X, \Delta y, \Delta Z)$ where ΔX is not symmetric. Therefore we use only the symmetric part of ΔX, i.e.

$$\Delta X \leftarrow \frac{\Delta X^T + \Delta X}{2}.$$

Using this search direction we perform standard line search. We compute step lengths α_P and $\alpha_D \in (0,1)$, such that

$$X + \alpha_P \Delta X \in S(n)^{++} \quad \text{and} \quad Z + \alpha_D \Delta Z \in S(n)^{++}.$$

We denote by α the minimum of primal and dual step length parameter, i.e., $\alpha = \min\{\alpha_P, \alpha_D\}$. The next iterate is defined by

$$\begin{aligned} X &\leftarrow X + \alpha \Delta X \\ y &\leftarrow y + \alpha \Delta y \\ Z &\leftarrow Z + \alpha \Delta Z. \end{aligned}$$

Next we compute the current value of μ using $\mu = \frac{\langle Z, X\rangle}{n}$. Finally we choose a $\sigma \in (0,1)$ and start the next iteration. To start, we use scaled multiples of the identity matrix for X and Z and set $y = 0$.

We have no formal proof of convergence for this approach, if Slater's condition fails. In fact, setting $\delta < 1$, but too close to one may lead to convergence towards a non-optimal point as can be seen in the subsequent sections. Setting δ too close to zero increases the number of iterations drastically. A convergence argument therefore has to take the specific value of δ into account.

4.3 Test Examples

In this section we describe two classes of problems which fall into the problem class where Slater‘s constraint qualification fails for at least one of **(PSDP)** and **(DSDP)**. The first example is $\mathbf{(GP)}_{\boldsymbol{\alpha}}$ and was introduced in section 4.1.

Example 4.1

$$
\begin{array}{lll}
(\mathbf{GP})_{\boldsymbol{\alpha}} & & \min \ \langle C, X \rangle \\
 & s.t. & \operatorname{diag}(X) = e \\
 & & \langle J, X \rangle = \alpha \\
 & & X \in S(n), \ X \succeq 0.
\end{array}
$$

Here C is a randomly generated matrix in $S(n)$. Obviously, for $\alpha = 0$, $(\mathbf{GP})_{\boldsymbol{\alpha}}$ is feasible but the interior of the feasible set is empty. The associated $(\mathbf{DGP})_{\boldsymbol{\alpha}}$ has feasible interior points, and strong duality holds. The dual is

$$
\begin{array}{lll}
(\mathbf{DGP})_{\boldsymbol{\alpha}} & & \max \ \langle e, y \rangle + \alpha s \\
 & \text{s.t.} & Z + \operatorname{Diag}(y) + sJ = C \\
 & & Z \in S(n), \, Z \succeq 0
\end{array}
$$

where $y \in \mathbb{R}^n$, $s \in \mathbb{R}$ and $\operatorname{Diag}(y)$ denotes the diagonal matrix with y as diagonal.

Example 4.2

Next we generate a class of feasible problems having zero duality gap, where neither the primal nor the dual has interior feasible points. The input data are generated as follows. We start with an orthogonal $n \times n$ matrix $V = (v_1, v_2, \ldots, v_n)$. We define $A_i := v_i v_i^T, \ 1 \le i \le n$. Next we select the number of primal equations $m \in \{1, 2, \ldots, n\}$ arbitrarily. The primal constraints are, recall (1.1), $\operatorname{trace} A_i X = b_i, \ 1 \le i \le m$. Setting $b_i > 0, \ 1 \le i \le m$ implies that the primal is strictly feasible. Therefore wlog. we define

$$
b_i = \begin{cases} 0, & \text{if } \ i = 1 \\ 1, & \text{if } \ 1 < i \le m. \end{cases}
$$

Now clearly the primal is feasible, (take e.g. $X = \sum_{i=2}^{m} v_i v_i^T$), but has no strictly feasible solution, because $b_1 = 0$. We define C to be a linear combination of the A_i, $C := \sum_{i=1}^{n} \alpha_i A_i$. Thus dual feasibility implies

$$
Z \ = \ C - \boldsymbol{\mathcal{A}}^{\boldsymbol{t}}(y) \ = \ \sum_{i=1}^{m} (\alpha_i - y_i) A_i + \sum_{i > m} \alpha_i A_i. \tag{4.3}
$$

We define α as follows.

$$
\alpha_i \begin{cases} > 0, & \text{if } i < n \\ = 0, & \text{if } i = n. \end{cases}
$$

It is easily checked that for this setting **(PSDP)** and **(DSDP)** are feasible without strictly interior points and that strong duality holds. The optimal value of both, primal and dual problem is $\sum_{i=2}^{m} \alpha_i$.

Remark 4.3 *If we set more of the b_i's and α_i's to zero, then the resulting problem is "more degenerate". In a second experiment we set 5 of the b_i's and 5 of the α_i's to zero. The results are contained in Table 4.5 below.*

Remark 4.4 *Finally, let $A_{ij} := v_i v_j^T + v_j v_i^T$. Then it is easy to see that $\{A_{ij} : i \leq j\}$ spans the entire space of symmetric matrices. In our case we use only a subset of the generating matrices. We experimented also with the inclusion of constraints of the form* $\operatorname{trace} A_{ij} X = 0$, *for (some randomly selected) $i < j$. It had no significant influence on the performance of our algorithm. Thus, even though our construction looks quite specialized, it can in fact be extended quite easily to general problems, by including also a subset of the A_{ij}'s to generate further equality constraints.*

4.4 Computational Experiments

In this section we report some numerical results obtained from solving the test examples explained in the previous section. For example 4.1 we consider the cases $\alpha \in \{0, 10^{-7}\}$. The problems are now solved with the modified interior-point approach introduced in section 4.2. We restrict our numerical testing on problem sizes $n \in \{10, 20, \ldots, 50\}$. We have solved 20 randomly generated test examples for each of the five problem sizes. Tables 4.2 and 4.3 show the computational results of example 4.1. The columns labeled min and max denote the minimum and maximum number of iterations required by the 20 test runs. The average number of iterations is denoted by $iter$. Column $Failures$ gives the number of test runs which terminated prematurely due to ill-conditioned schur matrix. The other columns represent averages over 20 runs.

Even though the problem is strictly feasible for $\alpha = 10^{-7}$, we see from Table 4.1 that the standard methods run into serious difficulties. Our experiments also indicate that SeDuMi seems to be dealing better with this type of problem than the other variants. Table 4.2 on the other hand shows that in all cases we manage with the new method to reduce the error in the optimality conditions **(OPT)** to a value below the desired threshold of 10^{-5}. In the case $\alpha = 0$, see Table 4.3, we terminate in 2 out of the 100 test runs prematurely. In all other cases the overall error in **(OPT)** is reduced to a value below 10^{-4}. The tables also indicate that the damping factor destroys the nice (local superlinear) convergence behaviour of Newton's method.

In Tables 4.4 and 4.5 we get similar results for the problem class from Example 4.2. The interval in column m shows the minimum and maximum number of primal constraints we have selected randomly over the 20 test runs. Even though these problems have no strictly feasible solutions, our damping strategy still leads to acceptable solutions. The column *Failures* is omitted, as all instances terminated normally. We also note that there is no indication that a higher degree of degeneracy of the data (Table 4.5) has a negative influence on the computational performance.

Finally we would like to demonstrate that the damping factor δ is the crucial ingredient of our approach. In Table 4.6 we keep the problem size $n = 30$ fixed, and vary δ for a set of 20 instances. As one would expect, a damping factor δ close to one takes us to the boundary prematurely, and the algorithm has difficulties due to ill-conditioning. A small value of δ on the other hand will eventually slow down progress towards feasibility. The choice of δ also has to be tuned to the way in which μ is reduced.

Table 4.6 shows that δ should be selected from the interval $[0.2, 0.4]$. We should mention that the choice of δ and σ, the parameter reducing μ between subsequent iterations, require some fine-tuning. We observed that reducing μ by a factor close to 0.9 works well in conjunction with $\delta \in [0.2, 0.4]$. We performed the numerical experiments on a DEC Alpha workstation using Matlab.

4.5 Concluding Comments

We have presented a damped version of the standard HKM-direction for SDP. The damping strategy allows to solve larger classes of problems, namely those, where Slater's condition fails for one of the problems. Here are some conclusions that we draw from our computational tests.

(1) Our experience indicates that the damping strategy, with the damping parameter δ from the interval $[0.2, 0.4]$ works well, even if there are no strictly feasible points to the problem.

(2) The convergence is slowed down considerably in this case. In contrast to standard implementations, the choice of internal parameters such as the parameter σ reducing μ in each iteration, and the choice of the damping factor δ seriously affect the computational performance.

(3) The present approach will fail, if strong duality does not hold, because in this case it is impossible to reduce the duality gap to zero. From our computational experiments, we consider it fair to conclude that our approach is superior to standard methods in the absence of constraint qualifications for problems having no duality gap.

(4) The damping strategy could also be used in combination with other search directions. Our experiments with damping applied to other directions did not indicate a performance behaviour, significantly different from the damped HKM-direction. Since this direction is quite simple to program, and also competitive with respect to running time, we made our tests with this direction.

n	m	$iter$	min	max	$\|r_D\|$	$\|r_P\|$	gap	CPU time	$Failures$
10	11	117	82	247	3.4272e-08	6.1174e-08	8.8671e-06	1.2372	0
20	21	115	73	273	2.7046e-07	4.4243e-07	7.6857e-06	3.7748	0
30	31	117	92	146	1.2070e-08	5.9516e-08	8.2850e-06	11.0931	0
40	41	114	83	144	8.2768e-08	2.0688e-07	7.7886e-06	28.5649	0
50	51	118	94	126	2.0246e-08	1.3001e-07	7.3176e-06	58.5176	0

Table 4.2: *Numerical Results for Example 4.1,* $\alpha = 10^{-7}$, $\delta = 0.2$, *accuracy:* 10^{-5}

n	m	$iter$	min	max	$\|r_D\|$	$\|r_P\|$	gap	CPU time	$Failures$
10	11	120	69	250	1.0863e-06	4.1060e-06	7.2414e-05	1.3057	0
20	21	134	71	261	1.0249e-06	6.7748e-06	5.9814e-05	4.4448	0
30	31	135	65	282	4.5857e-06	1.3098e-05	4.8225e-05	12.8219	0
40	41	92	62	146	1.0351e-05	1.7913e-05	5.3267e-05	23.1120	0
50	51	128	69	282	3.9422e-06	1.1157e-05	6.0428e-05	63.5947	2

Table 4.3: *Numerical Results for Example 4.1,* $\alpha = 0$, $\delta = 0.2$, *accuracy:* 10^{-4}

n	m	*iter*	*min*	*max*	$\lVert r_D \rVert$	$\lVert r_P \rVert$	*gap*	CPU time
10	[5 9]	52	49	54	1.6673e-06	6.5274e-10	7.6699e-06	0.3949
20	[8 19]	52	47	54	1.7585e-06	2.6521e-08	8.0267e-06	1.1413
30	[13 28]	52	47	55	1.6791e-06	2.5154e-09	7.8175e-06	6.2863
40	[15 39]	53	47	55	1.7035e-06	2.5319e-09	7.9011e-06	16.4001
50	[18 49]	53	51	55	1.7218e-06	1.0924e-09	7.8686e-06	33.6318

Table 4.4: *Numerical Results for Example 4.2, $\delta = 0.4$, accuracy:* 10^{-5}

n	m	*iter*	*min*	*max*	$\lVert r_D \rVert$	$\lVert r_P \rVert$	*gap*	CPU time
30	[21 29]	51	46	54	1.6507e-06	2.4667e-09	7.6402e-06	3.9250
40	[21 37]	52	47	55	1.7076e-06	1.7593e-09	7.8318e-06	9.1128
50	[21 49]	52	47	55	1.6434e-06	2.3365e-09	7.5308e-06	17.5824

Table 4.5: *Numerical Results for Example 4.2, more degenerate problems, data generated according to Remark 4.3, $\delta = 0.4$, accuracy:* 10^{-5}

	Termination		
δ	*normal*	*Schur matrix is ill-conditioned*	*step length* $\alpha < 10e-4$
0.9	0	20	0
0.7	11	8	1
0.5	16	2	2
0.4	13	0	7
0.3	11	0	9
0.2	26	0	4
0.1	0	0	20

Table 4.6: *Numerical Results for Example 4.1, with* $n = 30$ *and* $m = 31$ *and different values for* δ, $\alpha = 0$, *reduction of* μ *by a factor* $\sigma = 0.8$, *accuracy:* 10^{-4}

Part Three –

Combinatorial Applications

Chapter 5

Combining Polyhedral and Non-Polyhedral Relaxations for the Stable Set Problem

This chapter presents work of Gruber and Rendl, see [46, 47]. We investigate relaxations for the maximum stable set problem based on the Lovász number $\vartheta(G)$ as an initial upper bound. We strengthen this relaxation by adding two classes of cutting planes, odd circuit and triangle inequalities. We present computational results using this tighter model on many classes of graphs.

5.1 The Stable Set Problem

Let $G = (V(G), E(G))$ denote an undirected graph with node set $V(G)$ and edge set $E(G)$, where $|V(G)| = n$ and $|E(G)| = m$. An edge e with endnodes i and j is represented by (ij).

A stable set S in G is by definition a subset of $V(G)$ such that no two vertices in S are joined by an edge in $E(G)$. (A synonym for "stable set" used by many authors is "independent set".) We focus on unweighted problems. Then the *maximum stable set problem* is the problem of finding a stable set of maximum cardinality. We will denote the size of a largest stable set in G by $\alpha(G)$. In the literature this number is called stability number or independence number.

An assignment of colors to the vertices of G, one color to each vertex, so that adjacent vertices receive different colors, is called a *coloring of* G. Hence a coloring may be defined alternatively as a partition of $V(G)$ into disjoint sets such that each of these sets is an independent set for G. A $k-$coloring of G is a coloring of G using

k colors or equivalently a partition of $V(G)$ into k disjoint independent sets. The chromatic number $\chi(G)$ of a graph G is the minimum value k for which a $k-$coloring of G exists. The complement of a graph G is denoted by $\bar{G}$.

It is well known that it is NP-complete to compute the numbers $\alpha(G)$ and $\chi(G)$. The Lovász number $\vartheta(G)$ separates $\alpha(G)$ and $\chi(\bar{G})$, $\alpha(G) \leq \vartheta(G) \leq \chi(\bar{G})$. This number is also called theta function ($\vartheta-$function) of G. $\vartheta(G)$ can be computed in polynomial time, see [41]. In this work we are primarily interested in improving the first inequality, i.e., in getting tight upper bounds for $\alpha(G)$.

The *stable set polytope* of G is defined by

$$STAB(G) := \text{conv}\Big\{\chi^S \in \mathbb{R}^n | S \subseteq \text{V(G) stable set}\Big\}.$$

where χ^S denotes the incidence vector of S, i.e., $\chi_i^S = \begin{cases} 1 & \text{for } i \in S \\ 0 & \text{otherwise} \end{cases}$. Thus the maximum stable set problem **(STAB)** can be formulated as

$$(\textbf{STAB}) \quad \alpha(G) = \max \left\{ \sum_{i=1}^{n} x_i | \, x \in STAB(G) \right\}.$$

Before we continue let us fix some notation which will be used throughout this chapter. By definition, a graph is complete if every two of its vertices are joined by an edge. A complete graph of order p is denoted by K_p. Any complete subgraph is called a clique. Stable sets in G correspond to cliques in $\bar{G}$. Hence an equivalent problem to the *maximum stable set problem* is the *maximum clique problem.* It is the problem of finding a largest clique in a given graph.

We use diag(X) to denote the vector formed from the diagonal elements of a square matrix X. For $x \in \mathbb{R}^n$, Diag(x) denotes the diagonal matrix with diagonal x. We use $A_{i.}$ $(A_{.j})$ to denote the $i-$th row ($j-$th column) of a matrix A, respectively. We write trace(X) to denote the trace of a square matrix X, i.e., trace$(X) = \sum_i X_{ii}$. e denotes the vector of arbitrary dimension with each component equal to one, J denotes the matrix of all ones and E_{ij} is defined by $E_{ij} := e_i e_j^t + e_j e_i^t$ where e_i stands for the $i-$th column of the identity matrix of appropriate size. If $\gamma \in \mathbb{R}$, we will use $\lfloor \gamma \rfloor$ to denote the largest integer not larger than γ.

The remaining sections are organized as follows. In section 5.2 we survey linear relaxations of the stable set problem and introduce the relaxation leading to the $\vartheta-$function. In section 5.3 we describe the lift-and-project procedure of Lovász and Schrijver. This is a method for constructing new inequalities valid for the $0-1$ integral points of a linear program. We then investigate two semidefinite versions to compute $\vartheta(G)$ in section 5.4. We will show that both models are on a par and how optimal solutions of them are related. In section 5.5 we propose further refinements by combining linear with semidefinite relaxations. Implementation details are described in section 5.6. The computational tests and its analysis appear in section 5.7. We conclude with comments in section 5.8.

5.2 Vertex Based Relaxations of the Stable Set Problem

In this section we review some important valid inequalities for the stable set polytope, used in polyhedral relaxations. We also recall the non-polyhedral relaxation $TH(G)$ of the stable set polytope.

A complete description of $STAB(G)$ by linear inequalities is currently not known and unlikely to be tractable, in view of the NP-completeness to compute $\alpha(G)$. The following sets of linear inequalities contain a partial description of $STAB(G)$.

$$\begin{aligned} (\textit{nonnegativity constraints}): \quad & x_i \geq 0 \quad \text{for all } i \in V(G) && (5.1)\\ (\textit{edge constraints}): \quad x_i + & x_j \leq 1 \quad \text{for all } (ij) \in E(G) && (5.2)\\ & x_i \leq 1 \quad \text{for isolated nodes} \end{aligned}$$

Obviously (5.1) and (5.2) hold for $x \in STAB(G)$. Moreover, the integral solutions x of (5.1) and (5.2) are the incidence vectors of stable sets in G. In other words, the vertices of $STAB(G)$ are the integral solutions of (5.1) and (5.2). The polytope defined by this set of inequalities is in general much larger than $STAB(G)$. The following result is well known.

Proposition 5.1 [43] *The inequalities (5.1) and (5.2) are enough to describe the set $STAB(G)$ if and only if G is bipartite and has no isolated nodes.*

If we take for instance a triangle T and set $x_i = \frac{1}{2}$ for $i \in V(T)$, then it is easy to see that x satisfies (5.1) and (5.2) but x is not contained in $STAB(G)$. Therefore other classes of inequalities are needed to describe $STAB(G)$. We recall several such classes of inequalities which are valid for $STAB(G)$. First, the *clique constraints* are given by

$$\sum_{i \in V(Q)} x_i \leq 1 \quad \text{for all cliques } Q \text{ in } G. \tag{5.3}$$

In particular, (5.3) contains all 2-cliques K_2 and therefore the constraints (5.2). The polytope defined by (5.1) and (5.3) is usually called *clique constrained stable set polytope* of G, denoted $QSTAB(G)$, i.e.,

$$QSTAB(G) := \Big\{ x \in I\!R^n \,|\, x \text{ satisfies (5.1) and (5.3)} \Big\}.$$

It is known that (5.1) and (5.3) are not enough to describe $STAB(G)$. $STAB(G)$ is in general a proper subset of $QSTAB(G)$. In the 1970s Fulkerson [33] and Chvátal [21] characterized the graphs for which $QSTAB(G) = STAB(G)$. They showed,

Theorem 5.2 *$STAB(G) = QSTAB(G)$ if and only if G is perfect.*

(For a formal definition of perfect graphs, we refer to [41].) The following results were shown by Grötschel, Lovász and Schrijver.

