Full-Newton Step Interior-Point Methods for Conic Optimization

Full-Newton Step Interior-Point Methods for Conic Optimization

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
maandag 16 juni 2008 om 12.30 uur

door

Hossein MANSOURI

Master of Science in Applied mathematics
Sharif University of Technology, Tehran, Iran
geboren te Zarand, Kerman, Iran.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. C. Roos

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. C. Roos	Technische Universiteit Delft, promotor
Prof. dr. A. Nemirovski	Georgia Institute of Technology, USA
Prof. dr. A. Yoshise	Tsukuba University, Japan
Prof. dr. F. Glineur	Université Catholique de Louvain
Prof. dr. K. Aardal	Centrum voor Wiskunde en Informatica
Prof. dr. Y. Bai	Shanghai University, China
Prof. dr. ir. C. Vuik	Technische Universiteit Delft
Prof. dr. ir. H. J. Sips	Technische Universiteit Delft, reservelid

Financial support was provided by the Iranian Ministry of Science, Research, and Technology, and by the Shahrekord University.

Dit proefschrift kwam tot stand onder auspiciën van:

THOMAS STIELTJES INSTITUTE
FOR MATHEMATICS

ISBN 978-90-9023179-2

Typeset by the author with the LaTeX Documentation system.

Author email: H.Mansouri@tudelft.nl / mansouri@sci.sku.ac.ir

To my wife, Maryam, and
our daughter, Kiana

Acknowledgements

I would like first to thank my promotor and supervisor, Prof. dr. ir. C. Roos, for his constant scientific support, supervision, guidance and enriching comments during my PhD research, especially for his patience on reading and commenting this thesis.

Special thanks to Dr. Marielba Rojas for proofreading the first version of Chapter 1 of this thesis. I would like to thank Prof. F. Glineur for his careful reading and helpful comments. I thank also the computer support group and secretaries at the Department of Software Technology.

My research was financially supported by the Iranian Ministry of Science, Research, and Technology, and Shahrekord University which is kindly acknowledged.

My deep appreciations and thanks also go to my colleagues of the optimization group, at Delft university of Technology for their precious friendship and contribution to my life as a human and as a researcher. I also greatly appreciate the help, co-operation and friendship of all my Iranian friends in Holland.

I would like to thank my family specially my parents for all that they have done for me, which is too much to write down. Last but not least, I sincerely thank my wife Maryam and our daughter Kiana, to whom I dedicate this thesis, for all of their love and compassion. They understood the many times I was physically or mentally absent from them. For this, I am in their debt.

Delft, June 2008

Hossein Mansouri

Contents

List of Figures

List of Tables

List of Abbreviations

IPC	:	Interior-Point Condition;
IPM(s)	:	Interior-Point Method(s);
IIPM(s)	:	Infeasible Interior-Point Method(s);
LO	:	Linear Optimization;
SDO	:	Semidefinite Optimization;
SOCO	:	Second-Order Cone Optimization;
SC	:	Self-Concordant;
SCB	:	Self-Concordant Barrier;

List of Notations

$\mathbf{R}^n$	:	n-dimensional Euclidean vector space;
$\mathbf{R}^n_+$	:	positive orthant of $\mathbf{R}^n$;
$\mathbf{R}^{n\times n}$	:	space of real n-by-n matrices;
A^T	:	transpose of $A \in \mathbf{R}^{m\times n}$;
A_{ij}	:	(i, j)-entry of $A \in \mathbf{R}^{m\times n}$;
e	:	all-one vector;
I	:	identity matrix in suitable space;
μ	:	barrier parameter (or duality gap parameter);
μ^0	:	initial value of μ;
x_i	:	i-th coordinate of the vector x;
x^T	:	transpose of vector x;
$\|x\|$	:	the 2-norm of vector x;
$\|x\|_\infty$	:	the infinity norm of vector x;
x_{max}	:	the maximal component of x;

x_{min}	:	the minimal component of x;
Δx	:	search direction in x-space;
Δs	:	search direction in s-space;
v	=	$\sqrt{\frac{xs}{\mu}}$;
d_x	:	primal search direction in the scaled v-space: $d_x = \frac{v\Delta x}{x}$;
d_s	:	dual search direction in the scaled v-space: $d_s = \frac{v\Delta s}{s}$;
ϵ	:	accuracy parameter;
τ	:	proximity parameter;
θ	:	used in updating factor for μ;
$\mathbf{S}^n$	=	$\left\{X : X \in \mathbf{R}^{n\times n},\, X = X^T\right\}$;
$A \succeq 0 (A \succ 0)$	:	A is symmetric positive semidefinite (positive definite);
$\mathbf{S}^n_+$	=	$\{X : X \in \mathbf{S}^n,\, X \succeq 0\}$;
$\lambda_i(A)$	:	i-th eigenvalue of matrix $A \in \mathbf{R}^{n\times n}$;
$\lambda_{\max}(A)$	=	$\max_i \lambda_i(A)$, if $\lambda_i \in \mathbf{R}$ for all i;
$\lambda_{\min}(A)$	=	$\min_i \lambda_i(A)$, if $\lambda_i \in \mathbf{R}$ for all i;
Tr (A)	=	$\sum_i A_{ii} = \sum_i \lambda_i(A)$: trace of $A \in \mathbf{R}^{n\times n}$;
det (A)	:	determinant of $A \in \mathbf{R}^{n\times n} = \prod_i \lambda_i(A)$;
$\|A\|^2$	=	Tr $\left(AA^T\right) = \sum_i \sum_j A_{ij}^2$ (Frobenius norm)
	=	$\sum_i \lambda_i^2(A)$ if $A \in \mathbf{S}^n$;
$A \sim B$	:	matrices A and B are similar;
$A \bullet B$	=	Tr $\left(A^T B\right)$;
$A^{\frac{1}{2}}$	:	unique positive semidefinite square root of $A \succeq 0$;

$\log(t)$: natural logarithm of t;

(P) : primal LO problem in standard form;

(D) : dual problem of (P);

(SDP) : primal SDO problem in standard form;

(SDD) : dual problem of (SDP);

$\text{relint}(\mathcal{C})$: relative interior of a convex set $\mathcal{C}$;

$\mathcal{P}$: feasible set of problem (SDP);

$\mathcal{D}$: feasible set of problem (SDD);

$\mathcal{P}^*$: optimal set of problem (SDP);

$\mathcal{D}^*$: optimal set of problem (SDD);

$\mathcal{F}^* = \{(X,\, y,\, S) \in \mathcal{P} \times \mathcal{D} : \text{Tr}\,(XS) = 0\}$;

$D = \left[X^{\frac{1}{2}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{-\frac{1}{2}}X^{\frac{1}{2}}\right]^{-\frac{1}{2}}$ (Nesterov-Todd scaling matrix);

$\mathbf{vec}(A) = (A_{11};\, A_{21};\, \ldots;\, A_{n1};\, A_{12};\, A_{22};\, \ldots;\, A_{nn})$ for $A \in \mathbf{R}^{n\times n}$;

Chapter 1

Introduction

1.1 Introduction

A linear optimization (LO) problem is a problem of minimizing or maximizing a linear function in the presence of linear constraints of the inequality and /or the equality type. The first method to solve LO problems was the simplex method developed by George Dantzig in 1947 [16]. This algorithm explores the combinatorial structure of the feasible set to identify a solution by moving from a vertex of the feasible region to an adjacent vertex while improving the value of the objective function. Not only did the simplex method prove to be very efficient in practice but it also facilitated the development of duality theory.

The simplex method remained the champion for solving LO problems even after Klee and Minty [43] established its non-polynomiality. They showed that the simplex method has exponential complexity in the worst case. This means that it is impossible to bound the computational time required for the simplex method to solve any given LO problem by a polynomial function of the "input size" of a given problem.[1] For LO, the first polynomial-time algorithm is due to Khachiyan in 1979 [42]. His approach is based on the ellipsoid method for nonlinear optimization, developed by Shor [87]. With this technique, Khachiyan proved that LO is in the class of polynomially solvable problems. In spite of having a polynomial-time worst-case complexity, Khachiyan's method fell far short of the expectation that it would be faster than the simplex method in practice.

A more important development was the publication of Karmarkar's paper in 1984 [41], describing a new polynomial-time interior-point method (IPM). After the publication of Karmarkar's algorithm, researchers developed many algorithms inspired by different features of that algorithm and its analysis. Gill et al. [24]

[1] The input size of a problem (denoted by L) indicates the length of a binary coding of the input data.

observed that some simple variants of Karmarkar's algorithm could be traced back to a very old algorithm in nonlinear optimization, namely, the logarithmic barrier method of Frisch [22] and Fiacco and McCormick [20], and the method of centers of Huard [36, 37]. Roos and Vial [85] proved that the basic logarithmic barrier method for LO has polynomial complexity. Renegar [80] proposed to use an upper bound on the optimal value of the objective function to form successively smaller subsets of the feasible region, and employ Newton's method to follow the so-called *analytic centers* of these subsets to get a primal optimal solution. Another very important concept in the IPM literature is the *central path*, which was first recognized by Sonnevend [88] and Meggido [57]. Almost all known polynomial-time variants of IPMs use the central path as a guideline to the optimal set and some variant of Newton's method to follow the central path approximately. These Newton-type methods fall into different groups with respect to the strategies used in the algorithms to follow the central path. For more details see [29, 33].

Besides having polynomial complexity, IPMs are also highly efficient in practice. One may distinguish between IPMs according to wether they are primal-only, dual-only or primal-dual, and whether they are feasible IPMs or infeasible IPMs (IIPMs). Primal IPMs are called primal because they keep track only of the primal variables. These methods do not give any information about the solution of the dual problem nor the duality gap. Dual IPMs keep track only of the dual variables and they do not give any information about the solution of the primal problem. Primal-dual IPMs iterate on both the primal and dual variables and find a primal-dual solution (x^*, y^*, s^*) such that x^* and (y^*, s^*) are optimal solutions of the primal and dual problem, respectively. Primal-dual IPMs have proven to be among the most efficient methods for linear optimization, and many polynomiality results exist for these methods. The first primal-dual IPM for LO was constructed by Kojima et al. [45] based on the work of Megiddo [57]. See also [62, 92].

Feasible IPMs start with a strictly feasible point and maintain feasibility during the solution process. For feasible IPMs, it is not at all trivial how to find an initial feasible interior point. One strategy to overcome this problem is to use the homogeneous embedding model as introduced first by Ye et al. [106] for LO, and further developed in [4, 84, 104]. On the other hand IIPMs start with an arbitrary positive point and feasibility is reached as optimality is approached. The choice of the starting point in IIPMs is crucial for the performance. Lustig [52] and Tanabe [93] were the first to present IIPMs for LO. The first theoretical result on primal-dual IIPMs was obtained by Kojima, Meggido and Mizuno [44]. They showed that an infeasible-interior-point variant of the primal-dual feasible IPM studied in [59] is globally convergent. The first polynomial-complexity result was obtained by Zhang [107] who proved that, with proper initialization, an IIPM has $O\left(n^2 \log \frac{1}{\epsilon}\right)$-iteration complexity. Shortly after that, Mizuno [58] proved that the Kojima-Meggido-Mizuno algorithm also has $O\left(n^2 \log \frac{1}{\epsilon}\right)$-iteration complexity. In

[58, 78], two predictor-corrector IIPMs with $O\left(n \log \frac{1}{\epsilon}\right)$-iteration complexity are proposed. Several other authors tried to improve the complexity bounds and relax some of the assumptions on the initial point used by the algorithm. See, e.g., [12, 13, 21, 46, 60, 67, 68, 77, 86, 95, 101, 102]. More details on IIPMs can be found in the books by Vanderbei [97], Wright [103] and Ye [105].

1.2 Preliminaries

This section is introductory in its nature. Its purpose is to provide the background material needed in the rest of this thesis as well as to introduce the notations used throughout this work. The development of most of the basic concepts of this chapter is based on [10, 76, 84, 105].

1.2.1 Primal-dual model for LO and duality theory

There are many different ways to present the LO problem. One particular formulation of the linear optimization problem, the standard form, is frequently used to describe and analyze algorithms. This form is

$$(P) \qquad \begin{array}{rcl} \min & c^T x & \\ \text{s.t} & Ax & = b, \\ & x & \geq 0, \end{array}$$

where c and x are vectors in $\mathbf{R}^n$, b is a vector in $\mathbf{R}^m$, and A is an $m \times n$ matrix, where $m \leq n$. Without loss of generality we assume that A has full row rank.

We can convert any LO problem into standard form by introducing additional variables, as slack variables and other artificial variables, into its formulation [27].

Associated with any linear optimization problem is another linear optimization problem called the *dual* problem, which consists of the same data objects arranged in a different way. The dual of (P) is

$$(D) \qquad \begin{array}{rcl} \max & b^T y & \\ \text{s.t} & A^T y + s & = c, \\ & s & \geq 0, \end{array}$$

where y is a vector in $\mathbf{R}^m$ and s is a vector in $\mathbf{R}^n$. We call the components of y the *dual variables*, while s is the vector of *dual slacks*. The dual problem could be stated more compactly by eliminating s from (D) and rewriting the constraints as $A^T y \leq c$. However, in many cases it turns out to be expedient for the analysis and implementation of IPMs to include the slack vector s.
The linear optimization problem (P) is often called the *primal* problem, and (D)

the *dual* problem, and together the two problems are referred to as the *primal-dual pair*.

Definition 1.2.1. *If x satisfies the constraints $Ax = b$, $x \geq 0$ (or $x > 0$), we call it a feasible (or strictly feasible) point for (P). A pair (y, s) is feasible (or strictly feasible) for (D), if $A^T y + s = c$ and $s \geq 0$ (or $s > 0$). The triple (x, y, s) is primal-dual feasible (or strictly feasible) if x and (y, s) are feasible (or strictly feasible) for the primal and dual problems, respectively.*

A duality theory exists that explains the relationship between the two problems (P) and (D). We recall below the main results of this theory.

Theorem 1.2.2 (Weak Duality Theorem)**.** *If x is feasible for (P) and (y, s) is feasible for (D), then*

$$c^T x \geq b^T y.$$

Proof. By the construction of the primal and dual, we immediately get

$$c^T x - b^T y = \left(A^T y + s\right)^T x - b^T y = y^T Ax + s^T x - b^T y = s^T x \geq 0, \quad (1.1)$$

where the inequality comes from the non-negativity of x, s. □

Theorem 1.2.3 (Strong Duality Theorem)**.** *If (P) has an optimal solution then so does its dual. Furthermore, then their optimal values are equal.*

By Theorem 1.2.2, if (x, y, s) is a primal-dual feasible pair then $b^T y \leq c^T x$. In other words, the objective value in (D) gives a lower bound for the objective in (P), and the objective in (P) provides an upper bound for that in (D). Using Theorem 1.2.2 and Theorem 1.2.3, the main duality results can be summarized by the following theorem [38, 84].

Theorem 1.2.4. *For (P) and (D) one of the following holds:*

(i) (P) and (D) are feasible and there exists a primal-dual feasible pair (x^*, y^*, s^*) such that

$$c^T x^* = b^T y^*.$$

(ii) (P) is infeasible and (D) is unbounded.

(iii) (D) is infeasible and (P) is unbounded.

(iv) Both (P) and (D) are infeasible.

Hence, solving LO amounts to detecting which of these four cases holds, and in case (i) an optimal solution (x^*, y^*, s^*) must be found. Note that in case (i),

the two objective values in (P) and (D) coincide with each other at the solution (x^*, y^*, s^*), that is $c^T x^* = b^T y^*$, which implies by (1.1) that

$$(s^*)^T x^* = 0.$$

Observe that since $x^*, s^* \geq 0$, the above equality can also be written as

$$x_i^* s_i^* = 0, \quad i = 1, ..., n.$$

An intrinsic property of LO is given by the following result first established by Goldman and Tuker [28]. This theorem plays an important role in the design and analysis of IPMs [84, 105].

Theorem 1.2.5. *Suppose that both (P) and (D) are feasible. Then there exists a primal-dual pair (x^*, y^*, s^*) such that $(x^*)^T s^* = 0$ and $x^* + s^* > 0$. A solution (x^*, s^*) with this property is called strictly complementary.*

1.3 Primal-dual IPMs for LO

In this section we proceed by describing primal-dual IPMs for LO. All these methods start from an initial point (x^0, y^0, s^0) with $x^0, s^0 > 0$ and generate a sequence of positive points such that the sequence converges to the optimal solution. If, at each iteration, the point also satisfies the linear equality constraints for the primal and dual problem (P) and (D), then the algorithm is called a *feasible interior-point* algorithm. Otherwise, it is called an *infeasible interior-point* algorithm.

1.3.1 Feasible primal-dual IPMs for LO

Feasible primal-dual IPMs start with a strictly feasible interior point and maintain feasibility during the solution process. As mentioned in Section 1.1 for LO problems we can assume, without loss of generality, that both (P) and (D) satisfy the *interior-point condition* (IPC). That is, there exists a triple (x^0, y^0, s^0) such that

$$Ax^0 = b, \quad x^0 > 0, \quad A^T y^0 + s^0 = c, \quad s^0 > 0. \tag{1.2}$$

By using the self-dual embedding model, we can further assume that $x^0 = s^0 = e$. Under the IPC, we are in case (i) of Theorem 1.2.4 which implies that there exists an optimal solution pair. Hence Theorem 1.2.4 states that finding an optimal solution of (P) and (D) is equivalent to solving the following system:

$$\begin{aligned} Ax &= b, \quad \text{(Primal feasibility)} \\ A^T y + s &= c, \quad \text{(Dual feasibility)} \\ xs &= 0, \quad \text{(Complementary slackness)} \\ x \geq 0 \ &, \ s \geq 0. \end{aligned} \tag{1.3}$$

Here xs denotes the coordinatewise product of the vectors x and s, i.e.,

$$xs = (x_1 s_1;\ \ldots;\ x_n s_n)\,.$$

We also use the notation

$$\frac{x}{s} = \left(\frac{x_1}{s_1};\ \ldots;\ \frac{x_n}{s_n}\right),$$

for each vector x and s such that $s_i \neq 0$, for all $i = 1, \ldots, n$.

The basic idea of primal-dual interior-point methods is to replace the third equation (which is nonlinear), called the complementary slackness condition, in (1.3) by the nonlinear equation $xs = \mu e$, where e is the all-one vector, and $\mu > 0$. Thus we consider the following system

$$\begin{aligned} Ax &= b, \quad x \geq 0 \\ A^T y + s &= c, \quad s \geq 0 \\ xs &= \mu e. \end{aligned} \tag{1.4}$$

The existence of a unique solution to the above system is well-known (see [30, 45, 56]). We denote the unique solution of system (1.4) by $(x(\mu),\, y(\mu),\, s(\mu))$ for each $\mu > 0$; $x(\mu)$ is called the μ-center of (P) and $(y(\mu),\, s(\mu))$ is the μ-center of (D). The set of μ-centers (with $\mu > 0$) defines a homotopy path, which is called *the central path* of (P) and (D) [57, 88]. If μ goes to zero then the limit of the central path exists. This limit satisfies the complementarity condition, and hence yields optimal solutions for (P) and (D) [84].

IPMs follow the central path approximately. Let us briefly indicate how this goes. Without loss of generality, we assume that $(x(\mu),\, y(\mu),\, s(\mu))$ is known for some positive μ. As already mentioned, without loss of generality we can assume that $x^0 = s^0 = e$. Since then $x^0 s^0 = 1.e$, we see that system (1.4) holds with $\mu^0 = 1$. We now update μ by setting $\mu := (1-\theta)\,\mu$, for some $\theta \in (0,\, 1)$. Then we solve the following well-known Newton system, which is obtained by linearizing (1.4),

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= \mu e - xs. \end{aligned} \tag{1.5}$$

Because A has full row rank, the above system uniquely defines a search direction $(\Delta x,\, \Delta y,\, \Delta s)$ for any $x > 0$ and $s > 0$; this is the so-called Newton direction and this direction is used in all existing implementations of (feasible) primal-dual methods. By taking a step along the search direction where the step size "α" is defined by some line search rules, one constructs a new triple $(x^+,\, y^+,\, s^+)$, with positive x^+ and s^+. If necessary, we repeat the procedure until we find the iterates $(x^+,\, y^+,\, s^+)$ that are close enough to $(x(\mu),\, y(\mu),\, s(\mu))$. The new iterates are

given by

$$\begin{aligned} x^+ &= x + \alpha \Delta x, \\ y^+ &= y + \alpha \Delta y, \\ s^+ &= s + \alpha \Delta s. \end{aligned} \tag{1.6}$$

Then μ is again reduced by the factor $1 - \theta$ and we apply Newton's method targeting at the new μ-centers, and so on. This process is repeated until μ is small enough, say until $n\mu \leq \epsilon$; at this stage we have found ϵ-solutions of the problems (P) and (D).

For notational convenience, and as in [8, 72, 73, 84, 105], we define

$$v := \sqrt{\frac{xs}{\mu}}, \quad v^{-1} := \sqrt{\frac{\mu\, e}{xs}}. \tag{1.7}$$

$$d_x := \frac{v \Delta x}{x}, \quad d_s := \frac{v \Delta s}{s}. \tag{1.8}$$

Note that (1.7) implies that if x is primal feasible and s is dual feasible then the pair $(x,\, s)$ coincides with the μ-center $(x(\mu),\, s(\mu))$ if and only if $v = v^{-1} = e$.

Using this notation, the system (1.5) can be rewritten as follows:

$$\begin{aligned} \bar{A} d_x &= 0, \\ \bar{A}^T \Delta y + d_s &= 0, \\ d_x + d_s &= v^{-1} - v, \end{aligned} \tag{1.9}$$

where $\bar{A} = AV^{-1}X$ and $V := \operatorname{diag}(v), X := \operatorname{diag}(x)$. Note that since d_x belongs to the null space of the matrix $\bar{A}$ and d_s to its row space, it follows that d_x and d_s are orthogonal vectors, i.e.,

$$d_x^T d_s = 0.$$

In the same way, it follows from (1.5) that Δx and Δs are orthogonal,

$$(\Delta x)^T \Delta s = 0.$$

Using the third equation in (1.9) we obtain

$$\|d_x\|^2 + \|d_s\|^2 = \|d_x + d_s\|^2 = \left\|v^{-1} - v\right\|^2.$$

Note that $d_x = d_s = 0$ if and only if $\left\|v^{-1} - v\right\| = 0$. In this case, $v = e$ and hence x and s satisfy $xs = \mu e$, which implies that x, s are on the μ-center $(x(\mu),\, s(\mu))$. Thus, we can use $\left\|v - v^{-1}\right\|$ as a quantity to measure closeness to the pair of μ-centers. Following [84], we therefore define

$$\delta(x,\, s; \mu) := \delta(v) := \frac{1}{2} \left\|v - v^{-1}\right\|. \tag{1.10}$$

General Feasible Primal-Dual IPM for LO

Input:
Accuracy parameter $\epsilon > 0$;
a threshold parameter $\tau > 0$;
barrier update parameter θ, $0 < \theta < 1$.

begin
$x := e;\ s := e;\ \mu = 1;$
while $x^T s \geq \epsilon$ **do**
begin
$\mu := (1-\theta)\mu;$
$v := \sqrt{\frac{xs}{\mu}};$
while $\delta(v) \geq \tau$ **do**
begin
$x := x + \alpha\Delta x;$
$y := y + \alpha\Delta y;$
$s := s + \alpha\Delta s;$
$v := \sqrt{\frac{xs}{\mu}};$
end
end
end

Figure 1.1: *General Feasible Primal-Dual IPM for LO*

Using this proximity measure a generic primal-dual IPM is described in Figure 1.1. The parameters θ and α are very important both in theory and practice of IPMs. Usually, if θ is a constant independent of n, for instance $\theta = 0.9$, then we call the algorithm a large-update (or long-step) method. If θ depends on n, such as $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$, then the algorithm is called a small-update (or short-step) method. The choice of the step size α $(0 < \alpha \leq 1)$ is another crucial issue in the analysis of the algorithm. It has to be taken such that the closeness of the iterates to the current μ-center improves by a sufficient amount. In the theoretical analysis the step size α is usually given a default value that depends on the closeness of the current iterates to the μ-center. It is now known that small-update methods have an $O\left(\sqrt{n}\log\left(n/\epsilon\right)\right)$ iteration bound and that large-update methods have an $O\left(n\log\left(n/\epsilon\right)\right)$ iteration bound [84, 103, 105].

In this thesis, we focus on so-called full-Newton step IPMs. In these algorithms

the new iterates are given by

$$\begin{aligned} x^+ &= x + \Delta x, \\ y^+ &= y + \Delta y, \\ s^+ &= s + \Delta s. \end{aligned} \tag{1.11}$$

So the step size α equals 1 in each iteration. As we will see below we then need to take $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$, which means that full-Newton step methods belong to the class of small-update methods. A generic primal-dual IPM with full-Newton steps is described in Figure 1.2. Before proceeding with the analysis of this algorithm

Primal-Dual IPM with full-Newton steps for LO

Input:
Accuracy parameter $\epsilon > 0$;
a threshold parameter $\tau > 0$;
barrier update parameter θ, $0 < \theta < 1$.
begin
$x := e$; $s := e$; $\mu = 1$;
while $x^T s \geq \epsilon$ **do**
begin
$\mu := (1 - \theta)\mu$;
$x := x + \Delta x$;
$y := y + \Delta y$;
$s := s + \Delta s$;
end
end

Figure 1.2: *Primal-Dual IPM with full-Newton steps for LO*

we present a graphical illustration. We consider the problem with A, b and c as follows

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 2 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 11 \end{bmatrix}.$$

Figures 1.3 and 1.4 show the behavior of IPMs in the xs-space, with full Newton steps and large updates, respectively. In both figures the scale of the axes is logarithmic and the xs-space is projected onto its first two coordinates. In the xs-space the central path is represented by the half-line μe, $\mu > 0$. As starting

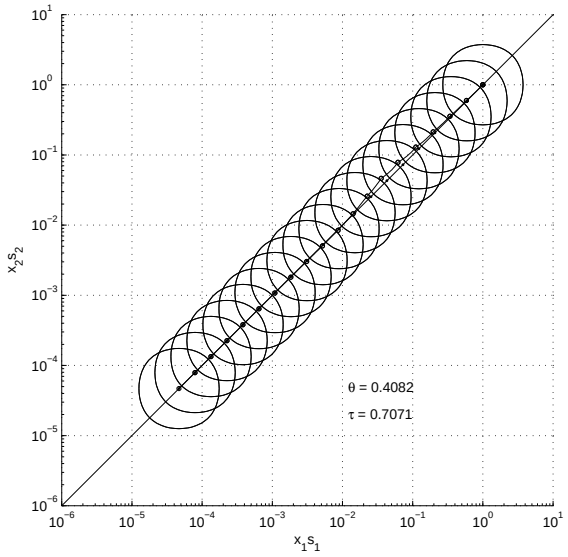

Figure 1.3: *Performance of the primal-dual IPM with full Newton steps*

point we use the vectors $x = s = (1;\ 1;\ 1)$ and $y = (0;\ 0)$ and we take the initial value of the barrier parameter μ equal to 1. Since $xs = (1;\ 1;\ 1) = 1 \cdot e$, it follows that the starting point is the μ-center for $\mu = 1$. In these figures we have drawn the iterates for the given problem and the level curves for $\delta(v) \leq \tau$ around the target points on the central path that are used during the algorithm.
We now recall some lemmas that are crucial in the analysis of the algorithm in Figure 1.2.

Lemma 1.3.1 (Lemma II.47 in [84])**.** *After a primal-dual full-Newton step one has* $(x^+)^T s^+ = n\mu$.

Proof. Using (1.11) and the fact that the vectors Δx and Δs are orthogonal, the duality gap after the Newton step can be written as follows:

$$\begin{aligned}\left(x^+\right)^T s^+ &= e^T\left((x+\Delta x)\left(s+\Delta s\right)\right)\\ &= e^T\left(xs + (x\Delta s + s\Delta x) + \Delta x\Delta s\right)\\ &= e^T\left(xs + (\mu e - xs) + \Delta x\Delta s\right)\\ &= \mu\, e^T e = n\mu,\end{aligned}$$

where we used the third equation in (1.5) for the third equality. This proves the lemma. □

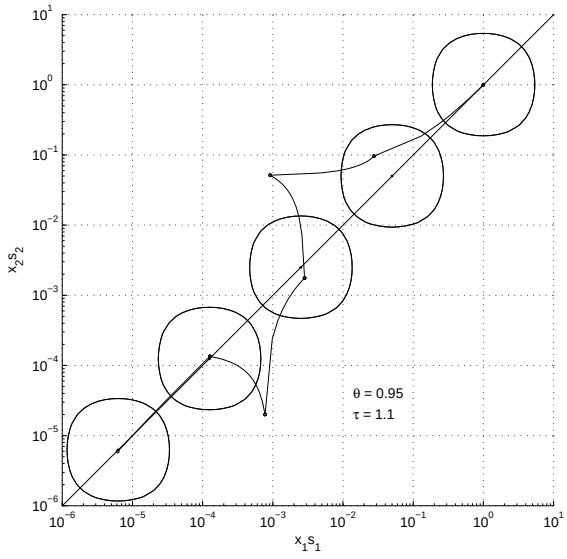

Figure 1.4: *Performance of a large-update primal-dual IPM*

Lemma 1.3.2 (Lemma II.54 in [84])**.** *Let* $(x,\, s)$ *be a positive primal-dual pair and* $\mu > 0$ *such that* $x^T s = n\mu$*. Moreover, let* $\delta := \delta\left(x,\, s; \mu\right)$ *and let* $\mu^+ = (1-\theta)\,\mu$*. Then*

$$\delta\left(x,\, s; \mu^+\right)^2 = (1-\theta)\,\delta^2 + \frac{\theta^2 n}{4\,(1-\theta)}.$$

Proof. We may write, using (1.10) and (1.7),

$$4\delta\left(x,\, s; \mu^+\right)^2 = \left\| \sqrt{1-\theta}\, v^{-1} - \frac{v}{\sqrt{1-\theta}} \right\|^2 = \left\| \sqrt{1-\theta}\left(v^{-1} - v\right) - \frac{\theta v}{\sqrt{1-\theta}} \right\|^2.$$

From $x^T s = \mu n$ it follows that $\|v\|^2 = n$. Since $v^T\left(v^{-1} - v\right) = n - \|v\|^2 = 0$ the vectors v and $v^{-1} - v$ are orthogonal. Therefore,

$$4\delta\left(x,\, s; \mu^+\right)^2 = (1-\theta)\left\|v^{-1} - v\right\|^2 + \frac{\theta^2 \|v\|^2}{1-\theta}.$$

Finally, since $\left\|v^{-1} - v\right\| = 2\delta$ and $\|v\|^2 = n$ the result follows. □

From (1.11) and the third equation in (1.5) we have

$$x^+ s^+ = \mu\, e + \Delta x \Delta s = \mu\left(e + d_x d_s\right), \tag{1.12}$$

where for the last equality we used (1.7) and (1.8). In the following we recall two lemmas, which are essential for our main Lemma 1.3.5.

Lemma 1.3.3 (Lemma II.48 in [84]). *The primal-dual Newton step is feasible if and only if*

$$e + d_x d_s \geq 0$$

and strictly feasible if and only if

$$e + d_x d_s > 0.$$

Lemma 1.3.4 (Lemma II.49 in [84]). *Let $(x,\, s)$ be any positive primal-dual pair and suppose $\mu > 0$. If $\delta := \delta(x,\, s;\, \mu)$, then $\|d_x d_s\|_\infty \leq \delta^2$ and $\|d_x d_s\| \leq \sqrt{2}\delta^2$.*

Lemma 1.3.5 (Theorem II.50 in [84]). *If $\delta(x,\, s; \mu) \leq 1$, then the primal-dual full-Newton step is feasible, i.e., x^+ and s^+ are nonnegative. Moreover, if $\delta < 1$, then x^+ and s^+ are positive and*

$$\delta\left(x^+,\, s^+; \mu\right) \leq \frac{\delta^2}{\sqrt{2\left(1-\delta^2\right)}}.$$

Proof. Lemma 1.3.3 and Lemma 1.3.4 imply the first part of the lemma. Now we want to prove the second statement. Let $\delta^+ := \delta\left(x^+,\, s^+;\, \mu\right)$ and

$$v^+ := \sqrt{\frac{x^+ s^+}{\mu}}.$$

Then by using $\delta(v)$ as defined in (1.10) we have

$$2\delta^+ = \left\| \left(v^+\right)^{-1} - v^+ \right\| = \left\| \left(v^+\right)^{-1} \left(e - \left(v^+\right)^2\right) \right\|.$$

Moreover, (1.7) and (1.12) imply that

$$v^+ = \sqrt{e + d_x d_s}.$$

Substitution gives

$$2\delta^+ = \left\| \frac{d_x d_s}{\sqrt{e + d_x d_s}} \right\| \leq \frac{\|d_x d_s\|}{\sqrt{1 - \|d_x d_s\|_\infty}}.$$

Now using the bounds in Lemma 1.3.4 we obtain

$$2\delta^+ \leq \frac{\delta^2 \sqrt{2}}{\sqrt{1-\delta^2}}.$$

Dividing the both sides of the last inequality by 2 gives the result in the lemma. □

Corollary 1.3.6. *If $\delta := \delta(x,\, s;\, \mu) \leq \frac{1}{\sqrt{2}}$, then $\delta\left(x^+,\, s^+; \mu\right) \leq \delta^2$.*

This corollary makes clear that if x, s and μ are such that $\delta(x,\, s;\, \mu) \leq \frac{1}{\sqrt{2}}$ then the Newton process converges quadratically fast to the μ-center. We say then that the pair $(x;\, s)$ lies in the region of quadratic convergence of the μ-center.

Theorem 1.3.7 (Theorem II.53 in [84]). *If $\tau = \frac{1}{\sqrt{2}}$ and $\theta = \frac{1}{\sqrt{2n}}$, then the primal-dual full-Newton IPM requires at most*

$$\sqrt{2n}\log\frac{n}{\epsilon}$$

iterations. The output is a primal-dual pair $(x;\, s)$ such that $x^T s \leq \epsilon$.

Proof. At the start of the algorithm we have $\delta(x,\, s;\, \mu) \leq \tau = \frac{1}{\sqrt{2}}$ and the duality gap corresponding to the starting point is $n\,\mu^0 = n$. After the primal-dual Newton step to the μ-center we have, by Lemma 1.3.5, $\delta(x^+,\, s^+;\, \mu) \leq \tau^2 = \frac{1}{2}$. Also, from Lemma 1.3.1, $(x^+)^T s^+ = n\mu$. Then, after the barrier parameter is updated to $\mu^+ = (1-\theta)\,\mu$, with $\theta = \frac{1}{\sqrt{2n}}$, Lemma 1.3.2 yields the following upper bound for $\delta(x^+,\, s^+;\, \mu^+)$:

$$\delta\left(x^+,\, s^+;\, \mu^+\right)^2 \leq \frac{1-\theta}{4} + \frac{1}{8\,(1-\theta)} \leq \frac{3}{8}.$$

Assuming $n \geq 2$, the last inequality follows since its left hand side is a convex function of θ, whose value is $3/8$ both in $\theta = 0$ and $\theta = 1/2$. Since $\theta \in [0,\, 1/2]$, the left hand side does not exceed $3/8$. Since $3/8 < 1/2$, we obtain $\delta(x^+,\, s^+;\, \mu^+) \leq \frac{1}{\sqrt{2}} = \tau$. Thus, after each iteration of the algorithm the property

$$\delta(x,\, s;\, \mu) \leq \tau$$

is maintained, and hence the algorithm is well defined. At the start of the algorithm duality gap is n and in each iteration duality gap is reduced by the factor $1-\theta$. Therefore, the duality gap, given by $x^T s = n\mu$, is smaller than, or equal to ϵ if

$$(1-\theta)^k\, n \leq \epsilon.$$

taking logarithms, this becomes

$$k\log(1-\theta) + \log(n) \leq \log\epsilon.$$

Since $-\log(1-\theta) \geq \theta$, this certainly holds if

$$k\theta \geq \log(n) - \log\epsilon = \log\frac{n}{\epsilon}.$$

Substituting $\theta = \frac{1}{\sqrt{2n}}$ the theorem follows. □

1.3.2 Infeasible Primal-dual IPMs for LO

So far, we have assumed that the starting point (x^0, y^0, s^0) is strictly feasible and, in particular, that it satisfies the linear equations $Ax^0 = b$, $A^T y^0 + s^0 = c$. All subsequent iterates also respect these constraints, because of the zero right-hand side terms in (1.5). In general, however, no strictly feasible starting point is known for a given primal-dual pair of problems (P) and (D). In such cases one might consider to use a so-called infeasible IPM (IIPM). In this section we briefly survey the ideas underlying IIPMs and we present the main complexity results for such methods.

IIPMs start at (x^0, y^0, s^0) with $x^0 > 0$, $s^0 > 0$ and usually $y^0 = 0$, but x^0 and (y^0, s^0) are not feasible for (P) and (D), respectively. This means that the initial primal and dual residual vectors, defined by

$$r_b^0 = b - Ax^0, \tag{1.13}$$

$$r_c^0 = c - A^T y^0 - s^0, \tag{1.14}$$

are nonzero. As we will see, the iterates will be infeasible as well, but during the course of the algorithm the residual vectors converge to zero, so that the final iterates are "almost" feasible. In case of IIPMs we call the triple (x, y, s) an ϵ-solution of (P) and (D) if the norms of the residual vectors do not exceed ϵ, and also $x^T s \leq \epsilon$.

As mentioned in Section 1.1 the first theoretical result on primal-dual IIPMs was obtained by Kojima et al. [44]. Then Zhang [107] for first time designed a primal-dual IIPM for complementary problems with polynomial complexity. In [58, 78, 103, 105], the authors also presented some primal-dual IIPMs with polynomial-complexity. In all algorithms with polynomial complexity the starting point is chosen as

$$(x^0, y^0, s^0) = \zeta\,(e, 0, e) \tag{1.15}$$

with $\zeta > 0$ such that

$$\|x^* + s^*\|_\infty \leq \zeta \tag{1.16}$$

for some optimal solution x^* of (P) and (y^*, s^*) of (D).

Now let us briefly describe the IIPM algorithms mentioned above. We first consider Wright's algorithm because it looks simpler than the others.