Proposition 5.3 [41] *The maximum stable set problem can be solved in polynomial time for perfect graphs.*

Theorem 5.4 [41] *It is NP-complete to optimize a linear objective function over $QSTAB(G)$.*

$QSTAB(G)$ is not enough to describe $STAB(G)$ for instance if G contains odd circuits. Suppose C is an odd circuit with $|V(C)| \geq 5$. Let $x_i = \frac{1}{2}$ for $i \in V(C)$, then it is not hard to verify that x is in $QSTAB(G)$ but not in $STAB(G)$. Here remedy is given by a further class of inequalities valid for $STAB(G)$, the *odd circuit constraints.* This class is given by

$$\sum_{i \in V(C)} x_i \leq \frac{|V(C)| - 1}{2} \quad \text{for } C \text{ odd circuit}, \tag{5.4}$$

yielding the *circuit constrained stable set polytope* of G, denoted $CSTAB(G)$.

$$CSTAB(G) := \Big\{ x \in I\!R^n |\, x \text{ satisfies } (5.1), (5.2) \text{ and } (5.4) \Big\}.$$

By definition, a graph is called $t-$perfect if $STAB(G)$ can be characterized by $CSTAB(G)$, i.e., the relaxation $CSTAB(G)$ gives $\alpha(G)$ for the class of $t-$perfect graphs, see [42, 43]. A graph is called $h-$perfect if (5.1), (5.3) and (5.4) suffice to describe $STAB(G)$. Note in the latter case we have $STAB(G) = CSTAB(G) \cap QSTAB(G)$. Grötschel, Lovász and Schrijver proved the two theorems which follow.

Theorem 5.5 [42] *Any linear function can be optimized over $CSTAB(G)$ in polynomial time.*

Theorem 5.6 [42] *The maximum stable set problem can be solved in polynomial time for $h-$perfect graphs.*

Clearly, it follows from Theorem 5.6 that the maximum stable set problem can be solved in polynomial time for $t-$perfect graphs.

The polytope $CSTAB(G)$ is still only a (polyhedral) relaxation of $STAB(G)$, i.e., $STAB(G) \subseteq CSTAB(G)$. We now turn to the non-polyhedral relaxation of $STAB(G)$, given by the set $TH(G)$, introduced by Lovász, see [76]. This relaxation can also be found in [78]. There are many equivalent ways to define this relaxation. Several were introduced already in [76], but see also Knuth [67]. We use the following definition, see [78],

$$TH(G) := \Big\{ d \in I\!R^n |\, \exists \in S(n) \text{ such that } d = \mathrm{diag}(X) \text{ and } X \text{ satisfies } (5.5) \Big\}$$

$$X = (x_{ij}) \text{ with } x_{ij} = 0 \text{ for } (ij) \in E(G),\ d = \mathrm{diag}(X),\ \bar{X} := \begin{pmatrix} X & d \\ d^T & 1 \end{pmatrix} \succeq 0. \tag{5.5}$$

($TH(G)$ stems from the Lovász number $\vartheta(G)$.) The set $TH(G)$ is a relaxation of the stable set polytope, because the characteristic vector χ of any stable set produces the matrix $\chi\chi^T$, which is feasible for (5.5), so $STAB(G) \subseteq TH(G)$.

Definition 5.7 *(convex corner)*

$$\text{We say that } K \text{ is a convex corner if } \begin{cases} K \subseteq \mathbb{R}^n_+ \\ K \text{ convex} \\ x \in K,\ 0 \leq y \leq x \Rightarrow y \in K. \end{cases}$$

$TH(G)$ enjoys the following properties.

Theorem 5.8 [42, 41]

(1) $STAB(G) \subseteq TH(G) \subseteq QSTAB(G)$, *i.e., every vector in* $TH(G)$ *satisfies (5.3).*

(2) $TH(G)$ *is a convex corner, but not a polytope in general.*

(3) $TH(G)$ *is a polytope if and only if* G *is perfect.*

(4) *Any linear objective function can be optimized over* $TH(G)$ *in polynomial time.*

As a consequence of Theorem 5.8 item (3) we get:

Corollary 5.9 $STAB(G) = TH(G) = QSTAB(G)$ *if and only if* G *is perfect.*

5.3 The Lift–and–Project Idea

Given an integer programming problem we shortly review a method for generating a series of relaxations for the convex hull of feasible solutions, becoming sharper and sharper. This general approach for strengthening the linear programming relaxation P of $0-1$ programs was proposed by Lovász and Schrijver [78]. To start with, we point out some details of their method. We then present results on the stable set polytope. We refer the reader to [117, 7] and [8] for similar work.

Let P_I denote the convex hull of all $0-1$ integral points of P. The key-note behind the method suggested by Lovász and Schrijver [78] is to lift the problem into a higher dimensional space. In fact, multiplying each inequality defining the linear relaxation P by x_j and $1-x_j, (j = 1, \ldots, n)$ gives a system of quadratic inequalities. This system is linearized by substituting y_{ij} for $x_i x_j$. $0-1$ solutions satisfy $x_i^2 = x_i$. Therefore x_i is replaced by y_{ii} as well. Imposing $y_{ij} = y_{ji}$, this method produces a system of linear matrix inequalities in the higher dimensional space $S(n)$. (The number of variables has been squared.) We will use $M(P)$ to denote the set of all matrices $Y = (y_{ij}),\ Y \in S(n)$ satisfying the matrix inequalities generated this way.

The reformulation of the problem in the lifted space possibly results in a tighter relaxation. One can investigate this relaxation in the higher dimensional space, or project it back onto the original space.

The projection $N(P)$ of $M(P)$ on the subspace $\mathbb{R}^n$ indexed by the node set of G is defined by

$$N(P) := \Big\{x \in \mathbb{R}^n | \, \exists Y \in M(P), x = \mathrm{diag}(Y)\Big\}.$$

In other words, the convex polyhedron $N(P)$ is the intersection of a family of halfspaces corresponding to that nonnegative linear combinations of the generated matrix inequalities which do not contain $y-$variables outside the diagonal of Y. By construction, $x \in P_I$ satisfies these inequalities. Hence, $P_I \subseteq N(P) \subseteq P$, which means that lifting and projecting gives a convex set that approximates the convex hull of all feasible $0-1$ points. That is to say, this method generates cutting planes valid for P_I in the original space.

We now outline some results proved in [78]. Lovász and Schrijver analyzed the repeated $N-$operator $N^k(P)$ defined recursively by $N^0(P) = P$ and $N^k(P) = N(N^{k-1}(P))$ for $k \geq 1$. They showed the following theorem.

Theorem 5.10 [78, Theorem 1.4] *Let $P \subseteq \mathbb{R}^n$. Then $N^n(P) = P_I$.*

We now consider instead of a general P the *fractional stable set polytope* $FRAC(G)$ defined as

$$FRAC(G) := \Big\{x \in \mathbb{R}^n | \, x \text{ satisfies (5.1) and (5.2)}\Big\}.$$

We have pointed out earlier that $STAB(G)$ is the convex hull of the $0-1$ integral points of $FRAC(G)$. Lovász and Schrijver introduced the $N-$index of an inequality valid for $STAB(G)$ as the least k such that this inequality is valid for $N^k(FRAC(G))$. They showed that the odd circuit constraints have $N-$index one. In fact, they proved

Theorem 5.11 [78, Theorem 2.3] *The polytope $N^1(FRAC(G))$ is exactly the solution set of $CSTAB(G)$.*

Proof. see [78], but here we give a simplified version of the if part. The line of argument illustrates the lift and project process just surveyed to $FRAC(G)$.

Consider an edge $(ij) \in E(G)$. For $x \in FRAC(G)$, $x_i + x_j \leq 1$. Multiplying this inequality by x_i and linearizing it yields $y_{ii} + y_{ij} \leq y_{ii}$, hence $y_{ij} \leq 0$. Since $y_{ij} \geq 0$ we have $y_{ij} = 0$ whenever $(ij) \in E(G)$ and therefore we can set $y_{ij} = 0$ if $(ij) \in E(G)$.

Suppose C is a cycle with $|V(C)| = 2k+1, k \in \mathbb{N}$. For the sake of simplicity we assume that $V(C) = \Big\{v_1, v_2, \ldots, v_{2k+1}\Big\}$ and $E(C) = \Big\{(v_1, v_2), (v_2, v_3), \ldots, (v_{2k-1}, v_{2k}), (v_{2k}, v_{2k+1})\Big\}$. Consider $C - v_1$. ($C - v_1$ is the graph obtained by deleting node v_1.)

Now, starting from node v_2 consider the edges on C and its edge constraints consecutively. The lifting process is applied by multiplying these edge constraints alternately by $1 - x_1$ and x_1, respectively. The initial condition of this process is, that the first occuring edge constraint (in our case $x_2 + x_3 \le 1$) is multiplied by $1 - x_1$. This way we get the following set of matrix inequalities.

$$\begin{array}{rcll}
y_{22} + y_{33} - y_{12} - y_{13} & \le & 1 - y_{11} & \ldots \quad (v_2, v_3) \\
y_{13} + y_{14} & \le & y_{11} & \ldots \quad (v_3, v_4) \\
y_{44} + y_{55} - y_{14} - y_{15} & \le & 1 - y_{11} & \ldots \quad (v_4, v_5) \\
y_{15} + y_{16} & \le & y_{11} & \ldots \quad (v_5, v_6) \\
 & \vdots & & \\
y_{1,2k-1} + y_{1,2k} & \le & y_{11} & \ldots \quad (v_{2k-1}, v_{2k}) \\
y_{2k,2k} + y_{2k+1,2k+1} - y_{1,2k} - y_{1,2k+1} & \le & 1 - y_{11} & \ldots \quad (v_{2k}, v_{2k+1})
\end{array} \tag{5.6}$$

For the sake of clarity we have specified the underlying edges to the inequalities. We have separated each by some points.

Clearly, the total number of inequalities is $2k - 1$. Note, k inequalities were multiplied by $1 - x_1$ and $k - 1$ inequalities were multiplied by x_1. From considerations above we know that $y_{12} = 0$ and $y_{1,2k+1} = 0$. Now it is easy to see that adding up the matrix inequalities nullifies all $y-$variables outside the diagonal of Y. In fact, we get the odd circuit constraint

$$\sum_{i=1}^{2k+1} y_{ii} = \sum_{i=1}^{2k+1} x_i \le k = \frac{|V(C)| - 1}{2}.$$

Thus, all odd circuit constraints can be obtained in one step of the lifting and projecting method applied to $FRAC(G)$.

□

Before we continue we introduce the *triangle inequalities* defining the metric polytope MET. By definition, a matrix $X \in S(n)$ satisfies

$$X \in MET \ :\Leftrightarrow \ \begin{cases} 0 \le X_{ij} \le X_{ii}, \\ X_{ii} + X_{jj} - X_{ij} \le 1, \\ -X_{kk} - X_{ij} + X_{ik} + X_{jk} \le 0, \\ X_{kk} + X_{ii} + X_{jj} - X_{ij} - X_{ik} - X_{jk} \le 1, \ \text{for } 1 \le i, j, k \le n. \end{cases}$$

These constraints are valid for the convex hull of all matrices of the form xx^T where x is some $0 - 1$ vector. So they have nothing particular to do with the stable set problem, but merely exploit the fact that we are ultimately interested in integer matrices.

The projection P_{MET} of MET on the subspace $I\!R^n$ indexed by the node set of G is defined by

$$P_{MET} := \Big\{ d \in I\!R^n \,|\, \exists\, X \in MET, d = \operatorname{diag}(X) \Big\}.$$

Given a graph G, let us consider the polytope $MET_1(G)$ contained in MET. We define it as

$$MET_1(G) := MET \cap S(G),$$

where

$$S(G) := \Big\{ X = (x_{ij}) \in S(n) |\, x_{ij} = 0 \text{ for } (ij) \in E(G) \Big\}.$$

The projection P_{MET_1} of MET_1 on the subspace $I\!R^n$ indexed by the node set of G is defined by

$$P_{MET_1(G)} := \Big\{ d \in I\!R^n \,|\, \exists\, X \in MET_1, d = \operatorname{diag}(X) \Big\}.$$

Consider also the set $P_{MET_2(G)}$, defined as

$$P_{MET_2(G)} := \Big\{ d \in I\!R^n \,|\, \exists\, X \in S(n) \text{ such that } d = \operatorname{diag}(X) \text{ and } X \text{ satisfies (5.7)} \Big\}.$$

$$\begin{array}{l} X = (x_{ij}),\ d = \operatorname{diag}(X),\ x_{ij} \geq 0,\ \ -x_{kk} + x_{ik} + x_{jk} \leq 0, \\ x_{kk} + x_{ii} + x_{jj} - x_{ik} - x_{jk} \leq 1,\ \text{for } 1 \leq i,j,k \leq n,\ \text{such that } (ij) \in E(G). \end{array} \tag{5.7}$$

From these definitions we see that $P_{MET_1(G)} \subseteq P_{MET}$ and $P_{MET_1(G)} \subseteq P_{MET_2(G)}$. It follows from the proof of Theorem 5.11, which is given in [78] that, if $x \in CSTAB(G)$, then $x \in P_{MET_2(G)}$. Note, all the lifted inequalities (5.6) are contained in $P_{MET_2(G)}$. Thus we have

Theorem 5.12 $N^1(FRAC(G)) = CSTAB(G) = P_{MET_2(G)}$.

Our extensive numerical experience indicates that optimizing over $CSTAB(G)$ and $P_{MET_1(G)}$, respectively is the same. On the theoretical side, however it is still open for us if the triangle inequalities in $P_{MET_1(G)} \backslash P_{MET_2(G)}$ can strengthen the relaxation $P_{MET_2(G)}$.

For a better understanding of the remainder of this section it is worth noting that Lovász and Schrijver homogenized the underlying problem by introducing a new variable x_{n+1} and considered $N^0(P)$ as a subset of the hyperplane H_{n+1} defined by $x_{n+1} = 1$. This way they could work with the cone spanned by the vectors $\binom{x}{1}$, where $x \in N^0(P)$. Hence $N(N^0(P)) = \Big\{ \binom{x}{1} | x \in N^0(P) \Big\} \cap H_{n+1}$.

Note, a further constraint valid for the convex hull of all matrices of the form xx^T where x is some $0-1$ vector is given by

$$xx^T - \operatorname{diag}(xx^T)\operatorname{diag}(xx^T)^T = 0.$$

Relaxing this quadratic constraint gives $xx^T - \mathrm{diag}(xx^T)\mathrm{diag}(xx^T)^T \succeq 0$. By Lemma A.9, this is equivalent to imposing

$$\begin{pmatrix} xx^T & \mathrm{diag}(xx^T) \\ \mathrm{diag}(xx^T)^T & 1 \end{pmatrix} \succeq 0.$$

This motivates to consider the more complicated set $M_+(P)$ defined as

$$M_+(P) := \left\{ Y \in M(P) | \, Y \text{ satisfies } \begin{pmatrix} Y & \mathrm{diag}(Y) \\ \mathrm{diag}(Y)^T & 1 \end{pmatrix} \succeq 0 \right\}.$$

and its projection $N_+(P)$ on the subspace $\mathbb{R}^n$ indexed by the node set of G defined as

$$N_+(P) := \left\{ x \in \mathbb{R}^n | \, \exists \, Y \in M_+(P), x = \mathrm{diag}(Y) \right\}$$

as done in [78]. Let the N_+-index of an inequality valid for $STAB(G)$ be defined in the way we have defined the $N-$index. It is shown in [78] that the clique constraints, the odd circuit constraints and several other classes have N_+-index one. From this result it follows that $P_I \subseteq N_+(P) \subseteq N(P) \subseteq P$. The following result is straightforward.

Lemma 5.13 $N_+^1(FRAC(G)) \subseteq TH(G)$.

Proof. Let $x \in N_+^1(FRAC(G))$. By definition, there exists $Y \in M_+(P)$, $x = \mathrm{diag}(Y)$. As shown above $y_{ij} = 0$ whenever $(ij) \in E(G)$. Hence $x \in TH(G)$. ($x \in N_+(FRAC(G))$ satisfies (5.7) but in general $x \in TH(G)$ does not.)

□

The results of this section suggest to go into the space of symmetric matrices to maintain the semidefiniteness condition on $\bar{X}$ in (5.5).

5.4 Semidefinite Programming and the ϑ-Function

In this section we take a closer look at two different semidefinite relaxations for the maximum stable set problem. For this we start with an earlier approach.

In 1956 Shannon [116] introduced the term Shannon capacity of a noisy channel. Suppose the vertices of a graph G represent letters in a finite alphabet which is used to send messages via a noisy channel. If the letters, assigned to nodes i and j can be confused then there is an edge (i, j) in G joining nodes i and j. Clearly, $\alpha(G)$ is the maximum number of single letters which can be transmitted without danger of confusion. Thus at least $\alpha(G)^k$ messages of length k can be sent. This number can be improved. The strong product of two graphs $G = (V(G), E(G))$ and $H = (W(H), F(H))$ is defined as the graph with the cartesian product $V(G) \times W(H)$ as vertex set and with edges $((i,u),(j,v))$ if and only if $(i,j) \in E(G)$ and $(u,v) \in F(H)$, or $(i,j) \in E(G)$ and $u = v$, or $i = j$ and $(u,v) \in F(H)$. Let G^k denote the strong product of k graphs equal to G. Then by construction the maximum number of messages of length k which can be sent without danger of confusion is exactly $\alpha(G^k)$. The Shannon capacity $\Theta(G)$ of a graph is defined as

$$\Theta(G) := \sup_{k \to \infty} \sqrt[k]{\alpha(G^k)}$$

and shows the rate of improvement of the information flow-rate (without confusion) if one exchanges from one-letter messages to longer messages. (By construction $\Theta(G) \geq \alpha(G)$.)

The $\vartheta-$function was originally introduced by Lovász [76] for estimating the Shannon capacity $\Theta(G)$. In this paper Lovász proved that $\Theta(G) \leq \vartheta(G)$ for every graph G. Hence $\vartheta(G)$ is also an upper bound on $\alpha(G)$. It is a very hard problem to determine $\Theta(G)$ and it is not known whether $\Theta(G)$ can be computed in polynomial time.

However, the Lovász number $\vartheta(G)$ can be computed in polynomial time. This follows from Theorem 5.8, item (4) and from the definition of $\vartheta(G)$, see (5.8). From a practical point of view one may wonder which of the several equivalent characterisations, given for instance in [78, 43, 67] lead to the most efficient algorithms in practice. We will use the following definition.

$$\vartheta(G) := \max\left\{\sum_{i=1}^{n} x_i | \, x \in TH(G)\right\}. \tag{5.8}$$

From consideration above we get that $\alpha(G) \leq \Theta(G) \leq \vartheta(G) \leq \chi(\bar{G})$. The reader is referred to [76] for additional properties of $\vartheta(G)$, and also to the survey [67].

Next we will see how $\vartheta(G)$ is related with SDP. In particular, we will show that the Lovász number can be characterized as the optimal objective solution from two distinct semidefinite formulations. To start, we note that **(STAB)** can alternatively be expressed as the quadratic problem in binary variables.

$$
\begin{array}{lll}
\textbf{(STAB)} \qquad & \alpha(G) = \max\ x^T x & \\
& \quad \text{s.t.} \qquad x_i x_j = 0 \quad (ij) \in E(G) & (5.9) \\
& \qquad\qquad\quad x_i^2 - x_i = 0 \quad i = 1, \ldots, n. & (5.10)
\end{array}
$$

Let y denote the characteristic vector for a (not necessarily largest) stable set. Suppose this set has cardinality k, i.e., $e^T y = k$ and $y^T y = k$. Hence $e^T y y^T e / y^T y = k$ which is, defining $Y = yy^T / y^T y$, equivalent to $\langle ee^T, Y \rangle = k$. Clearly, $Y \succeq 0$. Moreover, $x_i x_j = 0$ translates to $\langle E_{ij}, Y \rangle = 0$ for $(ij) \in E(G)$ and $\text{trace}(Y) = \text{trace}(yy^T)/y^T y = 1$. Hence,

$$
\begin{aligned}
(\mathbf{PSDP}\boldsymbol{\vartheta_S}) \qquad \alpha(G) \leq \vartheta^* = \max\ & \langle J, Y\rangle \\
\text{s.t. } & \langle E_{ij}, Y\rangle = 0 \quad \text{for all } (ij) \in E(G) \\
& \operatorname{trace}(Y) = 1 \\
& Y \in S(n),\ Y \succeq 0.
\end{aligned}
$$

Problem $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ is a semidefinite programming problem. $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ denotes the primal problem. Its associated dual problem denoted by $(\mathbf{DSDP}\boldsymbol{\vartheta_S})$ is

$$
\begin{aligned}
(\mathbf{DSDP}\boldsymbol{\vartheta_S}) \qquad & \min\ \lambda \\
& \text{s.t. } Z = \lambda I - J + \sum_{(ij)\in E(G)} y_{ij} E_{ij} \\
& Z \in S(n),\ Z \succeq 0,\ y \in \mathbb{R}^m, \lambda \in \mathbb{R}.
\end{aligned}
$$

Our original semidefinite problem, defining $\vartheta(G)$ uses the set $TH(G)$. The relaxation (5.8) can be written as the (primal) semidefinite program

$$
\begin{aligned}
(\mathbf{PSDP}\boldsymbol{\vartheta_L}) \qquad \vartheta(G) = \max\ & \operatorname{trace} X \\
\text{s.t. } & \langle E_{ij}, X\rangle = 0 \quad \forall (ij) \in E(G) \\
& d = \operatorname{diag}(X),\ \begin{pmatrix} X & d \\ d^T & 1 \end{pmatrix} \succeq 0.
\end{aligned}
$$

It follows from Lovász and Schrijver [78], Lemma 2.17, that actually $\vartheta^* = \vartheta(G)$. We will provide below a very simple argument for this fact. We include its proof, because our argument also has some interesting algorithmic implications, that will be used later on. We start with some technicalities.

Lemma 5.14 *Suppose that Y^* is optimal for $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$. Let ϑ^* denote the optimal objective value. Then there exists a matrix $\tilde{B}$ such that $Y^* = \tilde{B}^T\tilde{B}$ and $\tilde{B}e = (\sqrt{\vartheta^*}, 0, \ldots, 0)^T$.*

Proof. Since $Y^* \succeq 0$ there exists a matrix B such that $Y^* = B^T B$. Let Q be an orthogonal matrix having the first row parallel to $x := Be$. Since $x^T x = e^T Y^* e = \vartheta^*$, we get $Q = \begin{pmatrix} \frac{1}{\sqrt{\vartheta^*}} x^T \\ \tilde{Q} \end{pmatrix}$. Here $\tilde{Q}$ represents the orthogonal complement to x, $\tilde{Q}x = 0$. Defining $\tilde{B} = QB$ yields $\tilde{B}^T\tilde{B} = B^T Q^T Q B = B^T B = Y^*$ and $\tilde{B}e = QBe = Qx = \begin{pmatrix} \sqrt{\vartheta^*} \\ 0 \end{pmatrix}$.

□

Lemma 5.15 *Suppose that Y^* is optimal for $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$. Let ϑ^* denote the optimal objective value. Then*

$$\vartheta^* \operatorname{diag}(Y^*) = Y^* e.$$

Proof. Strong duality holds for $(\mathbf{PSDP}\boldsymbol{\vartheta}_{\mathbf{S}})$ and $(\mathbf{DSDP}\boldsymbol{\vartheta}_{\mathbf{S}})$, because both problems satisfy the Slater constraint qualification. In particular, the optimal $\lambda = \vartheta^*$. Let Z^* be an optimal solution for $(\mathbf{DSDP}\boldsymbol{\vartheta}_{\mathbf{S}})$. We have

$$Z^*_{ij} = \begin{cases} -1 & \text{for } i \neq j, (ij) \notin E(G) \\ -1 + y_{ij} & \text{for } i \neq j, (ij) \in E(G) \\ \vartheta^* - 1 & \text{for } i = j. \end{cases} \text{ and } Y^*_{ij} = \begin{cases} Y^*_{ij} & \text{for } i \neq j, (ij) \notin E(G) \\ 0 & \text{for } i \neq j, (ij) \in E(G) \\ Y^*_{ii} & \text{for } i = j. \end{cases}$$

Let $Y^* = (y^*_1, \dots, y^*_n)$ and $Z^* = (z^*_1, \dots, z^*_n)$, respectively. The optimal pair (Y^*, Z^*) satisfies $Y^* Z^* = 0$. So we get

$$(y^*_i)^T z^*_i = Z^*_{ii} Y^*_{ii} + \sum_{(ij) \in E(G)} Z^*_{ij} Y^*_{ij} + \sum_{(ij) \notin E(G)} Z^*_{ij} Y^*_{ij} = (\vartheta^* - 1) Y^*_{ii} - \sum_{(ij) \notin E(G)} Y^*_{ij}.$$

Since $(y^*_i)^T z^*_i = 0$ for all $i = 1, \dots, n$ we have that $\vartheta^* Y^*_{ii} = Y^*_{ii} + \sum_{(ij) \notin E(G)} Y^*_{ij}$ $= \sum_{j=1}^{n} Y^*_{ij} = (Y^* e)_i$.

□

Lemma 5.16 *Suppose that Y^* is optimal for* $(\mathbf{PSDP}\boldsymbol{\vartheta}_{\mathbf{S}})$. *Let ϑ^* denote the optimal objective value and let $d = \vartheta^* \operatorname{diag}(Y^*)$. Then*

$$\bar{Y} = \begin{pmatrix} \vartheta^* Y^* & d \\ d^T & 1 \end{pmatrix} \succeq 0.$$

Proof. Showing that $\bar{Y} \succeq 0$ is equivalent to showing that $\vartheta^* Y^* - dd^T \succeq 0$ by Schur complement, see Lemma A.9. Let $\tilde{B}$ be the matrix with the property from Lemma 5.14. Hence we get

$$Y^* e = \tilde{B}^T \tilde{B} e = \tilde{B}^T \begin{pmatrix} \sqrt{\vartheta^*} \\ 0 \end{pmatrix} = \sqrt{\vartheta^*} \tilde{B}^T_{.1}.$$

By Lemma 5.15 we have $d = Y^* e$. So we get

$$\begin{aligned} \vartheta^* Y^* - dd^T &= \vartheta^* Y^* - (Y^* e)(Y^* e)^T \\ &= \vartheta^* \tilde{B}^T \tilde{B} - \vartheta^* \tilde{B}^T_{.1} \tilde{B}_{1.} \\ &= \vartheta^* \left(\sum_{j=1}^{d} \tilde{B}^T_{.j} \tilde{B}_{j.} - \tilde{B}^T_{.1} \tilde{B}_{1.} \right) \\ &\quad \text{(by matrix multiplication via outer products)} \\ &= \vartheta^* \sum_{j=2}^{d} \tilde{B}^T_{.j} \tilde{B}_{j.}. \end{aligned}$$

Clearly, $\sum_{j=2}^{d} \tilde{B}_{.j}^T \tilde{B}_{j.} \succeq 0$ and hence $\bar{Y} \succeq 0$.

$\square$

In summary, we have shown the following result.

Corollary 5.17 *Suppose that Y^* is optimal for* $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$. *Let ϑ^* denote the optimal objective value. Then*

(i) $\vartheta^* \mathrm{diag}(Y^*) \in TH(G)$, *i.e.,* $\begin{pmatrix} \vartheta^* Y^* & \vartheta^* \mathrm{diag}(Y^*) \\ \vartheta^* \mathrm{diag}(Y^*)^T & 1 \end{pmatrix}$ *is feasible for* $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$.

(ii) $\mathrm{trace}(\vartheta^* Y^*) = \vartheta^*$.

Proof. Since Y^* is optimal for $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ it follows that $Y^*_{ij} = 0$ for $(ij) \in E(G)$. Clearly, this remains preserved for $\vartheta^* Y^*$. In view of (5.5) it remains to show that there exists a matrix $\tilde{Y} = (\tilde{y}_{ij})$ such that $\mathrm{diag}(\tilde{Y}) = \vartheta^* \mathrm{diag}(Y^*)$, $\tilde{y}_{ij} = 0$ for $(ij) \in E(G)$ and $\begin{pmatrix} \tilde{Y} & \mathrm{diag}(\tilde{Y}) \\ \mathrm{diag}(\tilde{Y})^T & 1 \end{pmatrix} \succeq 0$. We have shown that in Lemma 5.16. This proves (i).

Since Y^* is optimal for $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ we have $\mathrm{trace}(Y^*) = 1$. This completes the proof.

$\square$

So we have seen that an optimal solution to $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ provides a feasible solution of $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$ with the same objective function value. We show next the easier direction that feasible solutions of $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$ yield feasible solutions of $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ having no smaller cost.

Lemma 5.18 *Suppose that* $\bar{X} = \begin{pmatrix} X & d \\ d^T & 1 \end{pmatrix}$ *is feasible for* $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$ *and* $\mathrm{trace} X = \gamma > 0$. *Then*

$Y = \frac{1}{\gamma} X$ *is feasible for* $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ *and* $\langle J, Y \rangle \geq \gamma$.

Proof. Obviously $\begin{pmatrix} Y & \frac{1}{\gamma} d \\ \frac{1}{\gamma} d^T & \frac{1}{\gamma} \end{pmatrix} \succeq 0$. Hence $Y - \gamma \frac{1}{\gamma^2} dd^T \succeq 0$ by Schur complement, see Lemma A.9. From this we conclude that

$$e^T Y e \geq \frac{1}{\gamma} e^T dd^T e = \frac{1}{\gamma} (e^T d)^2 = \gamma.$$

Here the last equation follows from $d = \mathrm{diag}(X)$ and $\mathrm{trace}(X) = \gamma$. Since $x_{ij} = 0$ for $(ij) \in E(G)$ we have $y_{ij} = 0$ for $(ij) \in E(G)$. X as principal submatrix of the positive semidefinite matrix $\bar{X}$ implies that $Y \succeq 0$. We also see that $\operatorname{trace}(Y) = \frac{1}{\gamma}\operatorname{trace}(X) = 1$. This completes the proof.

□

We summarize the previous discussion in the following theorem, showing not only the well known fact that the two models are equal, but also how optimal solutions from one problem can be transformed into optimal solutions of the other problem.

Theorem 5.19 *Suppose that Y^* is optimal for* **(PSDP$\boldsymbol{\vartheta_S}$)** *and* $\bar{X}^* = \begin{pmatrix} X^* & d \\ d^T & 1 \end{pmatrix}$ *where $d = \operatorname{diag}(X^*)$ is optimal for* **(PSDP$\boldsymbol{\vartheta_L}$)***, respectively. Let ϑ^* and $\vartheta(G)$ denote the corresponding optimal objective values for* **(PSDP$\boldsymbol{\vartheta_S}$)** *and* **(PSDP$\boldsymbol{\vartheta_L}$)***, respectively. Then*

(i) $\bar{X} = \begin{pmatrix} \vartheta^* Y^* & \vartheta^* \operatorname{diag}(Y^*) \\ \vartheta^* \operatorname{diag}(Y^*)^T & 1 \end{pmatrix}$ *is optimal for* **(PSDP$\boldsymbol{\vartheta_L}$)**.

(ii) $Y = \frac{1}{\vartheta(G)} X^*$ *is optimal for* **(PSDP$\boldsymbol{\vartheta_S}$)**.

In particular, this gives $\vartheta^* = \vartheta(G)$.

Proof. Corollary 5.17 shows that $\bar{X}$ is feasible for **(PSDP$\boldsymbol{\vartheta_L}$)**. The corresponding objective value is $\operatorname{trace}(\vartheta^* Y^*) = \vartheta^*$. Note that **(PSDP$\boldsymbol{\vartheta_L}$)** is a maximization problem, hence $\vartheta^* \leq \vartheta(G)$. On the other hand, Lemma 5.18 implies that $\vartheta(G) \leq \langle J, Y \rangle \leq \vartheta^*$. Here the last inequality follows from the fact that **(PSDP$\boldsymbol{\vartheta_S}$)** is a maximization problem. This completes the proof.

□

From a theoretical point of view both models are equivalent. Model **(PSDP$\boldsymbol{\vartheta_S}$)** however can be solved much faster than **(PSDP$\boldsymbol{\vartheta_L}$)** (using interior-point methods). The running times from solving **(PSDP$\boldsymbol{\vartheta_S}$)** and **(PSDP$\boldsymbol{\vartheta_L}$)** are listed in Table 5.1. We further compare the performance with that from the publicly available software package CSDP, version 2.3., see [17], which also solves **(PSDP$\boldsymbol{\vartheta_S}$)**. This package is a C library for semidefinite programming. It contains a code exclusively developed for computing $\vartheta(G)$. The test problems are taken also from this package. Column n gives the number of vertices, column m the number of edges in the graph G.

Since we are interested in tightening the relaxation $TH(G)$ by adding further linear inequalities, we need a model, where those constraints can conveniently be included. In case of **(PSDP$\boldsymbol{\vartheta_L}$)** this is straightforward. It is not clear how the odd circuit constraints for instance could be included in **(PSDP$\boldsymbol{\vartheta_S}$)**. We therefore use the fast model **(PSDP$\boldsymbol{\vartheta_S}$)** to generate an initial relaxation Y which we transform (applying the above theorem) into an optimal solution $\bar{X}$ of **(PSDP$\boldsymbol{\vartheta_L}$)**. Starting from $\bar{X}$ we include further violated cutting planes, if there are any.

				CPU time [min/sec]		
File	n	m	$\vartheta(G)$	$(\mathbf{PSDP}\boldsymbol{\vartheta_S})$	$(\mathbf{PSDP}\boldsymbol{\vartheta_L})$	CSDP 2.3
g50	50	103	23.0000	0.36 s	1.41 s	0.36 s
g100	100	497	32.8792	5.07 s	31.64 s	5.99 s
g150	150	1105	42.1660	52.28 s	4 m 39.03 s	58.62 s
g200	200	1948	50.3212	5 m 41.26 s	26 m 29.84 s	5 m 45.24 s
g250	250	3027	57.2323	21 m 29.19 s	100 m 42.79 s	22 m 23.07 s
g300	300	4374	63.4771	64 m 06.25 s	372 m 37.16 s	65 m 56.15 s

Table 5.1: *Comparison of running times for computing* $\vartheta(G)$

5.5 Combining *TH(G)* with Linear Cutting Planes

A natural way to get improved bounds for $\alpha(G)$ consists in the combination of $TH(G)$ with linear cutting planes. We first observe that imposing the condition to $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$ that X is a $0-1$ matrix produces $\alpha(G)$. Therefore we may include in $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ any linear constraints that are satisfied by integer matrices. As already suggested in [52], one can include the triangle inequalities, $X \in MET$, to tighten the model $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$.

It should be mentioned that Schrijver [115] proposed a strengthening of the Lovász bound $\vartheta(G)$ by adding the nonnegativity constraints. In particular,

Theorem 5.20 [115] $\alpha(G) \leq \vartheta'(G) \leq \vartheta(G)$, *where*

$$\begin{array}{rcll} \vartheta'(G) & := & \max \langle J, Y\rangle & \\ & & s.t.\ \langle E_{ij}, Y\rangle = 0 & \textit{for all } (ij) \in E(G) \\ & & \langle E_{ij}, Y\rangle \geq 0 & \textit{for all } (ij) \notin E(G) \\ & & \operatorname{trace}(Y) = 1 & \\ & & Y \in S(n),\ Y \succeq 0. & \end{array}$$

The nonnegativity constraints are only a subclass of the constraints of MET, and there does not seem to be any reason to prefer them over the general constraints defining MET. Optimizing over $TH(G) \cap P_{MET}$ clearly dominates the Schrijver model. (In a slight abuse of notation we write $TH(G) \cap P_{MET}$ but mean the matrix space appendent to $TH(G) \cap P_{MET}$ or the projection of it and similar for other sets.)

A further refinement of $TH(G)$ with problem-specific cutting planes can be achieved by intersecting $TH(G)$ with $CSTAB(G)$. The following result was shown by Grötschel, Lovász and Schrijver.

Lemma 5.21 [42] *Let $G = (V(G), E(G))$ an edge weighted graph and let $w : E(G) \mapsto \mathbb{R}_+$ be a function that assigns the weight to its edges. Then a shortest odd circuit (with respect to the total weight of the circuit) can be found in polynomial time.*

Since the separation problem for $CSTAB(G)$ reduces to the shortest path problem in graphs with nonnegative edge weights, see Grötschel et al. [42], we can easily detect constraints violating $CSTAB(G)$ and incorporate them in our model.

It should be noted that $CSTAB(G)$ imposes constraints only on the main diagonal of the matrices in question. Directly solving this relaxation is computationally nontrivial, as there could be an exponential number of constraints. In contrast, MET has only $O(n^3)$ constraints.

All the preceding considerations give rise to the following relaxations, where (5.12) was already investigated in [52].

$$\begin{aligned} \max e^T d \quad \text{such that} \quad & d \in TH(G) \\ & d \in TH(G) \cap CSTAB(G) && (5.11) \\ & d \in TH(G) \cap P_{MET} && (5.12) \\ & d \in TH(G) \cap CSTAB(G) \cap P_{MET}. && (5.13) \end{aligned}$$

Starting from $TH(G)$ it is possible to proceed by either adding only odd ciruits or only constraints from P_{MET}, or by including the most violated constraints from both sets. From Theorem 5.12 we get the following result.

Lemma 5.22 $TH(G) \cap P_{MET} = TH(G) \cap CSTAB(G) \cap P_{MET}$,
$X \in TH(G) \cap P_{MET} \Rightarrow X \in TH(G) \cap CSTAB(G)$.

Ultimately we are interested in getting improved bounds on $\alpha(G)$ as quickly as possible. Practical experience has shown that optimizing over $TH(G) \cap P_{MET}$ is very expensive, see also [52]. At the beginning the inclusion of a moderate number of triangles and odd circuits induces a significant improvement of the bound on $\alpha(G)$ in comparison with adding the most violated triangles only. However, this favorable effect typically vanishes with increasing number of added triangles. (The more triangles are added the more odd circuits become inactive and therefore are unimportant.) If many triangles are incorporated in the model then the relaxations (5.13) and (5.12) are on a par. Concluding we re-emphasize that optimizing over (5.13) dominates optimizing over (5.12) at the beginning only. For us this is a crucial point and therefore the tightest relaxation that we consider in this work consists in intersecting $TH(G)$ with P_{MET} and $CSTAB(G)$. We therefore try to get a good (outer) approximation of $TH(G) \cap CSTAB(G) \cap P_{MET}$.

It is the main point of this work to compare the quality of the different relaxations and investigate their computational effort on a series of test problems. The computational effort to solve these relaxations is heavily influenced by the number of constraints included into the semidefinite program. Even if a constraint is inactive at the optimum, and hence could have been ignored without weakening the relaxation, it causes the same computational cost as active constraints. The computational effort is ultimately growing at least cubically with the number of constraints (because a linear system has to be solved for the dual variables), therefore it is crucial to identify constraints which are active at the optimum. For this reason we propose an iterative procedure and include only a few of the most violated inequalities and drop constraints as they become inactive. To get a feeling for the computational effort (at least approximately) to solve the above relaxations, we include in Table 5.2 some computation times on several randomly generated problems of various size.

This table compares computation times from optimizing over (5.13), (5.11) and $TH(G)$. Column n gives the respective order of the graphs. We separate the espected value and the number of edges of these graphs in column *density* by a dash. Each line lists the average CPU time to solve 10 problems of given size and density. The values in the parenthesis indicate the (average number of) active constraints at termination. It should be noted that optimizing over P_{MET} is not done exactly, but the iterations are stopped once the maximum violation of P_{MET} has fallen below a threshold (of 50 % of the maximum violation in our case). Consequently, adding triangles affects adversely the efficiency of the combined semidefinite and linear cutting plane algorithm.

On the positive side, optimizing over (5.11) is not much harder than computing $\vartheta(G)$ and as can be seen later leads to significantly improved bounds on $\alpha(G)$. This justifies our approach.

		CPU time [sec] from optimizing over				
n	*density* [%]	$TH(G)$	$TH(G) \cap CSTAB(G)$		(5.13)	
50	10 - 125	0.7125	4.3357	(10)	8.8111	(8/47)
100	10 - 500	5.0945	41.4115	(9)	141.8472	(2/180)
150	5 - 559	10.0519	150.7397	(42)	354.3603	(22/266)
200	3 - 597	16.1386	309.5576	(94)	958.8453	(63/383)
250	2 - 623	25.5563	424.7727	(95)	1458.3181	(98/297)

Table 5.2: *Average Running Times on randomly generated sparse graphs*

5.6 Implementation Details

In this section we discuss the essential ingredients of our cutting plane technique.