1.3.3 The IIPM of Wright

In this subsection we briefly describe Wright's algorithm [103]. In his algorithm he uses some constants $\sigma_{\min}$, $\sigma_{\max}$, γ and β such that $0 < \sigma_{\min} < \sigma_{\max} \leq \frac{1}{2}$, $\gamma \in (0, 1)$ and $\beta \geq 1$. This algorithm works with a neighborhood of the central path

that is defined as follows:

$$\mathcal{N}_{-\infty}(\gamma,\,\beta) := \left\{(x,\,y,\,s)\,:\,\frac{\|(r_b,\,r_c)\|}{\mu} \leq \beta\,\frac{\left\|\left(r_b^0,\,r_c^0\right)\right\|}{\mu^0},\;xs \geq \gamma\mu e,\;(x;\,s) > 0\right\},$$

where

$$\mu = \frac{x^T s}{n}, \quad \mu^0 = \frac{\left(x^0\right)^T s^0}{n}, \tag{1.17}$$

and r_b and r_c denote the residual vectors for the primal and dual problem, respectively:

$$r_b = b - Ax, \tag{1.18}$$

$$r_c = c - A^T y - s. \tag{1.19}$$

Recall that r_b^0 and r_c^0 are the initial primal and dual residual vectors as defined in (1.13) and (1.14). Without loss of generality, we suppose that the current point $(x,\,y,\,s)$ is in the neighborhood $\mathcal{N}_{-\infty}(\gamma,\,\beta)$ and μ is as defined in (1.17). Note that since $\beta \geq 1$, at the start of the first iteration this is certainly true.

The search directions are designed in such a way that they target at the $\sigma\mu$-centers of (P) and (D), where σ is an arbitrary number in $[\sigma_{\min},\,\sigma_{\max}]$. Ideally Δx, Δy and Δs would satisfy the following system,

$$\begin{aligned} A\,(x+\Delta x) &= b,\\ A^T\,(y+\Delta y) + s + \Delta s &= c,\\ (x+\Delta x)\,(s+\Delta s) &= \sigma\mu\,e. \end{aligned} \tag{1.20}$$

Since the third equation is nonlinear it is linearized, which leads to the following linear system of equations in Δx, Δy and Δs :

$$\begin{aligned} A\Delta x &= b - Ax,\\ A^T\Delta y + \Delta s &= c - A^T y - s,\\ x\Delta s + s\Delta x &= \sigma\mu\,e - xs. \end{aligned} \tag{1.21}$$

Using (1.18) and (1.19), the system (1.21) can be rewritten as follows

$$\begin{aligned} A\Delta x &= r_b,\\ A^T\Delta y + \Delta s &= r_c,\\ x\Delta s + s\Delta x &= \sigma\mu\,e - xs. \end{aligned} \tag{1.22}$$

Since the coefficient matrix in the linear system (1.22) is the same as in (1.5) and nonsingular, this system uniquely defines Δx, Δy and Δs. Using these search

directions, the new iterates are given by

$$x^+ = x + \bar{\alpha}\Delta x, \tag{1.23}$$
$$y^+ = y + \bar{\alpha}\Delta y, \tag{1.24}$$
$$s^+ = s + \bar{\alpha}\Delta s, \tag{1.25}$$

where the step length $\bar{\alpha}$ is the largest value of α in $[0,\, 1]$, such that the relations

$$(x,\, y,\, s) + \alpha\,(\Delta x,\, \Delta y,\, \Delta s) \in \mathcal{N}_{-\infty}\,(\gamma,\, \beta)\,,$$
$$(x + \alpha\Delta x)^T\,(s + \alpha\Delta s) \leq (1 - 0.01\alpha)\,x^T s$$

hold for every $\alpha \in [0,\, \bar{\alpha}]$. Wright shows [103, Lemma 6.7] that the value $\bar{\alpha}$ is always positive. Then he computes μ by using the new iterates and repeats this process until he obtains an ϵ-solution for the primal and the dual problems. From the definition of the neighborhood $\mathcal{N}_{-\infty}\,(\gamma,\, \beta)$ it is clear that for all points in the neighborhood, the norms of the residual vectors r_b and r_c are uniformly bounded by some multiple of the μ. This implies that when μ goes to zero then the residual vectors r_b and r_c also go to zero. Wright's algorithm is described more formally in Figure 1.5. According to [103, Theorem 6.2] the complexity of the algorithm is

Primal-Dual Infeasible IPM for LO

Input:
Accuracy parameter $\epsilon > 0$;
parameters $\sigma_{\min}$ and $\sigma_{\max}$; $0 < \sigma_{\min} < \sigma_{\max} \leq \frac{1}{2}$;
parameters $\gamma \in (0,\, 1)$ and $\beta \geq 1$;
parameter $\zeta > 0$;

begin
$x := \zeta e,\, s := \zeta e,\, y := 0$;
while $x^T s \geq \epsilon$ **do**
begin
choose $\sigma \in [\sigma_{\min},\, \sigma_{\max}]$
$x := x + \bar{\alpha}\Delta x$;
$y := y + \bar{\alpha}\Delta y$;
$s := s + \bar{\alpha}\Delta s$;
end
end

Figure 1.5: *Primal-dual IIPM algorithm*

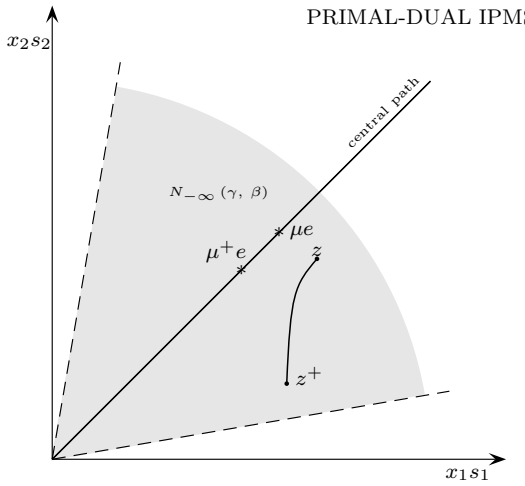

Figure 1.6: *One iteration of Wright's algorithm*

$$O\left(n^2 \log \frac{\max\left\{\left(x^0\right)^T s^0,\ \left\|r_b^0\right\|,\ \left\|r_c^0\right\|\right\}}{\epsilon}\right). \tag{1.26}$$

Figure 1.6 shows one iteration of Wright's algorithm. In this figure the xs-space is projected onto its first two coordinates. In the xs-space the central path of primal-dual pair of problems is a straight line consisting of μe, for $\mu > 0$ as defined in this algorithm. The shaded region in the figure is denoted the neighborhood of the central path $\mathcal{N}_{-\infty}(\gamma,\ \beta)$. In this algorithm he starts from a point $z = xs$ in the shaded region and computes the search directions $(\Delta x,\ \Delta y,\ \Delta s)$ from system (1.22). By using these search directions he obtains the new point $z^+ = x^+ s^+$ in this region again. The algorithms in [44, 58, 78, 105, 107] are almost the same as Wright's algorithm [103]. The differences between the algorithms are due to the use of different neighborhoods of the central path and the step lengths. By using the description of Wright's algorithm we now briefly describe the other algorithms.

1.3.4 The IIPM of Ye

In this subsection we briefly describe the IIPM algorithm presented by Ye in [105]. In this algorithm he uses two constants σ and η such that $\sigma \in (0,\ 1)$ and $\eta \in (0,\ 1)$. The neighborhood of the central path is defined as follows :

$$\mathcal{N}(\eta) := \left\{(x,\ y,\ s)\ :\ (x;\ s) > 0,\ \frac{\|r_b\|}{\left\|r_b^0\right\|} \leq \frac{\mu}{\mu^0},\ \frac{\|r_c\|}{\left\|r_c^0\right\|} \leq \frac{\mu}{\mu^0},\ xs \geq \eta\mu e\right\},$$

where μ, μ^0, r_b^0, r_c^0, r_b and r_c have the same meaning as before. He uses the search directions (1.22) to obtain new iterates as defined in (1.23)-(1.25), and as

step length $\bar{\alpha}$ the largest value of α in [0, 1] such that the following relations

$$\begin{aligned}(x,\, y,\, s) + \alpha\,(\Delta x,\, \Delta y,\, \Delta s) &\in \mathcal{N}\,(\eta)\\ (x + \alpha\Delta x)^T\,(s + \alpha\Delta s) &\leq (1 - \alpha\,(1 - \sigma))\,x^T s,\end{aligned}$$

hold for all $\alpha \in [0,\, \bar{\alpha}]$. He also shows [105, Lemma 5.15] that the value $\bar{\alpha}$ is always positive. He proved that the complexity of his algorithm is polynomial and the same as (1.26).

1.3.5 The IIPM of Kojima et al.

In this subsection we study Kojima et al.'s algorithm [44]. In this algorithm they use some constants γ, γ_P, γ_D, σ, β_2 and β_3 such that $0 < \gamma < 1$, $\gamma_P > 0$, $\gamma_D > 0$ and $0 < \sigma < \beta_2 < \beta_3 < 1$. The neighborhood of the central path used in this algorithm is defined as follows :

$$\begin{aligned}\mathcal{N} := \big\{(x,\, y,\, s) \,:\, & x^T s \geq \gamma_P \,\|r_b\| \;\text{ or }\; \|r_b\| \leq \epsilon_P,\\ & x^T s \geq \gamma_D \,\|r_c\| \;\text{ or }\; \|r_c\| \leq \epsilon_D,\\ & (x;\; s) > 0,\; xs \geq \gamma\mu e\big\},\end{aligned}$$

where ϵ_P and ϵ_D are accuracy parameters for the primal and dual problems, respectively. An iteration in Kojima et al.'s algorithm goes as follows. Given a triple $(x,\, y,\, s) \in \mathcal{N}$, they compute search directions from (1.22) and find the largest step length $\bar{\alpha}$ such that the relations

$$\begin{aligned}(x,\, y,\, s) + \alpha\,(\Delta x,\, \Delta y,\, \Delta s) &\in \mathcal{N}\\ (x + \alpha\Delta x)^T\,(s + \alpha\Delta s) &\leq (1 - \alpha\,(1 - \beta_2))\,x^T s,\end{aligned}$$

hold for every $\alpha \in [0,\, \bar{\alpha}]$. They show [44, Section 3] that the value $\bar{\alpha}$ is always positive. But the actual step length in the algorithm is not $\bar{\alpha}$, since they use different step length α_P and α_D for the primal and the dual problem, respectively. These step lengths are such that

$$\begin{aligned}(x^+,\, y^+,\, s^+) &= (x + \alpha_P\Delta x,\, y + \alpha_D\Delta y,\, s + \alpha_D\Delta s) \in \mathcal{N}\\ (x^+)^T\, s^+ &\leq (1 - \bar{\alpha}\,(1 - \beta_3))\,x^T s.\end{aligned} \tag{1.27}$$

In principle they take for α_P a fixed fraction (0.9995) of the maximal primal step length, and for α_D the same fraction of the maximal dual step length. But they do this only if α_P and α_D satisfy (1.27). Otherwise they take $\alpha_P = \alpha_D = \bar{\alpha}$. They proved that their algorithm is globally convergent. This result was the first theoretical result on primal-dual IIPMs.

1.3.6 The IIPM of Zhang

The algorithm proposed by Zhang is a modification of Kojima et al.'s algorithm. It is designed for solving the wider class of linear complementarity problems. In the description we restrict ourselves to LO problem. He uses a decreasing parameter ν and a constant γ. At the start of the first iteration $\nu = 1$, and $\gamma \in (0,\, 1)$ satisfies

$$\gamma \leq \frac{\min\left(x^0 s^0\right)}{\mu^0}.$$

The main difference with Kojima et al.'s algorithm is that the primal and dual step lengths are always equal to $\bar{\alpha}$, which is the largest value of α in $(0,\, 1]$ such that for all $\alpha \in (0,\, \bar{\alpha}]$

$$\begin{aligned}(x + \alpha\Delta x)\,(s + \alpha\Delta s) &\geq \gamma\,\frac{(x + \alpha\Delta x)^T\,(s + \alpha\Delta s)}{n}\,e,\\ (x + \alpha\Delta x)^T\,(s + \alpha\Delta s) &\geq \nu\left(x^0\right)^T s^0,\end{aligned}$$

where ν updated after each iteration according to $\nu := (1 - \bar{\alpha})\,\nu$. In this algorithm the size of the residual vectors controls by the second inequality above by using parameter ν. He shows [107, Lemma 5.1] that $\bar{\alpha}$ is positive. The complexity of the algorithm is polynomial, and the same as (1.26).

1.3.7 The first IIPM of Mizuno

Shortly after Zhang's paper, Mizuno proposed two different algorithms in [58]. In this subsection we briefly describe the first algorithm, which is a modification of Kojima et al.'s algorithm. He changed the neighborhood of the central path as follows :

$$\mathcal{N}(\gamma) := \left\{(x,\, y,\, s)\,:\,(x;\; s) > 0,\, \frac{\|r_b\|}{\left\|r_b^0\right\|} \leq \frac{\mu}{\mu^0},\, \frac{\|r_c\|}{\left\|r_c^0\right\|} \leq \frac{\mu}{\mu^0},\; xs \geq \gamma\mu e\right\},$$

where γ is a constant such that $0 < \gamma < 1$. This neighborhood was used later also by Ye in [105]. In each iteration of the algorithm he finds a positive step length $\bar{\alpha}$ in the same way as in Kojima et al.'s algorithm. He proved that the complexity of his algorithm is the same as (1.26).

1.3.8 The second IIPM of Mizuno

In this subsection we want to describe the second algorithm in [58], which was a breakthrough since it improved the complexity with a factor n, which gives currently the best known iteration bound for IIPMs (see (1.29)). This algorithm is predictor-corrector IIPM which based on Kojima et al.'s algorithm [44] and Mizuno-Todd-Ye [59]. The algorithm uses some constants γ_1, β_2 and σ and a parameter ν such that $\gamma_1 = \frac{1}{4}$ and $0 < \sigma < \beta_2 < 1$ and ν as in Zhang's algorithm.

Mizuno works with a new neighborhood of the central path that has become very popular since then, namely

$$\mathcal{N}_2(\gamma_1) = \left\{ (x,\, y,\, s) \,:\, (x,\, s) > 0,\, \left\| \frac{xs}{\mu} - e \right\| \leq \gamma_1 \right\}.$$

From this definition it is clear that $\mathcal{N}_2(\gamma_1) \subset \mathcal{N}_2(2\gamma_1)$. These two neighborhoods play a crucial role in the algorithm. Each iteration in Mizuno's algorithm consists of two steps. The first step starts from a triple $(x,\, y,\, s) \in \mathcal{N}_2(\gamma_1)$. Using search directions from (1.22) he moves along these directions and finds iterates $(x',\, y',\, s')$ as defined for $(x^+,\, y^+,\, s^+)$ in (1.23)-(1.25). The step length $\bar{\alpha}$ is the largest value of α in $[0,\, 1]$, such that the following relations hold for every $\alpha \in [0,\, \bar{\alpha}]$:

$$\begin{aligned} &(x,\, y,\, s) + \alpha\,(\Delta x,\, \Delta y,\, \Delta s) \in \mathcal{N}_2(2\gamma_1), \\ &(x + \alpha\Delta x)^T (s + \alpha\Delta s) \leq (1 - \alpha\,(1 - \beta_2))\, x^T s, \\ &(x + \alpha\Delta x)^T (s + \alpha\Delta s) \geq (1 - \alpha)\, \nu \left(x^0\right)^T s^0. \end{aligned}$$

Again it is shown [58, Lemma 4.3] that $\bar{\alpha}$ is positive. Moreover, the new iterates belong to $\mathcal{N}(2\gamma_1)$. To understand the second step we need to refer to Subsection 1.3.1, where we described the Newton step targeting at the μ-center of a primal-dual problem pair. Mizuno's second step is a Newton step of this form for the primal-dual problem pair given by $(A',\, b',\, c')$ with

$$\begin{aligned} A' &= A, \\ b' &= Ax', \\ c' &= A^T y' + s'. \end{aligned} \tag{1.28}$$

He shows that after such a "centering" step the iterates again belong to the neighborhood $\mathcal{N}_2(\gamma_1)$ (see Lemma 4.2 in [58]). He repeats this process until he obtains an ϵ-solution for the primal and dual problems. The resulting complexity is

$$O\left(n \log \frac{\max\left\{ \left(x^0\right)^T s^0,\, \left\|r_b^0\right\|,\, \left\|r_c^0\right\| \right\}}{\epsilon} \right). \tag{1.29}$$

We conclude this subsection with a graphical illustration of one iteration of the algorithm, in Figure 1.7. This figure is drawn in xs-space. In this figure the xs-space is projected onto its first two coordinates. In the xs-space the central path is represented by half line μe, for $\mu > 0$ as defined in this algorithm. The neighborhoods $\mathcal{N}_2(\gamma_1)$ and $\mathcal{N}_2(2\gamma_1)$ are denoted by shaded regions. In this algorithm he starts from a point $z = xs$ in the inner neighborhood $\mathcal{N}_2(\gamma_1)$, and calculates the search directions $(\Delta x,\, \Delta y,\, \Delta s)$ from the system (1.22). He moves along these directions until he finds a point $z' = x's'$ in the outer neighborhood $\mathcal{N}_2(2\gamma_1)$. After the second step of the algorithm he has a point z^1 inside the inner neighborhood $\mathcal{N}_2(\gamma_1)$.

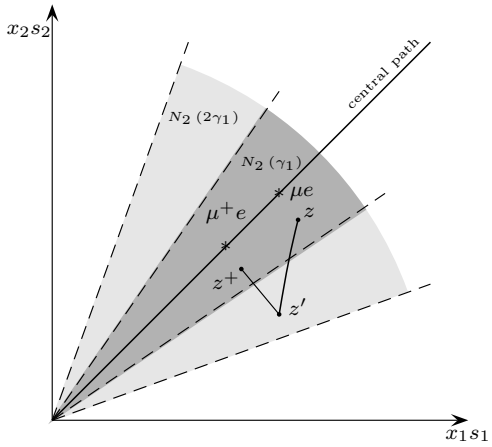

Figure 1.7: *One iteration of the second IIPM of Mizuno*

1.3.9 The IIPM of Potra

Potra [78] also presented a predictor-corrector IIPM. This algorithm is based on Mizuno-Todd-Ye [59]. In this algorithm he uses two constants γ and β such that $\gamma \in (0, 1)$ (he takes $\gamma = \frac{1}{4}$) and $0 < \gamma < \beta \leq 2\gamma \leq \frac{\sqrt{2}}{1+\sqrt{2}}$. The neighborhood of the central path is the same as the neighborhood in Mizuno's second algorithm, which is now denoted by $\mathcal{N}(\gamma)$. He also uses a second neighborhood $\mathcal{N}(\beta)$ with $\gamma < \beta$. This implies $\mathcal{N}(\gamma) \subset \mathcal{N}(\beta)$. Each iteration of Potra's algorithm consists of two steps. The first step starts from a triple $(x, y, s) \in \mathcal{N}(\gamma)$ and it serves to get iterates that are in the neighborhood $\mathcal{N}(\beta)$. In this step he obtains the affine-scaling directions $(\Delta x^a, \Delta y^a, \Delta s^a)$ that are the solution of the following system:

$$\begin{aligned} A\Delta x^a &= 0, \\ A^T \Delta y^a + \Delta s^a &= 0, \\ x\Delta s^a + s\Delta x^a &= -xs. \end{aligned}$$

He also computes search directions $(\Delta x, \Delta y, \Delta s)$, by solving

$$\begin{aligned} A\Delta x &= r_b, \\ A^T \Delta y + \Delta s &= r_c, \\ x\Delta s + s\Delta x &= 0. \end{aligned}$$

Now he moves along these directions and finds the iterates (x', y', s') as follows:

$$\begin{aligned} x' &= x + \alpha_1 \Delta x^a + \alpha_2 \Delta x, \\ y' &= y + \alpha_1 \Delta y^a + \alpha_2 \Delta y, \\ s' &= s + \alpha_1 \Delta s^a + \alpha_2 \Delta s, \end{aligned}$$

where the step lengths α_1 and α_2 are in $(0, 1]$ such that $(x', y', s') \in \mathcal{N}(\beta)$ (see [78, Section 2]). The second step is a centering step, just as the second step of Mizuno's second algorithm, which finds the iterates again belong to the neighborhood $\mathcal{N}(\gamma)$. He repeats this process until he obtains an ϵ-solution for primal and dual problems. He proved that the complexity of his algorithm coincides with iteration bound (1.29).

1.4 The scope of this thesis

As mentioned in Subsection 1.3.1, small-update IPMs theoretically have the best iteration bound and IPMs with full-Newton step belong to this class of methods. Recently, Roos [82] designed the first primal-dual IIPM with full Newton steps for LO problems. In his algorithm the starting point is chosen as in (1.15) and $\mu^0 = \zeta^2$. As established in [82] if ζ is chosen large enough (e.g., $\zeta = 2^L$), where L is the input size of the problem, the algorithm finds an ϵ-solution or detects that the primal and dual problems (P) and (D) are infeasible or unbounded. He also proved that the complexity of his algorithm coincides with the best iteration bound for the IIPM algorithms as given in (1.29). In the next two chapters of this thesis we further explore the algorithm [82]. In Chapter 2, we present an algorithm obtained by slightly changing the definition of the search direction used in [82]. Compared with the algorithm in [82], the analysis of the new algorithm is simpler, whereas the iteration bound of the algorithm is the same as before. This chapter is based on [55]. In Chapter 3, we investigate the generalization to semidefinite optimization (SDO) of the IIPM algorithm presented in [82]. The chapter is based on [54].

In Chapter 4 we change topic by considering full-Newton step primal-dual IPMs that are defined by barrier functions based on kernel functions. As mentioned in Subsection 1.3.1 the theoretical iteration bound for large-update methods is $O\left(n \log \frac{n}{\epsilon}\right)$, where n denotes the number of inequalities in the problem and ϵ is the desired accuracy of the solution. In practice, large-update methods are much more efficient than small-update methods for which the theoretical iteration bound is only $O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$. So the theoretical bounds differ by a factor $\sqrt{n}$, in favor of the small-update methods. This significant gap between theory and practice has been referred to as the irony of IPMs [81]. Some progress has recently been made in this respect by Peng et al. [74, 76] and Bai et al. [6–9]. In these results they introduced a class of new barrier functions that are defined by univariate functions, so-called kernel functions. In these results the "gap factor" was reduced from $\sqrt{n}$ to $\log n$. But the iteration bounds for small-update methods based on these barrier functions always were $O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$, which is the best known bound for IPMs for LO. In Chapter 4, we prove that the class of barrier functions introduced in [6–9, 74, 76] are locally self-concordant. By using the properties of self-concordance functions we show why all the small-update methods have the iteration bound $O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$. This chapter is based on [83].

Finally, Chapter 5 contains some conclusions and comments on further research.

Chapter 2

A full-Newton step primal-dual IIPM for LO

2.1 Introduction

As pointed out in Section 1.4, the algorithm presented in [82] is a primal-dual IIPM that uses full-Newton steps to solve LO problems. In this chapter we use a slightly different algorithm which is obtained by changing the definition of the search direction in the algorithm [82]. As we show, the analysis of our algorithm is easier than the analysis of the algorithm in [82] at some places, whereas the iteration bound essentially remains the same.

In the sections to follow we assume that (P) and (D) are feasible. Only at the end, in Section 2.4, we will discuss how our algorithm can be used to detect infeasibility or unboundedness of (P) and (D).

2.2 A primal-dual IIPM for LO

As usual for IIPMs, we use the starting point as in (1.15). So,

$$\left(x^0,\, y^0,\, s^0\right) = \zeta\,(e,\, 0,\, e) \tag{2.1}$$

with $\zeta > 0$ such that

$$\|x^* + s^*\|_\infty \leq \zeta \tag{2.2}$$

for some optimal solution x^* of (P) and $(y^*,\, s^*)$ of (D). Our aim is to show that, under this assumption, our algorithm obtains an ϵ-solution.

The initial residual vectors are denoted by r_b^0 and r_c^0, as defined in (1.13) and (1.14). So we have

$$r_b^0 = b - Ax^0 = b - \zeta Ae \tag{2.3}$$
$$r_c^0 = c - A^T y^0 - s^0 = c - \zeta e. \tag{2.4}$$

We start by recalling the main ideas underlying the algorithm in [82]. After that, in Subsection 2.2.3, we point out where our algorithm differs from the algorithm in [82].

2.2.1 The perturbed problems

Let $\left(x^0,\, y^0,\, s^0\right)$ be as defined in (2.1). For any ν with $0 < \nu \leq 1$ we consider the perturbed problem (P_ν), defined by

$$(P_\nu) \qquad \min \left\{ \left(c - \nu r_c^0\right)^T x : Ax = b - \nu r_b^0, \quad x \geq 0 \right\},$$

and its dual problem (D_ν), which is given by

$$(D_\nu) \qquad \max \left\{ \left(b - \nu r_b^0\right)^T y : A^T y + s = c - \nu r_c^0, \quad s \geq 0 \right\},$$

where r_b^0 and r_c^0 are the initial residual vectors as defined in (2.3) and (2.4). Note that if $\nu = 1$ then $x = x^0$ yields a strictly feasible solution of (P_ν), and $(y, s) = (y^0, s^0)$ a strictly feasible solution of (D_ν). Due to the choice of the initial iterates we may conclude that if $\nu = 1$ then (P_ν) and (D_ν) each have a strictly feasible solution, which means that both perturbed problems satisfy the IPC. More generally one has the following lemma (see also [82, Lemma 3.1]).

Lemma 2.2.1 (Theorem 5.13 in [105])**.** *The original problems, (P) and (D), are feasible if and only if for each ν satisfying $0 < \nu \leq 1$ the perturbed problems (P_ν) and (D_ν) satisfy the IPC.*

Proof. Suppose that (P) and (D) are feasible. Let $\bar{x}$ be feasible solution of (P) and $(\bar{y}, \bar{s})$ a feasible solution of (D). Then $A\bar{x} = b$ and $A^T\bar{y} + \bar{s} = c$, with $\bar{x} \geq 0$ and $\bar{s} \geq 0$. Now let $0 < \nu \leq 1$, and consider

$$x = (1-\nu)\,\bar{x} + \nu\, x^0, \quad y = (1-\nu)\,\bar{y} + \nu\, y^0, \quad s = (1-\nu)\,\bar{s} + \nu\, s^0.$$

One has

$$\begin{aligned} Ax &= A\left((1-\nu)\,\bar{x} + \nu\, x^0\right) \\ &= (1-\nu)\,A\bar{x} + \nu Ax^0 \\ &= (1-\nu)\,b + \nu Ax^0 = b - \nu\left(b - Ax^0\right), \end{aligned}$$

showing that x is feasible for (P_ν). Similarly,

$$\begin{aligned} A^T y + s &= (1-\nu)\left(A^T \bar{y} + \bar{s}\right) + \nu\left(A^T y^0 + s^0\right) \\ &= (1-\nu)\, c + \nu\left(A^T y^0 + s^0\right) = c - \nu\left(c - A^T y^0 - s^0\right), \end{aligned}$$

showing that (y, s) is feasible for (D_ν). Since $\nu > 0$, x and s are positive, thus proving that (P_ν) and (D_ν) satisfy the IPC.

To prove the inverse implication, suppose that (P_ν) and (D_ν) satisfy the IPC for each ν satisfying $0 < \nu \leq 1$. Obviously, then (P_ν) and (D_ν) are feasible for these values of ν. Letting ν go to zero it follows that (P) and (D) are feasible. □

2.2.2 Central path of the perturbed problems

We assumed that (P) and (D) are feasible. Letting $0 < \nu \leq 1$, Lemma 2.2.1 implies that the problems (P_ν) and (D_ν) satisfy the IPC, and hence their central paths exist. This means that the system

$$Ax = b - \nu r_b^0, \qquad x \geq 0 \tag{2.5}$$

$$A^T y + s = c - \nu r_c^0, \qquad s \geq 0. \tag{2.6}$$

$$xs = \mu e, \tag{2.7}$$

has a unique solution, for every $\mu > 0$. We denote this unique solution as $(x(\mu, \nu), y(\mu, \nu), s(\mu, \nu))$. These are the μ-centers of the perturbed problems (P_ν) and (D_ν).

Note that since $x^0 s^0 = \mu^0 e$ with $\mu^0 = \zeta^2$, x^0 is the μ^0-center of the perturbed problem (P_1) and (y^0, s^0) the μ^0-center of (D_1). In other words,

$$(x(\mu^0, 1), y(\mu^0, 1), s(\mu^0, 1)) = (x^0, y^0, s^0). \tag{2.8}$$

In the sequel we will always have $\mu = \nu\, \mu^0$, and we will accordingly denote $(x(\mu, \nu), y(\mu, \nu), s(\mu, \nu))$ simply as $(x(\nu), y(\nu), s(\nu))$.

2.2.3 An iteration of our algorithm

We just established that if $\nu = 1$ and $\mu = \mu^0$, then $x = x^0$ and $(y, s) = \left(y^0, s^0\right)$ are the μ-centers of (P_ν) and (D_ν) respectively. These are our initial iterates.

We measure proximity to the μ-center of the perturbed problems by the quantity $\delta(x, s; \mu)$ as defined in (1.10). So, initially we have $\delta(x, s; \mu) = 0$, due to (2.8). In the sequel we assume that at the start of each iteration, just before a μ-update, $\delta(x, s; \mu)$ is smaller than or equal to a (small) threshold value $\tau > 0$. So this is certainly true at the start of the first iteration.

Now we describe one iteration of our algorithm. Suppose that for some $\mu \in (0, \mu^0]$ we have x, y and s satisfying the feasibility conditions (2.5) and (2.6) for $\nu = \frac{\mu}{\mu^0}$, and such that $x^T s = n\mu$ and $\delta(x, s; \mu) \leq \tau$. We reduce μ to $\mu^+ = (1-\theta)\, \mu$,

with $\theta \in (0,\,1)$, and find new iterates x^+, y^+ and s^+ that satisfy (2.5) and (2.6), with ν replaced by $\nu^+ = \frac{\mu^+}{\mu^0}$, and such that $x^T s = n\mu^+$ and $\delta\left(x^+,\, s^+;\, \mu^+\right) \leq \tau$. Note that $\nu^+ = (1-\theta)\,\nu$.

To be more precise, this is achieved as follows. Each main iteration consists of a feasibility step and a few centering steps. The feasibility step serves to get iterates $\left(x^f,\, y^f,\, s^f\right)$ that are strictly feasible for (P_{ν^+}) and (D_{ν^+}), and close to their μ-centers $\left(x\left(\nu^+\right),\, y\left(\nu^+\right),\, s\left(\nu^+\right)\right)$ such that $\delta\left(x^f,\, s^f;\, \mu^+\right) \leq \frac{1}{\sqrt{2}}$. Since the triple $\left(x^f,\, y^f,\, s^f\right)$ is strictly feasible for (P_{ν^+}) and (D_{ν^+}), we can perform a few centering steps starting at $\left(x^f,\, y^f,\, s^f\right)$, targeting at the μ^+-centers of (P_{ν^+}) and (D_{ν^+}), and obtain iterates $(x^+,\, y^+,\, s^+)$ that are feasible for (P_{ν^+}) and (D_{ν^+}) and such that $\delta\left(x^+,\, s^+;\, \mu^+\right) \leq \tau$. For the analysis of the centering steps we can just use the theory presented in Section 1.3.1 of Chapter 1, since the iterates are feasible for the current pair of the perturbed problems. For the analysis of the feasibility step, which we define below, much more work is needed.

From the definition of the perturbed problems (P_ν) and (D_ν), it is clear that the feasibility equations for these problems are:

$$Ax = b - \nu r_b^0, \quad x \geq 0, \tag{2.9}$$

$$A^T y + s = c - \nu r_c^0, \quad s \geq 0, \tag{2.10}$$

and those of (P_{ν^+}) and (D_{ν^+}):

$$Ax = b - \nu^+ r_b^0, \quad x \geq 0, \tag{2.11}$$

$$A^T y + s = c - \nu^+ r_c^0, \quad s \geq 0. \tag{2.12}$$

So, assuming that x, y and s satisfy (2.9) and (2.10), to get iterates that are feasible for (P_{ν^+}) and (D_{ν^+}) we need search directions $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$ such that

$$\begin{aligned} A\left(x + \Delta^f x\right) &= b - \nu^+ r_b^0, \\ A^T\left(y + \Delta^f y\right) + \left(s + \Delta^f s\right) &= c - \nu^+ r_c^0. \end{aligned}$$

Since x is feasible for (P_ν) and $(y,\, s)$ is feasible for (D_ν), it follows that $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$ should satisfy

$$A\Delta^f x = (b - Ax) - \nu^+ r_b^0 = \nu r_b^0 - \nu^+ r_b^0 = \theta \nu r_b^0,$$

and

$$A^T \Delta^f y + \Delta^f s = \left(c - A^T y - s\right) - \nu^+ r_c^0 = \nu r_c^0 - \nu^+ r_c^0 = \theta \nu r_c^0.$$

In [82] the following system is used to define $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$:

$$\begin{aligned} A\Delta^f x &= \theta \nu r_b^0 \\ A^T \Delta^f y + \Delta^f s &= \theta \nu r_c^0 \\ s\Delta^f x + x\Delta^f s &= \mu e - xs. \end{aligned} \tag{2.13}$$

Since matrix A is full row rank, the system (2.13) uniquely defines $(\Delta^f x,\ \Delta^f y,\ \Delta^f s)$ for any $x > 0$ and $s > 0$.

We follow a different approach and replace the third equation in (2.13) by the equation

$$s\Delta^f x + x\Delta^f s = 0. \tag{2.14}$$

This leads to the following system

$$A\Delta^f x = \theta\nu r_b^0 \tag{2.15}$$
$$A^T\Delta^f y + \Delta^f s = \theta\nu r_c^0 \tag{2.16}$$
$$\Delta^f x + x\Delta^f s = 0, \tag{2.17}$$

which, by the same arguments as before, also has a unique solution. After the feasibility step the iterates are given by

$$x^f = x + \Delta^f x, \tag{2.18}$$
$$y^f = y + \Delta^f y, \tag{2.19}$$
$$s^f = s + \Delta^f s. \tag{2.20}$$

We conclude that after the feasibility step we have iterates $\left(x^f,\ y^f,\ s^f\right)$ that satisfy the affine equations of the new perturbed problem pair (P_{ν^+}) and (D_{ν^+}). In the analysis we should also guarantee that x^f and s^f are positive and moreover belong to the region of quadratic convergence of their μ^+-centers $(x(\nu^+),\ y(\nu^+),\ s(\nu^+))$. In other words, we must have $\delta\left(x^f,\ s^f;\ \mu^+\right) \leq \frac{1}{\sqrt{2}}$. Proving this is the crucial part in the analysis of the algorithm.

After the feasibility step we perform centering steps in order to get iterates $(x^+,\ y^+,\ s^+)$ that satisfy ${x^+}^T s^+ = n\mu^+$ and $\delta\left(x^+,\ s^+;\ \mu^+\right) \leq \tau$, where $\tau > 0$. By using Corollary 1.3.6, the required number of centering steps can easily be obtained. Indeed, assuming $\delta = \delta\left(x^f,\ s^f;\ \mu^+\right) \leq \frac{1}{\sqrt{2}}$, after k centering steps we will have iterates $(x^+,\ y^+,\ s^+)$ that are still feasible for (P_{ν^+}) and (D_{ν^+}) and that satisfy

$$\delta\left(x^+,\ s^+;\ \mu^+\right) \leq \left(\frac{1}{\sqrt{2}}\right)^{2^k}.$$

Therefore, $\delta\left(x^+,\ s^+;\ \mu^+\right) \leq \tau$ will hold if k satisfies

$$\left(\frac{1}{\sqrt{2}}\right)^{2^k} \leq \tau,$$

which gives

$$k \geq \log_2\left(\log_2 \frac{1}{\tau^2}\right). \tag{2.21}$$

We conclude this section with a graphical illustration of one iteration of the algorithm, in Figure 2.1. This figure is drawn in xs-space. In the xs-space the central path of primal-dual pair of problems is a straight line consisting of μe, for $\mu > 0$. Hence, the central paths of the perturbed problem pair (P_ν) and (D_ν) and also (P_{ν^+}) and (D_{ν^+}) are represented by two straight lines. We have drawn the level curves for $\delta(v) = \tau$ around target points on the central paths that are used during the algorithm. The starting point $z^0 = xs$ of an inner iteration is close to the μ-center μe of the perturbed problem pair (P_ν) and (D_ν). After the feasibility step we have a feasible point for the new perturbed problem pair (P_{ν^+}) and (D_{ν^+}), which is denoted by $z^1 = x^f s^f$ in the figure, and which lies in the region of quadratic convergence of their μ^+-centers. This is the shaded region in the figure. By performing centering steps, starting at z^1, we obtain iterates, which are denoted by z^2, z^3 and z^4 in xs-space, that are still feasible for the new perturbed problem pair (P_{ν^+}) and (D_{ν^+}) and z^4 is sufficiently close to their μ^+-center.

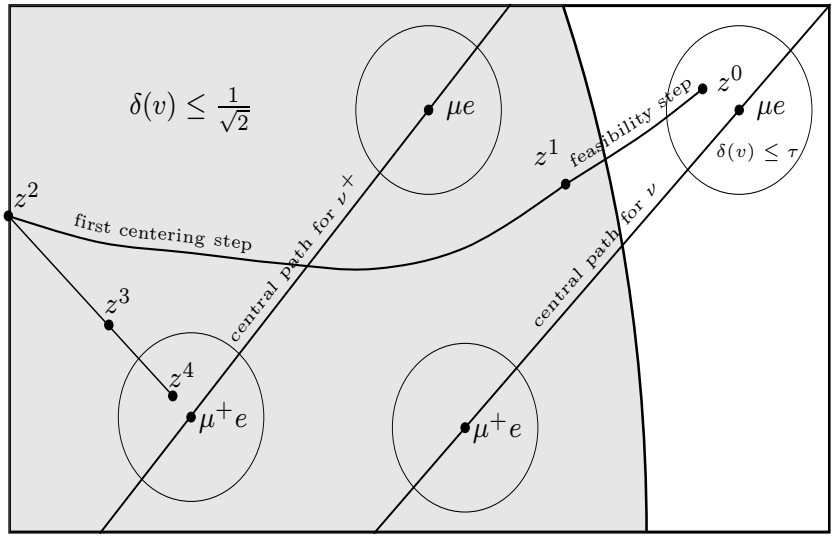

Figure 2.1: *Performance of one iteration of primal-dual IIPM*

2.2.4 The algorithm

A more formal description of the algorithm is given in Figure 2.2. Note that since in each iteration μ and ν are reduced by the factor $1 - \theta$, it follows that after each iteration the residuals and the duality gap are also reduced by the factor $1 - \theta$. The algorithm stops if the norms of the residuals and the duality gap are less than the accuracy parameter ϵ.