The implementation of our algorithm involves solving a number of semidefinite problems. These programs are solved using primal-dual path-following interior-point methods. The search direction used in our implementation was proposed independently by Helmberg et al. [54], Kojima et al. [73] and later based on a quite different approach by Monteiro [95]. Therefore this search direction is called HKM-direction, see also section 1.4.1. All our codes are implemented in Matlab with interfaces in C.

5.6.1 Computing $\vartheta(G)$

We first compute the optimal solution of **(PSDPϑ_S)**. Let Y^* denote this solution. From Theorem 5.19, it follows that the Lovász number $\vartheta(G)$ is the corresponding optimal objective value. Moreover also from Theorem 5.19 we know that

$$\bar{X} = \begin{pmatrix} X & d \\ d^T & 1 \end{pmatrix}, \text{ where } X = \vartheta(G)Y^* \text{ and } d = \operatorname{diag}(X),$$

is optimal for **(PSDPϑ_L)**. The optimal objective values for the models **(PSDPϑ_S)** and **(PSDPϑ_L)** coincide (see Theorem 5.19). We take it as initial upper bound on $\alpha(G)$.

5.6.2 Practical Formulation of the Model defining $\vartheta(G)$

In this section we reformulate **(PSDPϑ_L)**. This formulation is used in our implementation of the interior-point algorithm. Recall, our original semidefinite problem, defining $\vartheta(G)$ uses the set $TH(G)$. We express the constraints from (5.5) through $\bar{X} \in S(n+1)$. Then $\langle E_{ij}, \bar{X}\rangle = 0$ corresponds to $\bar{X}_{ij} = 0$ and $\operatorname{diag}(X) = d$ is equivalent to $\bar{X}_{ii} = \frac{1}{2}\bar{X}_{i,n+1} + \frac{1}{2}\bar{X}_{n+1,i}$ for $1 \le i \le n$. Setting $a_i := e_i - \frac{1}{2}e_{n+1}$, and using $\bar{X}_{n+1,n+1} = 1$, the preceding equation translates to $\langle a_i a_i^T, \bar{X}\rangle = \frac{1}{4}$. The relaxation (5.8) can therefore be written as the (primal) semidefinite program

$$(\mathbf{PSDP}\boldsymbol{\vartheta_L}) \qquad \begin{aligned} \vartheta(G) = \max\ & \operatorname{trace}\bar{X} - 1 \\ \text{s.t. } & \langle E_{ij}, \bar{X}\rangle = 0 \quad \forall (ij) \in E(G) \\ & \bar{X}_{n+1,n+1} = 1 \\ & \langle a_i a_i^T, \bar{X}\rangle = \frac{1}{4} \quad 1 \le i \le n \\ & \bar{X} \in S(n+1),\ \bar{X} \succeq 0. \end{aligned}$$

The number of constraints in this model is $m+n+1$. If $b \in \mathbb{R}^{m+n+1}$ denotes the right hand side of the above equalities, then the associated dual problem **(DSDPϑ_L)** to **(PSDPϑ_L)** is

$$(\mathbf{DSDP}\boldsymbol{\vartheta_L}) \qquad \begin{aligned} \min\ & \langle b, y\rangle - 1 \\ \text{s.t. } & Z = \sum_{(ij)\in E(G)} y_{ij}E_{ij} + y_{m+1}E_{n+1} + \sum_{k=1}^{n} y_{m+1+k}a_k a_k^T - I \\ & Z \in S(n),\ Z \succeq 0,\ y \in \mathbb{R}^{m+n+1}. \end{aligned}$$

The size of any stable set in G is a lower bound on $\alpha(G)$. We use the following rounding method to get a "good" lower bound for the problem.

5.6.3 Rounding

The main diagonal of the optimal $X \in S(n)$ is used to generate a stable set as follows. A large entry x_{ii} indicates that vertex i is likely to be contained in a large stable set. Hence we run through all the vertices of G in the order given by the main diagonal of X sorted nonincreasingly. We select the vertices to be contained in a stable set with probability proportional to x_{ii}. If vertex j is accepted to be in a stable set we exclude all the neighbours of j. The final stable set is locally improved in a second phase of the rounding heuristic. The complete process is repeated several times. We include the size of the largest stable set found this way in the tables of section 5.7. It is obvious that this number is a lower bound on $\alpha(G)$.

Let γ denote any upper bound on $\alpha(G)$. Clearly, $\lfloor\gamma\rfloor$ remains an upper bound on $\alpha(G)$. If $\lfloor\gamma\rfloor$ and the lower bound which we have found using the method sketched in the last preceding paragraph do not coincide we try to tighten the current relaxation. The next section describes the way how we do that.

5.6.4 Tightening the Relaxation

The initial semidefinite bound $\vartheta(G)$ (in a general phase of our algorithm the upper bound obtained from solving the current model) can be improved now by identifying and including violated odd circuit constraints and/or triangle inequalities. These extra constraints are called cutting planes, see section 5.5. In general there are a lot more violated cutting planes than we are willing to add to the current relaxation. The computational effort to reoptimize is growing drastically with the number of constraints. For this reason we follow an iterative procedure and add only the most violated constraints. (Hence we solve a series of stronger and stronger relaxations.) Since interior-point techniques are used for solving the relaxations we have to assure that the initial point is an interior point (not necessarily feasible point). Hence the current point has to be modified for restarting. A detailed description of this modification is given in section 5.6.5.

Now let us turn to the process of adding cutting planes and reoptimizing. For practical reasons we do not solve to optimality after each phase of adding cutting planes. While iterating we search for cutting planes which are violated by the current iterate once primal-dual feasibility is reached approximately. Once a significant violation is detected we perform some more Newton steps before strengthening the model. This is being done with the aim of increasing the number of violated cutting planes. It also should be noted that solving to optimality causes additional computational effort but does not provide additional information. Another advantage is that the non-optimal solution provides a well centered starting point which is not the case for optimal solutions. Starting from points close to the boundary (optimal solutions are such points) forces the iterates towards the center of the feasible region and only now the iterates move towards optimality.

On the negative side though, the detected cutting planes may be inactive at the optimum of the tightened relaxation or even not violated by the optimal solution of the current relaxation. So these cutting planes at hand are completely useless. In this case, if we had solved the updated relaxation to optimality we had dropped the inactive constraints (see below). Secondly, if we had solved the current relaxation to optimality we either had observed that no cutting plane had been violated or the search for cutting planes had provided important inequalities. Nevertheless, the stated advantages dominates these drawbacks. Actually, these limitations can be minimized. That will become clear subsequently.

After collecting inequalities performing the described process repeatedly we solve the model at present to optimality. At this point we have several options to proceed. One is that we can stop. If we do that we can take the current optimal objective value as an improved upper bound on $\alpha(G)$ (which it is in general). However, we can repeat the complete process as well. A major difference in starting from an optimal solution is that inactive constraints, which clearly are unimportant, can be dropped out from the model before separating further violated cutting planes. (In practice we drop a constraint if its corresponding dual variable has fallen below a given threshold value.) Clearly, this particular purge is not possible in the phase where we do not solve to optimality.

The code is implemented in such a way, that, whenever we want solve an improved relaxation, we can restart with the so far tightest relaxation model as initial relaxation and do the steps described above. In our experiments we found that computing the relaxation (5.11) can be done efficiently with computational effort of the same order of magnitude as computing $\vartheta(G)$. On the other hand we found it impractical to optimize over all constraints P_{MET}, because there was a typical tailing off effect. The biggest improvement of the bound is obtained in the first few iterations of adding violated inequalities. In subsequent iterations, the bound is hardly reduced, the largest violation of P_{MET} is already quite small, but there are still a lot of inequalities violating P_{MET}. In this situation it does not seem to be advisable to continue the iterations, because of the excessive computational effort. Further details are contained in section 5.7.

5.6.5 Restarting

Let the linear operator $\mathcal{B}(\cdot)$ be defined such that all added cutting planes (including active ones from previous adds and new violated ones) with complying right hand side b may be written as $\mathcal{B}(X) \leq b$. Clearly, the current optimal X is no longer an interior point, i.e., it violates $\mathcal{B}(X) \leq b$, but the interior-point machinery requires an initial $X \succ 0$ which strictly satisfies these inequalities. We use a simple trick in formulating the current semidefinite model. We introduce a primal slack variable $s = b - \mathcal{B}(X) \geq 0$. We then introduce an artificial dual slack variable $t \geq 0$ for the dual variable r corresponding to s. In fact, this way we get the additional dual constraint $t = r$. This restatement of the semidefinite model simplifies the update of the variables used in the interior-point code. To get the initial variables $(X^\circ, y^\circ, Z^\circ)$ we implemented a simple backtracking strategy from the current (X, y, Z) in the direction to a well centered point in the interior of the feasible set of $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$. This is a standard technique, therefore we do not restate it here precisely. The initial variables r°, s° and t° are constructed as follows. We set $(r_i, s_i, t_i) = (0, 0.1, 0.1)$ for every component i corresponding to a just added inequality. We take the current variables r, s and t for updating the remaining components. In fact,

$$r_i^\circ = \begin{cases} r_i & \text{if } r_i \geq 0.1 \\ 0 & \text{otherwise.} \end{cases}, \quad s_i^\circ = \begin{cases} s_i & \text{if } s_i \geq 0.1 \\ 0.1 & \text{otherwise.} \end{cases} \quad \text{and} \quad t_i^\circ = \begin{cases} t_i & \text{if } t_i \geq 0.1 \\ 0.1 & \text{otherwise.} \end{cases}$$

The initial point constructed this way is a strictly interior but not feasible point of the semidefinite model at the moment.

5.6.6 Optimizing over *CSTAB(G)*

We also report results from optimizing over the linear relaxation $CSTAB(G)$ of the stable set polytope. ILOG CPLEX 6.5 is used to solve the linear programs.

Note that $x = \frac{1}{3}e,\ x \in \mathbb{R}^n$ is contained in $CSTAB(G)$. This is true because for this x the constraints (5.4) can be expressed as $\frac{1}{3}|V(C)| \leq \frac{(|V(C)|-1)}{2}$ for every odd circuit C or equivalently $\frac{1}{3}|V(C)| - 1 \geq 0$. This inequality is satisfied for all odd circuits with $|V(C)| \geq 3$. Since optimizing over $CSTAB(G)$ is a maximization problem it follows that the optimal objective value is at least $\frac{n}{3}$. (In our extensive experiments it has been observed that optimizing over $CSTAB(G)$ often yields $x = \frac{1}{3}e$ if $\alpha(G) < \frac{n}{3}$.) Therefore we present numerical results from optimizing over $CSTAB(G)$ only if $\alpha(G) > \frac{n}{3}$.

5.7 Analysis of Numerical Results

In this section we report computational experience on different classes of test problems. It is the primary objective of this section to compare the quality of the different relaxations described above. We have already discussed the computational effort in section 5.5. Therefore running times are presented scantily. All our computational tests were performed on a DEC Alpha workstation.

5.7.1 Test Problems

Our numerical experiments were carried out on the following sets of graphs.

(i) Random graphs with given density,

(ii) Sanchis graphs, see [114, 50],

(iii) Mannino graphs, see [84],

(iv) Johnson graphs, see [50],

(v) Triangulated planar graphs,

(vi) Graphs with high girth,

(vii) Geometric graphs.

Sanchis graphs are hard instances for the minimum vertex cover problem (and hence for the equivalent maximum clique problem). Originally they were designed to test algorithms for these problems. These graphs are constructed in such a way that $\alpha(G)$ is known, see [114]. The independence number is also known for the Mannino graphs and the Johnson graphs. The Mannino graphs are obtained from the set covering formulation of the Steiner Triple problem [84] and the Johnson graphs emerge in coding problems [50]. For our tests on Sanchis, Mannino and Johnson graphs we take throughout the complement graphs and indicate this by suffix _c in the tables. These classes of graphs are some of the DIMACS benchmark instances and can be downloaded from the web.

Triangulated planar graphs are perfect, thus $\vartheta(G) = \alpha(G)$, see [43]. To generate graphs with high girth, we start with a randomly generated graph G and randomly replace edges of G by odd paths of prescribed maximum length.

Geometric graphs are constructed as follows. First we randomly generate points in the unit square. If the euclidean distance between two points is less than a prescribed distance then an edge connecting these two points is inserted.

5.7.2 Computations

We have already stated in section 5.6.1 that all experiments were performed starting from solving the semidefinite model $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ as initial relaxation. Thereafter we switch to the larger semidefinite formulation $(\mathbf{PSDP}\boldsymbol{\vartheta_L})$. This is done for performance reasons but primarily because a representation of the tightened relaxations using the small sized model $(\mathbf{PSDP}\boldsymbol{\vartheta_S})$ is not known, see also section 5.4.

The computational results on the seven problem sets are reported in Tables 5.3 to 5.9. Column n gives the number of vertices, column m the number of edges in the graph G. The column labeled *lbd on* $\alpha(G)$ stands for "lower bound on $\alpha(G)$". This column gives the size of the largest stable set which we have found using our heuristic method sketched in the previous section. The remaining columns contain the upper bounds found by our code. In addition to the value of the upper bound, we also provide the gap in % with respect to the values given in column *lbd on* $\alpha(G)$, separated from the upper bound by a dash. The numbers given in parenthesis represent the number of active inequality constraints of the final relaxation. Two numbers separated by a slash within one parenthesis show remaining odd circuit and triangle inequalities, active at termination, separately.

We first note that for triangulated planar graphs we are able to compute $\alpha(G)$, see Table 5.7. That is not surprising since these graphs are perfect, and hence

$\alpha(G) = \vartheta(G)$. It should be mentioned that the running times to obtain $\vartheta(G)$ are reasonably moderate for the tested graphs (even for large graphs). Nevertheless, this computational effort compares unfavourably with that required for optimizing over $CSTAB(G)$. By the way, note that already the linear bound on $\alpha(G)$ is tight. These findings show the limit of the semidefinite approach for this class of graphs. (It should be added that $\alpha(G) > \frac{n}{3}$ for all tested triangulated planar graphs.)

The results in Table 5.3 and in Table 5.8 illustrate that optimizing over (5.11) leads to substantial improvements of $\vartheta(G)$. The inclusion of additional inequalities from P_{MET} leads to more improvements, as can be seen from the last columns in these tables, however associated with an enormous increase of computational cost (see Table 5.2). Next, further notes the final solutions concerning are added on. They do not satisfy all the triangle inequalities $X \in MET$, but their violation has gone down from typically 0.1 to 0.03 (in the test sets of dimension 50 to 300 in Table 5.3). It is for the reasons stated in section 5.5 and in section 5.6.4 that we only did a limited number of rounds of adding inequalities on the remaining test data. No odd circuit constraints are violated, though. It should also be noted that only a few odd circuit constraints are active in the final solution. In more detail we have discussed this effect in section 5.5. Note that we have found already the optimal solution of many of the test problems with this approach. This is denoted by an asterisk in column *lbd on* $\alpha(G)$.

It is also interesting to see that the class of Sanchis graphs seems to be easy to approximate through $\vartheta(G)$. This is in contrast to several heuristic approaches, which have difficulties finding large stable sets in these graphs, see [16]. Table 5.5 shows our experiments on the class of Mannino graphs. We observed that no odd circuits get violated by the optimal $\bar{X}$ to $\mathbf{(PSDP\boldsymbol{\vartheta}_L)}$. Hence the relaxation $TH(G)$ cannot be further strenghtened by considering the odd circuits. The violation of the triangle inequalities is such small that a substantial improvement of the upper bound on $\alpha(G)$ through including the violated ones may not be expected. Therefore our experiments on the class of Mannino graphs are restricted on optimizing over $TH(G)$. Anyhow, in the leading two cases our rounding heuristic applied to the optimal $\bar{X}$ to $\mathbf{(PSDP\boldsymbol{\vartheta}_L)}$ yields a stable set with cardinality $\alpha(G)$. The size of the largest stable set in the remaining instance is also very close to $\alpha(G)$. This is again inconsistent with the heuristic approaches proposed in [16]. In the Tables 5.4 and 5.5 the column headed with BPG (stands for Bomze, Pellino and Giacomini) contains the size of the largest clique obtained by Bomze et al. [16].

Table 5.6 shows that optimizing over $TH(G)$ gives tight bounds on $\alpha(G)$. Note that we have found maximum independend sets for any tested ohnson graph.

We ran extensive tests on geometric graphs of different sizes and density. It turned out that $\vartheta(G)$ was tight in all instances. We present a choice of test results for this class of graphs in Table 5.9. The column headed by δ in this table contains the distances needed for the construction of these graphs, see section 5.7.1.

		lbd on	optimizing over				
n	m	$\alpha(G)$	$TH(G)$	$TH(G) \cap CSTAB(G)$		$TH(G) \cap P_{MET} \cap CSTAB(G)$	
50	251	14*	15.0016 - 7.15	14.9990 - 7.14	(1)	14.6960 - 4.97	(1/130)
80	320	27*	28.1924 - 4.42	28.1053 - 4.09	(20)	27.5501 - 2.04	(8/315)
100	393	35	37.0451 - 5.84	36.8770 - 5.36	(21)	36.2010 - 3.43	(7/320)
100	434	33	35.1823 - 6.61	35.0532 - 6.22	(21)	34.3682 - 4.15	(3/361)
150	451	59	62.2523 - 5.51	61.4944 - 4.23	(77)	60.6595 - 2.81	(37/694)
150	552	55	58.0287 - 5.51	57.6335 - 4.79	(49)	56.5200 - 2.76	(21/710)
170	383	77*	79.3290 - 3.02	77.6756 - 0.88	(60)	77.5463 - 0.71	(44/318)
170	423	75*	76.7735 - 2.36	75.6992 - 0.93	(56)	75.4973 - 0.66	(40/305)
200	418	93*	94.5638 - 1.68	93.2256 - 0.24	(65)	93.0111 - 0.01	(52/401)
200	563	81	85.5002 - 5.56	84.3313 - 4.11	(99)	83.1449 - 2.65	(39/1059)
300	894	121	129.3327 - 6.89	127.7402 - 5.57	(153)	126.4218 - 4.48	(55/1195)
350	994	146	152.5927 - 4.52	150.3067 - 2.95	(183)	150.1405 - 2.84	(122/213)
400	835	191*	193.4866 - 1.30	191.0000 - 0.00	(132)	191.1058 - 0.06	(72/5)

Table 5.3: *Numerical Results on randomly generated sparse graphs. The relative error in % with respect to the values given in column lbd on $\alpha(G)$ is separated from the upper bound by a dash. The numbers given in parenthesis represent the number of active inequality constraints of the final relaxation. Two numbers separated by a slash within one parenthesis show remaining odd circuit and triangle inequalities, active at termination, separately. Column lbd on $\alpha(G)$ gives the size of the largest stable set found by rounding.*

File	n	m	$\alpha(G)$	BPG	*lbd on* $\alpha(G)$	optimizing over $TH(G)$	optimizing over (5.13)	
san200_0.9_1.clq_c	200	1990	70	46	70	70.0000	-	
san200_0.9_2.clq_c	200	1990	60	36	60	60.0000	-	
san200_0.9_3.clq_c	200	1990	44	33	44	44.0000	-	
sanr200_0.9.clq_c	200	2037	-	41	42	49.2735	48.8689	(0/294)

Table 5.4: *Numerical Results on Sanchis graphs including the results obtained by Bomze, Pellino and Giacomini in column BPG. Column lbd on $\alpha(G)$ gives the size of the largest stable set found by rounding.*

File	n	m	$\alpha(G)$	BPG	*lbd on* $\alpha(G)$	optimizing over $TH(G)$
MANN_a9.clq_c	45	72	16	16	16	17.4750
MANN_a27.clq_c	378	702	126	121	126	132.7629
MANN_a45.clq_c	1035	1980	345	336	343	356.0488

Table 5.5: *Numerical Results on Mannino graphs including the results obtained by Bomze, Pellino and Giacomini in column BPG. Column lbd on $\alpha(G)$ gives the size of the largest stable set found by rounding.*

File	n	m	$\alpha(G)$	*lbd on* $\alpha(G)$	optimizing over $TH(G)$
johnson8-2-4.clq_c	28	168	4	4	4.0
johnson8-4-4.clq_c	70	560	14	14	14.0
johnson8-16-2-4.clq_c	120	1680	8	8	8.0

Table 5.6: *Numerical Results on Johnson graphs. Column lbd on $\alpha(G)$ gives the size of the largest stable set found by rounding.*

			optimizing over			
n	m	$\alpha(G)$	$TH(G)$	CPU time [sec]	$CSTAB(G)$	CPU time [sec]
100	294	44	44.0000	2.2698	44.6667	0.9380
200	594	87	87.0000	16.5672	87.3333	4.5259
300	894	126	126.0000	53.6591	127.0000	13.0489
400	1194	174	174.0000	155.8248	174.6667	29.7529
500	1494	220	220.0000	335.4370	220.3333	64.3262
600	1794	270	270.0000	606.9892	270.0000	59.0361

Table 5.7: *Numerical Results on triangulated planar graphs*

		lbd on	optimizing over				
n	m	$\alpha(G)$	$TH(G)$	$TH(G) \cap CSTAB(G)$		$TH(G) \cap P_{MET} \cap CSTAB(G)$	
118	152	57*	58.6107 - 2.83	57.4000 - 0.70	(18)	57.3252 - 0.57	(14/83)
292	365	141*	144.2935 - 2.34	142.6667 - 1.18	(86)	141.9742 - 0.69	(10/181)
280	394	134*	137.1680 - 2.36	136.6476 - 1.98	(94)	134.4599 - 0.34	(3/215)
282	339	136	139.5288 - 2.59	137.6923 - 1.24	(19)	137.7490 - 1.29	(16/120)
372	534	179*	181.2776 - 1.27	179.0000 - 0.00	(96)	179.0000 - 0.00	(104/370)
448	465	220	223.8245 - 1.74	222.3333 - 1.06	(20)	222.3333 - 1.06	(19/1)
520	947	245*	250.1892 - 2.12	250.1473 - 2.10	(7)	245.9738 - 0.40	(0/354)
544	688	264	269.0532 - 1.91	267.0000 - 1.14	(104)	266.1425 - 0.81	(24/254)

Table 5.8: *Numerical Results on graphs with long odd circuits.The relative error in % with respect to the values given in column lbd on* $\alpha(G)$ *is separated from the upper bound by a dash. The numbers given in parenthesis represent the number of active inequality constraints of the final relaxation. Two numbers separated by a slash within one parenthesis show remaining odd circuit and triangle inequalities, active at termination, separately. Column lbd on* $\alpha(G)$ *gives the size of the largest stable set found by rounding.*

			lbd on	optimizing over
n	m	δ	$\alpha(G)$	$TH(G)$
50	128	0.2	16	16.0000
50	476	0.4	7	7.0000
50	698	0.6	4	4.0000
100	185	0.1	39	39.0000
100	561	0.2	21	21.0000
100	1065	0.3	13	13.0000
150	327	0.1	53	53.0000
150	1137	0.2	22	22.0000
150	2358	0.3	13	13.3710

Table 5.9: *Numerical Results on Geometric graphs. Column δ gives a fixed distance needed for constructing these graphs. Column lbd on $\alpha(G)$ gives the size of the largest stable set found by rounding.*

5.8 Concluding Comments

In this chapter, we presented a semidefinite interior-point cutting plane approach for approximating the maximum stable set problem. In section 5.4 we have investigated two different semidefinite relaxations for the maximum stable set problem. On the theoretical side we have provided a simple argument that these two models are equivalent. In particular, we have shown that their objective values agree at the optimum and how optimal solutions from one formulation can be transformed into optimal solutions of the other formulation.

Our computational tests give rise to the following conclusions. The essential message is that the process of adding and dropping cutting planes works robust and can significantly improve the quality of the final solution. This was demonstrated on many different categories of test instances. However there are several drawbacks that the suggested approach has. Recall it was the primary objective of this work to compare not only the quality of the different relaxations but also to investigate their computational effort on a series of test problems. We shall illustrate the pros and cons of our approach.

Taking $\vartheta(G)$ as initial upper bound on $\alpha(G)$ we strenghtened it through incorporation of violated cutting planes in the current relaxation. We used two classes of cutting planes, the generic triangle inequalities and the problem-specific odd circuit constraints.

It should be pointed out that the improvement goes at the expense of computation time, which increases dramatically with growing dimension. Clearly, the computational effort depends also on the number of added constraints as expatiated in detail in section 5.5.

Numerical experiments have shown that the relaxations (5.12) and (5.13) are on a par. It should be noted that at the beginning of tightening the initial model $\mathbf{(PSDP\vartheta_L)}$ the relaxation (5.13) compares favourably with the relaxation (5.12). However this effect vanishes with increasing number of added triangles. (This is discussed in more detail in section 5.5.)

In our computational experiments a combination of triangles and odd circuits produces the widest progress. We observed that most of the improvement is obtained in the first few rounds of adding inequalities. The small rest of the improvement goes at the expense of the computational effort. The running times listed in Table 5.2 show that optimizing over (5.13) once the maximum violation of the cutting planes has fallen below a threshold is pretty inefficient.

These limitations justify the optimization over the set (5.11). Clearly, the quality of the final bound reduce somewhat but this debasement is marginal. The running times listed in Table 5.2 makes it easier to take this loss. Consequently, the characteristic feature of the relaxation (5.11) is that it yields improved upper bounds on $\alpha(G)$ in moderate computation time. (Optimizing over (5.11) is not much harder than computing $\vartheta(G)$.)

Anyhow, it should be mentioned that for us it seems not to be clear which one of polyhedral methods and Semidefinite Programming should be prefered, see the numerical results in Table 5.7. Probably it depends on the structure of the graphs which one of these methods dominates the other.

A further disadvantage of our approach is the large memory requirement for large problems. This comes from the Cholesky factorization used in our implementation. We realized after the experiments that for large scale instances there is a need of finding another approach for solving the relaxations of the stable set polytope. A promising way for future research would be to examine the relaxations in the context of Branch and Bound methods for the stable set problem.

In chapter 6 we consider another NP-hard combinatorial optimization problem, the maximum cut problem. A linear programming relaxation which is strengthened by including various classes of cutting planes is presented. This approach would also be an interesting issue for future research in context with the maximum stable set problem.

Chapter 6

Relaxations for the Maximum Cut Problem

In this chapter we investigate polyhedral and non-polyhedral relaxations for the maximum cut problem. In general a complete linear description of the cut polyhedron is not known. But large classes of valid inequalities for this polyhedron are known. We propose to strengthen a linear approximation of a well known semidefinite relaxation through adding various classes of cutting planes. We include numerical experiments and confront the results from solving linear models with the results from solving nonlinear models.

6.1 The Maximum Cut Problem

Let $G = (V(G), E(G))$ denote an edge-weighted undirected graph with vertex set $V(G)$ and edge set $E(G)$, where $|V(G)| = n$ and $|E(G)| = m$. An edge e with endnodes i and j is represented by (ij) or if it fits more by e itself. We assume throughout that the graphs are loopless, that is, $(ii) \notin E(G)$ for any $i \in V(G)$. Given a graph $G = (V(G), E(G))$ and $U \subseteq V(G)$, the subgraph of G induced by U, denoted by $G[U]$, has node set U and edge set $\left\{(ij) \in E(G) | i, j \in U\right\}$. If $G[U]$ is a complete graph, then it is called a clique.

Let $S \subseteq V(G)$ and suppose that $P = \left\{S, V(G) \backslash S\right\}$ is a partition of $V(G)$. Consider the set $\delta(S)$ of edges joining S and its complement $V(G) \backslash S$; such a set of edges is called a cut. Given a weight function $w : E(G) \mapsto \mathbb{R}$, the weight w of a cut $\delta(S)$ is defined by $w(\delta(S)) := \sum_{e \in \delta(S)} w_e$, where w_e denotes the weight on the edge e. The *maximum cut problem*, denoted **(MC)**, is the problem of finding a cut for which the total weight is maximum, hence **(MC)** can be formulated as

$$(\mathbf{MC}) \quad \max\Big\{w(\delta(S)) \,|\, S \subseteq V(G), \delta(S) \subseteq E(G) \text{ cut}\Big\}.$$

The maximum cut problem is known to be NP-complete, see Karp [65, 66]. However, there are several classes of graphs for which the maximum cut problem can be solved in polynomial time; for instance for graphs with no long odd cycles, (see [44]), for planar graphs, (see [49, 101]), or more generally for graphs with no K_5-minor (see [9]). Readers interested in the problem should consult for instance the survey paper [106] which contains a comprehensive treatment of this subject.

Before we continue let us fix further terminology which will be used throughout this chapter. Instead of w_e we often use w_{ij} to denote the weight associated to the edge (ij) in G. In case we use a weight vector $w \in \mathbb{R}^{\binom{n}{2}}$ indexed by the edges of the complete graph, we assume that $w_{ij} = 0$ whenever (ij) is not an edge of G and $w_{ij} = w_e$ otherwise. We define $y(S) := \sum_{e \in S} y_e$. Here $y \in \mathbb{R}^m$ denotes a vector indexed by the edges of G with components y_e (or y_{ij}) and $S \subseteq E(G)$. Given a subset $S \subseteq E(G)$ the incidence vector of S is the vector χ^S indexed by the edges of G defined as $\chi^S_e = \begin{cases} 1 & \text{if } e \in S \\ 0 & \text{if } e \in E(G) \backslash S \end{cases}$.

The rest of this chapter is organized as follows. Section 6.2 presents some fundamental results on the maximum cut problem. In particular, we derive the well known semidefinite relaxation for the maximum cut problem and then approximate it through a linear relaxation. Large classes of inequalities are known to be valid for the cut polyhedron. In section 6.3 we survey the types of inequalities used as cutting planes. More recently, several authors published strengthened non-polyhedral relaxations for **(MC)**, based on the semidefinite relaxation we talked a few lines above. In section 6.4 we survey some of these papers. Given any graph G we will not change its support. However it turns out that, tightening the current relaxation e.g. through adding hypermetric inequalities, requires changing the given graph. We shall discuss this crucial point in section 6.5. For the time our way out is focusing on graphs whose cliques are known in advance. (In particular, we consider the class of chordal graphs.) In section 6.6 we discuss the essential practical aspects of our implementation. The computational experiments are reported in section 6.7. We conclude with comments in section 6.8.

6.2 Polyhedral and Non-Polyhedral Relaxations

Polyhedral relaxations for the maximum cut problem were considered e.g. in [102, 11, 9, 105]. Therein the cut polytope of G, denoted by $P_c(G)$,

$$P_c(G) = \text{conv}\Big\{\chi^{\delta(S)} \in \mathbb{R}^m \,|\, S \subseteq V(G), \delta(S) \subseteq E(G) \text{ cut}\Big\}$$

is studied in some detail. For instance, the facial structure of $P_c(G)$ has been investigated by Barahona and Mahjoub [11]. In particular, they characterize facet defining inequalities of $P_c(G)$ and show different methods for constructing such inequalities from known ones.

Further results presented in the papers listed above are given in the subsequent sections, because it is easier to formulate it using the results of the current section.

Let $c \in \mathbb{R}^m$ denote the vector having the weights on the edges as components. It is quite obvious that **(MC)** can be reformulated as

$$(\mathbf{MC}) \qquad w^* := \max\, c^T y,\ y \in P_c(G). \tag{6.1}$$

There are several other ways, different from (6.1), to model the maximum cut problem.

We now recall an integer programming model and starting from this formulation we derive the well known semidefinite relaxation of **(MC)**. Given a subset $S \subseteq V(G)$, a vector naturally associated with the partition $\left\{S, V(G)\backslash S\right\}$ is the cut vector $y \in \mathbb{R}^n$, defined by $y_i = -1$ if $i \in S$ and $y_i = +1$ if $i \in V(G)\backslash S$. (A synonym for "cut vector" used in literature is "partition vector".) Clearly, $(ij) \in \delta(S)$ if and only if $y_i y_j = -1$, hence $w(\delta(S)) = \sum_{(ij)\in E(G),\ i<j} w_{ij}\frac{1}{2}(1 - y_i y_j)$. Hence **(MC)** can be modeled as the following integer programming model.

$$(\mathbf{MC}) \quad w^* = \max\, \frac{1}{2} \sum_{(ij)\in E(G),\ i<j} w_{ij}(1 - x_i x_j) \ \text{ s.t. } \ x_i \in \left\{-1, +1\right\} \ \text{ for all } i \in V(G).$$

Let $X = (x_{ij})$ be defined as $X := xx^T$. Clearly $X \in S(n)^+$ and X is a rank one matrix. Hence **(MC)** may be restated as follows.

$$(\mathbf{MC}) \quad w^* = \max\, \frac{1}{2} \sum_{(ij)\in E(G),\ i<j} w_{ij}(1 - x_{ij}) \ \text{ s.t. } \ \operatorname{diag}(X) = e,\ \operatorname{rank}(X) = 1,\ X \in S(n)^+.$$

Let $\mathcal{W} = (w_{ij})$ be the weighted adjacency matrix of G and let $D = \operatorname{Diag}(\mathcal{W}e)$. Then the Laplacian matrix $L = (l_{ij})$ associated with the graph G is defined as $L := D - \mathcal{W}$. We then see that

$$\begin{aligned}
\langle L, X\rangle &= \langle D, X\rangle - \langle \mathcal{W}, X\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_{ii} - 2 \sum_{(ij)\in E(G),\ i<j} w_{ij} x_{ij} \\
&= \sum_{(ij)\in E(G),\ i<j} w_{ij}(x_{ii} + x_{jj}) - 2 \sum_{(ij)\in E(G),\ i<j} w_{ij} x_{ij} \\
&= \sum_{(ij)\in E(G),\ i<j} w_{ij}\left(x_{ii} + x_{jj} - 2x_{ij}\right) \\
&= 2 \sum_{(ij)\in E(G),\ i<j} w_{ij}(1 - x_{ij}).
\end{aligned}$$

Here the last equation follows from $\operatorname{diag}(X) = e$. Hence, $\frac{1}{4}\langle L, X\rangle = \frac{1}{2}\sum_{e\in E(G)} w_{ij}(1 - x_{ij})$ and the maximum cut problem can be formulated equivalently as

$$\textbf{(MC)} \qquad w^* = \max \frac{1}{4}\langle L, X\rangle \text{ such that } \operatorname{diag}(X) = e,\ \operatorname{rank}(X) = 1,\ X \in S(n)^+. \tag{6.2}$$

Now we drop the rank one condition and define $Q = \frac{1}{4}L$. Hence an upper bound on w^* is given by the optimal objective value μ^* of the following semidefinite program **(SDPMC)**.

$$\textbf{(SDPMC)} \qquad \mu^* = \max \langle Q, X\rangle \text{ such that } \operatorname{diag}(X) = e,\ X \in S(n)^+.$$

This relaxation is studied extensively, see e.g. [34, 36, 81, 25]. The quality of this bound was analyzed by Goemans and Williamson [36]. Assuming nonnegative weights they have shown that the ratio between w^* and the semidefinite upper bound μ^* is at least .878. For completeness we mention that it is NP-complete to approximate the maximum cut problem with a factor closer to one than .9412, see [59].

The semidefinite relaxation can be solved using interior-point mothods, see section 1.4. To get a feeling for the computational effort to solve **(SDPMC)** we list in Table 6.1 running times on several randomly generated problems of various size.

n	CPU time [sec]
100	0.5562
200	3.5255
300	11.9430
400	26.1739
500	54.1967
600	92.7256

Table 6.1: *Computing times on randomly generated problems for solving* **(SDPMC)**. *The experiments were performed on a DEC Alpha workstation; constant edge density 30 %.*

We see that the computational effort to solve **(SDPMC)** is growing drastically with the size n of the problem. This comes from the critical fact that, using interior-point mothods, the variable $X \in S(n)^+$ becomes completely dense in general. In this work we intend to solve large scale problems.

Our idea is to solve a linear programming relaxation for the maximum cut problem and then to tighten this relaxation by adding valid cutting planes. Our approach is based on the semidefinite formulation **(SDPMC)**. We first remove the positive semidefiniteness constraint and set $x_{ii} = 1$ for all $i = 1, \ldots, n$. Secondly we transform the problem into $(0, 1)-$variables y. The components of y corresponds to the upper triangle part of X (or equivalently to the edge set of the complete graph K_n), hence y is a vector of length $\binom{n}{2}$. The transformation used in this work is given by

$$X_{ij} = 1 - 2y_{ij}, \qquad y_{ij} = \frac{1 - X_{ij}}{2}. \tag{6.3}$$

Thus, using this transformation, it follows that we have

$$\langle Q, X \rangle = \sum_{i<j} w_{ij} y_{ij}.$$

Hence a linear programming relaxation, we denote it by **(LPMC)**, is given by

$$(\textbf{LPMC}) \qquad \max \sum_{i<j} w_{ij} y_{ij} \ \text{ such that } \ 0 \le y_{ij} \le 1.$$

It is easy to see that **(LPMC)** is simply a relaxation of (6.1). So our way yielding **(LPMC)** seems to be somewhat circumstantial but it shows how the trivial polyhedral relaxation is related to the non-polyhedral relaxation **(SDPMC)**. We use this transformation in practice.

In general the solution of **(LPMC)** corresponds not to a cut. One can strengthen this relaxation by including additional cutting planes. Valid classes for instance are triangle inequalities, cycle inequalities, semidefinite cuts or hypermetric inequalities. In the following section we proceed by describing these types of inequalities used as cutting planes.

6.3 Various Classes of Linear Inequalities as Cutting Planes

6.3.1 Triangle Inequalities

The inequalities defining the metric polytope of a graph G, denoted $MET(G)$, are the *triangle inequalities*. First we introduce the metric polytope $MET \subseteq S(n)$. By definition,

$$X \in MET \ :\Leftrightarrow \begin{cases} X_{ii} = 1 & \text{for } i = 1, \dots, n, \\ X_{ij} - X_{ik} - X_{jk} \ge -1 & \text{for } 1 \le i, j, k \le n, \\ X_{ij} + X_{ik} + X_{jk} \ge -1 & \text{for } 1 \le i, j, k \le n. \end{cases} \tag{6.4}$$

Given a graph G, $MET(G)$ is defined as the projection of MET on the subspace $\mathbb{R}^m$ indexed by the edge set of G. Obviously, considering MET is equivalent to considering $MET(K_n)$. Both sets contain the same information.

Using the transformation (6.3), $MET(K_n)$ can be represented as

$$MET(K_n) = \Big\{ y \in \mathbb{R}^{\binom{n}{2}} \,|\, y \text{ satisfies } \; y_{ij} - y_{ik} - y_{jk} \leq 0, \;\; -y_{ij} - y_{ik} + y_{jk} \leq 0, \\ -y_{ij} + y_{ik} - y_{jk} \leq 0, \; y_{ij} + y_{ik} + y_{jk} \leq 2. \Big\}$$

Note, combining the inequalities defining $MET(K_n)$ conveniently, already implies $0 \leq y_{ij} \leq 1$.

6.3.2 Cycle Inequalities

Suppose that $x \in \mathbb{R}^m$ is an incidence vector of an arbitrary cut. Then for each cycle C, $x(C) = \sum_{e \in C} x_e$ is an even number. This follows from the fact that a cut and a cycle intersect in an even number of edges. Hence we get the following class of inequalities.

$$x(F) - x(C \backslash F) \leq |F| - 1 \text{ for each cycle } C, \; F \subseteq C, \; |F| \text{ odd.} \tag{6.5}$$

These inequalities are known as the *cycle inequalities*. Barahona [10] proved that the cycle inequalities form a linear description of $MET(G)$. In particular, he showed that

$$MET(G) = \Big\{ y \in \mathbb{R}^m \,|\, y \text{ satisfies } 0 \leq y_{ij} \leq 1 \text{ for } (ij) \in E(G) \text{ and } (6.5) \Big\}.$$

Let $|C| = 3$ and $F = C$. Then (6.5) is equivalent to $x(F) \leq 2$. For $|C| = 3$ and $F \subseteq C, |F| = 1$ (6.5) is equivalent to $x(F) - x(C \backslash F) \leq 0$. These two inequalities represent exactly the triangle inequalities (6.4) for $(0,1)-$variables, hence the triangle inequalities are special cases of cycle inequalities. Thus

$$MET(K_n) = \Big\{ y \in \mathbb{R}^m \,|\, y \text{ satisfies } (6.5) \Big\}.$$

We note that Barahona and Mahjoub [11] present a polynomial algorithm to solve the separation problem for the cycle inequalities. (Given a point $x \in \mathbb{R}^m$ the separation problem consists in either verifying that $x \in P_c(G)$ or else finding a hyperplane that separates x from $P_c(G)$.)