Primal-Dual Infeasible IPM

Input:
Accuracy parameter $\epsilon > 0$;
barrier update parameter θ, $0 < \theta < 1$;
threshold parameter $\tau > 0$;
parameter $\zeta > 0$.

begin
$x^0 := \zeta e$, $s^0 := \zeta e$, $y^0 := 0$; $\mu^0 := \zeta^2$;
while $\max\left(x^T s,\ \|r_c\|,\ \|r_b\|\right) \geq \epsilon$ **do**
begin
feasibility step:
$(x,\ y,\ s) := (x,\ y,\ s) + \left(\Delta^f x,\ \Delta^f y,\ \Delta^f s\right)$;
μ-update:
$\mu := (1-\theta)\mu$;
centering steps:
while $\delta(x,\ s,\ \mu) \geq \tau$ **do**
begin
$(x,\ y,\ s) := (x,\ y,\ s) + (\Delta x,\ \Delta y,\ \Delta s)$;
end
end
end

Figure 2.2: *Infeasible full-Newton-step algorithm*

2.3 Analysis of the feasibility step

Let x, y and s denote the iterates at the start of an iteration, and assume $\delta(x, s; \mu) \leq \tau$. Recall that at the start of the first iteration this is certainly true, because then $\delta(x, s; \mu) = 0$.

2.3.1 The feasibility step; choice of θ and τ

As we established in Section 2.2.3, the feasibility step generates new iterates x^f, y^f and s^f that satisfy the feasibility conditions for (P_{ν^+}) and (D_{ν^+}). A crucial element in the analysis is to show that after the feasibility step we have that $\delta(x^f, s^f; \mu^+) \leq 1/\sqrt{2}$, i.e., that the new iterates are within the region where the Newton process, targeting at the μ^+-centers of (P_{ν^+}) and (D_{ν^+}), is quadratically convergent.

We define

$$d_x^f := \frac{v\Delta^f x}{x}, \qquad d_s^f := \frac{v\Delta^f s}{s}, \tag{2.22}$$

with v as defined in (1.7). Now, by (1.7) we have $xs = \mu v^2$. Hence we may write

$$d_x^f d_s^f = \frac{v^2\Delta^f x\Delta^f s}{xs} = \frac{\Delta^f x\Delta^f s}{\mu}.$$

Using this and (2.17) we may write

$$x^f s^f = xs + \left(s\Delta^f x + x\Delta^f s\right) + \Delta^f x\Delta^f s = \mu v^2 + \Delta^f x\Delta^f s = \mu\left(v^2 + d_x^f d_s^f\right). \tag{2.23}$$

Lemma 2.3.1. *The iterates $\left(x^f,\, y^f,\, s^f\right)$ are strictly feasible if and only if $v^2 + d_x^f d_s^f > 0$.*

Proof. Note that if x^f and s^f are positive then (2.23) makes clear that $v^2 + d_x^f d_s^f > 0$, proving the 'only if' part of the statement in the lemma. For the proof of the converse implication we introduce a step length $\alpha \in [0,\, 1]$, and define

$$x\left(\alpha\right) = x + \alpha\Delta^f x, \quad y\left(\alpha\right) = y + \alpha\Delta^f y, \quad s\left(\alpha\right) = s + \alpha\Delta^f s.$$

We then have $x\left(0\right) = x$, $x\left(1\right) = x^f$ and similar relations for y and s. Hence we have $x\left(0\right)s\left(0\right) = xs > 0$. We write

$$x\left(\alpha\right)s\left(\alpha\right) = \left(x + \alpha\Delta^f x\right)\left(s + \alpha\Delta^f s\right) = xs + \alpha\left(s\Delta^f x + x\Delta^f s\right) + \alpha^2\Delta^f x\Delta^f s.$$

Using (2.23) and $s\Delta^f x + x\Delta^f s = 0$ and $xs = \mu v^2$, we obtain

$$x\left(\alpha\right)s\left(\alpha\right) = xs + \alpha^2\Delta^f x\Delta^f s = \mu v^2 + \alpha^2\mu d_x^f d_s^f = \mu\left(v^2 + \alpha^2 d_x^f d_s^f\right).$$

Now suppose that $v^2 + d_x^f d_s^f > 0$. Then $d_x^f d_s^f > -v^2$, so we get

$$x\left(\alpha\right)s\left(\alpha\right) > \mu\left(v^2 - \alpha^2 v^2\right) = \mu\left(1 - \alpha^2\right)v^2 = \left(1 - \alpha^2\right)xs, \quad \alpha \in [0,\, 1].$$

Since $\left(1 - \alpha^2\right)xs \geq 0$ it follows that $x\left(\alpha\right)s\left(\alpha\right) > 0$ for $0 \leq \alpha \leq 1$. Hence, none of the entries of $x\left(\alpha\right)$ and $s\left(\alpha\right)$ vanishes for $0 \leq \alpha \leq 1$. Since $x\left(0\right)$ and $s\left(0\right)$ are positive, and $x\left(\alpha\right)$ and $s\left(\alpha\right)$ depend linearly on α, this implies that $x\left(\alpha\right) > 0$ and $s\left(\alpha\right) > 0$ for all $0 \leq \alpha \leq 1$. Hence, $x\left(1\right)$ and $s\left(1\right)$ must be positive, proving the 'if' part of the statement in the lemma. □

Using (2.22) we may also write

$$x^f = x + \Delta^f x = x + \frac{xd_x^f}{v} = \frac{x}{v}(v + d_x^f) \tag{2.24}$$

$$s^f = s + \Delta^f s = s + \frac{sd_s^f}{v} = \frac{s}{v}(v + d_s^f). \tag{2.25}$$

To simplify the presentation we will denote $\delta(x, s; \mu)$ below simply as δ. Recall that we assume that before the feasibility step one has $\delta \leq \tau$.

Lemma 2.3.2. *The new iterates are certainly strictly feasible if*

$$\left\|d_x^f\right\| < \frac{1}{\rho(\delta)} \quad \text{and} \quad \left\|d_s^f\right\| < \frac{1}{\rho(\delta)}, \tag{2.26}$$

where

$$\rho(\delta) := \delta + \sqrt{1+\delta^2}. \tag{2.27}$$

Proof. It is clear from (2.24) that x^f is strictly feasible if and only if $v + d_x^f > 0$. This certainly holds if $\left\|d_x^f\right\| < \min(v)$. Since $2\delta = \left\|v - v^{-1}\right\|$, the minimal value t that an entry of v can attain will satisfy $t \leq 1$ and $1/t - t = 2\delta$. The last equation implies $t^2 + 2\delta t - 1 = 0$, which gives $t = -\delta + \sqrt{1+\delta^2} = 1/\rho(\delta)$. This proves the first inequality in (2.26). The second inequality is obtained in the same way. □

The above proof makes clear that the elements of the vector v satisfy

$$\frac{1}{\rho(\delta)} \leq v_i \leq \rho(\delta), \quad i = 1, \ldots, n. \tag{2.28}$$

One may easily check that the system (2.15)–(2.17), which defines the search directions $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$, can be expressed in terms of the scaled search directions d_x^f and d_s^f as follows.

$$\bar{A} d_x^f = \theta\nu r_b^0, \tag{2.29}$$

$$\bar{A}^T \frac{\Delta^f y}{\mu} + d_s^f = \theta\nu v s^{-1} r_c^0, \tag{2.30}$$

$$d_x^f + d_s^f = 0, \tag{2.31}$$

where

$$\bar{A} = AV^{-1}X. \tag{2.32}$$

Hence, due to (2.31), we have $d_s^f = -d_x^f$, and therefore

$$z := -d_x^f d_s^f = \left(d_x^f\right)^2. \tag{2.33}$$

Assuming $v^2 + d_x^f d_s^f = v^2 - \left(d_x^f\right)^2 > 0$, which according to Lemma 2.3.1 holds if and only if the iterates $\left(x^f,\, y^f,\, s^f\right)$ are strictly feasible, we proceed by deriving an upper bound for $\delta\left(x^f,\, s^f; \mu^+\right)$. According to definition (1.10) it holds that

$$\delta\left(x^f,\, s^f; \mu^+\right) = \frac{1}{2}\left\|v^f - \frac{e}{v^f}\right\|, \text{ where } \quad v^f = \sqrt{\frac{x^f s^f}{\mu^+}}.$$

In the sequel we denote $\delta\left(x^f,\, s^f; \mu^+\right)$ also shortly by $\delta\left(v^f\right)$. We need to find an upper bound for $\delta\left(v^f\right)$. To this end, we need the following technical lemma.

Lemma 2.3.3. *For $i = 1, \ldots, m$, let $f_i : \mathbf{R}_+ \to \mathbf{R}$ denote a convex function. Then we have for any nonzero vector $z \in \mathbf{R}^n_+$ the following inequality:*

$$\sum_{i=1}^{n} f_i(z_i) \leq \frac{1}{e^T z} \sum_{j=1}^{n} z_j \left(f_j(e^T z) + \sum_{i \neq j} f_i(0) \right).$$

Proof. We define the function $F : \mathbf{R}^n_+ \to \mathbf{R}$ by

$$F(z) = \sum_{i=1}^{n} f_i(z_i), \quad z \geq 0.$$

Letting e_j denote the j-th unit vector in $\mathbf{R}^n$, we may write z as a convex combination of the vectors $\left(e^T z\right) e_j$, as follows.

$$z = \sum_{j=1}^{n} \frac{z_j}{e^T z} \left(e^T z\right) e_j,$$

Indeed, $\sum_{j=1}^{n} \frac{z_j}{e^T z} = 1$ and $z_j / e^T z \geq 0$ for each j. Since $F(z)$ is a sum of convex functions, $F(z)$ is convex in z, and hence we have

$$F(z) \leq \sum_{j=1}^{n} \frac{z_j}{e^T z} F\left(\left(e^T z\right) e_j\right) = \sum_{j=1}^{n} \frac{z_j}{e^T z} \sum_{i=1}^{n} f_i\left(\left(e^T z\right) (e_j)_i\right).$$

Since $(e_j)_i = 1$ if $i = j$ and zero if $i \neq j$, we obtain

$$F(z) \leq \sum_{j=1}^{n} \frac{z_j}{e^T z} \left(f_j(e^T z) + \sum_{i \neq j} f_i(0) \right).$$

Hence the inequality in the lemma follows ☐

Lemma 2.3.4. *Assuming $v^2 - \left(d_x^f\right)^2 > 0$, one has*

$$4\delta(v^f)^2 \leq \frac{\theta^2 n - \left\|d_x^f\right\|^2}{1-\theta} + (1-\theta) \left(4\delta^2 + \frac{\rho(\delta)^4 \left\|d_x^f\right\|^2}{1 - \rho(\delta)^2 \left\|d_x^f\right\|^2} \right),$$

with ρ as defined in (2.27).

Proof. After division of both sides in (2.23) by μ^+ we get, using (2.33),

$$\left(v^f\right)^2 = \frac{x^f s^f}{\mu^+} = \frac{\mu\left(v^2 + d_x^f d_s^f\right)}{\mu^+} = \frac{v^2 + d_x^f d_s^f}{1-\theta} = \frac{v^2 - z}{1-\theta}.$$

Hence we have

$$4\delta(v^f)^2 = \sum_{i=1}^{n} \left(\left(v_i^f\right)^2 + \left(v_i^f\right)^{-2} - 2 \right) = \sum_{i=1}^{n} \left(\frac{v_i^2 - z_i}{1-\theta} + \frac{1-\theta}{v_i^2 - z_i} - 2 \right)$$

For each i we define the function

$$f_i(z_i) := \frac{v_i^2 - z_i}{1-\theta} + \frac{1-\theta}{v_i^2 - z_i} - 2, \quad i = 1, \ldots, n.$$

Due to hypothesis in the lemma we have $v_i^2 - z_i > 0$, for each i. Using this one may easily verify that $f_i(z_i)$ is convex in z_i. We therefore may apply Lemma 2.3.3. A direct application, also using $z_i = \left(d_x^f\right)_i^2$ and $e^T z = \left\|d_x^f\right\|^2$, gives

$$4\delta(v^f)^2 \leq \tag{2.34}$$
$$\sum_{j=1}^{n} \frac{\left(d_x^f\right)_j^2}{\left\|d_x^f\right\|^2} \left(\frac{v_j^2 - \left\|d_x^f\right\|^2}{1-\theta} + \frac{1-\theta}{v_j^2 - \left\|d_x^f\right\|^2} - 2 + \sum_{i \neq j} \left(\frac{v_i^2}{1-\theta} + \frac{1-\theta}{v_i^2} - 2 \right) \right).$$

From Lemma 1.3.2 we know that

$$\sum_{i=1}^{n} \left(\frac{v_i^2}{1-\theta} + \frac{1-\theta}{v_i^2} - 2 \right) = 4\,(1-\theta)\,\delta^2 + \frac{\theta^2 n}{1-\theta}.$$

Hence, the last term in the right-hand side of (2.34) can be rewritten as

$$\begin{aligned} \sum_{i \neq j} \left(\frac{v_i^2}{1-\theta} + \frac{1-\theta}{v_i^2} - 2 \right) &= \sum_{i=1}^{n} \left(\frac{v_i^2}{1-\theta} + \frac{1-\theta}{v_i^2} - 2 \right) - \left(\frac{v_j^2}{1-\theta} + \frac{1-\theta}{v_j^2} - 2 \right) \\ &= 4(1-\theta)\delta^2 + \frac{\theta^2 n}{1-\theta} - \left(\frac{v_j^2}{1-\theta} + \frac{1-\theta}{v_j^2} - 2 \right). \end{aligned}$$

Substituting this in the right-hand side of (2.34) gives the following upper bound for $4\,\delta\left(v^f\right)^2$:

$$\sum_{j=1}^{n} \frac{\left(d_x^f\right)_j^2}{\left\|d_x^f\right\|^2} \left(\frac{v_j^2 - \left\|d_x^f\right\|^2}{1-\theta} + \frac{1-\theta}{v_j^2 - \left\|d_x^f\right\|^2} - 2 + 4(1-\theta)\delta^2 + \frac{\theta^2 n}{1-\theta} - \left(\frac{v_j^2}{1-\theta} + \frac{1-\theta}{v_j^2} - 2 \right) \right).$$

This expression can be simplified to

$$4\left(1-\theta\right)\delta^2+\frac{\theta^2 n}{1-\theta}$$
$$+\frac{1}{\left\|d_x^f\right\|^2}\sum_{j=1}^{n}\left(d_x^f\right)_j^2\left(\frac{v_j^2-\left\|d_x^f\right\|^2}{1-\theta}+\frac{1-\theta}{v_j^2-\left\|d_x^f\right\|^2}-\left(\frac{v_j^2}{1-\theta}+\frac{1-\theta}{v_j^2}\right)\right),$$

which can be further reduced to

$$4(1-\theta)\delta^2+\frac{\theta^2 n}{1-\theta}+\frac{1}{\left\|d_x^f\right\|^2}\sum_{j=1}^{n}\left(d_x^f\right)_j^2\left(\frac{-\left\|d_x^f\right\|^2}{1-\theta}+\frac{1-\theta}{v_j^2-\left\|d_x^f\right\|^2}-\frac{1-\theta}{v_j^2}\right)$$
$$=4(1-\theta)\delta^2+\frac{\theta^2 n}{1-\theta}-\frac{\left\|d_x^f\right\|^2}{1-\theta}+\frac{1-\theta}{\left\|d_x^f\right\|^2}\sum_{j=1}^{n}\left(d_x^f\right)_j^2\frac{\left\|d_x^f\right\|^2}{v_j^2\left(v_j^2-\left\|d_x^f\right\|^2\right)}$$
$$=4(1-\theta)\delta^2+\frac{\theta^2 n}{1-\theta}-\frac{\left\|d_x^f\right\|^2}{1-\theta}+(1-\theta)\sum_{j=1}^{n}\frac{\left(d_x^f\right)_j^2}{v_j^2\left(v_j^2-\left\|d_x^f\right\|^2\right)}.$$

From (2.28) we know that $v_j \geq \frac{1}{\rho(\delta)}$. Using this we get

$$4\delta(v^f)^2 \leq 4(1-\theta)\delta^2+\frac{\theta^2 n-\left\|d_x^f\right\|^2}{1-\theta}+\frac{(1-\theta)\,\rho(\delta)^4\left\|d_x^f\right\|^2}{1-\rho(\delta)^2\left\|d_x^f\right\|^2}.$$

This implies the lemma. □

We conclude this section by presenting a value that we do not allow $\left\|d_x^f\right\|$ to exceed. It may be worth noting that d_x^f is dependent on the value of θ, as is clear from (2.29) - (2.31). This fact will be explored later on. Recall that we want $\delta(v^f) \leq 1/\sqrt{2}$, and observe that Lemma 2.3.4 implies that this holds if

$$\frac{\theta^2 n-\left\|d_x^f\right\|^2}{1-\theta}+(1-\theta)\left(4\delta^2+\frac{\rho(\delta)^4\left\|d_x^f\right\|^2}{1-\rho(\delta)^2\left\|d_x^f\right\|^2}\right)\leq 2.$$

At this stage we decide to choose

$$\tau=\frac{1}{8},\quad \theta=\frac{\alpha}{\sqrt{2n}},\quad \alpha\leq 1. \tag{2.35}$$

Then, for $n\geq 1$ and $\delta\leq\tau$, one may easily verify that

$$\left\|d_x^f\right\|\leq\frac{1}{2}\quad\Rightarrow\quad\delta(v^f)\leq\frac{1}{\sqrt{2}}. \tag{2.36}$$

We proceed by considering the vector d_x^f more in detail.

2.3.2 Upper bound for $\left\|d_{\boldsymbol{x}}^{\boldsymbol{f}}\right\|$

Using (2.31), we can rewrite (2.29)-(2.30) simply as

$$\begin{aligned} \bar{A}d_x^f &= \theta\nu r_b^0, \\ \bar{A}^T\tfrac{\Delta^f y}{\mu} - d_x^f &= \theta\nu v s^{-1} r_c^0. \end{aligned} \tag{2.37}$$

Now let us denote the null space of the matrix $\bar{A}$ as $\mathcal{L}$. So,

$$\mathcal{L} := \left\{\xi \in \mathbf{R}^n \ : \ \bar{A}\xi = 0\right\}.$$

Then the affine space $\left\{\xi \in \mathbf{R}^n \ : \ \bar{A}\xi = \theta\nu r_b^0\right\}$ equals $d_x^f + \mathcal{L}$. Note that due to a well-known result from linear algebra the row space of $\bar{A}$ equals the orthogonal complement $\mathcal{L}^\perp$ of $\mathcal{L}$. Therefore, the second equation in (2.37) shows that the affine space $\left\{-\theta\nu v s^{-1} r_c^0 + \bar{A}^T\varsigma \ : \ \varsigma \in \mathbf{R}^m\right\}$ equals $d_x^f + \mathcal{L}^\perp$. Since $\mathcal{L} \cap \mathcal{L}^\perp = \{0\}$, it follows that the affine spaces $d_x^f + \mathcal{L}$ and $d_x^f + \mathcal{L}^\perp$ meet in a unique point. Therefore, system (2.37) determines d_x^f uniquely. Now using similar arguments as in [82, Section 4.4] we can prove the following lemma.

Lemma 2.3.5 (Lemma 4.7 in [82])**.** *One has*

$$\sqrt{\mu}\,\left\|d_x^f\right\| \leq \theta\nu\,\zeta\sqrt{e^T\left(\frac{x}{s} + \frac{s}{x}\right)}. \tag{2.38}$$

Proof. From the definition (2.32) we deduce that $\bar{A} = \sqrt{\mu}AD$, where

$$D = \operatorname{diag}\left(\frac{xv^{-1}}{\sqrt{\mu}}\right) = \operatorname{diag}\left(\sqrt{\frac{x}{s}}\right) = \operatorname{diag}\left(\sqrt{\mu}vs^{-1}\right).$$

To simplify notations, let us write

$$r_b = \theta\nu r_b^0, \quad r_c = \theta\nu r_c^0 \ \text{ and } \ q = -d_x^f.$$

Then the system defining q is equivalent to

$$\begin{aligned} \sqrt{\mu}ADq &= -r_b, \\ \sqrt{\mu}DA^T\xi + q &= \frac{1}{\sqrt{\mu}}Dr_c, \end{aligned}$$

where $\xi = \frac{\Delta^f y}{\mu}$. This implies

$$\mu AD^2A^T\xi = AD^2r_c + r_b,$$

whence

$$\xi = \frac{1}{\mu}\left(AD^2A^T\right)^{-1}\left(AD^2r_c + r_b\right).$$

Substitution gives

$$q = \frac{1}{\sqrt{\mu}} \left(Dr_c - DA^T \left(AD^2 A^T \right)^{-1} \left(AD^2 r_c + r_b \right) \right).$$

Observe that

$$q_1 := Dr_c - DA^T \left(AD^2 A^T \right)^{-1} AD^2 r_c = \left(I - DA^T \left(AD^2 A^T \right)^{-1} AD \right) Dr_c$$

is the orthogonal projection of Dr_c onto the null space of AD. Let $(\bar{y},\, \bar{s})$ be such that $A^T \bar{y} + \bar{s} = c$. Then we may write

$$r_c = \theta \nu r_c^0 = \theta \nu \left(c - A^T y^0 - s^0 \right) = \theta \nu \left(A^T \left(\bar{y} - y^0 \right) + \bar{s} - s^0 \right).$$

Since $DA^T \left(\bar{y} - y^0 \right)$ belongs to the row space of AD, which is orthogonal to the null space of AD, we obtain

$$\|q_1\| \leq \theta \nu \left\| D \left(\bar{s} - s^0 \right) \right\|.$$

On the other hand, let $\bar{x}$ be such that $A\bar{x} = b$. Then

$$r_b = \theta \nu r_b^0 = \theta \nu \left(b - Ax^0 \right) = \theta \nu A \left(\bar{x} - x^0 \right),$$

and the vector

$$q_2 := \theta \nu \, DA^T \left(AD^2 A^T \right)^{-1} AD \left(D^{-1} \left(\bar{x} - x^0 \right) \right)$$

is the orthogonal projection of $\theta \nu D^{-1} \left(\bar{x} - x^0 \right)$ onto the row space of AD. Hence it follows that

$$\|q_2\| \leq \theta \nu \left\| D^{-1} \left(\bar{x} - x^0 \right) \right\|.$$

Since $\sqrt{\mu} q = q_1 - q_2$ and q_1 and q_2 are orthogonal, we may conclude that

$$\sqrt{\mu}\, \|q\| = \sqrt{\|q_1\|^2 + \|q_2\|^2} \leq \theta \nu \sqrt{\|D \left(\bar{s} - s^0 \right)\|^2 + \|D^{-1} \left(\bar{x} - x^0 \right)\|^2}, \quad (2.39)$$

where, as always, $\mu = \mu^0 \nu$. We are still free to choose $\bar{x}$ and $\bar{s}$, subject to constraints $A\bar{x} = b$ and $A^T \bar{y} + \bar{s} = c$. Taking $\bar{x} = x^* \geq 0$ and $\bar{s} = s^* \geq 0$, where x^* and $(y^*,\, s^*)$ are optimal solutions for (P) and (D), respectively and satisfy (1.16). This implies that the entries of the vectors $x^0 - \bar{x}$ and $s^0 - \bar{s}$ satisfy

$$0 \leq x^0 - \bar{x} \leq \zeta e, \quad 0 \leq s^0 - \bar{s} \leq \zeta e.$$

Thus it follows that

$$\sqrt{\|D \left(\bar{s} - s^0 \right)\|^2 + \|D^{-1} \left(\bar{x} - x^0 \right)\|^2} \leq \zeta \sqrt{\|De\|^2 + \|D^{-1} e\|^2} = \zeta \sqrt{e^T \left(\frac{x}{s} + \frac{s}{x} \right)}.$$

Substitution into (2.39) gives

$$\sqrt{\mu}\, \|q\| \leq \theta \nu \zeta \sqrt{e^T \left(\frac{x}{s} + \frac{s}{x} \right)}.$$

Since $q = -d_x^f$, this completes the proof. □

To proceed we need upper and lower bounds for the elements of the vectors x/s and s/x.

2.3.3 Bounds for x/s and s/x and the choice of α

Let x be feasible for (P_ν) and (y, s) for (D_ν) and, moreover $\delta(x, s; \mu) \leq \tau$, i.e., these iterates are close to the μ-centers of (P_ν) and (D_ν). Based on this information we need to estimate the sizes of the entries of the vectors x/s and s/x. To this end we recall some lemmas from [82]. Let

$$\Phi(xs; \mu) := \Psi(v) := \sum_{i=1}^{n} \psi(v_i), \quad v_i := \sqrt{\frac{x_i s_i}{\mu}}, \quad \psi(t) := \frac{1}{2}\left(t^2 - 1 - \log t^2\right),$$

where $\mu = \mu^0 \nu = \nu\zeta^2$. It is well known that $\psi(t)$ is the kernel function of the primal-dual logarithmic barrier function, which, up to some constant, is the function $\Phi(xs; \mu)$ (see, e.g., [8]).

Lemma 2.3.6. *One has*

$$\Phi(xs; \mu) = \Phi(xs(\nu); \mu) + \Phi(x(\nu)s; \mu).$$

Proof. The equality in the lemma is equivalent to

$$\begin{aligned}\sum_{i=1}^{n}\left(\frac{x_i s_i}{\mu} - 1 - \log\frac{x_i s_i}{\mu}\right) &= \sum_{i=1}^{n}\left(\frac{x_i(\nu)s_i}{\mu} - 1 - \log\frac{x_i(\nu)s_i}{\mu}\right) \\ &\quad + \sum_{i=1}^{n}\left(\frac{x_i s_i(\nu)}{\mu} - 1 - \log\frac{x_i s_i(\nu)}{\mu}\right).\end{aligned}$$

Since

$$\log\frac{x_i s_i}{\mu} = \log\frac{x_i s_i}{x_i(\nu)s_i(\nu)} = \log\frac{x_i}{x_i(\nu)} + \log\frac{s_i}{s_i(\nu)} = \log\frac{x_i s_i(\nu)}{\mu} + \log\frac{x_i(\nu)s_i}{\mu},$$

the lemma holds if and only if

$$x^T s - n\mu = \left(x^T s(\nu) - n\mu\right) + \left(s^T x(\nu) - n\mu\right).$$

Using that $x(\nu)s(\nu) = \mu e$, whence $x(\nu)^T s(\nu) = n\mu$, this can be written as $(x - x(\nu))^T(s - s(\nu)) = 0$, which holds if the vectors $x - x(\nu)$ and $s - s(\nu)$ are orthogonal. This is indeed the case, because $x - x(\nu)$ belongs to the null space and $s - s(\nu)$ to the row space of A. This proves the lemma. □

Lemma 2.3.7. *One has*

$$\psi\left(\sqrt{\frac{x_i}{x_i(\nu)}}\right) \leq \Psi(v), \quad \psi\left(\sqrt{\frac{s_i}{s_i(\nu)}}\right) \leq \Psi(v), \quad i = 1, \ldots, n.$$

Proof. By Lemma 2.3.6 we have $\Phi(xs;\mu) = \Phi(xs(\nu);\mu) + \Phi(x(\nu)s;\nu)$. Since $\Phi(xs;\mu)$ is always nonnegative, also $\Phi(xs(\nu);\mu)$ and $\Phi(x(\nu)s;\mu)$ are nonnegative. Thus it follows that $\Phi(xs(\nu);\mu) \leq \Psi(v)$ and $\Phi(x(\nu)s;\mu) \leq \Psi(v)$. The first of these two inequalities gives

$$\Phi(xs(\nu);\mu) = \sum_{i=1}^{n} \psi\left(\sqrt{\frac{x_i s_i(\nu)}{\mu}}\right) \leq \Psi(v).$$

Since $\psi(t) \geq 0$, for every $t > 0$, it follows that

$$\psi\left(\sqrt{\frac{x_i s_i(\nu)}{\mu}}\right) \leq \Psi(v), \quad i = 1, \ldots, n.$$

Due to $x(\nu)s(\nu) = \mu e$, we have

$$\frac{x_i s_i(\nu)}{\mu} = \frac{x_i s_i(\nu)}{x_i(\nu)s_i(\nu)} = \frac{x_i}{x_i(\nu)},$$

whence we obtain the first inequality in the lemma. The second inequality follows in the same way. □

Note that $\psi(t)$ is monotonically decreasing for $t \leq 1$ and monotonically increasing for $t \geq 1$. In the sequel we denote by $\varrho : [0,\infty) \to [1,\infty)$ the inverse function of $\psi(t)$ for $t \geq 1$ and by $\chi : [0,\infty) \to (0,1]$ the inverse function of $\psi(t)$ for $t \leq 1$. So we have

$$\varrho(s) = t \quad \Leftrightarrow \quad s = \psi(t), \quad s \geq 0,\ t \geq 1. \tag{2.40}$$

and

$$\chi(s) = t \quad \Leftrightarrow \quad s = \psi(t), \quad s \geq 0,\ 0 < t \leq 1. \tag{2.41}$$

Lemma 2.3.8. *let $t > 0$ and $\psi(t) \leq s$. Then $\chi(s) \leq t \leq \varrho(s)$.*

Proof. This is almost obvious. Since $\psi(t)$ is strictly convex and minimal at $t = 1$, with $\psi(1) = 0$, $\psi(t) \leq s$ implies that t belongs to a closed interval whose extremal points are $\chi(s)$ and $\varrho(s)$. □

The above lemma is illustrated in Figure 2.3.

Corollary 2.3.9. *One has*

$$\chi(\Psi(v)) \leq \sqrt{\frac{x_i}{x_i(\nu)}} \leq \varrho(\Psi(v)), \quad \chi(\Psi(v)) \leq \sqrt{\frac{s_i}{s_i(\nu)}} \leq \varrho(\Psi(v)).$$

Proof. This is immediate from Lemma 2.3.8 and Lemma 2.3.7. □

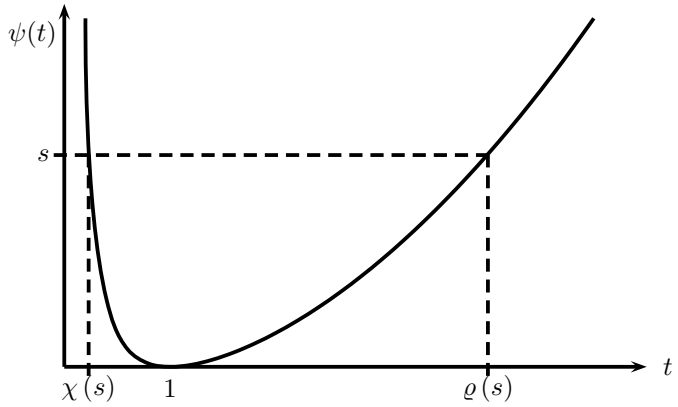

Figure 2.3: *Graphical illustration of the functions $\chi(s)$ and $\varrho(s)$.*

Lemma 2.3.10. *One has*

$$\sqrt{\frac{x}{s}} \leq \frac{\varrho\left(\Psi(v)\right)}{\chi\left(\Psi(v)\right)} \frac{x(\nu)}{\sqrt{\mu}}$$
$$\sqrt{\frac{s}{x}} \leq \frac{\varrho\left(\Psi(v)\right)}{\chi\left(\Psi(v)\right)} \frac{s(\nu)}{\sqrt{\mu}}.$$

Proof. Using that $x_i(\nu)s_i(\nu) = \mu$, for each i, Corollary 2.3.9 implies

$$\sqrt{\frac{x_i}{s_i}} \leq \frac{\varrho\left(\Psi(v)\right)\sqrt{x_i(\nu)}}{\chi\left(\Psi(v)\right)\sqrt{s_i(\nu)}} = \frac{\varrho\left(\Psi(v)\right)}{\chi\left(\Psi(v)\right)} \sqrt{\frac{x_i(\nu)}{s_i(\nu)}}$$
$$= \frac{\varrho\left(\Psi(v)\right)}{\chi\left(\Psi(v)\right)} \sqrt{\frac{x_i^2(\nu)}{\mu}} = \frac{\varrho\left(\Psi(v)\right)}{\chi\left(\Psi(v)\right)} \frac{x_i(\nu)}{\sqrt{\mu}},$$

which implies the first inequality. The second inequality is obtained in the same way. □

Lemma 2.3.11. *Let $t \geq 1$. Then $\psi(t) - \psi\left(\frac{1}{t}\right) \geq 0$.*

Proof. Define $f(t) := \psi(t) - \psi\left(\frac{1}{t}\right)$ for $t > 0$. Then

$$f'(t) = t - \frac{1}{t} - \left(\frac{1}{t} - t\right)\frac{-1}{t^2} = \left(t - \frac{1}{t}\right)\left(1 - \frac{1}{t^2}\right) = \frac{\left(t^2-1\right)^2}{t^3} \geq 0.$$

It follows that $f(t)$ is monotonically increasing for $t > 0$. Since $f(1) = 0$ this proves that $f(t) \geq 0$ for $t \geq 1$, and hence the lemma follows. □

Theorem 2.3.12. *Let $\delta(v)$ be as defined in (1.10) and $\rho(\delta)$ as in (2.27). Then $\Psi(v) \leq \psi\left(\rho(\delta(v))\right)$.*

Proof. The statement in the lemma is obvious if $v = e$ since then $\delta(v) = \Psi(v) = 0$ and since $\rho(0) = 1$ and $\psi(1) = 0$. Otherwise we have $\delta(v) > 0$ and $\Psi(v) > 0$. To deal with the nontrivial case we consider, for $\tau > 0$, the problem

$$z_\tau = \max_v \left\{ \Psi(v) = \sum_{i=1}^n \psi(v_i) \ : \ \delta(v)^2 = \tfrac{1}{4} \sum_{i=1}^n \psi'(v_i)^2 = \tau^2 \right\}.$$

The first order optimality conditions are

$$\psi'(v_i) = \lambda \psi'(v_i) \psi''(v_i), \quad i = 1, \ldots, n,$$

where $\lambda \in \mathbf{R}$. From this we conclude that we have either $\psi'(v_i) = 0$ or $\lambda \psi''(v_i) = 1$, for each i. Since $\psi''(t)$ is monotonically decreasing, this implies that all v_i's for which $\lambda \psi''(v_i) = 1$ have the same value. Denoting this value as t, and observing that all other coordinates have value 1 (since $\psi'(v_i) = 0$ for these coordinates), we conclude that for some k, and after reordering the coordinates, v has the form

$$v = (\underbrace{t, \ldots, t}_{k \text{ times}}, \underbrace{1, \ldots, 1}_{n-k \text{ times}}).$$

Since $\psi'(1) = 0$, $\delta(v) = \tau$ implies $k\psi'(t)^2 = 4\tau^2$. Since $\psi'(t) = t - 1/t$, it follows that

$$t - \frac{1}{t} = \pm \frac{2\tau}{\sqrt{k}},$$

which gives $t = \rho(\tau/\sqrt{k})$ or $t = 1/\rho(\tau/\sqrt{k})$. By Lemma 2.3.11, the first value, which is greater than 1, gives the largest value of $\psi(t)$. This follows because

$$\psi(t) - \psi\left(\frac{1}{t}\right) = \frac{1}{2}\left(t^2 - \frac{1}{t^2}\right) \geq 0, \quad t \geq 1.$$

Since we are maximizing $\Psi(v)$, we conclude that $t = \rho(\tau/\sqrt{k})$, whence we have

$$\Psi(v) = k\, \psi\left(\rho\left(\frac{\tau}{\sqrt{k}}\right)\right).$$

The question remains which value of k maximizes $\Psi(v)$. To investigate this we take the derivative of $\Psi(v)$ with respect to k. To simplify the notation we write

$$\Psi(v) = k\, \psi\,(t)\,, \quad t = \rho\,(s), \quad s = \frac{\tau}{\sqrt{k}}.$$

The definition of t implies $t = s + \sqrt{1 + s^2}$. This gives $(t - s)^2 = 1 + s^2$, or $t^2 - 1 = 2st$, whence we have

$$2s = t - \frac{1}{t} = \psi'(t).$$

Some straightforward computations now yield

$$\frac{d\,\Psi(v)}{dk} = \psi\,(t) - \frac{s^2\rho(s)}{\sqrt{1+s^2}} =: f(\tau).$$

We consider this derivative as a function of τ, as indicated. One may easily verify that $f(0) = 0$. We proceed by computing the derivative with respect to τ. This gives

$$f'(\tau) = -\frac{1}{\sqrt{k}}\frac{s^2}{(1+s^2)\sqrt{1+s^2}}$$

This proves that $f'(\tau) \leq 0$. Since $f(0) = 0$, it follows that $f(\tau) \leq 0$, for each $\tau \geq 0$. Hence we conclude that $\Psi(v)$ is decreasing in k. So $\Psi(v)$ is maximal when $k = 1$, which gives the result in the theorem. □

Corollary 2.3.13. *Let $\tau \geq 0, \delta(v) \leq \tau$ and $\rho\,(\delta)$ as defined in (2.27). Then $\Psi(v) \leq \tau'$, where*

$$\tau' := \psi\,(\rho(\tau))\,. \tag{2.42}$$

Proof. Since $\rho(s)$ is monotonically increasing in s, and $\rho(s) \geq 1$ for all $s \geq 0$, and, moreover, $\psi(t)$ is monotonically increasing if $t \geq 1$, the function $\psi\,(\rho(\delta))$ is increasing in δ, for $\delta \geq 0$. Thus the result is immediate from Theorem 2.3.12. □

Theorem 2.3.14. *Let $\tau \geq 0$ and $\delta(v) \leq \tau$ and τ' as defined in (2.42). Then*

$$\sqrt{\frac{x}{s}} \leq \frac{\varrho\,(\tau')}{\chi\,(\tau')}\frac{x(\nu)}{\sqrt{\mu}}$$
$$\sqrt{\frac{s}{x}} \leq \frac{\varrho\,(\tau')}{\chi\,(\tau')}\frac{s(\nu)}{\sqrt{\mu}}.$$

Proof. Since $\varrho(t)$ is monotonically increasing and $\chi(t)$ monotonically decreasing this is an immediate consequence of Lemma 2.3.10 and Corollary 2.3.13. □

Corollary 2.3.15. *Let $\tau = \frac{1}{8}$, $\delta\,(v) \leq \tau$ and $\rho\,(\delta)$ as defined in (2.27). Then*

$$\sqrt{\frac{x}{s}} \leq \sqrt{2}\frac{x\,(\nu)}{\sqrt{\mu}}, \qquad \sqrt{\frac{s}{x}} \leq \sqrt{2}\frac{s\,(\nu)}{\sqrt{\mu}}.$$

Proof. If $\tau = \frac{1}{8}$, then $\tau' \approx 0.016921$, $\varrho\,(\tau') \approx 1.13278$ and $\chi\,(\tau') \approx 0.872865$, whence

$$\frac{\varrho\,(\tau')}{\chi\,(\tau')} \approx 1.29777 < \sqrt{2}.$$

Thus the result follows. □

Based on the above analysis we choose

$$\tau = \frac{1}{8}. \tag{2.43}$$

By Corollary 2.3.15 we then have

$$\sqrt{\frac{x}{s}} \leq \sqrt{2}\frac{x\,(\nu)}{\sqrt{\mu}},$$
$$\sqrt{\frac{s}{x}} \leq \sqrt{2}\frac{s\,(\nu)}{\sqrt{\mu}},$$

Substitution into (2.38) gives

$$\sqrt{\mu}\,\left\|d_x^f\right\| \leq \theta\nu\,\zeta\sqrt{2e^T\left(\frac{x(\nu)^2}{\mu}+\frac{s(\nu)^2}{\mu}\right)}.$$

This implies

$$\mu\,\left\|d_x^f\right\| \leq \theta\nu\zeta\sqrt{2}\,\sqrt{\|x(\nu)\|^2+\|s(\nu)\|^2}.$$

Therefore, also using $\mu = \nu\zeta^2$ and $\theta = \frac{\alpha}{\sqrt{2n}}$, we obtain the following upper bound for $\left\|d_x^f\right\|$:

$$\left\|d_x^f\right\| \leq \frac{\alpha}{\zeta\sqrt{n}}\,\sqrt{\|x(\nu)\|^2+\|s(\nu)\|^2}.$$

We define

$$\kappa(\zeta,\nu) = \frac{\sqrt{\|x(\nu)\|^2+\|s(\nu)\|^2}}{\zeta\sqrt{2n}}, \quad 0<\nu\leq 1.$$

Now we may write

$$\left\|d_x^f\right\| \leq \alpha\,\bar{\kappa}(\zeta)\sqrt{2} \quad \text{where} \quad \bar{\kappa}(\zeta) = \max_{0<\nu\leq 1}\,\kappa(\zeta,\nu).$$

Note that since $x(1) = s(1) = \zeta e$, we have $\kappa(\zeta,1) = 1$. Hence it follows that $\bar{\kappa}(\zeta) \geq 1$. We found in (2.36) that in order to have $\delta(v^f) \leq 1/\sqrt{2}$, we should have $2\left\|d_x^f\right\| \leq 1$. This certainly holds if $2\alpha\,\bar{\kappa}(\zeta)\sqrt{2} \leq 1$. We conclude that if we take

$$\alpha = \frac{1}{2\sqrt{2}\,\bar{\kappa}(\zeta)}, \tag{2.44}$$

then we will certainly have $\delta(v^f) \leq 1/\sqrt{2}$. In the next section we will find an upper bound for $\bar{\kappa}\,(\zeta)$. The next section is based on Section 4.6 in [82].