6.3.3 Hypermetric Inequalities and Semidefinite Cuts

A fairly general class of valid inequalities are the so called hypermetric inequalities, see Deza and Laurent [26, 27]. For completeness we review the essential facts. Let b be an integer vector such that $\sum_{i=1}^n b_i$ is odd. This condition guarantees that $|x^T b| \geq 1$ for all vectors $x \in \Big\{-1, +1\Big\}^n$. Moreover, we see that

$$|x^T b| \geq 1 \Leftrightarrow (x^T b)(b^T x) \geq 1 \Leftrightarrow \text{trace}(xx^T)(bb^T) \geq 1.$$

Note, every matrix $X \in S(n)^+$ such that $\mathrm{diag}(X) = e$, $\mathrm{rank}(X) = 1$ has the simple form $X = xx^T$, where x is some $(-1, +1)-$vector. Recall, **(MC)** can be modeled such that exactly these matrices form the feasible set.

Hence, if X is feasible for (6.2) then

$$\mathrm{trace} X(bb^T) \geq 1, \text{ where } b \text{ is some integer vector and } \sum_{i=1}^{n} b_i \text{ odd} \tag{6.6}$$

is satisfied. Thus, the inequalities (6.6) can be used for strengthening the model **(SDPMC)**. The set of inequalities generated this way are known as the *hypermetric inequalities.*

For the sake of completeness we note that the triangle inequalities (6.4) are a very special case of hypermetric inequalities. (Let b be the incidence vector of the triangle (i, j, k), then (6.6) can be rewritten as $X_{ij} + X_{ik} + X_{jk} \geq -1$; Setting $b_k = -1$ leads to $X_{ij} - X_{ik} - X_{jk} \geq -1$.)

Clearly, any feasible X for **(SDPMC)** violates (6.6) by at most one. (Obviously X satisfies the positive semidefiniteness characterization $\mathrm{trace} X(vv^T) \geq 0$ for all $v \in \mathbb{R}^n$.) Therefore the homogeneous formulation of (6.6) for all $b \in \mathbb{R}^n$ (note, here the components must not sum up to an odd number) is called the class of *semidefinite cuts.* In fact, this class of inequalities is given by

$$\mathrm{trace} X(vv^T) \geq 0, \quad \text{for all } v \in \mathbb{R}^n. \tag{6.7}$$

6.4 Tighter Non-Polyhedral Relaxations

Poljak and Rendl [103] were the first who proposed to strengthen the positive semidefinite relaxation **(SDPMC)** through including the triangle inequalities. Computational results are presented for instance in [54, 112]. The semidefinite bound obtained from solving **(SDPMC)** can also be improved by adding hypermetric inequalities to the semidefinite model, see [53].

In [7] Balas, Ceria and Cornuejols suggested to lift a mixed $0-1$ problem in a higher dimensional space, where the resulting relaxation can be tightened by imposing the $0-1$ condition on one or more variables. The tightened formulation is then projected back onto the original space. Lovász and Schrijver [78] were the first who proposed to formulate $0-1$ integer programs in the higher dimensional space of symmetric positive semidefinite matrices. Based on the ideas from Lovász and Schrijver [78] and from Balas et al. [7, 8] a general framework to get semidefinite relaxations for constrained $0-1$ quadratic programming was presented by Helmberg et al. [52]. This approach is used in [51, 54] to obtain tighter bounds for the maximum cut problem.

More recently Anjos and Wolkowicz [4] presented a strengthened semidefinite programming relaxation for the maximum cut problem. This relaxation is based on **(SDPMC)**. In particular, first they add certain redundant constraints from **(MC)** and then their tighter semidefinite relaxation comes from the Lagrangian dual of the Lagrangian dual. This procedure was already suggested in [104, 136]. It is shown that the new relaxation yields a strengthened bound for the maximum cut problem relative to the bound from **(SDPMC)**. This is shown through numerical experiments and also theoretically. On the negative side though, they observed that solving the new relaxation is very expensive compared with solving **(SDPMC)**.

6.5 Tightening the Relaxation (LPMC)

Several polyhedral relaxations of **(MC)** were surveyed by Poljak [102]. One, he considered, is defined by $0 \leq y_{ij} \leq 1$ and (6.5). We will denote the optimal objective value of this relaxation by σ^*. Assuming the graph is nonnegative weighted Poljak observed that the ratio between the polyhedral upper bound σ^* and the optimum value of the maximum cut problem can be arbitrarily close to two.

In this work we consider general weighted graphs. Taking **(LPMC)** as initial relaxation we suggest to consider strengthened LP relaxations for the maximum cut problem, which are obtained by including various classes of cutting planes (see section 6.3).

6.5.1 Dealing with Unspecified Entries of X

Our goal is to solve the relaxation **(LPMC)** and also the strengthened ones only on the support of the given graph G. However, we have to separate the inequalities after lifting the current solution from the subspace $\mathbb{R}^m$ into the $(-1,+1)-$representation in $S(n)$. This is done using the transformation (6.3). Note that this representation (we will denote it by X) has unspecified entries on the positions corresponding to the nonedges in G; the diagonal entries of X are set all equal to one.

We have no problems in separating the cycle inequalities in $S(n)$ and projecting it back to the subspace $\mathbb{R}^m$, indexed by the edges of G, because this class of cutting planes operates on the support of G. However, if we separate (violated) semidefinite cuts or (violated) hypermetric inequalities then it is very likely that unspecified entries become covered. In particular, if that is we have to insert absent edges into the graph. But it is not clear which value should be assigned to the unspecified entries in X.

This is the matrix completition problem, see e.g. [74, 61]. In general this problem asks whether a partially given matrix can be completed to a full matrix by defining the unspecified entries in a way such that the completed matrix satisfies a desired property. It is a promising direction for future research to allow changing the graph by adding new edges, if required. Clearly, for computational reasons this extension of the support of G should confine to adding only a few new edges. In this context the matrix completition problem consists in finding the unspecified entries such that the matrix becomes roughly positive semidefinite and the corresponding cuts become preferably deep.

In this thesis, as we have already mentioned, we are not interested in changing the given problem. However, for general sparse problems here arise a series of open questions.

(i) How can we find small submatrices which correspond to complete subgraphs of the given graph G and which are indefinite? (Clearly, if we can find an indefinite submatrix $X_k \subseteq X$ which corresponds to a complete subgraph this proves that X itself is indefinite.)

(ii) Is there an efficient method that gives us these submatrices?

(iii) Which one of these submatrices produce deep cuts?

For these reasons we consider (if we include semidefinite and hypermetric cuts, respectively) the class of chordal graphs. The answer to the leading two questions is known for this special class of graphs.

6.5.2 Restriction on Chordal Graphs

By definition, a graph is said to be chordal if every cycle of $C \subseteq G$ of length ≥ 4 has a chord. (By definition, a chord is an edge joining two nonconsecutive nodes of C.) Dirac [28] has shown that a graph is chordal if and only if it can be divided by means of clique sums into cliques. In other words, every chordal graph can be build from cliques by taking clique sums.

Grone, Johnson, Sá and Wolkowicz [40] studied the problem of which classes of partial given Hermitian matrices may be completed to positive definite matrices. We will present the main result of this work. Therefore we will have need of the following notation.

For a graph G the elliptope $\varepsilon(G)$ is defined as the projection of the feasible set of the relaxation **(SDPMC)**. Suppose y is indexed by the edges of G, then

$$\varepsilon(G) = \Big\{ y \in \mathbb{R}^m \,|\, \exists X \in S(n)^+ \text{ s.t. } \operatorname{diag}(X) = e \text{ and } x_{ij} = y_{ij} \text{ for all } (ij) \in E(G) \Big\}.$$

We now present a condition, introduced first in [40], which permits to characterize the elliptope $\varepsilon(G)$ for the class of chordal graphs. An easy intelligible presentation is given in [74]. For completeness we include it here.

Let $y \in \mathbb{R}^m$ and suppose that X is the partial matrix which corresponds to this y. Clearly, if $y \in \varepsilon(G)$ then every principal submatrix of X whose entries are all specified is positive semidefinite. In other words, if $K \subseteq V(G)$ induces a clique in G then the projection y_k of y on the edge set of $G[K]$ belongs to the elliptope $\varepsilon(K)$ of the clique. (Because here is no confusion the meaning of K is twofold; K denotes the node set and the graph itself.) Hence a necessary condition for $y \in \varepsilon(G)$ is

$$y_k \in \varepsilon(K) \text{ for each clique } K \text{ in } G. \tag{6.8}$$

Condition (6.8) is called usually the clique condition. Grone, Johnson, Sá and Wolkowicz [40] have shown that the class of graphs for which (6.8) is sufficient to describe $\varepsilon(G)$ is precisely the class of chordal graphs.

Theorem 6.1 [40] Given a graph $G = (V(G), E(G))$, we have

$$\varepsilon(G) = \left\{ y \in \mathbb{R}^m |\, y_k \in \varepsilon(K) \text{ for each clique } K \text{ in } G \right\}$$

if and only if G is chordal.

Similarly, a necessary condition for $y \in \text{int}(\varepsilon(G))$ is $y_k \in \varepsilon(K)$ for each clique K in G and Theorem 6.1 has the following analogue.

Theorem 6.2 [40] Given a graph $G = (V(G), E(G))$, we have

$$\text{int}(\varepsilon(G)) = \left\{ y \in \mathbb{R}^m |\, y_k \in \text{int}(\varepsilon(K)) \text{ for each clique } K \text{ in } G \right\}$$

if and only if G is chordal.

Hence, a partial given matrix X corresponding to a chordal graph satisfying the clique condition (6.8) can be completed to a positive semidefinite matrix.

In practice we separate the cutting planes only on the support (including the semidefinite and hypermetric cuts; separating these cutting planes we simply take the cliques which build the chordal graph) and strengthen the current model as proposed above. This way the violation of (6.8) by the corresponding partial matrix X shrinks towards zero. The amount of violation depends on the number of added cuts. Practical details of this technique are described in the next section.

6.6 Implementation Details

The purpose of this section is to give a detailed description of selected parts of our implementation. All our codes are implemented in Matlab with interfaces in C. The implementation of our algorithm involves solving a number of linear programs. We use ILOG CPLEX 6.5 for solving it. We proceed with practical aspects of our implementation.

6.6.1 The Initial Bound

We take the optimal objective value of **(LPMC)**, restricted to the support of G, as initial upper bound on w^*. **(LPMC)** can be restated as

$$(\textbf{LPMC}) \qquad \max c^T y \text{ such that } 0 \leq y_i \leq 1 \text{ for all } i = 1, \ldots, m.$$

As we have already introduced in section 6.2, c denotes the vector indexed by the edges of G and having the weights on the edges as components. For the sake of simplicity we denote the components of y by y_i instead of y_{ij}. It is easy to see that the optimal solution of **(LPMC)** is given by

$$y_i = \begin{cases} 1 & \text{if } c_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Applying the transformation (6.3) leads a partial matrix $X \in S(n)$. We then separate various violated cutting planes (which do not change the support), project it back to the subspace $\mathbb{R}^m$ and include it to the linear model in order to tighten it.

It should be noted that we add only such constraints which are violated by a significant amount. However, we perform this process several times.

6.6.2 Cutting Planes in Practice

In this section we explain in more detail the separation of the cycle inequalities, semidefinite cuts and hypermetric inequalities. We also present the representation of the cuts used in our implementation.

Let x_e denote the weight of the edge $e \in E(G)$. Then the inequalities (6.5) can be written alternatively as

$$\sum_{e \in C \setminus F} x_e + \sum_{e \in F} (1 - x_e) \geq 1 \text{ for each cycle } C,\ F \subseteq C,\ |F| \text{ odd.}$$

For separating the cycle inequalities we use the polynomial algorithm given by Barahona and Mahjoub. We refer the reader to [11] and [62] for a detailed description of this algorithm.

Suppose $X \in S(n)$ denotes the partial covered matrix obtained from lifting the optimal solution $y \in \mathbb{R}^m$ using (6.3). For separating the semidefinite cuts or hypermetric inequalities we choose submatrices $X_k \subseteq X$ whose entries are all specified. We will denote the corresponding complete subgraphs by K_k. (Note, we work on chordal graphs if we separate these types of cutting planes. Hence, by construction we know the complete subgraphs of G.) If $X_k \notin S(n)^+$ then $X \notin S(n)^+$. Suppose $X_k \notin S(n)^+$. Let $\lambda_{\min}(X_k)$ and v_k denote the minimum eigenvalue of X_k and the corresponding eigenvector, respectively. Clearly, $\lambda_{\min}(X_k) < 0$ and $v_k^T X_k v_k < 0$.

Suppose v_k denotes the projection of $v \in \mathbb{R}^n$ (elements of v which corresponds to the vertices of K_k are labeled by the entries of v_k, all remainig components are set to zero) on the edge set of K_k. It is easy to see that v violates (6.6) and (6.7).

Note, adding hypermetric inequalities or semidefinite cuts in general lead to dense rows in the system matrix compared to the density of the rows in the current consisting system matrix. For reasons of computational efficiency we want avoid dense rows.

Now we briefly explain how we construct a family of violated hypermetric inequalities leading sparse rows. v_k contains real numbers. We get an integer vector h by rounding these numbers. In our computational experiments we derive this vector h by multiplying v_k by a scalar $\alpha \geq 1$ and then rounding the scaled v_k towards ∞. Because

$v_k^T X_k v_k = \lambda_{\min}(X_k) \leq 0$ it follows that $(\alpha v_k)^T X_k(\alpha v_k) = \alpha^2 \lambda_{\min}(X_k) \to -\infty$ for $\alpha \to \infty$. Moreover, if α is large then h generated this way is dense. Since we require additional cutting planes which lead to sparse rows in the system matrix (which is equivalent to a sparse h) we first fix the desired number of components different from zero in h. We then choose $\alpha \geq 1$ such that precisely those components in h, which correspond to the largest components in $|v_k|$ are not zero. (Suppose I denotes the index array of those components, then $\alpha = \max\Big\{1/(2v_i)|i \in I\Big\}$.) After that we ensure that $\sum_{i=1}^n h_i$ is odd. The final step is checking if $h^T X_k h$ violates (6.6).

If that is we have produced a valid cutting plane which gives a sparse row in the system matrix. At present we do not have powerful alternatives in generating valid sparse cuts. However, it seems to work well in practice.

We next derive the representation of (6.6) used in our implementation. For $X \in S(n)$ and $h \in \mathbb{R}^n$ we have

$$\begin{aligned}
\operatorname{trace} X(hh^T) &= h^T X h \\
&= \sum_{i=1}^n h_i^2 + 2\sum_{i<j} h_i h_j X_{ij} \\
&= \sum_{i=1}^n h_i^2 + 2\sum_{i<j} h_i h_j (1 - 2y_{ij}) \quad \text{(by (6.3))} \\
&= \sum_{i=1}^n h_i^2 + 2\sum_{i<j} h_i h_j - 4\sum_{i<j} h_i h_j y_{ij} \\
&= \sum_{i=1}^n \sum_{j=1}^n h_i h_j - 4\sum_{i<j} h_i h_j y_{ij}.
\end{aligned}$$

We see that

$$\sum_{i=1}^n \sum_{j=1}^n h_i h_j = \sum_{i=1}^n h_i \sum_{j=1}^n h_j = \left(\sum_{i=1}^n h_i\right)^2.$$

Hence we get

$$\operatorname{trace} X(hh^T) \geq 1 \Leftrightarrow 4\sum_{i<j} h_i h_j y_{ij} \leq \left(\sum_{i=1}^n h_i\right)^2 - 1.$$

In a further experiment we used semidefinite cuts as additional cutting planes. Here we simply use $v_k^T X_k v_k \geq 0$ as valid cut.

The computational results are presented in the next section.

6.7 Analysis of Numerical Results

In this section we report preliminary computational experience. All our computational tests were performed on a DEC Alpha workstation. We carried it out on randomly generated test examples (random weighted undirected graphs with varying edge probability) and on the class of randomly generated chordal graphs. As we have remarked earlier all experiments using Linear Programming were performed starting from the initial relaxation **(LPMC)**, see section 6.6.1.

The computational results on the two problem sets are reported in Tables 6.2 to 6.4. Note, Tables 6.3 and 6.4 belong together. These two tables present results for the same set of test examples. It is merely a matter of space we have on hand, why we report the results in two tables. The column headed P (for Problem) enumerates the test examples. This raises the readability of these tables.

In the tables column n gives the number of vertices, column m the number of edges in the graph G. For the randomly generated test instances we strengthened the initial linear bound by including exclusively the cycle inequalities as cutting planes. (We have stated the arising problems in the context of separating semidefinite cuts or hypermetric inequalities, respectively on general given problems.)

We did several experiments on the test set of chordal graphs. In the first one we investigated the strength of the relaxations by including cycle inequalities without using semidefiniteness in the cut generation. Secondly, we present numerical results obtained by exploiting the positive semidefiniteness postulation of the complete subgraphs of the given graphs. In particular, we report results obtained from strengthening the initial bound by including cycle inequalities together with semidefinite cuts and hypermetric inequalities, respectively.

In the tables we use the notation $\mathbf{(LPMC)}^{\boldsymbol{S}}$ where the superscript $\boldsymbol{S}$ is replaced by one of $\{\boldsymbol{C}, \boldsymbol{CH}, \boldsymbol{CS}\}$. The individual superscript indicates which class of cutting planes is used to strengthen the linear relaxation **(LPMC)**. In particular, the meaning is as follows

- $\boldsymbol{C}$ cycle inequalities
- $\boldsymbol{CH}$ cycle inequalities together with hypermetric inequalities
- $\boldsymbol{CS}$ cycle inequalities together with semidefinite cuts.

We compare the upper bound on w^* at each case with that obtained from solving the semidefinite relaxation **(SDPMC)**. In our experiments with chordal graphs we also strengthened the semidefinite bound through including triangle inequalities, as proposed in [103]. The upper bounds on w^* obtained from solving this non-polyhedral relaxations are presented in the columns **(SDPMC)** and $\mathbf{(SDPMC)}^{\boldsymbol{T}}$, respectively. Here the superscript $\boldsymbol{T}$ indicates that triangle inequalities have been used for strengthening **(SDPMC)**. In Table 6.3 the column headed $G = (V(G), E(G))$ shows the structure of the graph G. The meaning of this column is twofold. The leading number gives the number of cliques which sum up to G and the interval indicates that the size of the individual cliques vary within this interval.

In all tables the column labeled *lbd on* w^* stands for "lower bound on w^*". This column gives the weight of the best cut we have found by rounding the optimal solution corresponding to the (final) relaxation.

We ran extensive tests on randomly generated graphs of different size and density. Our observation is that solving the relaxation $\mathbf{(LPMC)}^{\boldsymbol{C}}$ do not yield an improved upper bound on w^* compared with that obtained from solving the non-polyhedral relaxation **(SDPMC)** in most cases. Table 6.2 presents a small selection of our test results. Note that there are only two out of seven instances for which the final bound could be improved somewhat. Because of this poor effect it seems that this linear approach is not warrentable.

This is in contrast to the results presented in Table 6.4. For the class of chordal graphs already the inclusion of cycle inequalities did improve the upper bound on w^* compared with the semidefinite bound from the non-polyhedral relaxation **(SDPMC)** in many instances. Moreover, the inclusion of additional hypermetric inequalities or semidefinite cuts substantially improved this bounds (in all instances). The linear approach yields bounds which are proven to be optimal in many instances which is indicated by the values given in column *lbd on* w^*. (This is denoted by an asterisk in this column.)

Suppose $\lambda_{\min}(X_k)$ denotes the minimum eigenvalue of any $X_k \subseteq X$ whose entries are all specified. (Here $X \in S(n)$ denotes the partial covered matrix obtained from lifting the optimal solution $y \in \mathbb{R}^m$ of $\mathbf{(LPMC)}^{\boldsymbol{S}}$, $\boldsymbol{S} \in \{\boldsymbol{C}, \boldsymbol{CH}, \boldsymbol{CS}\}$ using (6.3).) X_k runs through those matrices which correspond to the cliques which sum up to G. The column headed $\lambda^\circ_{\min}$ in Table 6.3 gives the minimum of all those minimum eigenvalues corresponding to the initial X° and the columns headed $\lambda_{\min}$ in Table 6.4 gives the minimum of all those minimum eigenvalues corresponding to the X obtained from the different linear relaxations at termination. These values show that the amount of violation of the semidefinite cuts (corresponding to the known cliques in G) could be reduced significantly.

Note that, if $\lambda_{\min}$ is positive (nonnegative), then the partial covered matrix X (at termination) can be completed to a positive definite (positive semidefinite) matrix, respectively. The completion process is discussed in detail in [40]. If $\lambda_{\min}$ is nonnegative, the completed matrix is strictly feasible for the non-polyhedral relaxation **(SDPMC)** and the corresponding objective value dominates the optimal value μ^*.

6.8 Concluding Comments

We have presented a cutting plane technique for approximating the maximum cut problem. This linear approach allows to solve large scale problems. We derive the following conclusions from our preliminary computational tests. We have mainly discussed the class of chordal graphs. Clearly, for this class we took advantage of the very special structure of the graphs. Our numerical results show that solving the strengthened linear relaxations $\mathbf{(LPMC)}^{\boldsymbol{S}}$, $\boldsymbol{S} \in \{\boldsymbol{C}, \boldsymbol{CH}, \boldsymbol{CS}\}$ provides strictly

		lbd on w^*	Relaxation		
n	m	best cut found	**(SDPMC)**	**(LPMC)**^{C}	
40	324	1290	1533.04	1425.22	(390)
50	628	2897	3333.91	3219.43	(667)
60	902	3872	4341.53	4360.12	(983)
70	714	3253	3780.32	3821.49	(772)
80	795	3703	4285.27	4432.73	(907)
90	1211	4630	5299.96	5631.49	(1289)
100	1985	7047	8005.08	9044.74	(2256)

Table 6.2: *Numerical Results on randomly generated sparse graphs. The last two columns give the optimal objective values from solving the semidefinite relaxation* **(SDPMC)** *and the linear relaxation* **(LPMC)**^{C}*, respectively. The numbers given in parenthesis represent the number of active cycle inequalities at termination.*

improved upper bounds on w^* compared with the semidefinite bounds from the non-polyhedral relaxation **(SDPMC)** in almost all instances. Our experience is that very sparse hypermetric inequalities and semidefinite cuts substantially improved the upper bound on w^*. We also observed that identifying and including violated cutting planes and then solving the strenghtened system is done very efficient. Compared with SDP we obtained the bounds in moderate time. Nevertheless, since we follow an iterative procedure (as briefly sketched in section 6.6.1) we believe that the total computational effort is still relatively high. It will be an interesting issue for future to upgrade the method so as to reduce the complete computation time. A further challenge for future will be to attack the difficulties stated in section 6.5.1. We believe that (if we are able to handle these difficulties) our approach can be very effective also on more general classes of graphs than on chordal graphs. But this is future research and can be found in forthcoming work.

				lbd on $\mathcal{W}^*$	Relaxation		
P	n	m	$G = (V(G), E(G))$	best cut found	**(SDPMC)**	$\mathbf{(SDPMC)}^{\boldsymbol{T}}$	
1	20	76	3 [5,10]	410*	425.69	410.00	(193)
2	30	137	4 [8,12]	771*	806.45	771.00	(236)
3	40	217	4 [10,15]	1050*	1142.62	1050.00	(151)
4	50	282	5 [10,15]	1751*	1865.68	1751.00	(214)
5	60	288	7 [10,15]	1821*	1909.99	1821.00	(156)
6	70	542	5 [15,20]	2993*	3277.55	2997.09	(134)
7	80	533	8 [15,20]	2887*	3110.57	2894.56	(119)
8	90	793	6 [20,25]	3900*	4212.75	3912.84	(129)
9	100	878	7 [20,25]	4289*	4578.40	4309.81	(110)
10	125	935	10 [20,25]	5002*	5294.83	5018.76	(124)
11	150	1808	8 [25,30]	7875*	8567.94	7950.08	(193)
12	175	971	19 [10,15]	6150*	6516.73	6199.96	(162)
13	200	1566	15 [15,20]	8428	9109.32	8543.89	(192)

Table 6.3: *Numerical Results on chordal graphs, Part I. P enumerates the test examples. The last two columns give the optimal objective values from solving the semidefinite relaxations* **(SDPMC)** *and* $\mathbf{(SDPMC)}^{\boldsymbol{T}}$*, respectively. The numbers given in parenthesis represent the number of active triangle inequalities at termination.*

Table 6.3 continued

	Relaxation									
P	$\lambda^{\circ}_{\min}$	$\mathbf{(LPMC)}^{\boldsymbol{C}}$		$\lambda_{\min}$	$\mathbf{(LPMC)}^{\boldsymbol{CH}}$		$\lambda_{\min}$	$\mathbf{(LPMC)}^{\boldsymbol{CS}}$		$\lambda_{\min}$
1	-2.87	410.00	(71)	-0.00	410.00	(74)	-0.00	410.00	(76)	-0.00
2	-1.56	821.00	(176)	-0.00	771.00	(202)	-0.00	771.00	(225)	0.00
3	-2.60	1091.00	(290)	-2.00	1091.00	(311)	-2.00	1050.42	(311)	-0.02
4	-4.28	1764.00	(354)	-0.00	1751.00	(345)	-0.00	1751.21	(411)	-0.00
5	-2.43	1830.00	(351)	-0.00	1830.00	(353)	-0.00	1821.09	(371)	-0.00
6	-3.74	3004.48	(802)	-0.00	2993.00	(865)	-0.00	2993.18	(849)	0.00
7	-0.00	2887.00	(749)	-0.00	2887.00	(811)	-0.00	2887.06	(835)	-0.00
8	-1.24	3900.00	(1012)	-0.00	3900.00	(1203)	-0.00	3900.14	(1117)	0.00
9	-0.00	4308.32	(1049)	-0.00	4289.00	(1306)	-0.00	4289.31	(1340)	0.00
10	-4.11	5050.40	(1247)	-0.00	5002.00	(1316)	-0.00	5003.00	(1778)	-0.00
11	-2.72	8440.17	(2259)	-2.17	7875.47	(2646)	-0.00	7878.60	(4105)	0.00
12	-1.00	6217.00	(1324)	0.00	6150.00	(1313)	0.00	6150.11	(1354)	-0.00
13	-6.25	8738.98	(1807)	-0.00	8508.16	(2074)	-0.00	8439.93	(3531)	-0.02

Table 6.4: *Numerical Results on chordal graphs, Part II. P enumerates the test examples. The optimal objective values from solving the linear relaxations* $\mathbf{(LPMC)}^{\boldsymbol{C}}$, $\mathbf{(LPMC)}^{\boldsymbol{CH}}$ *and* $\mathbf{(LPMC)}^{\boldsymbol{CS}}$ *are presented. The numbers given in parenthesis represent the number of active cutting planes at termination.* $\lambda_{\min}$ *gives the minimum over all minimum eigenvalues corresponding to well known complete submatrices of X at termination. ($\lambda^{\circ}_{\min}$ refers to X°.)*

Appendix A

Elements in Linear Algebra

We give a brief survey of the concepts and results of Linear Algebra and Matrix Analysis which are used in this work. For a more complete account we refer the reader to standard literature in Linear Algebra and to the works of Horn and Johnson [57, 58] and Bellman [13] in Matrix Analysis.

Lemma A.1 *Let A and B denote real symmetric matrices of order n. If $A \succeq 0$, $B \succeq 0$, then*

$$\langle A, B\rangle = 0 \text{ if and only if } AB = 0.$$

Proof. For a positive semidefinite matrix A the eigenvalue-eigenvector factorization is given by $A = P\Lambda P^T$. Since Λ is a diagonal matrix and P is an orthogonal matrix an equivalent formulation is given by $A = P\Lambda^{1/2}P^TP\Lambda^{1/2}P^T$. We set $A^{1/2} = P\Lambda^{1/2}P^T$ and therefore we can write $A = A^{1/2}A^{1/2}$. For this reason $\langle A, B\rangle = \langle A^{1/2}A^{1/2}, B\rangle = \langle A^{1/2}, BA^{1/2}\rangle$. Moreover $A^{1/2}BA^{1/2} \succeq 0$ and $\langle A^{1/2}, BA^{1/2}\rangle = 0$ implies that $A^{1/2}BA^{1/2} = 0$, and hence $AB = 0$.

□

Lemma A.2 *Let $\|\cdot\|$ denote the Frobenius norm. For any symmetric positive semidefinite matrix A, we have*

$$\|A\| \le \operatorname{trace}(A).$$

Proof. Let $\lambda_j(A), (j = 1, \ldots, n)$ denote the eigenvalues of A. Then

$$\begin{aligned} \|A\|^2 &= \operatorname{trace}(A^TA) \\ &= \operatorname{trace}(A^2) \quad (\text{since } A = A^T) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} \lambda_j^2(A) \\
&\leq \Big(\sum_{j=1}^{n} \lambda_j(A)\Big)^2 \quad \text{(since } A \succeq 0\text{)} \\
&= \Big(\operatorname{trace}(A)\Big)^2.
\end{aligned}
$$

□

Lemma A.3 *Let $A \in S(n)$, $A \succ 0$ and $\alpha > 0$. Denote the minimum eigenvalue of A by $\lambda_{\min}(A)$. If $B \in S(n)^+$ and $\langle A, B\rangle \leq \alpha$, then*

$$\sum_{j=1}^{n} \lambda_j(B) \leq \frac{\alpha}{\lambda_{\min}(A)}. \qquad (\lambda_j(B), (j = 1, \ldots, n) \textit{ are the eigenvalues of } B.)$$

Proof. On one hand we have by assumption

$$\langle A, B\rangle = \operatorname{trace}(AB) = \sum_{j=1}^{n} \lambda_j(AB) \leq \alpha.$$

On the other hand by lemma A.4 we have

$$\langle A, B\rangle \geq \lambda_{\min}(A)\operatorname{trace}(B).$$

Combining these two inequalities completes the proof.

□

Lemma A.4 *Let A and B denote real symmetric matrices of order n. If $B \succeq 0$, then*

$$\lambda_{\min}(A)\operatorname{trace}(B) \leq \operatorname{trace}(AB) \leq \lambda_{\max}(A)\operatorname{trace}(B).$$

Proof. Any real symmetric matrix S can be factored into $S = V\Lambda V^T$ with the orthonormal eigenvectors $v_j, (j = 1, \ldots, n)$ in V and the eigenvalues $\lambda_j(S), (j = 1, \ldots, n)$ in Λ.

Hence there exists orthogonal matrices P and Q and diagonal matrices D and E such that

$$
\begin{aligned}
A &= PDP^T,\ PP^T = I,\ D = \operatorname{Diag}(\lambda) \\
B &= QEQ^T,\ QQ^T = I,\ E = \operatorname{Diag}(\mu),
\end{aligned}
$$

where we used $\lambda = (\lambda_1(A), \ldots, \lambda_n(A))$ and $\mu = (\mu_1(B), \ldots, \mu_n(B))$. We define $R := P^T Q$. From that it follows that $R^T R = I$. Hence we obtain

$$\begin{aligned}
\operatorname{trace}(AB) &= \operatorname{trace}(PDP^T QEQ^T) \\
&= \operatorname{trace}(DP^T QEQ^T P) \quad (\text{since } \operatorname{trace}(AB) = \operatorname{trace}(BA)) \\
&= \operatorname{trace}(DRER^T) \\
&= \sum_{i=1}^{n} \lambda_i(A)(RER^T)_{ii} \quad (\text{since } D = \operatorname{Diag}(\lambda)).
\end{aligned}$$

Now we consider the expression $(RER^T)_{ii}$ and get

$$\begin{aligned}
(RER^T)_{ii} &= R_{i.} E R_{i.}^T \\
&= \sum_{j=1}^{n} \mu_j(B) R_{ij}^2 \quad (\text{since } E = \operatorname{Diag}(\mu)).
\end{aligned}$$

We then see that

$$\begin{aligned}
\operatorname{trace}(AB) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i(A)\mu_j(B) R_{ij}^2 \\
&\leq \lambda_{\max}(A) \sum_{j=1}^{n} \mu_j(B) \sum_{i=1}^{n} R_{ij}^2 \\
&= \lambda_{\max}(A) \sum_{j=1}^{n} \mu_j(B) \quad \left(\text{since } \sum_{i=1}^{n} R_{ij}^2 = 1\right) \\
&= \lambda_{\max}(A) \operatorname{trace}(B).
\end{aligned}$$

The inequality $\operatorname{trace}(AB) \geq \lambda_{\min}(A)\operatorname{trace}(B)$ can be proven similarly.

□

Lemma A.5 *Suppose that $(Z, X) \in (S(n)^{++})^2$. Let $\lambda_j(ZX), (j = 1, \ldots, n)$ be the eigenvalues of ZX. Then*

$\lambda_j(ZX)$ is real and positive for every $j = 1, \ldots, n$.

Proof. Since Z is positive definite there exists always a unique nonsingular symmetric $\sqrt{Z}$ such that $Z = \sqrt{Z}\sqrt{Z}$. Hence the matrix $\sqrt{Z}X\sqrt{Z}$ is symmetric positive definite by Lemma A.6. In fact ZX is similar to $\sqrt{Z}X\sqrt{Z}$ and therefore ZX and $\sqrt{Z}X\sqrt{Z}$ have the same eigenvalues. This completes the proof.

□

Lemma A.6 [57, Theorem 4.5.8] *Let $\mathcal{M}_{m,n}(\mathbf{C})$ denote the vector space of $m \times n$ complex matrices and A^* denote the Hermitian adjoint of $A \in \mathcal{M}_{m,n}(\mathbf{C})$. Let $A,\ B \in (\mathcal{M}_{n,n}(\mathbf{C}))^2$ be Hermitian matrices. There is a nonsingular matrix $S \in \mathcal{M}_{n,n}(\mathbf{C})$ such that $A = SBS^*$ if and only if A and B have the same inertia, that is, the same number of positive, negative, and zero eigenvalues.*

Lemma A.7 *Let Λ be a diagonal matrix of real numbers $\lambda_j > 0, (j = 1, \ldots, n)$. Denote the maximum and the minimum value of all diagonal entries by $\lambda_{\max}$ and $\lambda_{\min}$, respectively. Suppose $E \in M(n)$. Then*

$$\|\Lambda^{-1}E\Lambda\| \leq \frac{\lambda_{\max}}{\lambda_{\min}}\|E\|.$$

Proof.

$$\begin{aligned}
\|\Lambda^{-1}E\Lambda\|^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(E_{ij}\frac{\lambda_j}{\lambda_i}\right)^2 \\
&\leq \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^2 \sum_{i=1}^{n}\sum_{j=1}^{n}(E_{ij})^2 \\
&= \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^2 \|E\|^2
\end{aligned}$$

□

Lemma A.8 *Let $A \in S(n)$. Suppose $A = Q\Lambda Q^T$ with the orthonormal eigenvectors $q_j, (j = 1, \ldots, n)$ in Q and the eigenvalues $\lambda_j(A), (j = 1, \ldots, n)$ in Λ. Then,*

(i) $A^{-1} = \sum_{j=1}^{n} \frac{1}{\lambda_j} q_j q_j^T$

(ii) $I + A = \sum_{j=1}^{n} (1 + \lambda_j)(q_j q_j^T)$.

Proof. We see that

$$\begin{aligned}
A &= Q\Lambda Q^T \\
&= (q_1, \ldots, q_n)\begin{pmatrix} \lambda_1(A) & & \\ & \ddots & \\ & & \lambda_n(A) \end{pmatrix}\begin{pmatrix} q_1^T \\ \vdots \\ q_n^T \end{pmatrix}
\end{aligned}$$

$$= (\lambda_1(A)q_1, \ldots, \lambda_n(A)q_n) \begin{pmatrix} q_1^T \\ \vdots \\ q_n^T \end{pmatrix}$$
$$= \sum_{j=1}^{n} \lambda_j(A) q_j q_j^T.$$

Hence

$$\begin{aligned} AA^{-1} &= \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \frac{1}{\lambda_i} q_j (q_j^T q_i) q_i^T \\ &= \sum_{j=1}^{n} q_j q_j^T \quad \text{(since Q orthogonal).} \end{aligned}$$

Since Q is orthogonal we get $\sum_{j=1}^{n} q_j q_j^T = QQ^T = I$. Thus we have shown (i). Now (ii) follows as direct consequence, i.e.,

$$I + A = \sum_{j=1}^{n} q_j q_j^T + \sum_{j=1}^{n} \lambda_j q_j q_j^T = \sum_{j=1}^{n} (1+\lambda_j) q_j q_j^T.$$

□

Lemma A.9 *Suppose $Y \in S(n)$ and $x \in \mathbb{R}^n$ are given. Let $a > 0$. Then*
$\begin{pmatrix} Y & x \\ x^T & a \end{pmatrix} \succeq 0$ *if and only if* $Y - \frac{1}{a}xx^T \succeq 0$.

Proof. Define $L \in M(n)$ as $L = \begin{pmatrix} I & -\frac{1}{a}x \\ 0 & 1 \end{pmatrix}$. Clearly, L is a nonsingular matrix and hence by Sylvesters law of inertia $\begin{pmatrix} Y & x \\ x^T & a \end{pmatrix} \succeq 0$ if and only if $L\begin{pmatrix} Y & x \\ x^T & a \end{pmatrix} L^T \succeq 0$. We see that

$$L\begin{pmatrix} Y & x \\ x^T & a \end{pmatrix} L^T = \begin{pmatrix} Y - \frac{1}{a}xx^T & 0 \\ 0 & a \end{pmatrix}.$$

This completes the proof.

□

Summary and Outlook

Chapter 1 in Part One covered the main aspects concerning duality and interior-point methods in Semidefinite Programming. After introducing the problem we mainly discussed the different duality states arising in Semidefinite Programming. We considered asymptotic refinements and investigated the embedding of a given semidefinite problem in a larger problem. The solution of this extended problem allowed to classify the duality states of the original primal-dual pair in many cases. It should be noted that there are still some cases which cannot be handled with this approach.

Chapter 2 dealt with convergence analysis in Semidefinite Programming. Our main contribution was a global convergence proof based on the concept of closed mappings. Moreover we presented a condition which enables us to decide if a given problem does have an optimal solution in a prescribed region.

In chapter 3 of Part Two we presented an interior-point algorithm for Semidefinite Programming with linear-quadratic functionals. Using our results from chapter 2 we established the global convergence of this algorithm. We implemented it for quadratic cost functions of the form $\|X\|^2 + \langle C, X\rangle$. It will be a challenge for the future to consider more general quadratic functions.

In chapter 4 we have presented a damped version of the standard HKM-direction for Semidefinite Programming. The damping strategy allows to solve ill-posed semidefinite problems, namely those, where Slater's condition fails for one of the original problems. We observed that damping works well. In contrast to standard implementations however, the convergence is slowed down considerably. We note that our approach does not cover the case for which strong duality does not hold. We also note that recognizing strict infeasibility can be handled by damping. Recognizing "weakly infeasible" problems is more subtle, and is also a challenge for the present approach.