2.3.4 Bound for $\bar{\kappa}(\zeta)$.

Due to the choice of the vectors $\bar{x}$, $\bar{y}$, $\bar{s}$ and the number ζ we have

$$\begin{aligned} A\bar{x} &= b, \quad 0 \leq \bar{x} \leq \zeta e \\ A^T\bar{y} + \bar{s} &= c, \quad 0 \leq \bar{s} \leq \zeta e \\ \bar{x}\bar{s} &= 0. \end{aligned}$$

To simplify notation in the rest of this section, we denote $x = x(\nu)$, $y = y(\nu)$ and $s = s(\nu)$. Then x, y and s are uniquely determined by the system

$$\begin{aligned} b - Ax &= \nu(b - A\zeta e), \quad x \geq 0 \\ c - A^T y - s &= \nu(c - \zeta e), \quad s \geq 0 \\ xs &= \nu\zeta^2 e. \end{aligned}$$

Hence we have

$$\begin{aligned} A\bar{x} - Ax &= \nu(A\bar{x} - \zeta Ae), \quad x \geq 0 \\ A^T\bar{y} + \bar{s} - A^T y - s &= \nu\left(A^T\bar{y} + \bar{s} - \zeta e\right), \quad s \geq 0 \\ xs &= \nu\zeta^2 e. \end{aligned}$$

We rewrite this system as

$$\begin{aligned} A(\bar{x} - x - \nu\bar{x} + \nu\zeta e) &= 0, \quad x \geq 0 \\ A^T(\bar{y} - y - \nu\bar{y}) &= s - \bar{s} + \nu\bar{s} - \nu\zeta e, \quad s \geq 0 \\ xs &= \nu\zeta^2 e. \end{aligned}$$

By using again orthogonality of the row space of a matrix and its null space, we get

$$(\bar{x} - x - \nu\bar{x} + \nu\zeta e)^T (s - \bar{s} + \nu\bar{s} - \nu\zeta e) = 0.$$

Defining

$$a := (1-\nu)\bar{x} + \nu\zeta e, \quad b := (1-\nu)\bar{s} + \nu\zeta e,$$

we have $(a - x)^T (b - s) = 0$. This gives

$$a^T b + x^T s = a^T s + b^T x.$$

Since $\bar{x}^T\bar{s} = 0$, $\bar{x} + \bar{s} \leq \zeta e$ and $xs = \nu\zeta^2 e$, we may have

$$\begin{aligned} a^T b + x^T s &= ((1-\nu)\bar{x} + \nu\zeta e)^T ((1-\nu)\bar{s} + \nu\zeta e) + \nu\zeta^2 n \\ &= \nu(1-\nu)(\bar{x} + \bar{s})^T \zeta e + \nu^2\zeta^2 n + \nu\zeta^2 n \\ &\leq \nu(1-\nu)(\zeta e)^T \zeta e + \nu^2\zeta^2 n + \nu\zeta^2 n \\ &= \nu(1-\nu)\zeta^2 n + \nu^2\zeta^2 n + \nu\zeta^2 n = 2\nu\zeta^2 n. \end{aligned}$$

Moreover, also using $a \geq \nu\zeta e$, $b \geq \nu\zeta e$, we get

$$\begin{aligned} a^T s + b^T x &= \left((1-\nu)\,\bar{x} + \nu\zeta e\right)^T s + \left((1-\nu)\,\bar{s} + \nu\zeta e\right)^T x \\ &= (1-\nu)\left(s^T\bar{x} + x^T\bar{s}\right) + \nu\zeta e^T(x+s) \\ &\geq \nu\zeta e^T(x+s) = \nu\zeta\left(\|x\|_1 + \|s\|_1\right). \end{aligned}$$

Hence we obtain $\|x\|_1 + \|s\|_1 \leq 2\zeta n$. Since $\|x\|^2 + \|s\|^2 \leq \left(\|x\|_1 + \|s\|_1\right)^2$, it follows that

$$\frac{\sqrt{\|x\|^2 + \|s\|^2}}{\zeta\sqrt{2n}} \leq \frac{\|x\|_1 + \|s\|_1}{\zeta\sqrt{2n}} \leq \frac{2\zeta n}{\zeta\sqrt{2n}} = \sqrt{2\,n},$$

thus proving

$$\bar{\kappa}(\zeta) \leq \sqrt{2\,n}.$$

Substitution of the upper bound for $\bar{\kappa}(\zeta)$ in (2.44) we obtain that we certainly have $\delta(v^+) \leq 1/\sqrt{2}$ if

$$\alpha = \frac{1}{4\sqrt{n}}. \tag{2.45}$$

According to (2.35) this gives the following value for θ:

$$\theta = \frac{1}{4n\sqrt{2}}. \tag{2.46}$$

2.4 Iteration bound

In the previous sections we have found that if at the start of an iteration the iterates satisfy $\delta(x,s;\mu) \leq \tau$, with τ as defined in (2.35), and θ as in (2.46), then after the feasibility step and the μ-update the iterates satisfy $\delta(x,s;\mu) \leq 1/\sqrt{2}$.

According to (2.21), at most

$$\log_2\left(\log_2\frac{1}{\tau^2}\right) = \log_2\left(\log_2 64\right) \leq 3$$

centering steps suffice to get iterates that satisfy $\delta(x,s;\mu) \leq \tau$. So each iteration consists of one feasibility step and at most 3 centering steps. In each iteration both the duality gap and the norms of the residual vectors are reduced by the factor $1-\theta$. Hence, using ${x^0}^T s^0 = n\zeta^2$, the total number of iterations is bounded above by

$$\frac{1}{\theta}\log\frac{\max\left\{n\zeta^2,\ \left\|r_b^0\right\|,\ \left\|r_c^0\right\|\right\}}{\epsilon}.$$

Hence, due to (2.46) the total number of inner iterations is bounded above by

$$16\,n\,\sqrt{2}\log\frac{\max\left\{n\zeta^2,\ \left\|r_b^0\right\|,\ \left\|r_c^0\right\|\right\}}{\epsilon}.$$

Thus we may state without further proof our main result as follows.

Theorem 2.4.1. *If (P) and (D) are feasible and $\zeta > 0$ is such that $\|x^* + s^*\|_\infty \leq \zeta$ for some optimal solutions x^* of (P) and (y^*, s^*) of (D) then after at most*

$$16\, n\, \sqrt{2} \log \frac{\max\left\{n\zeta^2,\, \left\|r_b^0\right\|,\, \left\|r_c^0\right\|\right\}}{\epsilon}.$$

inner iterations the algorithm finds an ϵ-solution of (P) and (D).

Note that the order of this bound is exactly the same as the order of the bound in (1.29). A basic question is of course how to choose the number ζ, which determines the initial iterates and has to be fixed before starting the algorithm. A related question that we did not yet deal with is wether or not our algorithm can detect infeasibility (or unboundedness) of (P) and (D). These issues can be dealt with if the data A, b and c are integral (or rational). In that case it is well known that if (P) and (D) are feasible then there exist optimal solutions x^* and (y^*, s^*) of (P) and (D) such that $\|x^* + s^*\|_\infty \leq 2^L$, where L denotes the binary input size of (P) and (D) (See, e.g., [103]). The number L can be computed straightforwardly from the input data A, b and c. Thus, when starting the algorithm with $\zeta = 2^L$, after at most

$$16\, n\, \sqrt{2} \log \frac{\max\left\{n4^L,\, \left\|b - 2^L Ae\right\|,\, \left\|c - 2^L e\right\|\right\}}{\epsilon}.$$

iterations the algorithm finds an ϵ-solution if it exists. Otherwise we must decide that (P) and (D) are infeasible or unbounded. Working with the number L may not be possible in practice, however, since this number can be very large. For such cases it may be worth noting that if (P) and (D) are infeasible or unbounded then Lemma 2.2.1 implies that P_ν and D_ν do not satisfy the IPC for all $\nu \in (0,\, \bar{\nu}]$ for some $\bar{\nu} \in (0, 1)$. If the iterates after the feasibility step satisfy $\delta\left(x^f, s^f; \mu^+\right) \leq 1/\sqrt{2}$ we are sure that the perturbed problems corresponding to $\nu = \mu^+/\mu^0$ satisfy the IPC, and hence are strictly feasible. So we then have $\nu \geq \bar{\nu}$. On the other hand, if $\nu < \bar{\nu}$ the algorithm will find that $\delta\left(x^f, s^f; \mu^+\right) > 1/\sqrt{2}$, which implies that there are no optimal solutions x^* and $(y^*,\, s^*)$ such that $\zeta \geq \|x^* + s^*\|_\infty$. We can then run the algorithm again with $\zeta = 2\,\zeta$. If necessary this can be repeated. When starting with $\zeta = 1$, after doubling the value of ζ at most L times the algorithm must have found optimal solutions of (P) and (D) if these exist. Otherwise (P) and (D) are infeasible or unbounded.

Remark 2.4.2. *It may be worthwhile to point out that we used our algorithm to solve 10^5 instances. Each instance consisted of a primal-dual pair of randomly generated problems with known optimal solutions x^* and $(y^*,\, s^*)$. We ran the algorithm for each problem pair with $\zeta = \|x^* + s^*\|_\infty$. To our surprise we found in all cases that $\bar{\kappa}(\zeta) \leq 1$. If one could prove that $\bar{\kappa}(\zeta) \leq 1$ holds for all LO problems it would reduce the currently best iteration bound for IIPMs (1.29) by a factor $\sqrt{2n}$.*

Chapter 3

A full-Newton step primal-dual IIPM for SDO

3.1 Introduction

Semidefinite optimization (SDO) problems are convex optimization problems over the intersection of an affine set and the cone of positive semidefinite matrices. SDO arises in many scientific and engineering fields. For applications in system and control theory we refer to [11, 14, 15] and for applications in statistics and combinatorial optimization to [2, 26, 40, 50, 51, 71, 91]. SDO also has been utilized in solving polynomial optimization problems [19, 49, 70].

In this chapter we consider the standard form of the SDO problem:

$$(SDP) \qquad \begin{array}{l} \min \ \mathbf{Tr}\,(CX) \\ s.t \ \mathbf{Tr}\,(A_i X) = b_i, \quad i = 1, 2, \ldots, m, \quad X \succeq 0, \end{array}$$

and its dual:

$$(SDD) \qquad \begin{array}{l} \max \quad b^T y \\ s.t \ \sum_{i=1}^{m} y_i \ A_i + S = C, \quad S \succeq 0, \end{array}$$

where each $A_i \in \mathbf{S}^n$, $b \in \mathbf{R}^m$, and $C \in \mathbf{S}^n$. Recall that $\mathbf{Tr}\,(CX)$ is the inner product of the matrices C and X. Without loss of generality we assume that the matrices A_i are linearly independent.

Interior-point methods provide a powerful approach for solving SDO problems. Most IPMs for SDO can be viewed as natural extensions of IPMs for LO, and have similar polynomial complexity results. For example Nesterov and Todd [66]

showed that the primal-dual algorithm for LO maintains its theoretical efficiency when the nonnegativity constraints in LO are replaced by constraints that restrict the variables to a convex cone, provided that the cone is homogeneous and self-dual, or in the terminology of Nesterov and Todd, the cone is self-scaled. The non-negative orthant and the cone of positive semidefinite matrices are special cases of such cones. Many other IPMs for solving SDO problems exist. See, e.g., [1, 48, 61, 96, 100].

In the case of IIPMs, as mentioned in Chapter 1, Zhang [107] was the first who obtained polynomial iteration complexity for LO. Later he extended his algorithm to SDO [108]. In his algorithm, the starting point is chosen as

$$\left(X^0,\, y^0,\, S^0\right) = \zeta\,(I,\, 0,\, I) \tag{3.1}$$

with $\zeta > 0$ such that

$$\|X^*\| \leq \zeta, \qquad \|S^*\| \leq \zeta \tag{3.2}$$

for some optimal solution X^* of (SDP) and $(y^*,\, S^*)$ of (SDD).

Kojima et al. [47] and Potra and Sheng [79] independently analyzed a generalization to SDO of the Mizuno-Todd-Ye algorithm [59] for infeasible starting points and they proved that the complexity of their algorithm is

$$O\left(n \log \frac{\max\left\{\mathbf{Tr}\left(X^0 S^0\right),\, \left\|R_c^0\right\|,\, \left\|r_b^0\right\|\right\}}{\epsilon}\right). \tag{3.3}$$

Here X^0 and S^0 are as defined in (3.1) and (3.2) and R_c^0 and r_b^0 denote the initial values of the primal and dual residuals:

$$\left(r_b^0\right)_i = b_i - \mathbf{Tr}\left(A_i X^0\right), \quad i = 1,\, \ldots,\, m, \tag{3.4}$$

$$R_c^0 = C - \sum_{i=1}^{m} y_i^0 A_i - S^0. \tag{3.5}$$

In this chapter we undertake the task of generalizing the algorithm for solving LO that was presented in the previous chapter to the case of SDO. We show that the iteration bound of our algorithm is the same as the best iteration bound for IIPMs, as given in (3.3). However, to obtain this complexity result is much more difficult than in the LO case. The main differences occur in the analysis of the feasibility step. Below we present our algorithm for SDO and its analysis. In Section 3.4 we present the analysis of the feasibility step.

3.2 Preliminaries

3.2.1 Duality theory and assumption

In this section we first introduce the duality theory for SDO problem and the assumption that will be used in the chapter. It is worth noting that the duality

theory of SDO is weaker than that of LO. Like in LO we have the weak duality property (see, Theorem 1.2.2): for any $(X,\, y,\, S) \in \mathcal{P} \times \mathcal{D}$, where $\mathcal{P}$ and $\mathcal{D}$ denote the feasible sets of the problems (SDP) and (SDD), respectively, we have

$$\mathbf{Tr}\,(CX) - b^T y = \mathbf{Tr}\left(\left(S + \sum_{i=1}^{m} y_i A_i\right) X\right) - \sum_{i=1}^{m} y_i \mathbf{Tr}\,(A_i X) = \mathbf{Tr}\,(SX) \geq 0,$$

where the inequality follows from $X \succeq 0$ and $S \succeq 0$ (see Lemma A.3.3). In other words, the duality gap is nonnegative for any feasible primal-dual pair. As a consequence, feasible solutions $(X,\, y,\, S)$ with zero duality gap are optimal. Let p^* and d^* denote the optimal value of the primal and the dual problem respectively. The optimal sets for (SDP) and (SDD) are denoted as follows:

$$\mathcal{P}^* := \{X \in \mathcal{P} \mid \mathbf{Tr}\,(CX) = p^*\} \text{ and } \mathcal{D}^* := \left\{(y,\, S) \in \mathcal{D} \mid b^T y = d^*\right\}.$$

A problem (SDP) (resp. (SDD)) is called solvable if $\mathcal{P}^*$ (resp. $\mathcal{D}^*$) is nonempty.

The duality properties in SDO are less simple than LO. Recall that an LO problem may be feasible or infeasible. If it is feasible then it is either unbounded or bounded. In case it is bounded it is solvable and the dual problem is solvable as well, with zero duality gap. If it is infeasible then its dual is either unbounded or infeasible . So, for a single problem there are 3 possibilities: a problem is either solvable, unbounded or infeasible (see Theorem 1.2.4).

For an SDO problem, however, the situation is less simple, since a problem that is feasible and bounded is not necessarily solvable. So for a single problem there are now four cases: a problem is solvable, feasible and not solvable, unbounded or infeasible. As a consequence, for a primal-dual pair of SDO problems there are much more possible situations than in the LO case. For example, a problem may be solvable, whereas its dual is not solvable. Also both problems can be solvable but with positive duality gap. In this chapter we assume that there exist optimal solutions with zero duality gap, i.e., we assume that the set

$$\mathcal{F}^* = \{(X,\, y,\, S) \in \mathcal{P} \times \mathcal{D} : \quad \mathbf{Tr}\,(XS) = 0\},$$

is nonempty.

Following the same outline as in the LO case we first present a feasible IPM that uses full-Newton steps.

3.2.2 The central path and feasible full-Newton IPM for SDO

In this section we assume that (SDP) and (SDD) satisfy the IPC, i.e., there exist $X \in \mathcal{P}$, $S \in \mathcal{D}$ with $X \succ 0$, $S \succ 0$. Under this assumption both problems are solvable and the optimality conditions for (SDP) and (SDD) can be written as

follows.

$$\begin{aligned} \mathbf{Tr}\,(A_i X) &= b_i, \quad i = 1, 2, \ldots, m, \; X \succeq 0 \\ \sum_{i=1}^{m} y_i A_i + S &= C, \quad S \succeq 0, \\ XS &= 0. \end{aligned} \tag{3.6}$$

The basic idea of primal-dual IPMs for SDO is to replace the above complementarity condition $XS = 0$ by the parameterized equation $XS = \mu I$. Then we get the following system:

$$\begin{aligned} \mathbf{Tr}\,(A_i X) &= b_i, \quad i = 1, 2, \ldots, m, \; X \succeq 0, \\ \sum_{i=1}^{m} y_i A_i + S &= C, \quad S \succeq 0, \\ XS &= \mu I, \end{aligned} \tag{3.7}$$

where I denotes the $n \times n$ identity matrix and $\mu > 0$. It is well known that the system (3.7) has a unique solution for every $\mu > 0$, denoted as $(X(\mu), y(\mu), S(\mu))$, and that the limit $\lim_{\mu \to 0} (X(\mu), y(\mu), S(\mu))$ exists and is a solution of system (3.6) (see e.g., [31, 48]). The set of all solutions $(X(\mu), y(\mu), S(\mu))$ with $\mu > 0$ is known as the central path [17, 75, 76]. The principal idea of (feasible) IPMs is to follow this central path and approach the optimal set of (SDP) as μ goes to zero.

To obtain a search direction for IPMs an ordinary method is to follow an idea of Newton method and to linearize (3.7). Thus we have the following system

$$\begin{aligned} \mathbf{Tr}\,(A_i \Delta X) &= 0, \quad i = 1, 2, \ldots, m, \\ \sum_{i=1}^{m} y_i A_i + \Delta S &= 0, \\ X\Delta S + \Delta X S &= \mu I - XS. \end{aligned} \tag{3.8}$$

However, this system may yield as a solution a search direction ΔX which is not symmetric (ΔS is automatically symmetric!). Since we want ΔX to be a symmetric matrix, one must "symmetrize" the linearization of the complementary equation. Based on different symmetrization schemes, several search directions have been proposed, as presented in [48, 66, 90, 94, 108]. In this thesis, we use Zhang's direction [108] which is obtained by the following system:

$$\begin{aligned} \mathbf{Tr}\,(A_i \Delta X) &= 0, \quad i = 1, 2, \ldots, m, \\ \sum_{i=1}^{m} \Delta y_i A_i + \Delta S &= 0, \\ \Delta X + P \Delta S P^T &= \mu S^{-1} - X, \end{aligned} \tag{3.9}$$

where P is a symmetric nonsingular matrix. The system (3.9) has a unique solution for each P [100]. Now ΔX obtained from this system is automatically a symmetric matrix. Some choices for the matrix P are listed in Table 3.1.

P	References
$X^{\frac{1}{2}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{-1}{2}}X^{\frac{1}{2}}$	[65]
$X^{\frac{-1}{2}}$ and $S^{\frac{1}{2}}$	[48, 61]
I	[3]

Table 3.1: *Choices for the scaling matrix P.*

In this thesis we use the matrix P proposed by Nesterov and Todd [65], which is as follows:

$$P := X^{\frac{1}{2}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{-\frac{1}{2}}X^{\frac{1}{2}} = S^{-\frac{1}{2}}\left(S^{\frac{1}{2}}XS^{\frac{1}{2}}\right)^{\frac{1}{2}}S^{-\frac{1}{2}}. \tag{3.10}$$

The corresponding solution (ΔX, Δy, ΔS) of system (3.8) is called NT-direction (also Newton-direction, see [94]). Let $D = P^{-\frac{1}{2}}$, where $P^{-\frac{1}{2}}$ denotes the inverse symmetric square root of P. Then D can be used to scale X and S to the same matrix V, namely [18, 90, 98]:

$$V := \frac{1}{\sqrt{\mu}}D^{-1}XD^{-1} = \frac{1}{\sqrt{\mu}}DSD. \tag{3.11}$$

It follows that

$$V^2 := \frac{1}{\mu}D^{-1}XSD. \tag{3.12}$$

Note that the matrices D and V are symmetric and positive definite. Let us further define

$$\bar{A}_i := \frac{1}{\sqrt{\mu}}DA_iD, \quad i = 1, 2, \ldots, m;$$

and

$$D_X := \frac{1}{\sqrt{\mu}}D^{-1}\Delta XD^{-1}; \quad D_S := \frac{1}{\sqrt{\mu}}D\Delta SD. \tag{3.13}$$

Then (3.9) can be written as follows

$$\begin{aligned} \mathbf{Tr}\left(\bar{A}_i D_X\right) &= 0, \quad i = 1, 2, \ldots, m, \\ \sum_{i=1}^{m} \Delta y_i \bar{A}_i + D_S &= 0, \\ D_X + D_S &= V^{-1} - V. \end{aligned} \tag{3.14}$$

The first two equations in the system (3.9) imply that ΔX and ΔS are orthogonal, i.e.,

$$\mathbf{Tr}\left(\Delta X \Delta S\right) = 0.$$

This implies that the directions D_X and D_S are also orthogonal:

$$\mathbf{Tr}\left(D_X D_S\right) = 0. \tag{3.15}$$

Hence, using the third equation in (3.14) and the Frobenius matrix norm we obtain

$$\|D_X\|^2 + \|D_S\|^2 = \left\|V^{-1} - V\right\|^2. \tag{3.16}$$

This implies that D_X and D_S are both zero if and only if $V^{-1} - V = 0$. In this case, X and S satisfy $XS = \mu I$, which indicates that X and S are the μ-centers. Thus, we can use $\left\|V^{-1} - V\right\|$ as a quantity to measure closeness to μ-centers. We define

$$\delta\left(X, S; \mu\right) := \delta\left(V\right) := \frac{1}{2}\left\|V^{-1} - V\right\|. \tag{3.17}$$

Note that for the special case of LO, V is a diagonal matrix and this proximity measure becomes the same as (1.10). It was used first in [39].

A feasible full-Newton step IPM for SDO is described in Figure 3.1. This is an extension of the algorithm in Figure 1.2 for LO.

Below the new iterates are denoted as X^+, y^+ and S^+, so we have

$$X^+ = X + \Delta X, \tag{3.18}$$

$$y^+ = y + \Delta y,$$

$$S^+ = S + \Delta S. \tag{3.19}$$

We continue this section by recalling some lemmas that are crucial in the analysis of the algorithm in Figure 3.1. Before doing this we first mention that by using (3.11), (3.13), (3.18) and (3.19) we have

$$\begin{aligned} X^+ &= X + \Delta X = \sqrt{\mu} D \left(V + D_X\right) D, \\ S^+ &= S + \Delta S = \sqrt{\mu} D^{-1} \left(V + D_S\right) D^{-1}. \end{aligned}$$

Primal-Dual Feasible IPM for SDO

Input:
Accuracy parameter $\epsilon > 0$;
threshold parameter $\tau < 1$;
barrier update parameter θ, $0 < \theta < 1$;
feasible pair $\left(X^0,\, S^0\right)$ and $\mu^0 > 0$ such that $\delta\left(X^0,\, S^0;\, \mu^0\right) \leq \tau$.

begin
$X := X^0;\; S := S^0;\; \mu := \mu^0;$
while $\mathbf{Tr}\,(XS) \geq \epsilon$ **do**
begin
update of μ:
$\mu := (1-\theta)\mu;$
centering step:
$(X,\, S) := (X,\, S) + (\Delta X,\, \Delta S);$
end
end

Figure 3.1: *Feasible full-Newton-step algorithm for SDO*

Therefore, we have

$$X^+S^+ = \mu D\,(V + D_X)\,(V + D_S)\,D^{-1},$$

which makes clear that the eigenvalues of two matrices X^+S^+ and $\mu(V+D_X)(V+D_S)$ are equal. In other words, these matrices are similar,

$$X^+S^+ \sim \mu\,(V + D_X)\,(V + D_S)\,. \tag{3.20}$$

To simplify the notation we introduce

$$D_{XS} := \frac{1}{2}\,(D_X D_S + D_S D_X)\,, \tag{3.21}$$

and

$$H := (D_X V - V D_X) + \frac{1}{2}\,(D_X D_S - D_S D_X)\,. \tag{3.22}$$

Note that the matrix D_{XS} is symmetric and H is skew-symmetric. By multiplying both sides of the third equation in (3.14) from the left with matrix V we obtain

$$VD_X + VD_S = I - V^2.$$

Using this we may write,

$$\begin{aligned}(V + D_X)(V + D_S) &= V^2 + D_X V + V D_S + D_X D_S \\ &= I - V D_X + D_X V + D_X D_S.\end{aligned}$$

By subtracting and adding $\frac{1}{2} D_S D_X$ to the last expression we obtain

$$\begin{aligned}(V + D_X)(V + D_S) &= I + \frac{1}{2}(D_X D_S + D_S D_X) \\ &\quad + (D_X V - V D_X) + \frac{1}{2}(D_X D_S - D_S D_X) \\ &= I + D_{XS} + H,\end{aligned}$$

where the last equality follows by using (3.21) and (3.22). Using (3.20) we thus obtain

$$X^+ S^+ \sim \mu (I + D_{XS} + H). \tag{3.23}$$

Lemma 3.2.1 (Cf. Corollary 7.1 in [18]). *Let* $(X,\ S) \in \text{relint}(\mathcal{P} \times \mathcal{D})$ *and* $\mu > 0$. *Then*

$$\mathbf{Tr}\left(X^+ S^+\right) = n\mu.$$

Proof. Due to (3.23) we have

$$\mathbf{Tr}\left(X^+ S^+\right) = \mu \mathbf{Tr}(I + D_{XS} + H) = \mu \mathbf{Tr}(I) = n\mu,$$

where the second equality follows since $\mathbf{Tr}(D_{XS}) = 0$, by (3.15), and $\mathbf{Tr}(H) = 0$ since matrix H is skew-symmetric. This proves the lemma. □

Lemma 3.2.2 (Lemma 7.5 in [18]). *Let* $(X,\ S) \in \text{relint}(\mathcal{P} \times \mathcal{D})$, $\mathbf{Tr}(XS) = n\mu$ *and* $\delta := \delta(X,\ S;\ \mu)$. *If* $\mu^+ = (1 - \theta)\mu$ *for* $0 < \theta < 1$, *then one has*

$$\delta\left(X,\ S;\ \mu^+\right)^2 = \frac{n\theta^2}{4(1 - \theta)} + (1 - \theta)\delta^2.$$

Proof. Let matrix V be as defined in (3.11). Then, by using the definition of $\delta(V)$ in (3.17), we may write

$$\begin{aligned}4\delta\left(X,\ S;\ \mu^+\right)^2 &= \left\| \sqrt{1 - \theta} V^{-1} - \frac{V}{\sqrt{1 - \theta}} \right\|^2 \\ &= \left\| \frac{\theta V}{\sqrt{1 - \theta}} - \sqrt{1 - \theta}\left(V^{-1} - V\right) \right\|^2.\end{aligned}$$

From $\mathbf{Tr}(XS) = n\mu$ it follows that $\|V\|^2 = n$. This implies that V and $V^{-1} - V$ are orthogonal:

$$\mathbf{Tr}\left(V\left(V^{-1} - V\right)\right) = n - \|V\|^2 = 0.$$

Therefore,

$$4\delta\left(X,\, S;\, \mu^{+}\right)^{2} = \frac{\theta^{2}\left\|V\right\|^{2}}{1-\theta} + (1-\theta)\left\|V^{-1} - V\right\|^{2}.$$

Finally, since $\left\|V^{-1} - V\right\| = 2\delta$ and $\left\|V\right\|^{2} = n$ the result follows. □

Based on Lemma 7.1 in [18] we have the following lemma which gives a sufficient condition for feasibility of the new iterates $(X^{+},\, y^{+},\, S^{+})$.

Lemma 3.2.3. *Let* $(X,\, S) \in \operatorname{relint}(\mathcal{P} \times \mathcal{D})$ *and* $\mu > 0$. *Then the iterates* $(X^{+},\, y^{+},\, S^{+})$ *are strictly feasible if*

$$I + D_{XS} \succ 0.$$

Proof. For the proof we consider a step length $\alpha \in [0,\, 1]$, and we define

$$X(\alpha) = X + \alpha\Delta X, \quad y(\alpha) = y + \alpha\Delta y, \quad S(\alpha) = S + \alpha\Delta S.$$

We then have $X(0) = X$, $X(1) = X^{+}$ and similar relations for y and S. Hence we have $\det\left(X(0)\, S(0)\right) = \det\left(\mu^{0} I\right) = \left(\mu^{0}\right)^{n} > 0$. We want to prove that the determinant of $X(\alpha)\, S(\alpha)$ is positive for all $\alpha \leq 1$. We write

$$\begin{aligned}
\frac{X(\alpha)\, S(\alpha)}{\mu} &\sim (V + \alpha D_X)(V + \alpha D_S) \\
&= V^{2} + \alpha(D_X V + V D_S) + \alpha^{2} D_X D_S \\
&= V^{2} + \alpha(V D_X + V D_S) + \alpha(D_X V - V D_X) + \alpha^{2} D_X D_S.
\end{aligned}$$

By using (3.21) and (3.23), and subtracting and adding $\frac{1}{2}\alpha^{2} D_S D_X$ and $\alpha^{2} I$ to the right hand side of the above equality we obtain

$$\begin{aligned}
\frac{X(\alpha)\, S(\alpha)}{\mu} &= (1-\alpha) V^{2} + \alpha(1-\alpha) I + \alpha^{2}\left(I + \frac{1}{2}(D_X D_S + D_S D_X)\right) \\
&\quad + \left[\frac{1}{2}\alpha^{2}(D_X D_S - D_S D_X) + \alpha(D_X V - V D_X)\right] \\
&= (1-\alpha) V^{2} + \alpha(1-\alpha) I + \alpha^{2}(I + D_{XS}) \\
&\quad + \left[\frac{1}{2}\alpha^{2}(D_X D_S - D_S D_X) + \alpha(D_X V - V D_X)\right].
\end{aligned}$$

The matrix in brackets in the last expression is skew-symmetric. Therefore due to Lemma A.3.1 the determinant of $X(\alpha)\, S(\alpha)$ will be positive if the symmetric matrix

$$(1-\alpha) V^{2} + \alpha(1-\alpha) I + \alpha^{2}(I + D_{XS}) \tag{3.24}$$

is positive definite. The latter is true for $0 \leq \alpha \leq 1$. Hence, the symmetric matrix in (3.24) is positive definite for $\alpha \in [0,\, 1]$. This implies that $X(\alpha)\, S(\alpha)$

has positive determinant for $\alpha \in [0, 1]$. Since $X(0)$ and $S(0)$ are positive definite and since $X(\alpha)$ and $S(\alpha)$ depend continuously on α, it follows that $X(1)$ and $S(1)$ are also positive definite. This completes the proof. □

Corollary 3.2.4. *The iterates* (X^+, y^+, S^+) *are strictly feasible if*

$$|\lambda_i(D_{XS})| < 1, \quad \text{for} \quad i = 1, \ldots, n.$$

Proof. By Lemma 3.2.3 X^+ and S^+ are strictly feasible if $I + D_{XS} \succ 0$. Since the last inequality certainly holds if $|\lambda_i(D_{XS})| < 1$, for $i = 1, \ldots n$, the corollary follows. □

Lemma 3.2.5. *One has*

$$|\lambda_i(D_{XS})| \leq \delta^2, \qquad \text{for} \quad i = 1, \ldots, n,$$

where D_{XS} *is as defined in (3.21) and* δ *is as in (3.17).*

Proof. It is trivial to verify that

$$D_{XS} = \frac{1}{2}(D_X D_S + D_S D_X) = \frac{1}{4}\left[(D_X + D_S)^2 - (D_X - D_S)^2\right] \tag{3.25}$$

which implies

$$-\frac{1}{4}(D_X - D_S)^2 \preceq D_{XS} \preceq \frac{1}{4}(D_X + D_S)^2.$$

It follows that

$$-\frac{1}{4}\|D_X - D_S\|^2 I \preceq D_{XS} \preceq \frac{1}{4}\|D_X + D_S\|^2 I.$$

Since $\mathbf{Tr}(D_X D_S) = 0$, the matrices $D_X + D_S$ and $D_X - D_S$ have the same norm. Consequently

$$-\frac{1}{4}\|D_X + D_S\|^2 I \preceq D_{XS} \preceq \frac{1}{4}\|D_X + D_S\|^2 I.$$

These inequalities and the definition of δ in (3.17) imply the lemma. □

Lemma 3.2.6 (Lemma 7.2 in [18])**.** *One has*

$$\lambda_{\min}\left(\left(V^+\right)^2\right) \geq 1 - \delta^2,$$

where $\lambda_{\min}$ *denotes the smallest eigenvalue.*

Proof. From (3.23) it follows that

$$\lambda_{\min}\left(\left(V^{+}\right)^{2}\right)=\lambda_{\min}\left(I+D_{XS}+H\right).$$

Since X^{+} and S^{+} are symmetric and positive semidefinite, Lemma A.3.3 implies that the eigenvalues of matrix $\left(V^{+}\right)^{2} \sim \frac{X^{+}S^{+}}{\mu^{+}}$ are real and nonnegative. So the equality above shows that $\lambda_{\min}\left(I+D_{XS}+H\right)$ is also real and nonnegative. Since matrix H is skew-symmetric and $\lambda_{\min}\left(I+D_{XS}+H\right)$ is real, Lemma A.3.1 implies

$$\lambda_{\min}\left(\left(V^{+}\right)^{2}\right)\geq\lambda_{\min}\left(I+D_{XS}\right)=1+\lambda_{\min}\left(D_{XS}\right)\geq 1-\left|\lambda_{\min}\left(D_{XS}\right)\right|.$$

Substitution of the bound for $\left|\lambda_{\min}\left(D_{XS}\right)\right|$ from Lemma 3.2.5 now yields:

$$\lambda_{\min}\left(\left(V^{+}\right)^{2}\right)\geq 1-\delta^{2}.$$

This proves the lemma. □

Lemma 3.2.7 (Lemma 7.3 in [18])**.** *One has*

$$\left\|D_{XS}\right\|^{2}\leq 2\delta^{4}.$$

Proof. Let $D_V := D_X+D_S$ and $Q_V := D_X-D_S$. Since $\mathbf{Tr}\left(D_X D_S\right)=0$, it follows that the matrices D_V and Q_V have the same norm. Moreover, since matrices D_V^2 and Q_V^2 are positive semidefinite, Lemma A.3.3 implies that the eigenvalues of the matrix $D_V^2 Q_V^2$ are nonnegative. So it follows that $\mathbf{Tr}\left(D_V^2 Q_V^2\right)\geq 0$. Now by using this and (3.25) we obtain

$$\begin{aligned}
\left\|D_{XS}\right\|^{2} &= \frac{1}{16}\left\|\left(D_X+D_S\right)-\left(D_X-D_S\right)\right\|^{2}\\
&= \frac{1}{16}\left\|D_V^2-Q_V^2\right\|^{2}\\
&= \frac{1}{16}\mathbf{Tr}\left(D_V^4+Q_V^4-2D_V^2Q_V^2\right)\\
&\leq \frac{1}{16}\left(\left\|D_V^2\right\|^{2}+\left\|Q_V^2\right\|^{2}\right)\\
&\leq \frac{1}{16}\left(\left\|D_V\right\|^{4}+\left\|Q_V\right\|^{4}\right)\\
&= \frac{1}{8}\left\|D_V\right\|^{4}=2\delta^{4},
\end{aligned}$$

where the last equality follows from the definition of δ in (3.17). This proves the lemma. □

Lemma 3.2.8 (Quadratic convergence, Lemma 7.4 in [18]). *If* $\delta := \delta(X, S; \mu) \leq 1$, *then the full NT step is feasible, i.e.,* X^+ *and* S^+ *are feasible. Moreover, if* $\delta < 1$, *then* X^+ *and* S^+ *are positive definite and*

$$\delta\left(X^+, S^+; \mu\right) \leq \frac{\delta^2}{\sqrt{2\left(1-\delta^2\right)}}.$$

Proof. Since $\delta \leq 1$, Lemma 3.2.5 implies that $|\lambda_i(D_{XS})| \leq 1$, which guarantees feasibility of the iterates (X^+, y^+, S^+) (see Corollary 3.2.4). This implies the first part of the lemma. For the proof of the second part of the lemma, by using (3.17) we may write

$$\begin{aligned}
\left(\delta^+\right)^2 &= \frac{1}{4}\left\|\left(V^+\right)^{-1} - V^+\right\|^2 \\
&= \frac{1}{4}\left\|\left(V^+\right)^{-1}\left(\mu I - \left(V^+\right)^2\right)\right\|^2 \\
&\leq \frac{1}{4}\lambda_{\max}\left(\left(V^+\right)^{-1}\right)^2\left\|I - \left(V^+\right)^2\right\|^2 = \frac{1}{4\,\lambda_{\min}\left(\left(V^+\right)^2\right)}\left\|I - \left(V^+\right)^2\right\|^2.
\end{aligned}$$

We now substitute the bound from Lemma 3.2.6 to obtain

$$\left(\delta^+\right)^2 \leq \frac{1}{4\left(1-\delta^2\right)}\left\|I - \left(V^+\right)^2\right\|^2. \tag{3.26}$$

Moreover, from (3.23) we have

$$\left(V^+\right)^2 \sim \frac{X^+S^+}{\mu} \sim I + D_{XS} + H.$$

Using (3.23) and definition of the Frobenius norm for symmetric matrix as defined in (A.6), we have

$$\begin{aligned}
\left\|I - \left(V^+\right)^2\right\|^2 &= \sum_{i=1}^{n}\left(\lambda_i\left(I + D_{XS} + H\right) - 1\right)^2 \\
&= \sum_{i=1}^{n}\lambda_i\left(D_{XS} + H\right)^2 \\
&= \mathbf{Tr}\left(\left(D_{XS} + H\right)^2\right).
\end{aligned}$$

Since the matrices $HD_{XS} + D_{XS}H$ and H are skew-symmetric and HH^T is positive semidefinite we obtain

$$\begin{aligned}
\mathbf{Tr}\left(\left(D_{XS} + H\right)^2\right) &= \mathbf{Tr}\left(D_{XS}^2 + HD_{XS} + D_{XS}H + H^2\right) \\
&= \mathbf{Tr}\left(D_{XS}^2 + H^2\right) = \mathbf{Tr}\left(D_{XS}^2 - HH^T\right) \\
&\leq \mathbf{Tr}\left(D_{XS}^2\right) = \|D_{XS}\|^2 \\
&\leq 2\delta^4,
\end{aligned}$$

where the last inequality follows from Lemma 3.2.7. Substitution in (3.26) gives

$$\left(\delta^+\right)^2 \leq \frac{\delta^4}{2\left(1-\delta^2\right)},$$

which implies the lemma. □

Corollary 3.2.9 (Corollary 7.2 in [18]). *If* $\delta\left(X,\, S;\, \mu\right) < \frac{1}{\sqrt{2}}$, *then* $\delta\left(X^+,\, S^+;\, \mu\right) < \delta\left(X,\, S;\, \mu\right)^2$.

This corollary implies that the Newton-process is quadratically convergent when started close to the μ-center.

Lemma 3.2.2 and Lemma 3.2.8 show that the behavior of the Newton-process and the feasibility condition for the new iterates $(X^+,\, y^+,\, S^+)$ are exactly the same as in the LO case. We therefore have the following result for SDO which establishes a polynomial iteration bound for the algorithm in Figure 3.1. The proof is exactly the same as the proof of Theorem 1.3.7 and is therefore omitted.

Theorem 3.2.10 (Theorem 7.1 in [18]). *If* $\tau = \frac{1}{\sqrt{2}}$ *and* $\theta = \frac{1}{\sqrt{2n}}$, *then the algorithm in Figure 3.1 requires at most*

$$\sqrt{2n}\log\frac{n\,\mu^0}{\epsilon}$$

iterations. The output is a primal-dual pair $(X,\, S)$ *such that* $\mathbf{Tr}\left(XS\right) \leq \epsilon$.

3.3 Infeasible full-Newton step IPM for SDO

In this section we present an extension of the IIPM for LO in Figure 2.2 to SDO. We follow the same outline as in the LO case. So we start by introducing perturbed primal-dual problem pairs and their central paths.

3.3.1 The perturbed problems for SDO

Let X^0 and y^0, S^0 be as in (3.1) and $\mu^0 = \zeta^2$. The initial values of the primal and dual residuals are denoted by r_b^0 and R_c^0, as defined in (3.4) and (3.5). So we have

$$\left(r_b^0\right)_i = b_i - \mathbf{Tr}\left(A_i X^0\right), \quad i = 1,\, \ldots,\, m, \tag{3.27}$$

$$R_c^0 = C - \sum_{i=1}^{m} y_i^0 A_i - S^0. \tag{3.28}$$

By these notations, for any ν with $0 < \nu \leq 1$ we consider the perturbed problem (SDP_ν), defined by

$$(SDP_\nu) \qquad \min\left\{\mathbf{Tr}\left(\left(C - \nu R_c^0\right)X\right) : \mathbf{Tr}\left(A_i X\right) = b_i - \nu\left(r_b^0\right)_i,\, X \succeq 0\right\},$$

and its dual problem (SDD_ν), which is given by

$$(SDD_\nu) \qquad \max\left\{\sum_{i=1}^{m}\left(b_i-\nu\left(r_b^0\right)_i\right)y_i : \sum_{i=1}^{m}y_iA_i+S=C-\nu R_c^0,\ S\succeq 0\right\}.$$

It is easy to see that if $\nu=1$, then $X=X^0$ and $(y,\,S)=\left(y^0,\,S^0\right)$ are strictly feasible for (SDP_ν) and (SDD_ν), respectively. We conclude that if $\nu=1$ then (SDP_ν) and (SDD_ν) satisfy the IPC.

Lemma 3.3.1. *Let the original problems, (SDP) and (SDD), be feasible. Then for each ν satisfying $0<\nu\leq 1$ the perturbed problems (SDP_ν) and (SDD_ν) satisfy the IPC.*

Proof. Suppose that (SDP) and (SDD) are feasible. Let $\bar{X}$ be feasible solution of (SDP) and $(\bar{y},\bar{S})$ a feasible solution of (SDD). Then $\mathbf{Tr}\left(A_i\bar{X}\right)=b_i\ i=1,\ldots,m$ and $\sum_{i=1}^{m}\bar{y}_iA_i+\bar{S}=C$, with $\bar{X}\succeq 0$ and $\bar{S}\succeq 0$. Now let $0<\nu\leq 1$, and consider

$$X=(1-\nu)\ \bar{X}+\nu\,X^0,\quad y=(1-\nu)\ \bar{y}+\nu\,y^0,\quad S=(1-\nu)\ \bar{S}+\nu\,S^0.$$

One has, for each $i=1,\ldots,m$,

$$\begin{aligned}\mathbf{Tr}\left(A_iX\right)&=\mathbf{Tr}\left(A_i\left((1-\nu)\ \bar{X}+\nu\,X^0\right)\right)\\&=(1-\nu)\,\mathbf{Tr}\left(A_i\bar{X}\right)+\nu\mathbf{Tr}\left(A_iX^0\right)\\&=b_i-\nu\left(b_i-\mathbf{Tr}\left(A_iX^0\right)\right),\end{aligned}$$

showing that X is feasible for (P_ν). Similarly,

$$\begin{aligned}\sum_{i=1}^{m}y_iA_i+S&=(1-\nu)\left(\sum_{i=1}^{m}\bar{y}_iA_i+\bar{S}\right)+\nu\left(\sum_{i=1}^{m}y_i^0A_i+S^0\right)\\&=(1-\nu)\,C+\nu\left(\sum_{i=1}^{m}y_i^0A_i+S^0\right)=C-\nu\left(C-\sum_{i=1}^{m}y_i^0A_i-S^0\right),\end{aligned}$$

showing that (y,S) is feasible for (SDD_ν). Since $\nu>0$, X and S are positive definite. This proves that (SDP_ν) and (SDD_ν) satisfy the IPC. □

3.3.2 Central path of the perturbed problems for SDO

Since we assumed that (SDP) and (SDD) are feasible, Lemma 3.3.1 implies that for any $0<\nu\leq 1$, the perturbed problems (SDP_ν) and (SDD_ν) satisfy the IPC, and hence their central paths exist. This means that the following system has a unique solution, for every $\mu>0$,

$$b_i-\mathbf{Tr}\left(A_iX\right)=\nu\left(r_b^0\right)_i,\qquad i=1,\,2,\,\ldots,\,m,\qquad X\succeq 0 \tag{3.29}$$

$$C-\sum_{i=1}^{m}y_iA_i-S=\nu R_c^0,\qquad S\succeq 0 \tag{3.30}$$

$$XS=\mu I. \tag{3.31}$$

We denote this unique solution as $(X(\mu,\,\nu),\, y(\mu,\,\nu),\, S(\mu,\,\nu))$. These are the μ-centers of the perturbed problems (SDP_ν) and (SDD_ν), respectively.

Since $X^0S^0 = \mu^0 I$, X^0 is the μ^0-center of the perturbed problem (SDP_1) and (y^0, S^0) the μ^0-center of (SDD_1). In the sequel we will always have

$$\mu = \nu\,\mu^0 = \nu\zeta^2,$$

and to simplify notations we denote $(X(\mu,\,\nu),\, y(\mu,\,\nu),\, S(\mu,\,\nu))$ shortly as $(X(\nu),\, y(\nu),\, S(\nu))$.

3.3.3 An iteration of our algorithm

In this subsection we describe one iteration of our algorithm. Each main iteration starts with a triple $(X,\, y,\, S)$ that is feasible for (SDP_ν) and (SDD_ν), such that

$$\delta\,(X,\, S;\, \mu) \leq \tau, \qquad \mathbf{Tr}\,(XS) = n\mu,$$

where $\mu = \nu\zeta^2$, $0 \leq \nu \leq 1$. Then we reduce μ to $\mu^+ = (1-\theta)\,\mu$, with $\theta \in (0,\,1)$.

Just as in the case of LO each main iteration of our algorithm consists of a feasibility step and a few centering steps. The feasibility step serves to get iterates $\left(X^f,\, y^f,\, S^f\right)$ that are strictly feasible for (SDP_{ν^+}) and (SDD_{ν^+}), where $\nu^+ = (1-\theta)\,\nu$, and such that $\delta\left(X^f,\, S^f;\, \mu^+\right) \leq \frac{1}{\sqrt{2}}$. Since the triple $\left(X^f,\, y^f,\, S^f\right)$ is strictly feasible for (SDP_{ν^+}) and (SDD_{ν^+}), we can perform a few centering steps starting at $\left(X^f,\, y^f,\, S^f\right)$ targeting at the μ^+-centers of (SDP_{ν^+}) and (SDD_{ν^+}) and obtain iterates $(X^+,\, y^+,\, S^+)$ that are still feasible for (SDP_{ν^+}) and (SDD_{ν^+}) such that

$$\delta\left(X^+,\, S^+;\, \mu^+\right) \leq \tau, \qquad \mathbf{Tr}\left(X^+S^+\right) = n\mu^+.$$

We now describe each main iteration in more detail. Due to the definition of the perturbed problem pair the feasibility conditions for (SDP_ν) and (SDD_ν) are

$$\mathbf{Tr}\,(A_iX) = b_i - \nu\left(r_b^0\right)_i, \quad i = 1,\,\ldots,\,m, \quad X \succeq 0,$$
$$\sum_{i=1}^{m} y_iA_i + S = C - \nu R_c^0, \quad S \succeq 0.$$

For finding iterates that are feasible for (SDP_{ν^+}) and (SDD_{ν^+}) we need search directions $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$ such that

$$\mathbf{Tr}\left(A_i\left(X + \Delta^f X\right)\right) = b_i - \nu^+\left(r_b^0\right)_i, \quad i = 1,\,\ldots,\,m,$$
$$\sum_{i=1}^{m}\left(y_i + \Delta^f y\right)A_i + \left(S + \Delta^f S\right) = C - \nu^+ R_c^0.$$

Since X and $(y,\, S)$ are feasible for (SDP_ν) and (SDD_ν) respectively, it follows that $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$ should satisfy

$$\mathbf{Tr}\left(A_i\Delta^f X\right) = (b_i - \mathbf{Tr}\left(A_i X\right)) - \nu^+\left(r_b^0\right)_i = \nu\left(r_b^0\right)_i - \nu^+\left(r_b^0\right)_i = \theta\nu\left(r_b^0\right)_i,$$

$$\sum_{i=1}^m \Delta^f y_i A_i + \Delta^f S = \left(C - \sum_{i=1}^m y_i A_i - S\right) - \nu^+ R_c^0 = \nu R_c^0 - \nu^+ R_c^0 = \theta\nu R_c^0.$$

Therefore, $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$ need to satisfy

$$\mathbf{Tr}\left(A_i\Delta^f X\right) = \theta\nu\left(r_b^0\right)_i, \quad i = 1, \ldots, m, \tag{3.32}$$

$$\sum_{i=1}^m \Delta^f y_i A_i + \Delta^f S = \theta\nu R_c^0. \tag{3.33}$$

We add to this system the equation

$$\Delta^f X + P\Delta^f S P^T = \mu S^{-1} - X, \tag{3.34}$$

where P is as defined in (3.10). The system (3.32)-(3.34) has a unique solution [100] and $\Delta^f X$ and $\Delta^f S$ are symmetric. After the feasibility step the iterates are given by

$$X^f = X + \Delta^f X, \tag{3.35}$$

$$y^f = y + \Delta^f y,$$

$$S^f = S + \Delta^f S. \tag{3.36}$$

By definition, after the feasibility step the iterates satisfy the affine equations in (3.29) and (3.30), with $\nu = \nu^+$. The main part in the analysis will be to guarantee that X^f and S^f are positive definite and satisfy $\delta\left(X^f, S^f; \mu^+\right) \leq \frac{1}{\sqrt{2}}$.
After the feasibility step, in order to get iterates (X^+, y^+, S^+) that satisfy

$$\mathbf{Tr}\left(X^+S^+\right) = n\mu^+, \quad \delta\left(X^+, S^+; \mu^+\right) \leq \tau,$$

we perform centering steps. Just as in the LO case, the required number of centering steps is at most

$$\log_2\left(\log_2\frac{1}{\tau^2}\right), \tag{3.37}$$

due to the quadratic behavior of the Newton process.

3.3.4 The algorithm

A more formal description of the algorithm is given in Figure 3.2. Like in LO, one may easily verify after each iteration the residuals and the duality gap are reduced by a factor $1-\theta$. The algorithm stops if the norms of residuals and the duality gap are less than the accuracy parameter ϵ.

Primal-Dual Infeasible IPM for SDO

Input:
Accuracy parameter $\epsilon > 0$;
barrier update parameter θ, $0 < \theta < 1$;
threshold parameter $\tau > 0$;
parameter $\zeta > 0$.

begin
$X^0 := \zeta I$, $S^0 := \zeta I$, $y^0 := 0$; $\mu^0 := \zeta^2$;
while $\max\left(\mathbf{Tr}\left(XS\right), \|r_b\|, \|R_c\|\right) \geq \epsilon$ **do**
begin
feasibility step:
$(X, y, S) := (X, y, S) + \left(\Delta^f X, \Delta^f y, \Delta^f S\right)$;
μ-update:
$\mu := (1 - \theta)\mu$;
centering steps:
while $\delta\left(X, S; \mu\right) \geq \tau$ **do**
begin
$(X, y, S) := (X, y, S) + (\Delta X, \Delta y, \Delta S)$;
end
end
end

Figure 3.2: *Infeasible full-Newton-step algorithm for SDO*

3.4 An analysis of the algorithm

As before, X, y and S denote the iterates at the start of an iteration and are such that $\mathbf{Tr}\left(XS\right) = n\mu$ and $\delta\left(X, S; \mu\right) \leq \tau$. Recall that we have this situation in the first iteration since $\mathbf{Tr}\left(X^0 S^0\right) = n\mu^0$ and $\delta\left(X^0, S^0; \mu^0\right) = 0$.

3.4.1 The effect of the feasibility step and the choice of θ

As we established in Section 3.3.3, in the feasibility step we obtain new iterates X^f, y^f and S^f that satisfy the affine equations for (SDP_{ν^+}) and (SDD_{ν^+}). The main part in the analysis is to show that X^f and S^f are positive semidefinite and $\delta\left(X^f, S^f; \mu^+\right) \leq \frac{1}{\sqrt{2}}$, i.e., that the new iterates are in the region where the Newton process targeting at the μ^+-centers of (SDP_{ν^+}) and (SDD_{ν^+}) is quadratically convergent.

We define

$$D_X^f := \frac{1}{\sqrt{\mu}} D^{-1}\Delta^f X D^{-1}, \quad D_S^f := \frac{1}{\sqrt{\mu}} D\Delta^f S D, \quad \left(V^f\right)^2 := \frac{1}{\mu} D^{-1} X^f S^f D, \tag{3.38}$$

with D as defined in Section 3.2.2. We can now rewrite (3.32)-(3.34) as follows.

$$\begin{aligned} \mathbf{Tr}\left(DA_i D\, D_X^f\right) &= \frac{1}{\sqrt{\mu}}\theta\nu\left(r_b^0\right)_i, \quad i = 1, \ldots, m, \\ \sum_{i=1}^{m} \frac{\Delta y_i}{\sqrt{\mu}} DA_i D + D_S^f &= \frac{1}{\sqrt{\mu}}\theta\nu D R_c^0 D, \\ D_X^f + D_S^f &= V^{-1} - V. \end{aligned} \tag{3.39}$$

From the third equation in (3.39) we obtain, by multiplying both sides from the left with V,

$$VD_X^f + VD_S^f = I - V^2. \tag{3.40}$$

Using (3.11), (3.35), (3.36) and (3.38), we obtain

$$\begin{aligned} X^f &= X + \Delta^f X = \sqrt{\mu} D\left(V + D_X^f\right) D, \\ S^f &= S + \Delta^f S = \sqrt{\mu} D^{-1}\left(V + D_S^f\right) D^{-1}. \end{aligned}$$

Therefore

$$X^f S^f = \mu D\left(V + D_X^f\right)\left(V + D_S^f\right) D^{-1}.$$

The last equality shows that the matrix $X^f S^f$ is similar to $\mu\left(V + D_X^f\right)\left(V + D_S^f\right)$. This means that we have

$$X^f S^f \sim \mu\left(V + D_X^f\right)\left(V + D_S^f\right). \tag{3.41}$$

To simplify the notation in the sequel we introduce

$$D_{XS}^f := \frac{1}{2}\left(D_X^f D_S^f + D_S^f D_X^f\right), \tag{3.42}$$

and

$$M := \left(D_X^f V - V D_X^f\right) + \frac{1}{2}\left(D_X^f D_S^f - D_S^f D_X^f\right). \tag{3.43}$$

Note that D_{XS}^f is symmetric and M is skew-symmetric.
Now we may write, using (3.40),

$$\begin{aligned} \left(V + D_X^f\right)\left(V + D_S^f\right) &= V^2 + VD_S^f + D_X^f V + D_X^f D_S^f \\ &= I - VD_X^f + D_X^f V + D_X^f D_S^f. \end{aligned}$$

By subtracting and adding $\frac{1}{2}D_S^f D_X^f$ to the last expression we get

$$I + \frac{1}{2}\left(D_X^f D_S^f + D_S^f D_X^f\right) + \left(D_X^f V - V D_X^f\right) + \frac{1}{2}\left(D_X^f D_S^f - D_S^f D_X^f\right).$$

Using (3.42) and (3.43) we obtain

$$X^f S^f \sim \mu\left(I + D_{XS}^f + M\right). \tag{3.44}$$

Lemma 3.4.1. *Let $X \succ 0$ and $S \succ 0$. Then the iterates $\left(X^f,\, y^f,\, S^f\right)$ are strictly feasible if*

$$I + D_{XS}^f \succ 0.$$

Proof. For the proof we introduce a step length $\alpha \in [0,\, 1]$, and we define

$$X^f(\alpha) = X + \alpha\Delta^f X, \quad y^f(\alpha) = y + \alpha\Delta^f y, \quad S^f(\alpha) = S + \alpha\Delta^f S.$$

We then have $X^f(0) = X$, $X^f(1) = X^f$ and similar relations for y and S. Obviously $\det\left(X^f(0)\, S^f(0)\right) = \det\left(\mu^0 I\right) = \left(\mu^0\right)^n > 0$. Our aim is to show that the determinant of $X^f(\alpha)\, S^f(\alpha)$ remains positive for all $\alpha \leq 1$. For proving this we continue in exactly the same way as in the proof of the Lemma 3.2.3. By using similar arguments as there, we conclude that under the assumption in the lemma the (symmetric) matrix

$$(1-\alpha)\, V^2 + \alpha(1-\alpha)\, I + \alpha^2\left(D_{XS}^f + I\right)$$

is positive definite for all $\alpha \leq 1$. This implies the lemma. □

Corollary 3.4.2. *The iterates $\left(X^f,\, y^f,\, S^f\right)$ are certainly strictly feasible if*

$$\left|\lambda_i\left(D_{XS}^f\right)\right| < 1, \quad \textit{for} \quad i = 1,\, \ldots,\, n.$$

Proof. The proof is similar to the proof of Corollary 3.2.4, and is therefore omitted. □

In the sequel we denote

$$\omega(V) := \frac{1}{2}\sqrt{\left\|D_X^f\right\|^2 + \left\|D_S^f\right\|^2}, \tag{3.45}$$

which implies $\left\|D_X^f\right\| \leq 2\omega(V)$ and $\left\|D_S^f\right\| \leq 2\omega(V)$.

By using (A.6), (A.7) and (3.45) we have

$$\left\|D_{XS}^f\right\| \leq \left\|D_X^f\right\| \left\|D_S^f\right\| \leq \frac{1}{2}\left(\left\|D_X^f\right\|^2 + \left\|D_S^f\right\|^2\right) = 2\omega(V)^2 \tag{3.46}$$

$$\left|\lambda_i\left(D_{XS}^f\right)\right| \leq \left\|D_{XS}^f\right\| \leq 2\omega(V)^2, \qquad i = 1,\, \ldots,\, n. \tag{3.47}$$

Lemma 3.4.3. *If* $\omega(V) < \frac{1}{\sqrt{2}}$ *then the iterates* $\left(X^f,\, y^f,\, S^f\right)$ *are strictly feasible.*

Proof. Let $\omega(V) < \frac{1}{\sqrt{2}}$. Then (3.47) implies that $\left|\lambda_i\left(D_{XS}^f\right)\right| < 1$, for $i = 1, \ldots, n$. By Corollary 3.4.2 this implies the statement in the lemma. □

Assuming $\omega(V) < \frac{1}{\sqrt{2}}$, which guarantees strict feasibility of the iterates $(X^f,\, y^f,\, S^f)$, we proceed by deriving an upper bound for $\delta\left(X^f,\, S^f;\, \mu^+\right)$. Recall from definition (3.17) that

$$\delta\left(X^f,\, S^f;\, \mu^+\right) = \frac{1}{2}\left\|\left(V^f\right)^{-1} - V^f\right\|, \tag{3.48}$$

with $\left(V^f\right)^2$ as defined in (3.38). In the sequel we denote $\delta\left(X^f,\, S^f;\, \mu^+\right)$ also shortly as $\delta\left(V^f\right)$. We need some technical results which give information on the eigenvalues and the norm of V^f.

Lemma 3.4.4. *One has, for each* $i = 1, \ldots, n$,

$$\lambda_i\left(\left(V^f\right)^2\right) \geq \frac{1}{1-\theta}\left(1 - 2\omega(V)^2\right).$$

Proof. Using (3.44), after division of both sides by $\mu^+ = (1-\theta)\,\mu$ we get

$$\left(V^f\right)^2 \sim \frac{\mu\left(I + D_{XS}^f + M\right)}{\mu^+} = \frac{I + D_{XS}^f + M}{1-\theta}. \tag{3.49}$$

It follows that

$$\lambda_i\left(\left(V^f\right)^2\right) = \frac{1}{1-\theta}\lambda_i\left(I + D_{XS}^f + M\right).$$

Since M is skew-symmetric, Lemma A.3.1 implies

$$\lambda_i\left(\left(V^f\right)^2\right) \geq \frac{1}{1-\theta}\lambda_i\left(I + D_{XS}^f\right) = \frac{1}{1-\theta}\left(1 + \lambda_i\left(D_{XS}^f\right)\right).$$

Substitution of the bound for $\left|\lambda_i\left(D_{XS}^f\right)\right|$ in (3.47) yields

$$\lambda_i\left(\left(V^f\right)^2\right) \geq \frac{1}{1-\theta}\left(1 - 2\omega(V)^2\right),$$

which completes the proof. □

Lemma 3.4.5. *One has*

$$\left\|I - \left(V^f\right)^2\right\| \leq \frac{2\omega(V)^2 + \theta\sqrt{n}}{1-\theta}.$$

Proof. Using (3.49) and properties of the Frobenius norm we have

$$\begin{aligned}\left\|I-\left(V^f\right)^2\right\|^2 &= \sum_{i=1}^{n}\left(\frac{\lambda_i\left(I+D_{XS}^f+M\right)}{1-\theta}-1\right)^2\\ &= \frac{1}{(1-\theta)^2}\sum_{i=1}^{n}\left(\lambda_i\left(I+D_{XS}^f+M\right)-1+\theta\right)^2\\ &= \frac{1}{(1-\theta)^2}\sum_{i=1}^{n}\left(\lambda_i\left(D_{XS}^f+M\right)+\theta\right)^2.\end{aligned}$$

The right hand-side of the last equality can be rewritten as

$$\frac{1}{(1-\theta)^2}\left(n\theta^2+\sum_{i=1}^{n}\left(\lambda_i\left(D_{XS}^f+M\right)\right)^2+2\theta\sum_{i=1}^{n}\lambda_i\left(D_{XS}^f+M\right)\right).$$

Since $\left(\lambda_i\left(D_{XS}^f+M\right)\right)^2=\lambda_i\left(\left(D_{XS}^f+M\right)^2\right)$, for each i, we obtain

$$\left\|I-\left(V^f\right)^2\right\|^2=\frac{1}{(1-\theta)^2}\left(n\theta^2+\mathbf{Tr}\left(\left(D_{XS}^f+M\right)^2\right)+2\theta\mathbf{Tr}\left(D_{XS}^f+M\right)\right). \tag{3.50}$$

Using the skew-symmetry of M we obtain $\mathbf{Tr}\left(D_{XS}^f+M\right)=\mathbf{Tr}\left(D_{XS}^f\right)$ and

$$\mathbf{Tr}\left(\left(D_{XS}^f+M\right)^2\right)=\mathbf{Tr}\left(\left(D_{XS}^f\right)^2+MD_{XS}^f+D_{XS}^fM-MM^T\right).$$

Since $MD_{XS}^f+D_{XS}^fM$ is skew-symmetric we obtain

$$\mathbf{Tr}\left(\left(D_{XS}^f+M\right)^2\right)=\mathbf{Tr}\left(\left(D_{XS}^f\right)^2-MM^T\right)\leq\mathbf{Tr}\left(\left(D_{XS}^f\right)^2\right)=\left\|D_{XS}^f\right\|^2,$$

where the inequality follows since the matrix MM^T is positive semidefinite. Substitution in (3.50) gives

$$\left\|I-\left(V^f\right)^2\right\|^2\leq\frac{1}{(1-\theta)^2}\left(n\theta^2+\left\|D_{XS}^f\right\|^2+2\theta\,\mathbf{Tr}\left(D_{XS}^f\right)\right).$$

Now let $\lambda\left(D_{XS}^f\right)$ be the vector consisting of the eigenvalues of D_{XS}^f. Using the Cauchy-Schwartz inequality and (3.46) we get

$$\begin{aligned}\mathbf{Tr}\left(D_{XS}^f\right)&=\sum_{i=1}^{n}\lambda_i\left(D_{XS}^f\right)=e^T\lambda\left(D_{XS}^f\right)\\ &\leq\|e\|\left\|\lambda\left(D_{XS}^f\right)\right\|=\|e\|\left\|D_{XS}^f\right\|\leq 2\sqrt{n}\,\omega\left(V\right)^2.\end{aligned}$$

Substitution gives, also using (3.47),

$$\left\|I-\left(V^f\right)^2\right\|^2 \leq \frac{1}{(1-\theta)^2}\left(n\theta^2+4\omega\left(V\right)^4+4\theta\sqrt{n}\,\omega\left(V\right)^2\right)$$
$$=\left(\frac{2\omega\left(V\right)^2+\theta\sqrt{n}}{1-\theta}\right)^2,$$

which implies the lemma. □

Lemma 3.4.6. *Let* $\omega\left(V\right)<\frac{1}{\sqrt{2}}$. *Then one has*

$$2\delta\left(V^f\right)\leq\frac{2\,\omega\left(V\right)^2+\theta\,\sqrt{n}}{\sqrt{(1-\theta)\left(1-2\omega\left(V\right)^2\right)}}.$$

Proof. We may write, using (3.48),

$$2\delta\left(V^f\right)=\left\|V^f-\left(V^f\right)^{-1}\right\|=\left\|\left(V^f\right)^{-1}\left(I-\left(V^f\right)^2\right)\right\|$$
$$\leq\lambda_{\max}\left(\left(V^f\right)^{-1}\right)\left\|I-\left(V^f\right)^2\right\|=\frac{1}{\lambda_{\min}\left(\left(V^f\right)\right)}\left\|I-\left(V^f\right)^2\right\|.$$

Using the bounds in Lemma 3.4.4 and Lemma 3.4.5 the lemma follows. □

Recall from Section 3.3.3 that we need to have $\delta\left(V^f\right)\leq\frac{1}{\sqrt{2}}$. By Lemma 3.4.6 it suffices for this that

$$\frac{2\,\omega\left(V\right)^2+\theta\,\sqrt{n}}{\sqrt{(1-\theta)\left(1-2\omega\left(V\right)^2\right)}}\leq\sqrt{2}. \tag{3.51}$$

Lemma 3.4.7. *Let* $\omega\left(V\right)\leq\frac{1}{2}$ *and*

$$\theta=\frac{\alpha}{2\left(\sqrt{n}+1\right)},\quad 0\leq\alpha\leq 1. \tag{3.52}$$

Then the iterates $\left(X^f,\,y^f,\,S^f\right)$ *are strictly feasible and* $\delta\left(V^f\right)\leq\frac{1}{\sqrt{2}}$.

Proof. Due to Lemma 3.4.3 and $\omega\left(V\right)\leq\frac{1}{2}$, the iterates $\left(X^f,\,y^f,\,S^f\right)$ are strictly feasible. We just established that if inequality (3.51) is satisfied then $\delta\left(V^f\right)\leq\frac{1}{\sqrt{2}}$. The left hand side in this inequality is monotonically increasing in $\omega\left(V\right)$. By substituting $\omega\left(V\right)=\frac{1}{2}$, the inequality (3.51) reduces to

$$\frac{\frac{1}{2}+\theta\,\sqrt{n}}{\sqrt{\frac{1}{2}\left(1-\theta\right)}}\leq\sqrt{2},$$

which is equivalent to

$$4n\theta^2 + 4\left(\sqrt{n}+1\right)\theta - 3 \leq 0.$$

The largest possible value of θ satisfing this inequality is given by

$$\begin{aligned}\theta &= \frac{3}{2\left(\sqrt{n}+1+\sqrt{\left(\sqrt{n}+1\right)^2+3n}\right)} \\ &\geq \frac{3}{2\left(\sqrt{n}+1+\sqrt{\left(\sqrt{n}+1\right)^2+3\left(\sqrt{n}+1\right)^2}\right)} = \frac{1}{2\left(\sqrt{n}+1\right)},\end{aligned}$$

which is in agreement with (3.52). Thus the lemma has been proved. □

3.4.2 An upper bound for $\omega(V)$

As became clear in (3.39), the system (3.32)-(3.34), which defines the search directions $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$, can be expressed in terms of the scaled search directions D_X^f and D_S^f. We define the linear space $\mathcal{N}$ as follows:

$$\mathcal{N} := \left\{\Gamma \in \mathbf{S}^n \,:\, \mathbf{Tr}\left(DA_iD\,\Gamma\right) = 0, \quad i = 1, \ldots, m\right\}.$$

Using the linear space $\mathcal{N}$, it is clear from the first equation in (3.39) that the affine space

$$\left\{\Gamma \in \mathbf{S}^n \,:\, \mathbf{Tr}\left(DA_iD\,\Gamma\right) = \frac{1}{\sqrt{\mu}}\theta\nu\left(r_b^0\right)_i,\ i = 1, \ldots, m\right\}$$

equals $D_X^f + \mathcal{N}$. From linear algebra we also know that the affine space

$$\left\{\sum_{i=1}^m \vartheta_i DA_iD \;:\; \vartheta \in \mathbf{R}^m\right\}.$$

is the orthogonal complement of $\mathcal{N}$, denoted by $\mathcal{N}^\perp$. By the second equation in system (3.39), we conclude that the affine space

$$\left\{\frac{1}{\sqrt{\mu}}\theta\nu DR_c^0D + \sum_{i=1}^m \vartheta_i DA_iD \;:\; \vartheta \in \mathbf{R}^m\right\}$$

equals $D_S^f + \mathcal{N}^\perp$. Since $\mathcal{N} \cap \mathcal{N}^\perp = \{0\}$, the spaces $D_X^f + \mathcal{N}$ and $D_S^f + \mathcal{N}^\perp$ meet in a unique matrix. This matrix is denoted below by Q.

Lemma 3.4.8. *Let Q be the (unique) matrix in the intersection of the affine spaces $D_X^f + \mathcal{N}$ and $D_S^f + \mathcal{N}^\perp$. Then*

$$2\omega(V) \leq \sqrt{\|Q\|^2 + \left(\|Q\| + 2\delta(V)\right)^2}.$$

Proof. The proof resembles the proof of Lemma 4.6 in [82] for the LO case. To simplify notation in this proof we denote $R = V^{-1} - V$. Since $\mathcal{N} + \mathcal{N}^{\perp} = \mathbf{S}^n$, there exist Q_1, $R_1 \in \mathcal{N}$ and Q_2, $R_2 \in \mathcal{N}^{\perp}$ such that

$$Q = Q_1 + Q_2, \qquad R = R_1 + R_2.$$

On the other hand, since $D_X^f - Q \in \mathcal{N}$ and $D_S^f - Q \in \mathcal{N}^{\perp}$ there must exist $L_1 \in \mathcal{N}$ and $L_2 \in \mathcal{N}^{\perp}$ such that

$$D_X^f = Q + L_1, \qquad D_S^f = Q + L_2.$$

Due to the third equation in system (3.39) it follows that $R = 2Q + L_1 + L_2$, which implies

$$(2Q_1 + L_1) + (2Q_2 + L_2) = R_1 + R_2.$$

Since the decomposition $\mathcal{N} + \mathcal{N}^{\perp} = \mathbf{S}^n$ is unique, we conclude that

$$L_1 = R_1 - 2Q_1, \qquad L_2 = R_2 - 2Q_2.$$

Hence we obtain

$$\begin{aligned} D_X^f &= Q + R_1 - 2Q_1 = (R_1 - Q_1) + Q_2, \\ D_S^f &= Q + R_2 - 2Q_2 = Q_1 + (R_2 - Q_2). \end{aligned}$$

Since the spaces $\mathcal{N}$ and $\mathcal{N}^{\perp}$ are orthogonal we conclude from this that

$$\begin{aligned} 4\omega(V)^2 &= \left\| D_X^f \right\|^2 + \left\| D_S^f \right\|^2 \\ &= \|R_1 - Q_1\|^2 + \|Q_2\|^2 + \|Q_1\|^2 + \|R_2 - Q_2\|^2 \\ &= \|Q - R\|^2 + \|Q\|^2. \end{aligned}$$

Assuming $Q \neq 0$, since $\|R\| = 2\delta(V)$ the right-hand side is maximal if $R = \frac{-2\delta(V)}{\|Q\|} Q$, and thus we obtain

$$4\omega(V)^2 \leq \left\| \left(1 + \frac{2\delta(V)}{\|Q\|} \right) Q \right\|^2 + \|Q\|^2 = \|Q\|^2 + (\|Q\| + 2\delta(V))^2,$$

which implies the inequality in the lemma if $Q \neq 0$. Since the inequality in the lemma holds with equality if $Q = 0$, this completes the proof. □

From Lemma 3.4.7 we know that we want to have $\omega(V) \leq \frac{1}{2}$ because then $\delta\left(V^f\right) \leq \frac{1}{\sqrt{2}}$. Due to Lemma 3.4.8 this will hold if $\|Q\|$ satisfies

$$\|Q\|^2 + (\|Q\| + 2\delta(V))^2 \leq 1. \tag{3.53}$$

3.4.3 An upper bound for $\|Q\|$

Recall from Lemma 3.4.8 that Q is the unique solution of the system

$$\begin{aligned} \mathbf{Tr}\,(DA_iDQ) &= \tfrac{1}{\sqrt{\mu}}\theta\nu\left(r_b^0\right)_i, \quad i = 1, \ldots, m, \\ \textstyle\sum_{i=1}^m \tfrac{\eta_i}{\sqrt{\mu}} DA_iD + Q &= \tfrac{1}{\sqrt{\mu}}\theta\nu DR_c^0 D. \end{aligned} \tag{3.54}$$

We proceed to finding an upper bound for $\|Q\|$. Before doing this we choose the initial iterates $\left(X^0,\, y^0,\, S^0\right)$ as follows:

$$X^0 = S^0 = \zeta I, \quad y^0 = 0, \quad \mu^0 = \zeta^2, \tag{3.55}$$

where $\zeta > 0$ is such that

$$X^* + S^* \preceq \zeta\, I, \tag{3.56}$$

for some $(X^*,\, y^*,\, S^*) \in \mathcal{F}^*$. For the moment, let us write

$$(r_b)_i = \theta\nu\left(r_b^0\right)_i, \quad i = 1,\, 2,\, \ldots,\, m, \quad R_c = \theta\nu R_c^0,$$

and let r_b be the vector $((r_b)_1\,;\,(r_b)_2\,;\,\ldots;\,(r_b)_m)$.