In chapter 5 we investigated various relaxations of the maximum stable set problem based on the Lovász number $\vartheta(G)$. On the theoretical side we have provided a simple argument that two different semidefinite relaxations are equivalent. In particular, we have shown that their objective values agree at the optimum and how optimal solutions from one formulation can be transformed into optimal solutions of the other formulation. We presented a semidefinite interior-point cutting plane approach for approximating the maximum stable set problem. In computational experiments we achieved significantly improved bounds on the stability number of the considered graphs. This was demonstrated on many different categories of test instances. However despite these promising results it seems not to be clear if Semidefinite Programming ever dominates polyhedral approaches in practical performance. In chapter 6 we have used Linear Programming to approximate semidefinite relaxations of the maximum cut problem and received very promising results on chordal graphs. We believe that our approach can be very effective also on more general classes of graphs. This approach would also be an interesting issue for future research in context with other problems in Combinatorial Optimization, including the maximum stable set problem.

Bibliography

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.

[2] F. Alizadeh, J-P. A. Haeberly, and M. L. 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, 1994.

[3] F. Alizadeh, J-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.

[4] M. F. Anjos and H. Wolkowicz. A strengthened SDP relaxation via a second lifting for the MAX-CUT problem. Technical Report CORR–95, Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada, 1999.

[5] K. M. Anstreicher and M. Fampa. A long-step path following algorithm for semidefinite programming problems. In *Topics in Semidefinite and Interior-Point Methods*, volume 18 of *The Fields Institute for Research in Mathematical Sciences, Communications Series*, pages 181–196, Providence, Rhode Island, 1998. American Mathematical Society.

[6] K. M. Anstreicher and J.-Ph. Vial. On the convergence of an infeasible primal–dual interior–point method for convex programming. *Optimization Methods and Software*, 3:273–283, 1994.

[7] E. Balas, S.Ceria, and G. Cornuéjols. A lift project cutting plane algorithm for mixed 0–1 programs. *Mathematical Programming*, 58(3):195–324, 1993.

[8] E. Balas, S.Ceria, G. Cornuéjols, and G. Pataki. Polyhedral methods for the maximum clique problem. In David S. Johnson and Michael A. Trick, editors, *Cliques, Coloring, and Satisfiability*, volume 26 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 11–28. AMS, 1996.

[9] F. Barahona. The max-cut problem on graphs not contractible to K_5. *Operations Research Letters*, 2(3):107–111, 1983.

[10] F. Barahona. On cuts and matchings in planar graphs. *Mathematical Programming*, 36(1):53–68, 1993.

[11] F. Barahona and A. R. Mahjoub. On the cut polytope. *Mathematical Programming*, 36:157–173, 1986.

[12] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms.* John Wiley, New York, second edition, 1993.

[13] R. Bellman. *Introduction to Matrix Analysis.* SIAM Publications, 1995. (First edition by McGraw-Hill, 1960).

[14] R. Bellman and K. Fan. On systems of linear inequalities in Hermitian matrix variables. In *Proceedings of Symposia in Pure Mathematics, Vol 7, AMS*, 1963.

[15] A. Ben-Israel, A. Charnes, and K. Kortanek. Duality and asymptotic solvability over cones. *Bulletin of American Mathematical Society*, 75(2):318–324, 1969.

[16] I. M. Bomze, M. Pelillo, and R. Giacomini. Evolutionary approach to the maximum clique problem: Empirical evidence on a larger scale. In I. M. Bomze, T. Csendes, R. Horst, and P. M. Pardalos, editors, *Developments in Global Optimization*, pages 95–108. Kluwer Academic Publishers, Dordrecht, 1997.

[17] B. Borchers. CSDP 2.3 user's guide. *Optimization Methods and Software*, 11:597–611, 1999.

[18] 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.

[19] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *SIAM Studies in Applied Mathematics.* Society of Industrial and Applied Mathematics (SIAM), Philadelphia, PA 19101, USA, 1994.

[20] S. Boyd and L. Vandenberghe. Primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69(1):205–236, 1995.

[21] V. Chvátal. On certain polytopes associated with graphs. *J. Combin. Theory B*, 18:138–154, 1975.

[22] E. de Klerk. *Interior point methods for semidefinite programming.* PhD thesis, Faculty of Technical Mathematics and Informatics, TU Delft, NL–2600 GA Delft, The Netherlands, 1997.

[23] 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.

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

[25] C. Delorme and S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Mathematical Programming*, 62(3):557–574, 1993.

[26] M. Deza and M. Laurent. Applications of cut polyhedra I. *Journal of Computational and Applied Mathematics*, 55:191–216, 1994.

[27] M. Deza and M. Laurent. Applications of cut polyhedra II. *Journal of Computational and Applied Mathematics*, 55:217–247, 1994.

[28] G. A. Dirac. On rigid circuit graphs. Abhandlungen aus dem Mathematischen Seminar, Universtät Hamburg, 1961.

[29] R. J. Duffin. Infinite programs. In A. W. Tucker, editor, *Linear Equalities and Related Systems*, pages 157–170. Princeton University Press, Princeton, NJ, 1956.

[30] L. Faybusovich. Semidefinite programming: A path-following algorithm for a linear-quadratic functional. *SIAM Journal on Optimization*, 6(4):1007–1024, 1996.

[31] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming, Sequential Unconstrained Minimization Techniques*, volume 4 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 1990. originally published by Research Analysis Corperation, and again in 1968 by John Wiley and Sons.

[32] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, Chichester, second edition, 1987.

[33] D. R. Fulkerson. On the perfect graph theorem. In T. C. Hu and S. M. Robinson, editors, *Mathematical Programming*, pages 69–76. Academic Press, New York, 1973.

[34] M. X. Goemans. Semidefinite programming in combinatorial optimization. *Mathematical Programming*, 79(1–3):143–161, 1997.

[35] M. X. Goemans and D. P. Williamson. .878-approximation algorithms for MAX CUT and MAX 2SAT. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 422–431, Montréal, Québec, Canada, 1994.

[36] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995. preliminary version, see [35].

[37] D. Goldfarb and K. Scheinberg. Interior point trajectories in semidefinite programs. *SIAM Journal on Optimization*, 8(4):871–886, 1998.

[38] A. J. Goldman and A. W. Tucker. Polyhedral convex cones. In *Linear equalities and related systems*, pages 19–40. Princeton University Press, Princeton, N. J., 1956. Annals of Mathematics Studies, no. 38.

[39] A. Graham. *Kronecker Products and Matrix Calculus with Applications*. Mathematics and its Applications, Ellis Horwood Limited, Chichester, 1981.

[40] B. Grone, C.R. Johnson, E. M. de Sá, and H. Wolkowicz. Positive definite completions of partial Hermitian matrices. *Linear Algebra and its Applications*, 58:109–124, 1984.

[41] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.

[42] M. Grötschel, L. Lovász, and A. Schrijver. Relaxations of vertex packing. *Journal of Combinatorial Theory*, 40:330–343, 1986.

[43] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, New York, 1988.

[44] M. Grötschel and G. L. Nemhauser. A polynomial algorithm for the max-cut problem on graphs without long odd cycles. *Mathematical Programming*, 29:28–40, 1984.

[45] G. Gruber and F. Rendl. Computational experience with ill-posed problems in semidefinite programming. Technical report, Department of Mathematics, University of Klagenfurt, 9020 Klagenfurt, Austria, 1999. (forthcoming).

[46] G. Gruber and F. Rendl. Approximating stable sets using the ϑ-function and cutting planes. In *Operations Research Proceedings 1999*, CON Operations Research Proceedings 1999, pages 69–74. Springer-Verlag, Berlin · Heidelberg, Germany, 2000.

[47] G. Gruber and F. Rendl. Combining polyhedral and nonpolyhedral relaxations for the stable set problem. Technical report, Department of Mathematics, University of Klagenfurt, 9020 Klagenfurt, Austria, 2000. (forthcoming).

[48] O. Güler and Y. Ye. Convergence behaviour of interior point algorithms. *Mathematical Programming*, 60(2):215–228, 1993.

[49] F. Hadlock. Finding a maximum cut of a planar graph in polynomial time. *SIAM Journal on Computing*, 4:221–225, 1975.

[50] J. Hasselberg, Panos M. Pardalos, and G. Vairaktarakis. Test case generators and computational results for the maximum clique problem. *Journal of Global Optimization*, 3:463–482, 1993.

[51] C. Helmberg. *An interior point method for semi–definite programming and max–cut bounds*. PhD thesis, Department of Mathematics, Graz, University of Technology, A–8010 Graz, Austria, 1994.

[52] C. Helmberg, S. Poljak, F. Rendl, and H. Wolkowicz. Combining semidefinite and polyhedral relaxations for integer programs. In E. Balas and J. Clausen, editors, *Integer Programming and Combinatorial Optimization (Proceedings of the 4th International IPCO Conference, Copenhagen, Denmark, May 1995)*, volume 920 of *Lecture Notes in Computer Science*, pages 124–134. Springer Verlag, Berlin, Germany, 1995.

[53] C. Helmberg and F. Rendl. Solving quadratic (0,1)–problems by semidefinite programs and cutting planes. *Mathematical Programming*, 82(3):291–315, 1998.

[54] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz. An interior point method for semidefinite programming. *SIAM Journal on Optimization*, 6(2):342–361, 1996.

[55] D. Den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming*. Kluwer Academic Publishers, Norwall, MA, 1994.

[56] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming, Algorithms and Complexity*, volume 277 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[57] R. A. Horn and Charles A. Johnson. *Matrix Analysis*. Cambridge University press, Cambridge, 1985.

[58] R. A. Horn and Charles A. Johnson. *Topics in Matrix Analysis*. Cambridge University press, Cambridge, 1991.

[59] J. Håstad. Some optimal inapproximability results. In *Proceedings of the 29th ACM Symposium on Theory of Computing (STOC)*, pages 1–10, 1997.

[60] F. Jarre. An interior–point method for minimizing the largest eigenvalue of a linear combination of matrices. *SIAM Journal on Control and Optimization*, 31:1360–1377, 1993. See also *NATO ASI Series E Vol. 232 (1993), pp. 395.*

[61] C. R. Johnson. Matrix completition problems: a survey. In *Matrix Theory and Applications*, volume 40 of *Proceedings of Symposia in Applied Mathematics*, pages 171–198. American Mathematical Society, 1990.

[62] M. Jünger, F. Barahona, and G. Reinelt. Experiments in quadratic 0-1 programming. *Mathematical Programming*, 44(2):127–137, 1989.

[63] S. E. Karisch and F. Rendl. Semidefinite programming and graph equipartition. In *Topics in Semidefinite and Interior-Point Methods*, volume 18 of *The Fields Institute for Research in Mathematical Sciences, Communications Series*, pages 77–95, Providence, Rhode Island, 1998. American Mathematical Society.

[64] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[65] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103, New York, 1972. Plenum Press.

[66] R. M. Karp. On the computational complexity of combinatorial problems. *Networks*, 5(1):45–68, 1975.

[67] D. E. Knuth. The sandwich theorem. *The Electronic Journal of Combinatorics*, 1(A1), 1994.

[68] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial–time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[69] M. Kojima, S. Mizuno, and A. Yoshise. An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. *Mathematical Programming*, 50(3):331–342, 1991.

[70] M. Kojima, M. Shida, and S. Shindoh. Global and local convergence of predictor–corrector infeasible–interior–point algorithms for semidefinite programs. Research Reports on Information Sciences, Ser. B : Operations Research B–305, Department of Information Sciences, Tokyo Institute of Technology, Oh–Okayama, Meguro–ku, Tokyo 152, Japan, 1995.

[71] M. Kojima, M. Shida, and S. Shindoh. Local convergence of predictor–corrector infeasible–interior–point algorithms for SDPs and SDLCPs. *Mathematical Programming*, 80(2):129–160, 1998.

[72] 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. *SIAM Journal on Optimization*, 9(2):444–465, 1999.

[73] M. Kojima, S. Shindoh, and S. Hara. Interior point methods for the monotone linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7(1):86–125, 1997.

[74] M. Laurent. A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. Technical report, LIENS, Paris, France, 1997.

[75] C.-J. Lin and R. Saigal. An infeasible start predictor–corrector method for semidefinite linear programming. Technical Report, Department of Industrial and Operational Engineering, University of Michigan, Ann Arbor, MI 48109–2117, USA, 1995.

[76] L. Lovász. On the shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[77] L. Lovász. Semidefinite programs and combinatorial optimization. Lecture Notes, 1995.

[78] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal of Optimization*, 1(2):166–190, 1991.

[79] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley Publishing Co., Reading, second edition, 1984. Reprinted with corrections in May, 1989.

[80] 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, The Netherlands, 1996.

[81] M. Laurent, S. Poljak, F. Rendl. Connections between semidefinite relaxations of the max-cut and stable set problems. *Mathematical Programming*, 77(2):225–246, 1997.

[82] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley and Sons, New York, 1988.

[83] O. L. Mangasarin. *Nonlinear Programming.* McGraw-Hill, New York, 1969.

[84] C. Mannino and A. Sassano. Edge projection and the maximum cardinality stable set problem. In David S. Johnson and Michael A. Trick, editors, *Cliques, Coloring, and Satisfiability*, volume 26 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 205–219. AMS, 1996.

[85] S. Mizuno. Polynomiality of infeasible interior–point algorithms for linear programming. *Mathematical Programming*, 67(1):109–119, 1994.

[86] S. Mizuno, M. Kojima, , and M. J. Todd. Infeasible–interior–point primal–dual potential–reduction algorithms for linear programming. *SIAM Journal on Optimization*, 5(1):52–67, 1995.

[87] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive–step primal–dual interior–point algorithms for linear programming. *Mathematics of Operations Research*, 18(4):964–981, 1993.

[88] R. D. C. Monteiro. Polynomial convergence of primal–dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions. *SIAM Journal on Optimization*, 8(3):797–812, 1998.

[89] R. D. C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part I : Linear programming. *Mathematical Programming*, 44(1):27–41, 1989.

[90] R. D. C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part II : Convex quadratic programming. *Mathematical Programming*, 44(1):43–66, 1989.

[91] R. D. C. Monteiro and T. Tsuchiya. Polynomiality of primal–dual algorithms for semidefinite linear complementarity problems based on the Kojima–Shindoh–Hara family of directions. *Sūrikaisekikenkyūsho Kōkyūroku*, 1004:138–152, 1997. See also Monteiro and Tsuchiya [93].

[92] R. D. C. Monteiro and T. Tsuchiya. Polynomial convergence of a new family of primal–dual algorithms for semidefinite programming. *SIAM Journal on Optimization*, 9(3):551–577, 1999.

[93] R. D. C. Monteiro and T. Tsuchiya. Polynomiality of primal–dual algorithms for semidefinite linear complementarity problems based on the Kojima–Shindoh–Hara family of directions. *Mathematical Programming*, 84(1):39–53, 1999. See also Monteiro and Tsuchiya [91].

[94] R. D. C. Monteiro and Y. Zhang. A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Mathematical Programming*, 81(3, Ser. A):281–299, 1998.

[95] R. D.C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7(3):636–678, 1997.

[96] Y. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms.* SIAM Publications. SIAM, Philadelphia, USA, 1993.

[97] Yu. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming.* SIAM Publications. SIAM, Philadelphia, USA, 1994.

[98] Yu. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, 1997.

[99] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8(2):324–364, 1998.

[100] H. Neudecker. *Some Theorems on Matrix Differentiation with special reference to Kronecker Matrix Products*, volume 64. J. Amer. Statist. Assoc., 1969.

[101] G. I. Orlova and Ya.G. Dorfman. Finding the maximum cut in a graph. *Engineering Cybernetics*, 10(3):502–506, 1972.

[102] S. Poljak. Polyhedral and eigenvalue approximations of the max–cut problem. In G. Halász et al., editor, *Sets, Graphs and Numbers*, volume 60 of *Colloquia Mathematica Societatis Lános Bolyai*, pages 569–581. North Holland Publishing Company, Amsterdam, 1992.

[103] S. Poljak and F. Rendl. Nonpolyhedral relaxations of graph-bisection problems. *SIAM Journal on Optimization*, 5(3):467–487, 1995.

[104] S. Poljak, F. Rendl, and H. Wolkowicz. A recipe for semidefinite relaxation for $(0,1)$-quadratic programming. *Journal of Global Optimization*, 7(1):51–73, 1995.

[105] S. Poljak and Z. Tuza. On the expected relative error of the polyhedral approximation of the max-cut problem. *Operations Research Letters*, 16:191–198, 1994.

[106] S. Poljak and Z. Tuza. The max-cut problem – a survey. In *Special Year on Combinatorial Optimization*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science. AMS, 1995.

[107] F. A. Potra and R. Sheng. Homogeneous interior–point algorithms for semidefinite programming. Reports on Computational Mathematics 82, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, 1995.

[108] F. A. Potra and R. Sheng. A superlinearly convergent primal–dual infeasible–interior–point algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8(4):1007–1028, 1998.

[109] M. C. Pullan. A duality theory for separated continuous linear programs. *SIAM Journal on Control and Optimization*, 34(3):931–965, 1996.

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

[111] M. V. Ramana, L. Tuncel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM Journal on Optimization*, 7(3):641–662, 1997.

[112] F. Rendl. Semidefinite programming and combinatorial optimization. *Applied Numerical Mathematics*, 29:255–281, 1999.

[113] R. Tyrrell Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematics Series*. Princeton University Press, Princeton, 1970.

[114] L. A. Sanchis. Test case construction for the vertex cover problem. In *Computational Support for Discrete Mathematics*, volume 15 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 1052–1798. American Mathematical Society, 1994.

[115] A. Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Transactions on Information Theory*, IT-25:425–429, 1979.

[116] C. E. Shannon. The zero-error capacity of a noisy channel. *I.R.E. Transactions*, 3, 1956.

[117] H. D. Sherali and W. P. Adams. A hierarchy of relaxation between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal Discrete Mathematics*, 3(3):411–430, 1990.

[118] J. F. Sturm. *Primal–dual interior point approach to semidefinite programming*. PhD thesis, Tinbergen Institute, Amsterdam, The Netherlands, 1997.

[119] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11:625–653, 1999.

[120] 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.

[121] M. J. Todd and Y. Ye. Approximate Farkas Lemmas and stopping rules for iterative infeasible–point algorithms for linear programming. *Mathematical Programming*, 81(1):1–21, 1998.

[122] K. C. Toh, M. J. Todd, and R. Tutuncu. SDPT3 — a matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.

[123] P. Tseng. Analysis of infeasible path-following methods using the Alizadeh-Haeberley-Overton direction for the monotone semi-definite LCP. Technical report, Department of Mathematics, University of Washington, Seattle, WA, 1996.

[124] P. Tseng. Search directions and convergence analysis of some infeasible path-following methods for the monotone semi-definite LCP. *Optimization Methods and Software*, 9:245–268, 1998.

[125] A. W. Tucker. Dual systems of homogeneous linear relations. In *Linear equalities and related systems*, pages 3–18. Princeton University Press, Princeton, N. J., 1956.

[126] L. Vandenberghe and S. Boyd. A primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69(1, Ser. B):205–236, 1995. Nondifferentiable and large-scale optimization (Geneva, 1992).

[127] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[128] J.-Ph. Vial. Computational experience with a primal–dual interior–point method for smooth convex programming. In I. Maros, editor, *Symposium on Applied Mathematical Programming and Modeling (APMOD '93), Budapest, Hungary, January 1993, Volume of Extended Abstracts*, pages 583–586. Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, Hungary, 1992. Full version in Vial [129].

[129] J.-Ph. Vial. Computational experience with a primal–dual interior–point method for smooth convex programming. *Optimization Methods and Software*, 3:285–310, 1994. Condensed version in Vial [128].

[130] H. Wolkowicz. Some applications of optimization in matrix theory. *Linear Algebra and its Applications*, 40:101–118, 1981.

[131] S. J. Wright. *Primal-Dual Interior-Point Methods.* Society for Industrial and Applied Mathematics, 3600 University City Science Center, Philadelphia, 1997.

[132] X. J. Xu, P.-F. Hung, and Y. Ye. A simplified homogeneous and self–dual linear programming algorithm and its implementation. *Annals of Operations Research*, 62:151–171, 1996.

[133] 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(1):53–67, 1994.

[134] W. I. Zangwill. *Nonlinear Programming. A Unified Approach.* Prentice-Hall, Englewood Cliffs, NJ, 1969.

[135] 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.

[136] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *Journal of Combinatorial Optimization*, 2(1):71–109, 1998.

Index