Below we use the Kronecker product of two matrices and the **vec** operator on matrices. Their properties are given in Appendix A. The system (3.54) can be rewritten as follows:

$$\begin{aligned} \mathbf{vec}\,(A_i)^T\,(D \otimes D)\,\mathbf{vec}\,(Q) &= \tfrac{1}{\sqrt{\mu}}\,(r_b)_i, \quad i = 1, \ldots, m, \\ \textstyle\sum_{i=1}^m \tfrac{\eta_i}{\sqrt{\mu}}\,(D \otimes D)\,\mathbf{vec}\,(A_i) + \mathbf{vec}\,(Q) &= \tfrac{1}{\sqrt{\mu}}\,(D \otimes D)\,\mathbf{vec}\,(R_c)\,. \end{aligned} \tag{3.57}$$

Let $\mathcal{A}^T = [\mathbf{vec}\,(A_1) \quad \mathbf{vec}\,(A_2) \quad \ldots \quad \mathbf{vec}\,(A_m)]$ and $\eta = (\eta_1;\ \eta_2;\ \ldots;\ \eta_m)$. One may easily verify that we can rewrite the system (3.57) as follows:

$$\begin{aligned} \mathcal{A}\,(D \otimes D)\,\mathbf{vec}\,(Q) &= \tfrac{1}{\sqrt{\mu}} r_b, \\ (D \otimes D)\,\mathcal{A}^T \tfrac{\eta}{\sqrt{\mu}} + \mathbf{vec}\,(Q) &= \tfrac{1}{\sqrt{\mu}}\,(D \otimes D)\,\mathbf{vec}\,(R_c)\,. \end{aligned} \tag{3.58}$$

Lemma 3.4.9. *With $\left(X^0,\, y^0,\, S^0\right)$ as defined in (3.55) and (3.56), we have*

$$\|Q\| \le \theta\sqrt{\nu\,\mathbf{Tr}\,(P^2 + P^{-2})}, \tag{3.59}$$

where P is as defined in (3.10).

Proof. Replacing $\mathcal{A}$, $D \otimes D$ and $\mathbf{vec}\,(Q)$ in system (3.58) by A, D and q, respectively, yields exactly the same system as in the proof of Lemma 2.3.5 in Chapter 2. By using similar arguments as there, we obtain the following result:

$$\sqrt{\mu}\,\|\mathbf{vec}\,(Q)\| \le \theta\nu\sqrt{\left\|\mathbf{vec}\left(D\left(S^0 - \bar{S}\right)D\right)\right\|^2 + \left\|\mathbf{vec}\left(D^{-1}\left(X^0 - \bar{X}\right)D^{-1}\right)\right\|^2},$$

where $\bar{X}$, $\bar{y}$ and $\bar{S}$ are arbitrary solutions of the system

$$\begin{aligned} \mathcal{A}\mathbf{vec}\left(\bar{X}\right) &= b, \\ \mathcal{A}^T\bar{y} + \mathbf{vec}\left(\bar{S}\right) &= \mathbf{vec}\left(C\right). \end{aligned} \tag{3.60}$$

Using that $\|\mathbf{vec}\left(U\right)\| = \|U\|$ holds for any matrix U, we obtain

$$\sqrt{\mu}\,\|Q\| \leq \theta\nu\sqrt{\left\|\left(D\left(S^0 - \bar{S}\right)D\right)\right\|^2 + \left\|D^{-1}\left(X^0 - \bar{X}\right)D^{-1}\right\|^2}. \tag{3.61}$$

We are still free to choose $\bar{X}$ and $\bar{S}$, such that (3.60) is satisfied. We use $\bar{X} = X^*$ and $\bar{S} = S^*$, with X^* and S^* as in (3.56). Then we have

$$0 \preceq X^0 - \bar{X} = X^0 - X^* \preceq \zeta I, \quad 0 \preceq S^0 - \bar{S} = S^0 - S^* \preceq \zeta I.$$

It follows that

$$\left\|D\left(S^0 - \bar{S}\right)D\right\|^2 \leq \zeta^2\left\|D^2\right\|^2 = \zeta^2\|P\|^2 = \zeta^2\mathbf{Tr}\left(P^2\right).$$

where we used Lemma A.3.2 and $D = P^{\frac{1}{2}}$. In the same way it follows that

$$\left\|D^{-1}\left(X^0 - \bar{X}\right)D^{-1}\right\|^2 \leq \zeta^2\mathbf{Tr}\left(P^{-2}\right).$$

Substitution the last inequalities and $\mu = \nu\mu^0 = \nu\zeta^2$ into (3.61) gives

$$\|Q\| \leq \theta\sqrt{\nu\,\mathbf{Tr}\left(P^2 + P^{-2}\right)},$$

proving the lemma. □

Lemma 3.4.10. *With $\left(X^0, y^0, S^0\right)$ as defined in (3.55) and (3.56), we have*

$$\|Q\| \leq \frac{\theta}{\zeta\,\lambda_{\min}\left(V\right)}\mathbf{Tr}\left(X + S\right). \tag{3.62}$$

Proof. Using (3.10) and Lemma A.3.2 we have

$$\begin{aligned} \mathbf{Tr}\left(P^2\right) &= \mathbf{Tr}\left(X^{\frac{1}{2}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{-1}{2}}X\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{-1}{2}}X^{\frac{1}{2}}\right) \\ &= \mathbf{Tr}\left(X\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{-1}{2}}X\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{-1}{2}}\right) \\ &\leq \frac{1}{\lambda_{\min}\left(\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)}\mathbf{Tr}\left(X^2\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{-1}{2}}\right) \\ &\leq \frac{\mathbf{Tr}\left(X^2\right)}{\lambda_{\min}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)} = \frac{\mathbf{Tr}\left(X^2\right)}{\mu\lambda_{\min}\left(V^2\right)}, \end{aligned}$$

where for the last equality we used $V^2 \sim \frac{X^{\frac{1}{2}} S X^{\frac{1}{2}}}{\mu} \sim \frac{XS}{\mu}$. In the same way we get

$$\mathbf{Tr}\left(P^{-2}\right) \leq \frac{\mathbf{Tr}\left(S^2\right)}{\mu \lambda_{\min}\left(V^2\right)}.$$

Thus we obtain

$$\sqrt{\mathbf{Tr}\left(P^2 + P^{-2}\right)} \leq \frac{1}{\lambda_{\min}\left(V\right)} \sqrt{\frac{\mathbf{Tr}\left(X^2 + S^2\right)}{\mu}}. \tag{3.63}$$

Moreover, by the positive definiteness of X and S and Lemma A.3.3 it follows that

$$\mathbf{Tr}\left(X^2 + S^2\right) \leq \mathbf{Tr}\left(X^2 + S^2 + XS + SX\right) = \mathbf{Tr}\left(\left(X+S\right)^2\right) \leq \mathbf{Tr}\left(X+S\right)^2. \tag{3.64}$$

Substituting (3.63) and (3.64) in (3.59) gives

$$\|Q\| \leq \frac{\theta}{\lambda_{\min}\left(V\right)} \sqrt{\frac{\nu}{\mu} \mathbf{Tr}\left(X+S\right)^2}.$$

Since $\mu = \nu\mu^0 = \nu\zeta^2$, the lemma follows. □

3.4.4 Bounds for $\mathbf{Tr}\left(X+S\right)$ and $\lambda_{\min}\left(V\right)$. The choice of τ and α

Let X be feasible for (SDP_ν) and (y, S) for (SDD_ν), we need to find an upper bound for $\mathbf{Tr}\left(X+S\right)$ and lower bound on the eigenvalues of V as defined in (3.12). We can rewrite $\delta\left(V\right)$ in (3.17) as follows:

$$\begin{aligned}
4\delta\left(V\right)^2 &= \left\|V - V^{-1}\right\|^2 \\
&= \mathbf{Tr}\left(\left(V - V^{-1}\right)^T \left(V - V^{-1}\right)\right) \\
&= \mathbf{Tr}\left(V^2 - 2I + V^{-2}\right) \\
&= \sum_{i=1}^{n} \left(\lambda_i\left(V\right)^2 - 2 + \frac{1}{\lambda_i\left(V\right)^2}\right) \\
&= \sum_{i=1}^{n} \left(\lambda_i\left(V\right) - \frac{1}{\lambda_i\left(V\right)}\right)^2.
\end{aligned} \tag{3.65}$$

Using this we may state the following result.

Lemma 3.4.11 (Cf. Lemma II.60 in [84]). *Let $\delta = \delta\left(V\right)$ be given by (3.65). Then*

$$\frac{1}{\rho\left(\delta\right)} \leq \lambda_i\left(V\right) \leq \rho\left(\delta\right), \tag{3.66}$$

where

$$\rho(\delta) := \delta + \sqrt{1+\delta^2}. \tag{3.67}$$

Proof. The proof is exactly the same as the proof given in Lemma II.60 in [84]. We will repeat it here to make our presentation self contained. Since $\lambda_i(V)$ is positive for each $i = 1, \ldots, n$, we know by (3.65) that

$$-2\delta\lambda_i(V) \leq 1 - \lambda_i(V)^2 \leq 2\delta\lambda_i(V).$$

This implies that

$$\lambda_i(V)^2 - 2\delta\lambda_i(V) - 1 \leq 0 \leq \lambda_i(V)^2 + 2\delta\lambda_i(V) - 1.$$

Rewriting this as

$$(\lambda_i(V) - \delta)^2 - 1 - \delta^2 \leq 0 \leq (\lambda_i(V) + \delta)^2 - 1 - \delta^2$$

we obtain

$$(\lambda_i(V) - \delta)^2 \leq 1 + \delta^2 \leq (\lambda_i(V) + \delta)^2,$$

which implies

$$\lambda_i(V) - \delta \leq |\lambda_i(V) - \delta| \leq \sqrt{1+\delta^2} \leq \lambda_i(V) + \delta.$$

Thus we arrive at

$$-\delta + \sqrt{1+\delta^2} \leq \lambda_i(V) \leq \delta + \sqrt{1+\delta^2} = \rho(\delta).$$

For the left-hand expression we write

$$-\delta + \sqrt{1+\delta^2} = \frac{1}{\delta + \sqrt{1+\delta^2}} = \frac{1}{\rho(\delta)}.$$

This proves the theorem. □

Lemma 3.4.12. *Let X and (y, S) be feasible for the perturbed problems (SDP_ν) and (SDD_ν) respectively and $\left(X^0, y^0, S^0\right)$ as defined in (3.55). Then for any $(X^*, y^*, S^*) \in \mathcal{F}^*$, we have*

$$\begin{aligned}\nu\,\mathbf{Tr}\left(S^0X + X^0S\right) &= \mathbf{Tr}(SX) + \nu^2\,\mathbf{Tr}\left(S^0X^0\right)\\ &\quad +\nu(1-\nu)\,\mathbf{Tr}\left(S^0X^* + X^0S^*\right) - (1-\nu)\,\mathbf{Tr}\left(SX^* + S^*X\right).\end{aligned}$$

Proof. Let

$$\begin{aligned}X' &= X - \nu X^0 - (1-\nu)X^*,\\ y' &= y - \nu y^0 - (1-\nu)y^*,\\ S' &= S - \nu S^0 - (1-\nu)S^*.\end{aligned}$$

From (3.27), (3.28) and the definitions of the perturbed problems (SDP_ν) and (SDD_ν), it is easily seen that (X', y', S') satisfies

$$\begin{aligned} \mathbf{Tr}\,(A_i X') &= 0, \quad i = 1, \ldots, m, \\ \sum_{i=1}^{m} y'_i A_i + S' &= 0, \end{aligned}$$

which implies $\mathbf{Tr}\,(X'S') = 0$. i.e.,

$$\mathbf{Tr}\left(\left(X - \nu X^0 - (1-\nu)\,X^*\right)\left(S - \nu S^0 - (1-\nu)\,S^*\right)\right) = 0.$$

By expanding the last equality and using the fact that $\mathbf{Tr}\,(X^*S^*) = 0$ we obtain the desired result. □

Lemma 3.4.13. *Let X and (y, S) be feasible for the perturbed problems (SDP_ν) and (SDD_ν) respectively and $\delta(V)$ as defined in (3.65) and $X^0 = S^0 = \zeta I$, where $\zeta > 0$ is a constant such that $X^* + S^* \preceq \zeta I$ for some $(X^*, y^*, S^*) \in \mathcal{F}^*$. Then we have*

$$\mathbf{Tr}\,(X + S) \leq \left(\rho\,(\delta)^2 + 1\right) n\zeta, \tag{3.68}$$

where $\rho(\delta)$ as defined in (3.67).

Proof. Since X, S, X^* and S^* are positive semidefinite, Lemma A.3.3 implies that $\mathbf{Tr}\,(SX^*)$ and $\mathbf{Tr}\,(XS^*)$ are nonnegative. Therefore Lemma 3.4.12 implies

$$\mathbf{Tr}\left(S^0X + X^0S\right) \leq \frac{\mathbf{Tr}\,(SX)}{\nu} + \nu\,\mathbf{Tr}\left(S^0X^0\right) + (1-\nu)\,\mathbf{Tr}\left(S^0X^* + X^0S^*\right).$$

Since $X^0 = S^0 = \zeta I$ and $X^* + S^* \preceq \zeta I$, we have

$$\mathbf{Tr}\left(S^0X^* + X^0S^*\right) = \zeta\,\mathbf{Tr}\,(X^* + S^*) \leq \zeta^2\,\mathbf{Tr}\,(I) = n\zeta^2.$$

Also $\mathbf{Tr}\left(X^0S^0\right) = n\zeta^2$. Hence we get

$$\mathbf{Tr}\left(S^0X + X^0S\right) \leq \frac{\mathbf{Tr}\,(SX)}{\nu} + n\zeta^2 = \frac{\mu\mathbf{Tr}\left(V^2\right)}{\nu} + n\zeta^2 = \zeta^2\mathbf{Tr}\left(V^2\right) + n\zeta^2,$$

where the last equality follows because of $\nu = \frac{\mu}{\mu^0}$ and $\mu^0 = \zeta^2$. Using Lemma 3.4.11 we obtain

$$\mathbf{Tr}\left(S^0X + X^0S\right) \leq \left(\rho\,(\delta)^2 + 1\right) n\zeta^2$$

Since $X^0 = S^0 = \zeta I$ we have

$$\mathbf{Tr}\left(S^0X + X^0S\right) = \zeta\mathbf{Tr}\,(X + S).$$

Hence it follows that

$$\mathbf{Tr}\,(X + S) \leq \left(\rho\,(\delta)^2 + 1\right) n\zeta,$$

which proves the lemma. □

By substituting (3.66) and (3.68) into (3.62) we get

$$\|Q\| \leq n\theta\rho\left(\delta\right)\left(1+\rho\left(\delta\right)^2\right).$$

At this stage we choose

$$\tau = \frac{1}{8}. \tag{3.69}$$

Since $\delta \leq \tau = \frac{1}{8}$ and $\rho\left(\delta\right)$ is monotonically increasing in δ, we have

$$\|Q\| \leq n\theta\rho\left(\delta\right)\left(1+\rho\left(\delta\right)^2\right) \leq n\theta\rho\left(\frac{1}{8}\right)\left(1+\rho\left(\frac{1}{8}\right)^2\right) = 2.586\, n\, \theta.$$

By using $\theta = \frac{\alpha}{2\left(\sqrt{n}+1\right)}$ (see Lemma 3.4.7) we obtain the following upper bound for the norm of Q:

$$\|Q\| \leq \frac{2.586\, n\, \alpha}{2\left(\sqrt{n}+1\right)}.$$

In (3.53) we found that in order to have $\delta\left(V^f\right) \leq \frac{1}{\sqrt{2}}$, we should have $\|Q\|^2 + \left(\|Q\| + 2\delta\left(V\right)\right)^2 \leq 1$. Therefore, since $\delta\left(V\right) \leq \tau = \frac{1}{8}$, it suffices if Q satisfies $\|Q\|^2 + \left(\|Q\| + \frac{1}{4}\right)^2 \leq 1$. So we have $\delta\left(V^f\right) \leq \frac{1}{\sqrt{2}}$ if $\|Q\| \leq 0.57097$. Since $\|Q\| \leq \frac{2.586\, n\, \alpha}{2\left(\sqrt{n}+1\right)}$, the latter inequality is satisfied if we take

$$\alpha = \frac{2\left(\sqrt{n}+1\right)}{5\, n}, \tag{3.70}$$

because

$$\frac{1.14194}{2.586} = 0.442 \geq \frac{2}{5}.$$

3.5 Complexity

As already showed for LO in Section 2.4 of Chapter 2, with τ as defined in (3.69) according to (3.37), we need at most

$$\log_2\left(\log_2\frac{1}{\tau^2}\right) = \log_2\left(\log_2 64\right) \leq 3$$

centering steps to get iterates that satisfy $\delta(X, S; \mu^+) \leq \tau$. So each iteration consists of one feasibility step and at most 3 centering steps. In each iteration both the duality gap and the norms of the residual vectors are reduced by the

factor $1 - \theta$. Hence, using $\mathbf{Tr}\left(X^0 S^0\right) = n\zeta^2$, the total number of iterations is bounded above by

$$\frac{1}{\theta} \log \frac{\max\left\{n\zeta^2, \left\|r_b^0\right\|, \left\|R_c^0\right\|\right\}}{\epsilon}.$$

Due to (3.52) and (3.70) we have

$$\theta = \frac{\alpha}{2\left(\sqrt{n}+1\right)} = \frac{1}{5\,n}.$$

Hence the total number of inner iterations is bounded above by

$$20\,n \log \frac{\max\left\{n\zeta^2, \left\|r_b^0\right\|, \left\|R_c^0\right\|\right\}}{\epsilon}.$$

Note that the order of this bound is the same as the bound in (1.29) for LO. We may state without further proof our main result.

Theorem 3.5.1. *If (SDP) and (SDD) have optimal solutions $(X^*, y^*, S^*) \in \mathcal{F}^*$ such that $X^* + S^* \preceq \zeta\, I$, then after at most*

$$20\,n \log \frac{\max\left\{n\zeta^2, \left\|r_b^0\right\|, \left\|R_c^0\right\|\right\}}{\epsilon}.$$

iterations the algorithm finds an ϵ-solution of (SDP) and (SDD).

The above theorem gives a convergence result under the assumption (3.56). One might ask what if this condition is not satisfied. From Lemma 3.4.10 under conditions (3.55) and (3.56) we know that during the course of the algorithm the matrix Q satisfies $\|Q\| \leq 0.57097$. So, if during the execution of the algorithm $\|Q\| > 0.57097$, then we may conclude that there exist no optimal solutions (X^*, y^*, S^*) with zero duality gap such that

$$X^* + S^* \preceq \zeta\, I.$$

Chapter 4

Full-Newton step primal-dual IPMs for LO based on locally SCB functions

4.1 Introduction

As discussed in Chapter 1 after the seminal work of Karmarkar [41], many researchers have proposed and analyzed various IPMs for LO and a large amount of results have been reported. IPMs are among the most effective methods for solving wide classes of linear and nonlinear optimization problems. At present there is still a gap between the practical behavior of the algorithms and the theoretical performance results, in favor of the practical behavior. This is especially true for so-called large-update methods, which are the most efficient methods in practice (see, e.g. Andersen et al. [5]). Until very recently, the theoretical iteration bound for all large-update methods was $O(n \log \frac{n}{\epsilon})$, where n denotes the number of inequalities in the problem, and ϵ the desired accuracy in terms of the objective value. The best known iteration bound has been derived for so-called small-update IPMs and is given by $O(\sqrt{n} \log \frac{n}{\epsilon})$; small-update IPMs, however, are very inefficient in practice. This phenomenon is known as the *gap between theory and practice*, also called the *irony of IPMs* [81, page 51]. It is summarized in the following table:

	Large-update methods	Small-update methods
Iteration bound	$O(n) \log \frac{n}{\epsilon}$	$O(\sqrt{n}) \log \frac{n}{\epsilon}$
Practical performance	Highly efficient	Very poor

Recently some progress has been made by introducing a wide class of new barrier functions that are defined by univariate functions, so-called *kernel functions*. As a result the 'gap factor' was reduced from $\sqrt{n}$ to $\log n$. See [6–9, 74, 76]. The goal of these papers was to investigate large-update methods, and to improve their iteration bounds.

A surprising result in these works is that the iteration bounds for small-update methods based on these barrier functions always turned out to be $\mathcal{O}(\sqrt{n}\log\frac{n}{\epsilon})$ (see, e.g., [23]), which is the best known bound for IPMs for LO. It is the aim of this chapter to explain this surprising phenomenon by using the powerful results from the theory of *self-concordant functions*. The theory of self-concordant (SC) functions was developed by Nesterov and Nemirovski [64] and is generally considered to be a milestone in the theory of IPMs for linear and convex optimization. As we will see this theory does not directly apply to the new barrier functions, since these functions are not SC. The main purpose of this chapter is to show that the new barrier functions are 'locally SC' in a neighborhood of the central path. As we show this observation explains why under very weak conditions on the kernel function all small-update methods have the iteration bound $O(\sqrt{n}\log\frac{n}{\epsilon})$.

4.2 Self-concordancy

We start this section with a definition and some examples of self-concordant functions.

4.2.1 Definition and examples

We start by considering the case where ϕ is a univariate function. We assume for the moment that $n = 1$, and the domain $\mathcal{D}$ of $\phi : \mathcal{D} \to \mathbf{R}$ is just an open interval in $\mathbf{R}$. One calls ϕ a κ-SC function if κ is a nonnegative number and

$$|\phi'''(x)| \leq 2\kappa\left(\phi''(x)\right)^{\frac{3}{2}}, \qquad \text{for all} \qquad x \in \mathcal{D}. \tag{4.1}$$

Note that this definition implies that $\phi''(x)$ is nonnegative, whence ϕ is convex, and that ϕ is three times continuously differentiable. Obviously, linear and convex quadratic functions are 0-SC, because their third derivatives are zero. We proceed by presenting a well-known example of SC function on $\mathbf{R}$.

Example 4.2.1. *Consider the function*

$$f(x) = -\log x, \qquad \mathcal{D} = \mathbf{R}_+.$$

Then

$$f'(x) = \frac{-1}{x}, \quad f''(x) = \frac{1}{x^2}, \quad f'''(x) = \frac{-2}{x^3}.$$

Therefore,

$$\frac{|f'''(x)|}{2f''(x)^{\frac{3}{2}}} = \frac{2/x^3}{2\left(1/x^2\right)^{3/2}} = 1.$$

This implies that $f(x)$ is a 1-SC function. □

Now suppose that $n > 1$, so ϕ is a multivariate function. Then ϕ is called a κ-SC function if its restriction to an arbitrary line in its domain is κ-SC. In other words, ϕ is a κ-SC if and only if $\bar{\phi}(t) = \phi(x + th)$ is κ-SC for all $x \in \mathcal{D}$ and for all $h \in \mathbf{R}^n$. The domain of $\bar{\phi}(t)$ is defined in the natural way: given x and h it consists of all t such that $x + th \in \mathcal{D}$. For notational convenience and as in [11, 15, 63, 64] we define:

$$\begin{aligned}
\bar{\phi}'(t) &= \nabla\phi(x)[h] = \langle \nabla\phi(x), h\rangle,\\
\bar{\phi}''(t) &= \nabla^2\phi(x)[h, h] = \langle \nabla^2\phi(x)\, h, h\rangle,\\
\bar{\phi}'''(t) &= \nabla^3\phi(x)[h, h, h] = \langle \nabla^3\phi(x)[h]h, h\rangle.
\end{aligned}$$

By these notations we can give the following definition for SC-functions on $\mathbf{R}^n$.

Definition 4.2.2. *We call a function κ-SC if the inequality*

$$\left|\nabla^3\phi(x)[h, h, h]\right| \le 2\kappa\left(\nabla^2\phi(x)[h, h]\right)^{\frac{3}{2}}, \tag{4.2}$$

holds for any $x \in \mathcal{D}$ and $h \in \mathbf{R}^n$.

Example 4.2.3. *Consider the function*

$$f(x) = -\sum_{i=1}^{n} \log x_i,$$

with $0 < x \in \mathbf{R}^n$. Then

$$\begin{aligned}
\nabla f(x) &= \left(\frac{-1}{x_1};\ \frac{-1}{x_2};\ \ldots;\ \frac{-1}{x_n}\right),\\
\nabla^2 f(x) &= \operatorname{diag}\left(\frac{1}{x_1^2};\ \frac{1}{x_2^2};\ \ldots;\ \frac{1}{x_n^2}\right),\\
\nabla^3 f(x) &= \operatorname{diag}\left(\frac{-2}{x_1^3};\ \frac{-2}{x_2^3};\ \ldots;\ \frac{-2}{x_n^3}\right).
\end{aligned}$$

Therefore we have for any $h \in \mathbf{R}^n$

$$\nabla^2 f(x)[h, h] = \sum_{i=1}^{n} \frac{h_i^2}{x_i^2}, \qquad \nabla^3 f(x)[h, h, h] = -\sum_{i=1}^{n} \frac{2h_i^3}{x_i^3}.$$

For any $\xi \in \mathbf{R}^n$ we have

$$\left|\sum_{i=1}^{n} \xi_i^3\right| \le \sum_{i=1}^{n} |\xi_i|^3 \le \left(\sum_{i=1}^{n} \xi_i^2\right)^{\frac{3}{2}},$$

where the second inequality comes from the well-known relation $\|\cdot\|_3 \leq \|\cdot\|_2$ *applied to vector* $\xi \in \mathbf{R}^n$. *Hence, taking* $\xi_i := \frac{h_i}{x_i}$ *we get*

$$\left|\nabla^3 f(x)[h, h, h]\right| \leq 2\left(\nabla^2 f(x)[h, h]\right)^{\frac{3}{2}},$$

proving that f *is 1-SC.* □

4.2.2 Newton step and proximity measure

We assume henceforth that ϕ is a κ-SC function and $\mathcal{D} \subseteq \mathbf{R}^n$. The Newton step at x is defined by

$$\Delta x = -H(x)^{-1} g(x), \tag{4.3}$$

where $g(x)$ and $H(x)$ denote the gradient and the Hessian of $\phi(x)$ at x, respectively. So,

$$g(x) = \nabla\phi(x), \quad H(x) = \nabla^2\phi(x), \quad \forall x \in \mathcal{D}.$$

An important feature is that if $\phi(x)$ is strictly convex then $H(x)$ is positive definite everywhere on its domain; [1] $\phi(x)$ is certainly strictly convex if it is coercive and its domain does not contain a (complete) straight line (see Theorem 4.3.7 in [63]). In the sequel we always assume that ϕ is strictly convex. As a consequence, the quantity

$$\lambda(x) = \sqrt{\Delta x^T H(x) \Delta x} = \|\Delta x\|_{H(x)} = \sqrt{g(x)^T H(x)^{-1} g(x)},$$

i.e., the length of the Newton step with respect to the norm induced by the (local) Hessian matrix $H(x)$, can be used as a measure for the 'distance' of x to the minimizer of $\phi(x)$ (if it exists). This quantity plays a crucial role in the analysis of Newton's method for minimizing $\phi(x)$.

In the sequel we consider only barrier functions that have a minimizer, and for which the minimal value of ϕ on its domain $\mathcal{D}$ equals 0. Moreover, we have to deal only with points x that are close to the minimizer, in the sense that $\lambda(x)$ is small. For such a point the quadratic convergence of Newton's method to the minimizer is very nicely expressed by the following lemma.

Lemma 4.2.4 (Theorem 2.2.2 in [69])**.** *If* $\lambda(x) < \frac{1}{\kappa}$ *then* $x + \Delta x$ *is feasible. Moreover,*

$$\lambda(x + \Delta x) \leq \kappa \left(\frac{\lambda(x)}{1 - \kappa\lambda(x)}\right)^2.$$

Corollary 4.2.5. *If* $\lambda(x) \leq \frac{1}{3\kappa}$ *then* $x + \Delta x$ *is feasible and* $\lambda(x + \Delta x) \leq \left(\frac{3}{2}\lambda(x)\sqrt{\kappa}\right)^2$.

[1] This is not obvious! E.g., x^4 ($x \in \mathbf{R}$) is strictly convex, but its second derivative vanishes at $x = 0$.

One other result that we need estimates the value of ϕ at $x \in \mathcal{D}$ in terms of $\lambda(x)$. It makes use of the univariate (nonnegative) function $\omega(t)$ defined by

$$\omega(t) = t - \log(1+t), \quad t > -1. \tag{4.4}$$

We recall the following result.

Lemma 4.2.6 (Section 9.6.3 in [15]). *Let ϕ be κ-self-concordant on $\mathcal{D}$, $x \in \mathcal{D}$ and $\lambda = \lambda(x) < \frac{1}{\kappa}$. Then*

$$\phi(x) \leq -\frac{\kappa\lambda(x) + \ln(1-\kappa\lambda(x))}{\kappa^2} = \frac{\omega(-\kappa\lambda)}{\kappa^2}. \tag{4.5}$$

Hence, since $\omega(t)$ is monotonically decreasing if $t \in (-1, 0]$, we obtain

$$\lambda(x) \leq \frac{1}{4\kappa} \quad \Rightarrow \quad \phi(x) \leq \frac{\omega(-\frac{1}{4})}{\kappa^2} = \frac{0.0376821}{\kappa^2} \leq \frac{1}{26\kappa^2}. \tag{4.6}$$

4.2.3 Minimization of a linear function over a convex domain

We now consider the problem of minimizing a linear function over the closure $\bar{\mathcal{D}}$ of the domain $\mathcal{D}$ of a κ-SC function $\phi : \mathcal{D} \to \mathbf{R}$. So $\mathcal{D} = \operatorname{int}\bar{\mathcal{D}}$, and the given problem has the form

$$(P) \qquad \min\left\{c^T x \ : \ x \in \bar{\mathcal{D}}\right\}.$$

We assume that $H(x) = \nabla^2\phi(x)$ is positive definite for every $x \in \mathcal{D}$. For $\mu > 0$ we define

$$\phi_\mu(x) := \frac{c^T x}{\mu} + \phi(x), \quad x \in \mathcal{D}$$

and we consider the problem (P_μ) defined by

$$(P_\mu) \qquad \inf\left\{\phi_\mu(x) \ : \ x \in \mathcal{D}\right\}.$$

We now have

$$\begin{aligned} g_\mu(x) := \nabla\phi_\mu(x) &= \frac{c}{\mu} + \nabla\phi(x) = \frac{c}{\mu} + g(x), \\ H_\mu(x) := \nabla^2\phi_\mu(x) &= \nabla^2\phi(x) = H(x), \\ \nabla^3\phi_\mu(x) &= \nabla^3\phi(x). \end{aligned}$$

Note that the two higher derivatives do not depend on μ. It follows that $\phi_\mu(x)$ is also κ-SC. The minimizer of $\phi_\mu(x)$, if it exists, is denoted as $x(\mu)$. When μ runs through all positive numbers then $x(\mu)$ runs through the *central path* of (P). When μ approaches 0 then $x(\mu)$ converges to an optimal solution of (P). Therefore, in IPMs the central path is used as a guideline to the set of optimal

solutions of (P). This approach is likely to be feasible because since $\phi_\mu(x)$ is self-concordant, its minimizer can be computed efficiently.

The Newton step at $x \in \mathcal{D}$ with respect to $\phi_\mu(x)$ is given by

$$\Delta x = -H(x)^{-1} g_\mu(x)$$

and the distance of x to the μ-*center* $x(\mu)$ is measured by the quantity

$$\lambda_\mu(x) = \sqrt{\Delta x^T H(x) \Delta x} = \sqrt{g_\mu(x)^T H(x)^{-1} g_\mu(x)} = \|g_\mu(x)\|_{H(x)^{-1}}.$$

A major question is what the effect is on $\lambda_\mu(x)$ when μ is reduced to $\mu^+ = (1-\theta)\mu$. Let $\lambda = \lambda_\mu(x)$ and $\mu^+ = (1-\theta)\mu$. We then have

$$\begin{aligned} g_{\mu^+}(x) &= \frac{c}{\mu^+} + \nabla\phi(x) = \frac{c}{(1-\theta)\mu} + \nabla\phi(x) \\ &= \frac{1}{1-\theta}\left(\frac{c}{\mu} + \nabla\phi(x) - \theta\nabla\phi(x)\right) = \frac{1}{1-\theta}\left(g_\mu(x) - \theta\nabla\phi(x)\right). \end{aligned}$$

Hence, denoting $H(x)$ shortly as H,

$$\begin{aligned} \lambda_{\mu^+}(x) &= \tfrac{1}{1-\theta}\|g_\mu(x) - \theta\nabla\phi(x)\|_{H^{-1}} \\ &\leq \tfrac{1}{1-\theta}\left(\|g_\mu(x)\|_{H^{-1}} + \theta\|g(x)\|_{H^{-1}}\right) \\ &= \tfrac{1}{1-\theta}\left(\lambda_\mu(x) + \theta\|g(x)\|_{H^{-1}}\right) \\ &= \tfrac{1}{1-\theta}\left(\lambda_\mu(x) + \theta\lambda(x)\right). \end{aligned} \tag{4.7}$$

Definition 4.2.7. *Let $\nu \geq 0$. The self-concordant function ϕ is called a (κ,ν)-self-concordant barrier (SCB) function if ϕ is κ-SC and*

$$\lambda(x)^2 \leq \nu, \quad \forall x \in \mathcal{D}. \tag{4.8}$$

An immediate consequence of (4.7) and this definition is the following lemma.

Lemma 4.2.8. *If ϕ is a (κ, ν)-SCB then*

$$\lambda_{\mu^+}(x) \leq \frac{\lambda_\mu(x) + \theta\sqrt{\nu}}{1-\theta}.$$

For future use we recall the following result.

Lemma 4.2.9 (Theorem 2.2 in [25])**.** *If ϕ is κ-SC then ϕ is a (κ, ν)-SCB if and only if*

$$\left(\nabla\phi(x)[h]\right)^2 \leq \nu\nabla^2\phi(x)[h,h], \quad \forall x \in \mathcal{D}, \forall h \in \mathbf{R}^n. \tag{4.9}$$

The algorithm used in this chapter can now be presented, see Figure 4.1 (see [25]). For the purpose of this chapter the following convergence result is of utmost importance. We include part of the proof, as given in [25].

Primal-Dual IPM with Full-Newton step

Input:

Accuracy parameter $\epsilon > 0$;

proximity parameter $\tau > 0$;

update parameter θ, $0 < \theta < 1$;

$x = x^0 \in \mathcal{D}$ and $\mu = \mu^0 > 0$ such that $\lambda_\mu(x) \leq \tau < \frac{1}{\kappa}$.

begin

while $\mu\left(\nu + \frac{\tau(\tau+\sqrt{\nu})}{1-\kappa\tau}\right) \geq \epsilon$ **do**

begin

$\mu := (1-\theta)\mu$;

$x := x + \Delta x$;

end

end

Figure 4.1: *Full-Newton step algorithm.*

Theorem 4.2.10. *If* $\tau = \frac{1}{9\kappa}$ *and* $\theta = \frac{5}{9+36\kappa\sqrt{\nu}}$, *then the algorithm with full Newton steps requires not more than*

$$\left\lceil 2\left(1 + 4\kappa\sqrt{\nu}\right) \ln \frac{2\mu^0\nu}{\epsilon} \right\rceil$$

iterations. The output is a point $x \in \mathcal{D}$ *such that* $c^T x \leq c^T x^* + \epsilon$, *where* x^* *denotes an optimal solution of* (P).

Proof. At the start of the first iteration we have $x \in \mathcal{D}$ and $\mu = \mu^0$ such that $\lambda_\mu(x) \leq \tau$. When the barrier parameter is updated to $\mu^+ = (1-\theta)\mu$, Lemma 4.2.8 gives

$$\lambda_{\mu^+}(x) \leq \frac{\lambda_\mu(x) + \theta\sqrt{\nu}}{1-\theta} \leq \frac{\tau + \theta\sqrt{\nu}}{1-\theta}. \tag{4.10}$$

Then after the Newton step, the new iterate is $x^+ = x + \Delta x$ and, by Lemma 4.2.4,

$$\lambda_{\mu^+}(x^+) \leq \kappa \left(\frac{\lambda_{\mu^+}(x)}{1 - \kappa\lambda_{\mu^+}(x)}\right)^2. \tag{4.11}$$

The algorithm is well defined if we choose τ and θ such that $\lambda_{\mu^+}(x^+) \leq \tau$. To get the lowest iteration bound, we need at the same time to maximize θ. From (4.11) we deduce that $\lambda_{\mu^+}(x^+) \leq \tau$ certainly holds if

$$\frac{\lambda_{\mu^+}(x)}{1-\kappa\lambda_{\mu^+}(x)} \leq \frac{\sqrt{\tau}}{\sqrt{\kappa}},$$

which is equivalent to

$$\lambda_{\mu^+}(x) \leq \frac{\sqrt{\tau}}{\kappa\sqrt{\tau}+\sqrt{\kappa}}.$$

According to (4.10) this will hold if $\frac{\tau+\theta\sqrt{\nu}}{1-\theta} \leq \frac{\sqrt{\tau}}{\kappa\sqrt{\tau}+\sqrt{\kappa}}$. This leads to the following condition on θ:

$$\theta \leq \sqrt{\tau}\frac{1-\kappa\tau-\sqrt{\kappa\tau}}{\sqrt{\tau}+\sqrt{\nu\kappa}\,(1+\sqrt{\kappa\tau})}$$

Since $\tau = \frac{1}{9\kappa}$, this upper bound for θ gets the value $\frac{5}{9+36\kappa\sqrt{\nu}} \leq \frac{1}{2+8\kappa\sqrt{\nu}}$. This justifies the choice of the value of τ and θ in the theorem. The rest of the proof is not relevant for the purpose of this chapter, and is therefore omitted. We refer the interested reader to the relevant references, e.g., [25]. □

Note that the order of magnitude of the above iteration bound is dominated by the quantity $\kappa\sqrt{\nu}$. Following [25], we call this number the *complexity number* of ϕ.

It follows easily from the above proof that after the μ-update we have

$$\lambda_{\mu^+}(x) \leq \frac{1}{4\kappa}. \tag{4.12}$$

Since the Newton step decreases the proximity value, we may conclude that the following holds.

Lemma 4.2.11. *During the course of the algorithm, the iterates x always satisfy*

$$\lambda\,(x) \leq \frac{1}{4\kappa}.$$

The above analysis is based on the fact that the logarithmic barrier function is SC in interior of $\mathbf{R}^n_+$. After all it turns out that the iterates occur only in a small part of interior of $\mathbf{R}^n_+$, namely in a narrow neighborhood of the central path. Therefore it suffices to know the behavior of the barrier function in this narrow neighborhood, instead of the whole interior of $\mathbf{R}^n_+$. In the next subsection we introduce the concept of locally self-concordance.

4.2.4 Locally self-concordant barrier functions

Let $\mathcal{D}$ be an open convex subset of $\mathbf{R}^n$ and $\phi : \mathcal{D} \to \mathbf{R}$ three times continuously differentiable.

Definition 4.2.12. *The local κ-value at $x \in \mathcal{D}$ is defined by*

$$\kappa(x) := \min\left\{\kappa : \nabla^3\phi(x)[h,h,h] \leq 2\kappa\left(\nabla^2\phi(x)[h,h]\right)^{\frac{3}{2}}, \quad \forall h \in \mathbf{R}^n\right\} \tag{4.13}$$

and the local ν-value by

$$\nu(x) := \min\left\{\nu : \left(\nabla\phi(x)[h]\right)^2 \leq \nu\nabla^2\phi(x)[h,h], \quad \forall h \in \mathbf{R}^n\right\}. \tag{4.14}$$

Clearly ϕ is SC if $\kappa(x)$ is bounded above by some (finite) constant on the domain of ϕ. Furthermore, ϕ is a barrier function if moreover $\nu(x)$ is bounded above by some (finite) constant on the domain.

We define the local complexity number at x as

$$\gamma(x) = \kappa(x)\sqrt{\nu(x)}.$$

Definition 4.2.13. *Let $\mathcal{S}$ be a nonempty subset of $\mathcal{D}$. Then we say that ϕ is locally κ-SC on $\mathcal{S}$ if*

$$\kappa(x) \leq \kappa, \quad \forall x \in \mathcal{S}. \tag{4.15}$$

If ϕ is locally κ-SC on $\mathcal{S}$ then we call ϕ a local (κ,ν)-barrier on $\mathcal{S}$ if moreover

$$\nu(x) \leq \nu, \quad \forall x \in \mathcal{S}. \tag{4.16}$$

4.2.5 Composition rules

In Subsection 4.2.3 it has been made clear that we can find an ϵ-solution of (P) in polynomial time if we have a self-concordant barrier function for the interior of the domain. But how do we recognize, or obtain such functions. For this purpose we have some so-called composition rules.

First we recall that self-concordancy is preserved under scaling by a positive factor. This follows from the following lemma.

Lemma 4.2.14. *Let ϕ be κ-SC ν barrier (or shortly (κ,ν)-SCB) and $\lambda \in \mathbf{R}$, $\lambda > 0$. Then $\lambda\phi$ is a $(\frac{\kappa}{\sqrt{\lambda}}, \lambda\nu)$-SCB.*

Self-concordancy is also preserved under addition.

Lemma 4.2.15. *Let ϕ_i be (κ_i,ν_i)-SCB's on $\mathcal{D}_i$, for $i = 1,2$. Then $\phi_1 + \phi_2$ is a (κ,ν)-SCB for $\mathcal{D}_1 \cap \mathcal{D}_2$, where $\kappa = \max\{\kappa_1,\kappa_2\}$ and $\nu = \nu_1 + \nu_2$.*

4.3 Linear optimization based on kernel functions

It is well known that every linear optimization problem can be solved efficiently (i.e., in polynomial time) if we can find in polynomial time an optimal solution of problems of the form

$$(SP) \qquad \min\{q^T x : Mx \geq -q, x \geq 0\},$$

where the matrix M is skew-symmetric (i.e., $M^T = -M$) and the vector q is nonnegative (entry-wise) and nonzero. This can be achieved by embedding a given linear optimization problem and its dual problem in a problem of the above form. This embedding technique is due to [106] and described in detail in [84, Part I]. The problem (SP) is trivial in the sense that it has a trivial optimal solution, namely $x = 0$ which yields the objective value 0; this is optimal because $x \geq 0$ and $q \geq 0$ imply $q^T x \geq 0$ for every feasible solution. But this observation is not sufficient for our goal, since we need a strictly complementary solution of (SP). What this means requires some explanation.

We associate to any vector $x \in \mathbf{R}^n$ its slack vector $s(x)$ according to

$$s(x) = Mx + q.$$

In the sequel we simply denote $s(x)$ as s. Observe that since M is skew-symmetric we have $z^T Mz = 0$ for every vector $z \in \mathbf{R}^n$, where n denotes the order of M. Hence we have

$$q^T x = (s - Mx)^T x = s^T x + x^T Mx = s^T x.$$

Therefore, if x is feasible, then x is optimal if and only if $s^T x = 0$. Since x and s are nonnegative this holds if and only if $x_i s_i = 0$ for each i. This shows that x is optimal if and only if the vectors x and s are complementary vectors. As discussed in Section 1.2.1 in Chapter 1 we say that x is a strictly complementary solution if moreover $x_i + s_i > 0$ for each i. Summarizing these facts, we have that x is feasible if $x \geq 0$ and $s \geq 0$. A feasible x is optimal if $xs = 0$, and x is a strictly complementary solution if moreover $x + s > 0$.

4.3.1 The central path of (SP)

In this subsection we study the central path of the problem (SP). We already introduced central path in Section 4.2.3 as the set of minimizers of a barrier function.

Using the vector $s = s(x)$, the problem (SP) can be written as

$$\min\{q^T x : s - Mx = q,\ x \geq 0,\ s \geq 0\}.$$

So the feasible region is the intersection of the affine space $\{(x, s) \ : \ s - Mx = q\}$ with the cone $\left\{(x, s) \ : \ x \in \mathbf{R}^n_+,\ s \in \mathbf{R}^n_+\right\} = \mathbf{R}^n_+ \times \mathbf{R}^n_+$. Example 4.2.3 and Lemma

4.2.15 imply that

$$-\sum_{i=1}^{n}\log x_i-\sum_{i=1}^{n}\log s_i$$

is a SC function on the interior of the cone $\mathbf{R}^n_+\times\mathbf{R}^n_+$. According to Subsection 4.2.3 the function

$$\Phi_{LB}(x,s,\mu)=\frac{q^Tx}{\mu}-\sum_{i=1}^{n}\log x_i-\sum_{i=1}^{n}\log s_i \tag{4.17}$$

is a SC barrier function for (SP). This is the so-called logarithmic barrier function. The μ-center of (SP) is obtained by minimizing this function subject to the affine constraint $s-Mx=q$. The μ-center is (uniquely!) determined by the equations [84]

$$\begin{aligned} s&=Mx+q,\quad x\geq 0,\ s\geq 0,\\ xs&=\mu e. \end{aligned} \tag{4.18}$$

Clearly, any solution (x,s) of (4.18) will satisfy $x>0$ and $s>0$. So a solution exists only if (SP) satisfies the IPC, i.e., if the affine space $\{(x,s)\ :\ s-Mx=q\}$ has nonempty intersection with the interior of the cone $\mathbf{R}^n_+\times\mathbf{R}^n_+$. Surprisingly enough, if the IPC is satisfied then a solution of (4.18) exists, for each $\mu>0$, and this solution is unique since the logarithmic barrier function is strictly convex. We denote it as $x(\mu)$ and call $x(\mu)$ the μ-center of (SP); $s(\mu)$ is the corresponding slack vector. The set of μ-centers (with μ running through all positive real numbers) gives a homotopy path, which is called *the central path* of (SP).[2] If $\mu\to 0$ then the limit of the central path exists and since the limit point satisfies the complementarity condition, the limit yields an optimal solution for (SP). Moreover, this solution can be shown to be strictly complementary.

Without loss of generality we may assume that (SP) satisfies the IPC. In fact we may, and will assume that $x=e$ is feasible for (SP) and $s(x)=e$. So, we can start the algorithm with $x=e$, $s=e$ and $\mu=1$. For this and other properties mentioned above we refer to [84].

In this section we discussed the logarithmic barrier function for (SP). As discussed in Section 4.1, small-update methods theoretically have the best iteration bound, whereas large-update methods are in practice much more efficient than small-update methods. As we mentioned in Section 4.1 the iteration bound for large-update methods was recently improved by using a class of barrier functions. We continue this chapter by introducing this class of barrier functions.

[2]Note that the word 'central path' here has precisely the same meaning as in Section 4.2.

4.3.2 Kernel-function-based barrier functions

Kernel function of the logarithmic barrier function

For any x that is strictly feasible for (SP) we have $x > 0$ and also $s(x) > 0$. We define the vector v as in (1.7):

$$v := \sqrt{\frac{xs}{\mu}}, \qquad s = s\,(x)\,. \tag{4.19}$$

Note that if x is the μ-center $x(\mu)$ then $v = e$ and vice versa. The logarithmic barrier function, given by (4.17) can be nicely expressed in terms of v, by using that $q^T x = s^T x = \mu \sum_{i=1}^n v_i^2$:

$$\begin{aligned}
\Phi_{LB}(x, s, \mu) &= \frac{q^T x}{\mu} - \sum_{i=1}^n \log x_i s_i \\
&= \sum_{i=1}^n v_i^2 - \sum_{i=1}^n \log\left(\mu v_i^2\right) \\
&= \sum_{i=1}^n \left(v_i^2 - 1 - \log v_i^2\right) + n - n \log \mu \\
&= 2 \sum_{i=1}^n \psi_{LB}(v_i) + n - n \log \mu,
\end{aligned} \tag{4.20}$$

where ψ_{LB} denotes the univariate function defined by

$$\psi_{LB}(t) = \frac{1}{2}\left(t^2 - 1\right) - \log t, \quad t > 0. \tag{4.21}$$

Note that the above term $n - n \log \mu$ disappears when taking (partial) derivatives to x and s. So, when μ is fixed, which is the case when we apply Newton's method in the algorithm of Figure 4.1, this term is not relevant. We call the univariate function ψ_{LB} the *kernel function* of the logarithmic barrier function. It is clear from (4.20) that the logarithmic barrier function is based on the kernel function ψ_{LB}. As mentioned in Section 4.1 this idea has recently been used to define a new class of barrier functions based on other kernel functions. In the next subsection we will discuss this phenomenon in more detail.

Barrier functions based on other kernel functions

In the kernel-function-based approach we let $\psi(t)$ be any univariate function $\psi : (0, \infty) \to [0, \infty)$ that has the following properties in common with ψ_{LB}: $\psi(t)$ must be strictly convex, minimal at $t = 1$, and $\psi(1) = 0$. Any such function is called a *kernel function*, and gives rise to a barrier function $\phi_\mu(x, s)$ in the following way:

$$\phi_\mu(x, s) := 2\Psi(v) := 2 \sum_{i=1}^n \psi\,(v_i)\,, \quad x > 0, s > 0. \tag{4.22}$$

So the domain of ϕ is the interior of $\mathbf{R}^n_+ \times \mathbf{R}^n_+$. It is called the barrier function based on the kernel function $\psi(t)$. Note that if $x = x(\mu)$ then $v = e$, and hence $\psi(v_i) = \psi(1) = 0$, which implies $\phi_\mu(x, s) = 0$. Since $\phi_\mu(x, s) \geq 0$ for every pair (x, s) of positive vectors x and s, we see that ϕ_μ is minimal if and only if $x = x(\mu)$ and $s = s(\mu)$.

Let us write $\psi(t)$ as follows:

$$\psi(t) = \frac{1}{2}\left(t^2 - 1\right) + \psi_b(t). \tag{4.23}$$

We call $\psi_b(t)$ the *barrier term* of $\psi(t)$. Recall that if x is strictly feasible for (SP) and $s = s(x)$, then $q^T x = x^T s = \mu \sum_{i=1}^n v_i^2$. Hence if $s = s(x)$ with x feasible for (SP), then we have

$$\phi_\mu(x, s) = 2\sum_{i=1}^n \psi(v_i) = \sum_{i=1}^n \left(v_i^2 - 1 + 2\psi_b(v_i)\right) = \frac{q^T x}{\mu} - n + \sum_{i=1}^n 2\psi_b(v_i).$$

We denote this function as $\Phi_\mu(x, s)$. So,

$$\Phi_\mu(x, s) = \frac{q^T x}{\mu} - n + \sum_{i=1}^n 2\psi_b(v_i). \tag{4.24}$$

Obviously, $\Phi_\mu(x, s)$ and $\phi_\mu(x, s)$ coincide if x is strictly feasible for (SP) and $s = s(x)$, but otherwise their values may differ. The reason why we use $\Phi_\mu(x, s)$ instead of $\phi_\mu(x, s)$ is because the barrier function $\Phi_\mu(x, s)$ is strictly convex on the interior of $\mathbf{R}^n_+ \times \mathbf{R}^n_+$, as is shown in the next section, whereas one easily may verify that $\phi_\mu(x, s)$ is in general not convex on the interior of $\mathbf{R}^n_+ \times \mathbf{R}^n_+$.[3]

For proving that the barrier function $\Phi_\mu(x, s)$ is convex and self-concordant, we require that the barrier term of the kernel function $\psi(t)$ satisfies the following three conditions:

$$\psi_b'(t) < 0, \qquad t > 0, \tag{4.25}$$

$$\psi_b''(t) > 0, \qquad t > 0, \tag{4.26}$$

$$\psi_b'''(t) < 0, \qquad t > 0. \tag{4.27}$$

Since the barrier term of $\psi_{LB}(t)$ is $-\log t$, one easily understands that these conditions are satisfied by $\psi_{LB}(t)$.

4.3.3 Convexity of $\Phi_\mu(x, s)$

In this subsection we show that under conditions (4.25) and (4.26) the barrier function $\Phi_\mu(x, s)$ is strictly convex on interior $\mathbf{R}^n_+ \times \mathbf{R}^n_+$. Since the sum of

[3] If $n = 1$, the determinant of $\nabla^2 \phi_\mu(x, s)$ is equal to $-\frac{v^3}{4x^2 s^2}\left(v + \psi_b'(v)\right)\left(1 + \psi_b''(v)\right)$, as one may verify. Obviously this not always positive, even not if $\psi(t)$ is the kernel function of the logarithmic barrier function. This implies that $\phi_\mu(x, s)$ is not convex on interior of $\mathbf{R}^n_+ \times \mathbf{R}^n_+$.

convex function is convex, it suffices to show that each term of the barrier function $\Phi_\mu(x,\, s)$ is convex. Since $q^T x$ is linear, it is convex. Hence it suffices if $\psi_b(v_i)$ is convex, for each i. Since $\psi_b(v_i)$ depends only on x_i and s_i, it is enough if $\psi_b(v_i)$ is convex in x_i and s_i. To simplify notation we drop the subscript i. Then it remains to show that

$$f(x,s) := 2\psi_b(v), \qquad \text{where} \qquad v = \sqrt{\frac{xs}{\mu}}, \tag{4.28}$$

is convex in x and s, where the variables x and s are positive scalars. One has

$$\frac{\partial v}{\partial x} = \frac{s}{2\mu v} = \frac{v}{2x}, \quad \frac{\partial v}{\partial s} = \frac{v}{2s}.$$

Using this it follows that

$$\frac{\partial f(x,s)}{\partial x} = \psi_b'(v)\frac{v}{x}, \quad \frac{\partial f(x,s)}{\partial s} = \psi_b'(v)\frac{v}{s}, \tag{4.29}$$

and

$$\begin{aligned}
\frac{\partial^2 f(x,s)}{\partial^2 x} &= \psi_b''(v)\frac{v^2}{2x^2} - \psi_b'(v)\frac{v}{2x^2} = \frac{v}{2x^2}\left(\psi_b''(v)v - \psi_b'(v)\right), \\
\frac{\partial^2 f(x,s)}{\partial^2 s} &= \frac{v}{2s^2}\left(\psi_b''(v)v - \psi_b'(v)\right), \\
\frac{\partial^2 f(x,s)}{\partial x \partial s} &= \psi_b''(v)\frac{v^2}{2xs} + \psi_b'(v)\frac{v}{2xs} = \frac{v}{2xs}\left(\psi_b''(v)v + \psi_b'(v)\right).
\end{aligned} \tag{4.30}$$

Hence the Hessian matrix is given by

$$\nabla^2 f(x,s) = \begin{bmatrix} \frac{v}{2x^2}\left(\psi_b''(v)v - \psi_b'(v)\right) & \frac{v}{2xs}\left(\psi_b''(v)v + \psi_b'(v)\right) \\ \frac{v}{2xs}\left(\psi_b''(v)v + \psi_b'(v)\right) & \frac{v}{2s^2}\left(\psi_b''(v)v - \psi_b'(v)\right) \end{bmatrix}. \tag{4.31}$$

The conditions (4.25) and (4.26) imply that the diagonal elements are positive. The determinant is given by

$$\frac{v^2}{4x^2s^2}\left(\left(\psi_b''(v)v - \psi_b'(v)\right)^2 - \left(\psi_b''(v)v + \psi_b'(v)\right)^2\right) = -\frac{v^3}{x^2s^2}\psi_b'(v)\psi_b''(v),$$

which by the same reason is also positive. Thus we have shown that the Hessian matrix is positive definite, which implies that the barrier function $\Phi_\mu(x,\, s)$ is strictly convex. Therefore, we can state the following lemma without further proof.

Lemma 4.3.1. *If $\psi(t)$ satisfies the conditions (4.25) and (4.26) then $\Phi_\mu(x,\, s)$ is strictly convex.*

We now have to deal with the crucial question whether or not $\Phi_\mu(x,\, s)$ is self-concordant and if so, if it is a ν-barrier for some appropriate value of ν. Because then we can use it in our generic algorithm, to solve (SP) in polynomial time. The disappointing answer to this question is 'no', with the logarithmic barrier function being the only (positive) exception. But this is not the end of our story. As we show in the next section, when applying our generic algorithm with one of the new barrier functions, then in the region where the iterates live, which is a rather small neighborhood of the central path, the new barrier functions are locally self-concordant. In the next two subsections we compute the local values of κ and ν for the kernel-function-based barrier function $\Phi_\mu(x,\, s)$ as given by (4.24). Because of the composition rules in Lemma 4.2.15 we may start by considering the case where $n = 1$, and then apply these rules to obtain results for higher dimensions.

4.3.4 Computation of $\nu\,(x,\, s)$

In this subsection we compute the local value $\nu\,(x,\, s)$. We start with considering the function $f(x, s)$ given by (4.28). We can easily compute the value of $\nu(x, s)$, since we already computed the first and second derivatives of $f(x, s)$. From (4.29) we deduce that the gradient is given by

$$\nabla f(x,s) = \begin{bmatrix} \psi_b'(v)\dfrac{v}{x} \\ \psi_b'(v)\dfrac{v}{s} \end{bmatrix} = \psi_b'(v)v \begin{bmatrix} \dfrac{1}{x} \\ \dfrac{1}{s} \end{bmatrix}, \tag{4.32}$$

and the Hessian by (4.31). According to (4.14) in Definition 4.2.12 the value of $\nu(x, s)$ is given by

$$\nu(x,s) = \max_{h\in\mathbf{R}^2} \frac{\left(\nabla f(x,s)^T h\right)^2}{h^T \nabla^2 f(x,s) h},$$

where $h = (h_1,\, h_2)$ and $h_1,\, h_2 \in \mathbf{R}$. Using the expressions (4.32) and (4.31) it follows that

$$\nabla f(x,s)^T h = \psi_b'(v)v\left(\frac{h_1}{x} + \frac{h_2}{s}\right)$$

and

$$\begin{aligned} h^T\nabla^2 f(x,s)h &= \frac{1}{2}v\left(\psi_b''(v)v - \psi_b'(v)\right)\left(\frac{h_1^2}{x^2} + \frac{h_2^2}{s^2}\right) + v\left(\psi_b''(v)v + \psi_b'(v)\right)\frac{h_1h_2}{xs} \\ &= \frac{1}{2}v\left(\psi_b''(v)v\left(\frac{h_1}{x} + \frac{h_2}{s}\right)^2 - \psi_b'(v)\left(\frac{h_1}{x} - \frac{h_2}{s}\right)^2\right). \end{aligned}$$

Denoting

$$\sigma = \frac{h_1}{x}, \quad \tau = \frac{h_2}{s}, \tag{4.33}$$

we thus obtain

$$\nu(x,s) = \max_{\sigma,\tau} \frac{\left(\psi_b'(v)v\left(\sigma+\tau\right)\right)^2}{\frac{1}{2}v\left(\psi_b''(v)v\left(\sigma+\tau\right)^2 - \psi_b'(v)\left(\sigma-\tau\right)^2\right)}.$$

Changing to new variables $y = \sigma + \tau$ and $z = \sigma - \tau$ we get

$$\nu(x,s) = \max_{y,z} \frac{2\psi_b'(v)^2 v y^2}{\psi_b''(v)vy^2 - \psi_b'(v)z^2}.$$

Since $\psi_b'(v) < 0$ the maximum occurs for $z = 0$. Therefore, we immediately obtain that

$$\nu(x,s) = \frac{2\psi_b'(v)^2}{\psi_b''(v)}.$$

Thus we have found the local value $\nu(x,s)$ if $n = 1$. Now using Lemma 4.2.15 we can state the next lemma without further proof.

Lemma 4.3.2. *The local value $\nu\,(x,\,s)$ is given by*

$$\nu(x,s) = \sum_{i=1}^{n} 2\frac{\psi_b'(v_i)^2}{\psi_b''(v_i)}.$$

4.3.5 Computation of $\kappa\,(x,\,s)$

In this subsection we compute the local value $\kappa(x,s)$. For this we first need to find the third order derivatives of $f(x,s)$. Straightforward computations yield the following expressions:

$$\begin{aligned}
\frac{\partial^3 f(x,s)}{\partial^3 x} &= \frac{v}{4x^3}\left[3\psi_b'(v) - 3v\psi_b''(v) + v^2\psi_b'''(v)\right],\\
\frac{\partial^3 f(x,s)}{\partial^2 x \partial s} &= \frac{v}{4x^2 s}\left[-\psi_b'(v) + v\psi_b''(v) + v^2\psi_b'''(v)\right],\\
\frac{\partial^3 f(x,s)}{\partial x \partial^2 s} &= \frac{v}{4xs^2}\left[-\psi_b'(v) + v\psi_b''(v) + v^2\psi_b'''(v)\right],\\
\frac{\partial^3 f(x,s)}{\partial^3 s} &= \frac{v}{4s^3}\left[3\psi_b'(v) - 3v\psi_b''(v) + v^2\psi_b'''(v)\right].
\end{aligned}$$

It will be convenient to use the following short hand notation:

$$\xi_b(t) = \psi_b''\left(t\right) - \frac{\psi_b'\left(t\right)}{t}, \quad t > 0. \tag{4.34}$$

Then we may write these derivatives as follows.

$$\begin{aligned}
\frac{\partial^3 f(x,s)}{\partial^3 x} &= \frac{1}{4x^3}\left[-3v^2\xi_b(v) + v^3\psi_b'''(v)\right],\\
\frac{\partial^3 f(x,s)}{\partial^2 x \partial s} &= \frac{1}{4x^2 s}\left[v^2\xi_b(v) + v^3\psi_b'''(v)\right],\\
\frac{\partial^3 f(x,s)}{\partial x \partial^2 s} &= \frac{1}{4xs^2}\left[v^2\xi_b(v) + v^2\psi_b'''(v)\right],\\
\frac{\partial^3 f(x,s)}{\partial^3 s} &= \frac{1}{4s^3}\left[-3v^2\xi_b(v) + v^3\psi_b'''(v)\right].
\end{aligned}$$

Therefore, using the variables σ and τ introduced in (4.33), we may write

$$\begin{aligned}
4\nabla^3 f(x,s)[h,h,h] &= \left[-3v^2\xi_b(v) + v^3\psi_b'''(v)\right]\left(\sigma^3+\tau^3\right)\\
&\quad +3\left[v^2\xi_b(v) + v^3\psi_b'''(v)\right]\left(\sigma^2\tau+\sigma\tau^2\right)\\
&= \left[-3v^2\xi_b(v) + v^3\psi_b'''(v)\right]\left(\sigma^2-\sigma\tau+\tau^2\right)(\sigma+\tau)\\
&\quad +3\left[v^2\xi_b(v) + v^3\psi_b'''(v)\right]\sigma\tau\,(\sigma+\tau)\\
&= \left[v^3\psi_b'''(v)\left(\sigma^2-\sigma\tau+\tau^2\right) + 3v^3\psi_b'''(v)\sigma\tau\right](\sigma+\tau)\\
&\quad +\left[-3v^2\xi_b(v)\left(\sigma^2-\sigma\tau+\tau^2\right) + 3v^2\xi_b(v)\sigma\tau\right](\sigma+\tau)\\
&= \left[\psi_b'''(v)\,v^2\,(\sigma+\tau)^2 - 3\,\xi_b(v)\,v\,(\sigma-\tau)^2\right]v\,(\sigma+\tau).
\end{aligned}$$

Hence, according to (4.13) in Definition 4.2.12 the value of $\kappa(x,s)$ satisfies

$$2\kappa(x,s) = \max_{\sigma,\tau}\frac{\frac{1}{4}\left[\psi_b'''(v)\,v^2\,(\sigma+\tau)^2 - 3\,\xi_b(v)\,v\,(\sigma-\tau)^2\right]v\,(\sigma+\tau)}{\left(\frac{1}{2}v\left[\psi_b''(v)v\,(\sigma+\tau)^2 - \psi_b'(v)\,(\sigma-\tau)^2\right]\right)^{\frac{3}{2}}}.$$

Changing again to variables $y=\sigma+\tau$ and $z=\sigma-\tau$ we get

$$2\kappa(x,s) = \max_{y,z}\frac{\frac{1}{4}\left[\psi_b'''(v)\,v^2y^2 - 3\,\xi_b(v)\,vz^2\right]vy}{\left(\frac{1}{2}v\left[\psi_b''(v)vy^2 - \psi_b'(v)z^2\right]\right)^{\frac{3}{2}}}.$$

The last expression is homogeneous in (y,z) and the denominator is positive. It follows that

$$8\kappa(x,s) = \max\left\{\psi_b'''(v)\,v^3y^3 - 3\,\xi_b(v)\,v^2yz^2 \;:\; \psi_b''(v)\,v^2y^2 - \psi_b'(v)\,vz^2 = 2\right\}. \tag{4.35}$$

The optimality conditions for this problem are, for some suitable Lagrange multiplier γ,

$$\begin{aligned}
3\psi_b'''(v)\,v^3y^2 - 3\,\xi_b(v)\,v^2z^2 &= 2\gamma\,\psi_b''(v)\,v^2y,\\
-6\,\xi_b(v)\,v^2yz &= -2\gamma\,\psi_b'(v)\,vz.
\end{aligned}$$

Replacing γ by 3λ, this simplifies to

$$\psi_b'''(v)\, vy^2 - \xi_b(v)\, z^2 = 2\lambda\, \psi_b''(v)\, y, \tag{4.36}$$

$$\xi_b(v)\, vyz = \lambda\, \psi_b'(v)\, z. \tag{4.37}$$

From (4.37) we see that either $z = 0$ or

$$\xi_b(v)\, vy = \lambda\, \psi_b'(v), \tag{4.38}$$

If $z = 0$ then the constraint in (4.35) implies that $\psi_b''(v)\, v^2y^2 = 2$. Since we are maximizing, and $\psi_b''(v) > 0$ and $\psi_b'''(v) < 0$, the maximal value of $8\kappa(x,s)$ is in this case given by

$$-2\sqrt{2}\frac{\psi_b'''(v)}{(\psi_b''(v))^{\frac{3}{2}}}. \tag{4.39}$$

Next we deal with the case where $z \neq 0$. It is convenient to introduce two new functions, namely

$$\rho(t) = \frac{\psi_b'(t)\,\psi_b'''(t)}{\xi_b(t)\psi_b''(t)}, \quad \bar{\rho}(t) = \min\left[2, \rho(t)\right]. \tag{4.40}$$

Note that $\bar{\rho}(t) \in (0, 2]$. We can eliminate λ from (4.36) by using (4.38), which gives

$$\psi_b'(v)\left[\psi_b'''(v)\, vy^2 - \xi_b(v)\, z^2\right] = 2\lambda\, \psi_b'(v)\, \psi_b''(v)\, y = 2\,\xi_b(v)\psi_b''(v)\, vy^2.$$

Rearranging the terms, and using (4.40) we obtain

$$-\psi_b'(v)\xi_b(v)\, z^2 = \left[2\,\xi_b(v)\psi_b''(v) - \psi_b'(v)\psi_b'''(v)\right]\, vy^2 = \left[2 - \rho(v)\right]\, \xi_b(v)\psi_b''(v)\, vy^2,$$

yielding

$$-\psi_b'(v)\, z^2 = (2 - \rho(v))\, \psi_b''(v)\, vy^2, \tag{4.41}$$

Since $-\psi_b'(v) > 0$ and $\psi_b''(v) > 0$, this equation has no nonzero solution if $\rho(v) > 2$, and hence the maximal value is then given by (4.39).

We proceed by assuming $\rho(v) \leq 2$. The optimality condition (4.41) together with the constraint of (4.35) yields the following system:

$$\begin{aligned}\left[2 - \rho(v)\right]\, \psi_b''(v)\, v^2y^2 + \psi_b'(v) v\, z^2 &= 0,\\ \psi_b''(v)\, v^2y^2 - \psi_b'(v)\, vz^2 &= 2.\end{aligned}$$

We may consider this as a linear system in the unknowns v^2y^2 and vz^2. By solving this system we obtain

$$v^2y^2 = \frac{2}{\left[3 - \rho(v)\right]\psi_b''(v)}, \quad vz^2 = \frac{2\left[\rho(v) - 2\right]}{\left[3 - \rho(v)\right]\psi_b'(v)}.$$

Substitution of these values yields:

$$\begin{aligned}
8\kappa(x,s) &= \psi_b'''(v)\, v^3 y^3 - 3\,\xi_b(v)\, v^2 y z^2 \\
&= vy\left(\psi_b'''(v)\, v^2 y^2 - 3\,\xi_b(v)\, z^2\right) \\
&= vy\left(\psi_b'''(v)\,\frac{2}{[3-\rho(v)]\,\psi_b''(v)} - 3\,\xi_b(v)\,\frac{2\,[\rho(v)-2]}{[3-\rho(v)]\,\psi_b'(v)}\right) \\
&= vy\left(\frac{\psi_b'''(v)}{\psi_b''(v)}\,\frac{2}{[3-\rho(v)]} - 3\,\frac{\xi_b(v)}{\psi_b'(v)}\,\frac{2\,[\rho(v)-2]}{[3-\rho(v)]}\right).
\end{aligned}$$

Now using the definition of $\rho(v)$, we obtain

$$\begin{aligned}
8\kappa(x,s) &= vy\left(\frac{\psi_b'''(v)}{\psi_b''(v)}\,\frac{2}{[3-\rho(v)]} - 3\,\frac{\psi_b'''(v)}{\rho(v)\psi_b''(v)}\,\frac{2\,[\rho(v)-2]}{[3-\rho(v)]}\right) \\
&= vy\frac{\psi_b'''(v)}{[3-\rho(v)]\,\psi_b''(v)}\left(2 - 6\,\frac{1}{\rho(v)}\,[\rho(v)-2]\right) \\
&= vy\frac{\psi_b'''(v)}{[3-\rho(v)]\,\psi_b''(v)}\,\frac{2\rho(v) - 6[\rho(v)-2]}{\rho(v)} \\
&= vy\frac{\psi_b'''(v)}{[3-\rho(v)]\,\psi_b''(v)}\,\frac{12 - 4\rho(v)}{\rho(v)} \\
&= vy\frac{4\psi_b'''(v)}{\rho(v)\psi_b''(v)}.
\end{aligned}$$

Since we are maximizing, and $\psi_b'''(v) < 0$, we have $y < 0$. Thus we obtain

$$\begin{aligned}
8\kappa(x,s) = vy\frac{4\psi_b'''(v)}{\rho(v)\psi_b''(v)} &= \frac{-4\sqrt{2}}{\sqrt{3-\rho(v)}\sqrt{\psi_b''(v)}}\,\frac{\psi_b'''(v)}{\rho(v)\psi_b''(v)} \\
&= \frac{-4\sqrt{2}}{\rho(v)\sqrt{3-\rho(v)}}\,\frac{\psi_b'''(v)}{\psi_b''(v)^{\frac{3}{2}}}.
\end{aligned}$$

Note that for $\rho(v) = 2$ this gives precisely the same value as in (4.39). Hence, also using Lemma 4.2.15 we may state without further proof the following result.

Lemma 4.3.3. *The local value $\kappa(x,s)$ is given by*

$$\kappa(x,s) = \max_i \frac{-1}{\sqrt{2}\,\bar{\rho}(v_i)\sqrt{3-\bar{\rho}(v_i)}}\,\frac{\psi_b'''(v_i)}{\psi_b''(v_i)^{\frac{3}{2}}}.$$

4.3.6 Application to barrier functions

In this subsection we use Lemmas 4.3.2 and 4.3.3 to compute the local values of κ and ν for some barrier functions induced by some well-known kernel functions. See Table 4.1. The second line in Table 4.1 establishes the well-known fact that

$\psi(t)$	$\psi_b(t)$	$\psi_b'(t)$	$\psi_b''(t)$	$\psi_b'''(t)$	$\xi_b(t)$	$\rho(t)$	$\kappa(x,s)$	$\nu(x,s)$	$\gamma\left(x(\mu),s(\mu)\right)$
$\frac{t^2-1}{2}-\log t$	$-\log t$	$-\frac{1}{t}$	$\frac{1}{t^2}$	$-\frac{2}{t^3}$	$\frac{2}{t^2}$	1	1	$2n$	$\sqrt{2n}$
$\frac{1}{2}\left(t-\frac{1}{t}\right)^2$	$\frac{t^{-2}-1}{2}$	$-\frac{1}{t^3}$	$\frac{3}{t^4}$	$-\frac{12}{t^5}$	$\frac{4}{t^4}$	1	$\frac{2}{\sqrt{3}}\|v\|_\infty$	$\frac{2}{3}\sum_{j=1}^n v_j^{-2}$	$\frac{2}{3}\sqrt{2n}$
$\frac{t^2-1}{2}+\frac{t^{1-q}-1}{q-1}$	$\frac{t^{1-q}-1}{q-1}$	$-\frac{1}{t^q}$	$\frac{q}{t^{q+1}}$	$-\frac{q(q+1)}{t^{q+2}}$	$\frac{q+1}{t^{q+1}}$	1	$\frac{q+1}{2\sqrt{q}}\|v\|_\infty^{\frac{q-1}{2}}$	$\frac{2}{q}\sum_{j=1}^n v_j^{1-q}$	$\frac{q+1}{2q}\sqrt{2n}$

Table 4.1: *Local values of κ and ν for some well-known kernel functions.*

the logarithmic barrier function is a $(1, 2n)$-SCB on the cone $\mathbf{R}^n_+ \times \mathbf{R}^n_+$. One may easily understand that the logarithmic barrier function is also a $(1, 2n)$-SCB for the feasible region of (SP), which is the intersection of this cone with the affine space $\{(x, s) : s = Mx + q\}$. Hence, by Theorem 4.2.10, the generic algorithm of Figure 4.1 will solve (SP) in at most

$$\left\lceil 2\left(1+4\sqrt{2n}\right)\,\ln\frac{4n}{\epsilon}\right\rceil = O\left(\sqrt{n}\,\ln\frac{n}{\epsilon}\right)$$

iterations.

For the other two kernel functions the situation is quite different. Their values of κ are not bounded above if v runs through $\mathbf{R}^n$. Hence the corresponding barrier functions are not self-concordant. Note, however, that if v is the all-one vector, i.e., if $x = x(\mu)$ and $s = Mx + q$, then, e.g., for the third kernel function the local complexity number is given by

$$\kappa\sqrt{\nu} = \frac{q+1}{2\sqrt{q}}\cdot\sqrt{\frac{2n}{q}} = \frac{q+1}{2q}\sqrt{2n} = \frac{q+1}{2q}\sqrt{2n}.$$

So for $q > 1$ the local complexity number $\gamma\left(x(\mu), s(\mu)\right)$ of this barrier function on the central path is even smaller than that for the logarithmic barrier function.

Since κ and ν depend continuously on v, and since during the algorithm the iterates stay very close to the central path, we can exploit this fact as we show in the next section.

4.4 Exploiting locally self-concordance

In the previous sections we found expressions for the local κ and ν values of the barrier function determined by a given kernel function ψ under the assumption that the conditions (4.25)–(4.27) are satisfied. Let us point out the surprising fact that these values depend only on the variance vector v at x. In this section we want to show that such a barrier function is locally self-concordant in the region

where the iterates of the algorithm occur. In other words, we need to obtain upper bounds for the local κ and ν values in these iterates.

First we recall from Lemma 4.2.11 that during the course of the algorithm we have

$$\lambda_\mu(x,s) \leq \frac{1}{4\kappa}. \tag{4.42}$$

According to (4.6) and (4.22) this implies

$$\phi_\mu(x,s) = 2\sum_{i=1}^{n} \psi(v_i) \leq \frac{\omega(-\frac{1}{4})}{\kappa^2} = \frac{0.0376821}{\kappa^2} \leq \frac{1}{26\kappa^2}, \tag{4.43}$$

where we used that $\omega(t)$ is monotonically decreasing on $(-1,0]$ as defined in (4.4). Since $\psi(t)$ is nonnegative for all $t > 0$ it follows that

$$\psi(v_i) \leq \frac{1}{52\kappa^2}, \quad i = 1, \ldots, n. \tag{4.44}$$

Let us define

$$\nu_\psi(t) := \frac{2\psi_b'(t)^2}{\psi_b''(t)}, \quad \kappa_\psi(t) := \frac{1}{\sqrt{2}\,\bar{\rho}(t)\sqrt{3-\bar{\rho}(t)}} \frac{-\psi_b'''(t)}{\psi_b''(t)^{\frac{3}{2}}}.$$

By Lemmas 4.3.2 and 4.3.3 the local values of κ and ν at the pair (x,s) are then given by

$$\kappa(x,s) = \max_i \kappa_\psi(v_i), \quad \nu(x,s) = \sum_{i=1}^{n} \nu_\psi(v_i).$$

We need to find values of κ and ν such that

$$\kappa(x,s) \leq \kappa, \quad \nu(x,s) \leq \nu \tag{4.45}$$

for all iterates x and s that occur during the course of the algorithm. This certainly holds if κ and ν are such that

$$\phi_\mu(x,s) = 2\sum_{i=1}^{n} \psi(v_i) \leq \frac{1}{26\kappa^2} \quad \Rightarrow \quad \max_i \kappa_\psi(v_i) \leq \kappa, \quad \sum_{i=1}^{n} \nu_\psi(v_i) \leq \nu.$$

The left-hand side of this implication implies

$$\psi(v_i) \leq \frac{1}{52\kappa^2}, \quad i = 1, \ldots, n.$$

Therefore, it suffices if κ and ν are such that

$$\psi(t) \leq \frac{1}{52\kappa^2} \quad \Rightarrow \quad \kappa_\psi(t) \leq \kappa, \quad n\nu_\psi(t) \leq \nu. \tag{4.46}$$

According to Lemma 2.3.8 we have

$$\psi(t) \leq \frac{1}{52\kappa^2} \quad \Leftrightarrow \quad \chi\left(\frac{1}{52\kappa^2}\right) \leq t \leq \varrho\left(\frac{1}{52\kappa^2}\right),$$

where the functions χ and ϱ are as in Chapter 2. So, $\varrho : [0,\infty) \to [1,\infty)$ is the inverse function of $\psi(t)$ for $t \geq 1$ and $\chi : [0,\infty) \to (0,1]$ the inverse function of $\psi(t)$ for $t \leq 1$.

If we choose κ such that

$$\max\left\{\kappa_\psi(t) : \chi\left(\frac{1}{52\kappa^2}\right) \leq t \leq \varrho\left(\frac{1}{52\kappa^2}\right)\right\} \leq \kappa,$$

then the barrier function is locally κ-self concordant in each of the iterates of the algorithm. The above inequality certainly has a solution, because if κ goes to infinity then the left-hand side in the last inequality approaches $\kappa_\psi(1)$, which is finite, whereas the right-hand side goes to infinity. Let $\bar{\kappa}$ be not smaller than the smallest solution of the above inequality:

$$\bar{\kappa} \geq \min\left\{\kappa : \max\left\{\kappa_\psi(t) : \chi\left(\frac{1}{52\kappa^2}\right) \leq t \leq \varrho\left(\frac{1}{52\kappa^2}\right)\right\} \leq \kappa\right\}, \tag{4.47}$$

Finally, given $\bar{\kappa}$, if we take $\bar{\nu}$ such that

$$\bar{\nu} := n \times \max\left\{\nu_\psi(t) : \chi\left(\frac{1}{52\bar{\kappa}^2}\right) \leq t \leq \varrho\left(\frac{1}{52\bar{\kappa}^2}\right)\right\}, \tag{4.48}$$

then the barrier function $\Phi_\mu(x,s)$ is locally $(\bar{\kappa},\bar{\nu})$-self concordant barrier function in the region where the iterates generated by the algorithm occur. Substitution of these values, also using that $\mu^0 = 1$, yields the following iteration bound for the algorithm:

$$\left\lceil 2\left(1 + 4\bar{\kappa}\sqrt{\bar{\nu}}\right) \ln\frac{2\bar{\nu}}{\epsilon}\right\rceil.$$

Note that apart from the factor n in the expression for $\bar{\nu}$ this expression depends only on the kernel function ψ. Thus we may safely state that for every kernel function satisfying our conditions the iteration bound is

$$\mathcal{O}\left(\sqrt{n}\log\frac{n}{\epsilon}\right).$$

We conclude this section by illustrating the use of the above analysis for the kernel functions in Table 4.1.

Example: $\psi(t) = \frac{t^2-1}{2} - \log t$

We deduce from Table 4.1 that $\kappa(x,s) = 1$ and $\nu(x,s) = 2n$. Therefore, we may use simply $\bar{\kappa} = 1$ and $\bar{\nu} = 2n$ in this case, which yields the well-known iteration

bound for IPMs based on the logarithmic barrier function, namely

$$\left\lceil 2\left(1+4\sqrt{2n}\right)\,\ln\frac{2n}{\epsilon}\right\rceil.$$

Example: $\psi(t)=\frac{1}{2}\left(t-\frac{1}{t}\right)^2$

We need to compute the inverse functions of $\psi(t)$ for $t\geq 1$ and for $t\leq 1$. One has

$$s=\frac{1}{2}\left(t-\frac{1}{t}\right)^2$$

if and only if

$$t^2=1+s\pm\sqrt{s^2+2s}.$$

Hence we have

$$\begin{aligned}\chi(s)&=\sqrt{1+s-\sqrt{s^2+2s}},\\ \varrho(s)&=\sqrt{1+s+\sqrt{s^2+2s}}.\end{aligned}$$

Moreover, from Table 4.1 we deduce that

$$\nu_\psi(t):=\frac{2}{3t^2},\quad \kappa_\psi(t):=\frac{2t}{\sqrt{3}}.$$

Hence, by (4.47), $\bar\kappa$ should be taken such that

$$\begin{aligned}\bar\kappa&\geq\min\left\{\kappa\ :\ \max\left\{\kappa_\psi\left(t\right)\ :\ \chi\left(\frac{1}{52\kappa^2}\right)\leq t\leq\varrho\left(\frac{1}{52\kappa^2}\right)\right\}\leq\kappa\right\}\\ &=\min\left\{\kappa\ :\ \max\left\{\frac{2t}{\sqrt{3}}\ :\ \chi\left(\frac{1}{52\kappa^2}\right)\leq t\leq\varrho\left(\frac{1}{52\kappa^2}\right)\right\}\leq\kappa\right\}\\ &=\min\left\{\kappa\ :\ \frac{2}{\sqrt{3}}\,\varrho\left(\frac{1}{52\kappa^2}\right)\leq\kappa\right\}.\end{aligned}$$

The inequality $f\left(\kappa\right):=\frac{2}{\sqrt{3}}\,\varrho\left(\frac{1}{52\kappa^2}\right)-\kappa\leq 0$ is satisfied if $\kappa=1.24892$ (see Figure 4.2). Hence we may take $\bar\kappa=1.24892$. Then

$$\begin{aligned}\bar\nu&=n\times\max\left\{\nu_\psi\left(t\right)\ :\ \chi\left(\frac{1}{52\bar\kappa^2}\right)\leq t\leq\varrho\left(\frac{1}{52\bar\kappa^2}\right)\right\}\\ &=n\times\max\left\{\frac{2}{3t^2}\ :\ \chi\left(\frac{1}{52\bar\kappa^2}\right)\leq t\leq\varrho\left(\frac{1}{52\bar\kappa^2}\right)\right\}\\ &=\frac{2n}{3\chi\left(\frac{1}{52\bar\kappa^2}\right)^2}=\frac{2n}{3}\varrho\left(\frac{1}{52\bar\kappa^2}\right)^2\leq\frac{n}{2}\bar\kappa^2,\end{aligned}$$

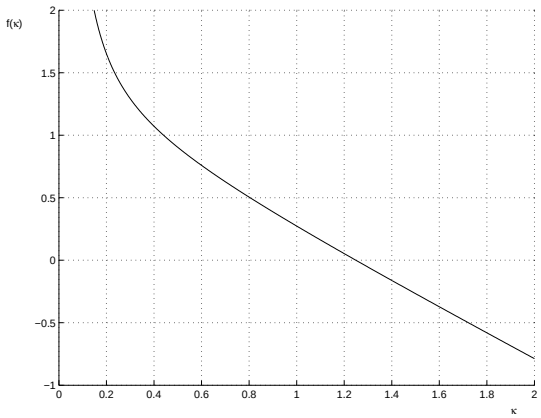

Figure 4.2: *Graphical illustration of the function $f(\kappa)$*

where we used that $\chi(s)\varrho(s) = 1$. The complexity number satisfies

$$\bar{\kappa}\sqrt{\nu} \leq \bar{\kappa}\sqrt{\frac{n}{2}\bar{\kappa}^2} = \tfrac{1}{2}\bar{\kappa}^2\sqrt{2n} \approx 0.779901\sqrt{2n}.$$

Note that this is about 22% smaller than the complexity number of the logarithmic barrier function, which is $\sqrt{2n}$.

Example: $\psi(t) = \dfrac{t^2-1}{2} + \dfrac{t^{1-q}-1}{q-1}$, $q \geq 3$

To obtain the inverse functions of $\psi(t)$ for $t \geq 1$ and for $t \leq 1$ we need to solve t from the equation

$$s = \frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}.$$

For general values of $q > 1$ it is hard to find a closed form solution of this equation. At this stage the following lemmas are useful.

Lemma 4.4.1 (cf. Lemma 6.1 in [8])**.** *For each kernel function $\psi(t)$ satisfying (4.25) one has*

$$\sqrt{1+2s} \leq \varrho(s) \leq \sqrt{1+s+\sqrt{s^2+2s}} \leq 1+\sqrt{2s}.$$

Proof. The inverse function $\varrho(s)$ of $\psi(t)$ for $t \in [1,\infty)$ is obtained by solving t from the equation $\psi(t) = s$, for $t \geq 1$. Since $\psi_b(1) = 0$, and $\psi_b'(t) < 0$ for all $t > 0$, we have $\psi_b(t) \leq 0$ for $t \geq 1$. Hence, if $t \geq 1$ then

$$s = \psi(t) = \frac{t^2-1}{2} + \psi_b(t) \leq \frac{t^2-1}{2}.$$

It follows that

$$t = \varrho(s) \geq \sqrt{1+2s}.$$

For the second inequality we use that $\psi''(t) = 1 + qt^{-q-1} \geq 1$. Since $\psi(1) = \psi'(1) = 0$ it follows that

$$\begin{aligned} s = \psi(t) &= \int_1^t \int_1^\xi \psi''(\zeta)\, d\zeta d\xi = \int_1^t \int_1^\xi \left(1 + \frac{q}{\zeta^{q+1}}\right) d\zeta d\xi \\ &= \int_1^t \left(\zeta - \frac{1}{\zeta^q}\right)\Big|_{\zeta=1}^{\xi} d\xi = \int_1^t \left(\xi - \frac{1}{\xi^q}\right) d\xi \\ &\geq \int_1^t \left(\xi - \frac{1}{\xi^3}\right) d\xi = \frac{1}{2}\left(\xi^2 + \frac{1}{\xi^2}\right)\Big|_{\xi=1}^{t} = \frac{1}{2}\left(t - \frac{1}{t}\right)^2, \end{aligned}$$

which, since $t \geq 1$, implies

$$t = \varrho(s) \leq \sqrt{1 + s + \sqrt{s^2+2s}}.$$

Since the last inequality in the lemma easily follows, this proves the lemma. □

Lemma 4.4.2. *If $q \geq 3$ and $t \in (0,1]$ then*

$$\psi(t) \geq \psi\left(\frac{1}{t}\right).$$

Proof. Consider

$$f(t) := \psi(t) - \psi\left(\frac{1}{t}\right).$$

One has

$$f'(t) = \frac{1}{t^3} + t - \frac{1}{t^q} - \frac{1}{t^{q-2}} = \frac{1}{t^q}\left(t^{q-3} - 1\right)\left(1 + t^2\right).$$

Since $f(1) = 0$ and $f'(t) \leq 0$ for $t \in (0,1]$, the lemma follows. □

Lemma 4.4.3. *If $q \geq 3$ then*

$$\chi(s)\varrho(s) \geq 1.$$

Proof. Let $t = \chi(s)$, for some $s \geq 0$. Then $s = \psi(t)$ and $t \leq 1$. By Lemma 4.4.2 this implies $s \geq \psi\left(\frac{1}{t}\right)$. Since ϱ is monotonically increasing it follows that

$$\varrho(s) \geq \varrho\left(\psi\left(\frac{1}{t}\right)\right) = \frac{1}{t} = \frac{1}{\chi(s)},$$

proving the lemma. □

From Table 4.1 we deduce that

$$\nu_\psi(t) := \frac{2}{qt^{q-1}}, \quad \kappa_\psi(t) := \frac{(q+1)t^{\frac{q-1}{2}}}{2\sqrt{q}}.$$

Hence, by (4.47), $\bar{\kappa}$ should be taken such that

$$\begin{aligned}\bar{\kappa} &\geq \min\left\{\kappa \;:\; \max\left\{\frac{(q+1)t^{\frac{q-1}{2}}}{2\sqrt{q}} \;:\; \chi\left(\frac{1}{52\kappa^2}\right) \leq t \leq \varrho\left(\frac{1}{52\kappa^2}\right)\right\} \leq \kappa\right\} \\ &= \min\left\{\kappa \;:\; \frac{q+1}{2\sqrt{q}}\,\varrho\left(\frac{1}{52\kappa^2}\right)^{\frac{q-1}{2}} \leq \kappa\right\}.\end{aligned}$$

By Lemma 4.4.1, the inequality

$$\frac{q+1}{2\sqrt{q}}\,\varrho\left(\frac{1}{52\kappa^2}\right)^{\frac{q-1}{2}} \leq \kappa \tag{4.49}$$

is satisfied if

$$\frac{q+1}{2\sqrt{q}}\left(1+\frac{1}{\kappa\sqrt{26}}\right)^{\frac{q-1}{2}} \leq \kappa.$$

This is equivalent to

$$\left(1+\frac{1}{\kappa\sqrt{26}}\right)^{\frac{q-1}{2}} \leq \frac{2\sqrt{q}}{q+1}\,\kappa,$$

which gives, by taking logarithms at both sides,

$$\frac{q-1}{2}\log\left(1+\frac{1}{\kappa\sqrt{26}}\right) \leq \log\left(\frac{2\sqrt{q}}{q+1}\,\kappa\right).$$

Since $\log(1+x) \leq x$, this certainly holds if

$$\frac{q-1}{2}\frac{1}{\kappa\sqrt{26}} \leq \log\left(\frac{2\sqrt{q}}{q+1}\,\kappa\right).$$

Putting

$$\kappa = \frac{q+1}{2\sqrt{q}}\,(1+\sigma)$$

this leads to the inequality

$$\frac{q-1}{2}\frac{1}{\frac{q+1}{2\sqrt{q}}\,(1+\sigma)\,\sqrt{26}} \leq \log\,(1+\sigma)\,,$$

which gives

$$\frac{q-1}{q+1}\frac{\sqrt{q}}{\sqrt{26}} \leq (1+\sigma)\log\,(1+\sigma)\,.$$

Since $(1+\sigma)\log\,(1+\sigma) \geq \sigma$ this inequality is satisfied if

$$\sigma = \frac{q-1}{q+1}\frac{\sqrt{q}}{\sqrt{26}},$$

and hence

$$\kappa = \frac{q+1}{2\sqrt{q}}\left(1+\frac{q-1}{q+1}\frac{\sqrt{q}}{\sqrt{26}}\right)$$

satisfies the inequality (4.49). Let $\bar{\kappa}$ be given this value. Then, using Lemma 4.4.3, we may write

$$\begin{aligned}\bar{\nu} &= n \times \max\left\{\frac{2}{qt^{q-1}} \;:\; \chi\left(\frac{1}{52\bar{\kappa}^2}\right) \le t \le \varrho\left(\frac{1}{52\bar{\kappa}^2}\right)\right\} \\ &= \frac{2n}{q\chi\left(\frac{1}{52\bar{\kappa}^2}\right)^{q-1}} \le \frac{2n}{q}\,\varrho\left(\frac{1}{52\bar{\kappa}^2}\right)^{q-1} \le \frac{2n}{q}\,\frac{4q}{(q+1)^2}\bar{\kappa}^2 = \frac{8n}{(q+1)^2}\bar{\kappa}^2.\end{aligned}$$

Hence, the complexity number satisfies

$$\begin{aligned}\bar{\kappa}\sqrt{\bar{\nu}} = \frac{2\bar{\kappa}^2}{q+1}\sqrt{2n} &= \frac{2}{q+1}\left(\frac{q+1}{2\sqrt{q}}\left(1+\frac{q-1}{q+1}\frac{\sqrt{q}}{\sqrt{26}}\right)\right)^2\sqrt{2n} \\ &= \frac{q+1}{2q}\left(1+\frac{q-1}{q+1}\frac{\sqrt{q}}{\sqrt{26}}\right)^2\sqrt{2n}.\end{aligned}$$

The smallest value is obtained for $q \approx 2.13136$, and then the coefficient of $\sqrt{2n}$ equals 0.89443. For $q = 3$ the coefficient of $\sqrt{2n}$ equals 0.912353, and this coefficient becomes 1 for $q \approx 5$.

Chapter 5

Conclusions

5.1 Conclusions and remarks

In [82], a new primal-dual infeasible interior-point algorithm was presented that uses full-Newton steps. Each iteration of his algorithm consists of a step that restores the feasibility for an intermediate problem (the so-called feasibility step) and few (usual) centering steps. He proved that no more than $O\left(n \log \frac{n}{\epsilon}\right)$ iterations are required for getting an ϵ-solution of the problem at hand, which coincides with the best-known bound for infeasible interior-point algorithms.

In Chapter 2, a slightly different infeasible primal-dual interior-point algorithm is presented for linear optimization problems. This algorithm is obtained by changing the definition of the search directions in the feasibility step in the algorithm [82]. We show that the same complexity result can be obtained with a relatively simpler analysis.

The results of the Chapter 2 for LO can be extended to other conic optimization problem classes like second order cone and semidefinite optimization. We show this in Chapter 3, where we deal with semidefinite optimization and show that the iteration bound of the algorithm has the same order as in [55, 82].

It looks that similar results hold for the case of second order cone optimization (SOCO), but this is not considered in this thesis.

Chapter 4 was inspired by recent work on so-called kernel function based barrier function for linear optimization [6–9, 74, 76]. These barrier functions have been used to improve the iteration bound for large-update IPMs from $O\left(n \log \frac{n}{\epsilon}\right)$ to $O\left(\sqrt{n}\,(\log n) \log \frac{n}{\epsilon}\right)$. It was observed that the iteration bounds for small-update methods based on these barrier functions were always $O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$. In this chapter we explained this phenomenon. This is achieved by using the powerful tools for analyzing Newton's method provided by the theory of self-concordant functions.

It is shown that barrier functions based on kernel functions are self-concordant in the region where the iterates of a full-Newton step algorithm occur. We called these functions locally self-concordant because they are not self-concordant on the whole domain of the problem under consideration. A surprising result is that the complexity number for the kernel functions analyzed in this chapter are smaller than the complexity number for logarithmic barrier method. This is quite surprising, because it gives evidence to the claim that the classical logarithmic barrier function may not always be the best barrier function to use, at least not from a theoretical point of view.

5.2 Further research

We mention some interesting research topics related to the work presented in this thesis.

- Let us recall from Chapter 2 that based on extensive computational evidence, we conjecture that if ζ is large enough, then $\bar{\kappa}(\zeta) = 1$. How can we prove that this conjecture is true or not?
- By modifying the algorithm in Chapter 2, how we can design a large-update IIPM for LO problems?
- How can we extend the algorithms in Chapter 2 and 3 to other symmetric optimization problems?
- It is possible to design primal (or dual) IIPMs using full-Newton steps for LO? If the result is positive, how we can extend these results to SDO and SOCO problems?
- Can we design primal-dual IIPMs for SDO based on scaling techniques other than the Nesterov-Todd scaling?
- Is it possible to extend the algorithm in Chapter 2 to more general nonlinear optimization problems?
- How can we generalize the results of Chapter 4 to other (symmetric) cone optimization problems, like SOCO and SDO?
- Can we use the approach presented in Chapter 4 to obtain the improved bounds as mentioned in this chapter for large-update methods?

Appendix A

Some concepts in linear algebra

In this appendix we recall some concepts in linear algebra that are used in this thesis. The sources of this appendix are [18, 32, 34, 35, 53, 89].

A.1 Vectors

A vector x is always a column vector, denoted as $x = (x_1; \ldots; x_n) \in \mathbf{R}^n$. Corresponding to each vector x, there exists a row vector called the transpose of the vector x denoted by $x^T = (x_1, \ldots, x_n)$.

Definition A.1.1 (Norm)**.** *A norm $\|\cdot\|$ on a vector space $\mathcal{L}$ over $\mathbf{R}$ is a real-valued function with the following properties:*

(1) $\|\alpha x\| = |\alpha| \, \|x\|$ *for all* $\alpha \in \mathbf{R}$, $x \in \mathcal{L}$,

(2) $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in \mathcal{L}$,

(3) $\|x\| > 0$ *for all nonzero* $x \in \mathcal{L}$ *and* $\|0\| = 0$.

Definition A.1.2 (Inner product)**.** *Let $\mathcal{L}$ be a vector space over $\mathbf{R}$. An inner product on $\mathcal{L}$ is a real-valued function $\langle \cdot, \cdot \rangle$ defined on $\mathcal{L} \times \mathcal{L}$ with the following properties:*

(1) $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ *and* $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ *for all* $\alpha \in \mathbf{R}$ *and* $x, y, z \in \mathcal{L}$,

(2) $\langle x, y\rangle = \langle y, x\rangle$ *for all* $x, y \in \mathcal{L}$,

(3) $\langle x, x\rangle \geq 0$ *for all* $x \in \mathcal{L}$ *and* $\langle x, x\rangle = 0$ *implies that* $x = 0$.

If $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{L}$ then $\|\cdot\|$, defined by $\|x\| = \langle x, x \rangle^{\frac{1}{2}}$, is a norm on $\mathcal{L}$.
A very useful relation between the norm and the inner product is the *Cauchy-Schwartz inequality*,

$$|\langle x, y \rangle| \leq \|x\| \, \|y\| \, . \tag{A.1}$$

For all $x, y \in \mathbf{R}^n$,

$$\langle x, y \rangle := y^T x = \sum_{i=1}^{n} y_i x_i \tag{A.2}$$

defines the standard inner product on $\mathbf{R}^n$. This inner product induces the well-known Euclidean norm on $\mathbf{R}^n$:

$$\|x\| = \langle x, x \rangle^{\frac{1}{2}}.$$

Definition A.1.3. *Two vectors $x, y \in \mathbf{R}^n$ are orthogonal if $\langle x, y \rangle = 0$. A vector x is called normal or unitary if $\|x\| = 1$.*

We denote by $\mathbf{R}_+$ the set of nonnegative real numbers and by $\mathbf{R}_{++}$ the set of positive real numbers. In $\mathbf{R}^n$ there exists a standard partial ordering. Take $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^n$, then this partial ordering is defined by

$$x \leq y \quad \Leftrightarrow \quad y - x \in \mathbf{R}^n_+.$$

By e_i we denote the i-th standard basis element of $\mathbf{R}^n$, that is,

$$e_i = (0; \, \ldots; \, 0; \, 1; \, 0; \, \ldots; \, 0) \, ,$$

where 1 is in the i-th position. By $e \in \mathbf{R}^n$ we denote the all one vector of length n, that is,

$$e = (1; \, 1; \, \ldots; \, 1) \, .$$

A.2 Matrices

We usually denote matrices by capital letters. Let A be an $m \times n$ matrix. Then A^T, the transpose of A, is the matrix obtained by interchanging the rows and columns of the matrix A.

A *symmetric* matrix is a matrix that equals its own transpose, i.e. $A^T = A$.

Definition A.2.1. *If*

$$Ax = \lambda x,$$

where $A \in \mathbf{R}^{n \times n}$, $x \in \mathbf{R}^n$ and λ is a number, then the number λ and vector x are called an eigenvalue and an eigenvector of the matrix A, respectively.

Definition A.2.2. *A matrix A is called nonsingular or invertible if and only if there exists a matrix B such that*

$$AB = I,$$

where I is the identity matrix. The matrix B is called the inverse matrix of A and denoted by A^{-1}.

Definition A.2.3. *Two $n \times n$ matrices A and B are similar, if there exists a non-singular $n \times n$ matrix P such that*

$$B = P^{-1}AP.$$

The trace of a square $n \times n$ matrix A is defined to be the sum of the elements on the main diagonal (diagonal from the upper left to the lower right) of A, i.e.

$$\mathbf{Tr}\,(A) = \sum_{i=1}^{n} A_{ii}, \tag{A.3}$$

where A_{ij} represents the (i, j)-entry of A.

The trace is clearly a linear operator and has the following properties.

Theorem A.2.4. *Let A, $B \in \mathbf{R}^{n\times n}$. Then the following holds:*

(i) $\mathbf{Tr}\,(A) = \sum_{i=1}^{n} \lambda_i\,(A)$, *where $\lambda_i\,(A)$ is the i-th eigenvalue of matrix A;*

(ii) $\mathbf{Tr}\,(A) = \mathbf{Tr}\left(A^T\right)$;

(iii) $\mathbf{Tr}\,(AB) = \mathbf{Tr}\,(BA)$.

(iv) If A and B are similar. Then, $\mathbf{Tr}\,(A) = \mathbf{Tr}\,(B)$.

Let $\mathbf{R}^{n\times n}$ be the space of all square $n \times n$ matrices with real entries and let

$$\mathbf{S}^n := \left\{X \in \mathbf{R}^{n\times n} : X^T = X\right\}$$

be the space of symmetric matrices. A matrix $X \in \mathbf{S}^n$ is positive definite, denoted by $X \succ 0$, if for all nonzero $z \in \mathbf{R}^n$ we have $z^T Xz > 0$, and positive semidefinite if for all z, $z^T Xz \geq 0$. We use the notations $\mathbf{S}^n_+$ and $\mathbf{S}^n_{++}$ to denote the sets of positive semidefinite and positive definite matrices, respectively, i.e.,

$$\begin{aligned}\mathbf{S}^n_+ &= \{X \in \mathbf{S}^n : X \succeq 0\},\\ \mathbf{S}^n_{++} &= \{X \in \mathbf{S}^n : X \succ 0\}.\end{aligned}$$

$X \succeq Y$ and $X \succ Y$ are used to denote the facts $X - Y \succeq 0$ and $X - Y \succ 0$, respectively.

Given two matrices $A = (A_{ij})$, $B = (B_{ij})$ in $\mathbf{R}^{m\times n}$, the standard inner product for matrices is defined by

$$A \bullet B = \sum_{j=1}^{n}\sum_{i=1}^{m} A_{ij}B_{ij} = \mathbf{Tr}\left(A^T B\right). \tag{A.4}$$

The norm associated with this inner product is the Frobenius norm,

$$\|A\|^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2 = \mathbf{Tr}\left(A^T A\right). \tag{A.5}$$

Obviously one has for any symmetric matrix $A \in \mathbf{S}^n$

$$\|A\|^2 = \mathbf{Tr}\left(A^2\right) = \sum_{i=1}^{n} \lambda_i\left(A^2\right) = \sum_{i=1}^{n} \lambda_i\left(A\right)^2. \tag{A.6}$$

We recall from [99] that this norm is sub-multiplicative, i.e.

$$\|AB\| \leq \|A\|\,\|B\|, \quad \text{for any} \quad A,\, B \in \mathbf{R}^{n\times n} \tag{A.7}$$

We conclude this section with a well-known lemma from linear algebra. That provide several useful characterizations of positive semidefinite matrices.

Lemma A.2.5 (Characterizations of positive semidefinte matrices)**.** *The following statements are equivalent for a symmetric matrix $X \in \mathbf{S}^n$:*

(1) X is positive semidefinite.

(2) The eigenvalues of X are nonnegative.

(3) There exists a matrix C such that $X = C^T C$.

(4) X can be represented as $X = Q\Lambda Q^T$, where Q is an orthogonal matrix i.e., $QQ^T = Q^T Q = I$, where I is identity matrix, and Λ is a diagonal matrix with nonnegative entries on the diagonal.

Also, matrix X is positive definite if and only if all its eigenvalues are positive.

We continue by recalling some useful lemmas from linear algebra.

A.3 Linear algebra lemmas

Lemma A.3.1 (Lemma A.1 in [18])**.** *Let $Q \in \mathbf{S}^n_{++}$, and let $M \in \mathbf{R}^{n\times n}$ be skew-symmetric $\left(M = -M^T\right)$. One has $\det(Q+M) > 0$. Moreover, if $\lambda_i(Q+M) \in \mathbf{R}$, $(i = 1, \ldots, n)$, then*

$$0 < \lambda_{\min}(Q) \leq \lambda_{\min}(Q+M) \leq \lambda_{\max}(Q+M) \leq \lambda_{\max}(Q).$$

Proof. First note that $Q + M$ is nonsingular since for all nonzero $x \in \mathbf{R}^n$

$$x^T (Q + M) x = x^T Q x > 0,$$

where for equality we use the skew symmetry of M. Since matrix tM remains skew symmetric for all $t \in \mathbf{R}$, we have

$$\psi(t) := \det (Q + tM) \neq 0 \qquad \forall t \in \mathbf{R}.$$

One may easily see that $\psi(t)$ is a continuous function which is nowhere zero and strictly positive for $t = 0$ as $\det(Q) > 0$. This shows that $\det(Q + M) > 0$.

To prove the second part of the lemma, assume $\lambda > 0$ is such that $\lambda > \lambda_{\max}(Q)$. It then follows that $Q - \lambda I \prec 0$. By the same argument as above we then have $(Q + M) - \lambda I$ nonsingular, or

$$\det ((Q + M) - \lambda I) \neq 0.$$

Which implies that λ cannot be an eigenvalue of $Q + M$. In the same way we can prove that $Q + M$ cannot have an eigenvalue smaller than $\lambda_{\min}(Q)$. This gives the required result. □

Lemma A.3.2 (Lemma 1.2.4 in [32])**.** *Let A, $B \in \mathbf{S}^n_+$. Then we have following inequalities*

$$\lambda_{\min}(A)\lambda_{\max}(B) \leq \lambda_{\min}(A)\mathbf{Tr}(B) \leq \mathbf{Tr}(AB) \leq \lambda_{\max}(A)\mathbf{Tr}(B) \leq n\lambda_{\max}(A)\lambda_{\max}(B).$$

Proof. Since matrix A is positive semidefinite, Lemma A.2.5 implies that there exists an orthonormal matrix P such that $A = P\Lambda P^T$. Then

$$\begin{aligned}
\mathbf{Tr}(AB) &= \mathbf{Tr}\left(P\Lambda P^T B\right) \\
&= \mathbf{Tr}\left(\Lambda P^T B P\right) \\
&\geq \lambda_{\min}(A)\,\mathbf{Tr}\left(P^T B P\right) = \lambda_{\min}(A)\,\mathbf{Tr}(B) \\
&\geq \lambda_{\min}(A)\lambda_{\max}(B).
\end{aligned}$$

Which proves the two left hand side inequalities and the proof of the inequalities in the right hand side is the same. □

Lemma A.3.3 (Theorem A.4 in [18])**.** *Let $X \in \mathbf{S}^n_{++}$ and $S \in \mathbf{S}^n_{++}$. Then all the eigenvalues of XS are real and positive.*

Proof. Since

$$XS \sim \left(X^{\frac{1}{2}}\right)^{-1} XS \left(X^{\frac{1}{2}}\right) = X^{\frac{1}{2}} S X^{\frac{1}{2}} \succ 0,$$

the lemma immediately proved. □

A.4 The vec operator and the Kronecker product

Definition A.4.1 (**vec** Operator)**.** . *The* **vec** *operator maps an* $m \times n$ *matrix into an* $mn \times 1$ *vector. The* **vec** *of matrix* $A\,(m \times n)$*, denoted by* $\mathbf{vec}\,(A)$*, is the vector formed by stacking the columns of* A *into an* $mn \times 1$ *vector.*

$$\mathbf{vec}\,(A) = (A_{11};\ \ldots;\ A_{m1};\ A_{12};\ \ldots;\ A_{m2};\ \ldots;\ A_{1n};\ \ldots;\ A_{mn})$$

Definition A.4.2 (The Kronecker product)**.** *The Kronecker product of* $A\,(m \times n)$ *and* $B\,(p \times q)$ *is denoted by* $A \otimes B$ *and is defined to be the* $(mp \times nq)$ *block matrix*

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{bmatrix}.$$

Notice that $A \otimes B \neq B \otimes A$ *in general.*

A.4.1 Properties of the Kronecker product and the vec operator

In the following it is assumed that A, B, C and D are real valued matrices. Some identities only hold for appropriately dimensioned matrices.

(1) For any $\alpha \in \mathbf{R}$

$$(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha\,(A \otimes B)\,.$$

(2) The Kronecker product distributes over addition

$$\begin{aligned}(A + B) \otimes C &= (A \otimes C) + (B \otimes C) \\ A \otimes (B + C) &= (A \otimes B) + (A \otimes C)\,.\end{aligned}$$

(3) The Kronecker product is associative:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)\,.$$

(4) Transpose distributes over the Kronecker product (does not invert order)

$$(A \otimes B)^T = A^T \otimes B^T.$$

(5) Matrix multiplication, when dimensions are appropriate,

$$(A \otimes B)\,(C \otimes D) = (AC \otimes BD)\,.$$

(6) When A and B are square and full rank

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

(7) For any two matrices $A\,(m \times m)$ and $B\,(n \times n)$

$$\det\,(A \otimes B) = \det\,(A)^m \,.\det\,(B)^n\,.$$

(8) The trace of a Kronecker product is

$$\mathbf{Tr}\,(A \otimes B) = \mathbf{Tr}\,(A)\,.\mathbf{Tr}\,(B)\,.$$

(9) **vec** of a matrix product ABC, when dimensions are appropriate for the product to be well defined, is given by

$$\mathbf{vec}\,(ABC) = \left(C^T \otimes A\right)\mathbf{vec}\,(B)\,.$$

(10) $\mathbf{Tr}\,(AB) = \mathbf{vec}\,(A)^T\,\mathbf{vec}\,(B)\,.$

Bibliography

[1] F. Alizadeh. *Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices*. PhD thesis, University of Minnesota, Minneapolis, Minnesota, USA, 1991. Ph. D. thesis.

[2] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.

[3] F. Alizadeh, J. P. Haeberly, and M. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.

[4] E. D. Andersen and Y. Ye. A computational study of the homogeneous algorithm for large-scale convex optimization. *Computational Applications and Optimization*, 10:243–269, 1998.

[5] E.D. Andersen, J. Gondzio, Cs. Mészáros, and X. Xu. Implementation of interior point methods for large scale linear programming. In T. Terlaky, editor, *Interior Point Methods of Mathematical Programming*, pages 189–252. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[6] Y. Q. Bai, M. El Ghami, and C. Roos. A primal-dual interior-point algorithm for linesr optimization based on a new proximity function. *Optimization Methods and Software*, 17(6):985–1008, 2002.

[7] Y. Q. Bai, M. El Ghami, and C. Roos. A new efficient large-update primal-dual interior-point method based on a finite barrier. *SIAM Journal on Optimization*, 13(3):766–782, 2003.

[8] Y. Q. Bai, M. El Ghami, and C. Roos. A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. *SIAM Journal on Optimization*, 15(1):101–128, 2004.

[9] Y. Q. Bai and C. Roos. A plynomial-time algorithm for linear optimization based on a new simple kernel function. *Optimization Methods and Software*, 18(6):631–646, 2003.

[10] M. S. Bazaraa and J. J. Jarvis. *Linear Programming and Network Flows*. John Wiley & Sons, New York, 1990.

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

[12] J. F. Bonnans and M. C. Ferris. Convergence of an infeasible interior-point algorithm from arbitrary positive starting points. *SIAM Journal on Optimization*, 6(2):316–325, 1996.

[13] J. F. Bonnans and F. A. Potra. On the convergence of the iteration sequence of infeasible path following algorithms for linear complementarity problems. *Mathematics of Operations Research*, 22(2):378–407, 1997.

[14] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics. SIAM, Philadelphia, USA, 1994.

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

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

[17] E. de Klerk. *Interior Point Methods for Semidefinite Programming*. PhD thesis, Faculty of Mathematics and Computer Science, TU Delft, NL–2628 CD Delft, The Netherlands, December 1997. Ph. D. thesis.

[18] E. de Klerk. *Aspects of Semidefintie Programming*, volume 65 of *Applied Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[19] G. Elabwabi. *Topics in global optimization using semidefinite optimization*. PhD thesis, Faculty of Mathematics and Computer Science, TU Delft, NL–2628 CD Delft, The Netherlands, March 2007. Ph. D. thesis.

[20] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, New York, 1968. Reprint: Volume 4 of *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, PA 19104–2688, USA, 1990.

[21] R. M. Freund. A potential-function reduction algorithm for solving a linear program directly from an infeasible 'warm start'. *Mathematical Programming*, 52:441–466, 1991.

[22] K. R. Frisch. The logarithmic potential method for convex programming. Unpublished manuscript, Institute of Economics, University of Oslo, Oslo, Norway, May 1955.

[23] M. El Ghami. *New Primal-Dual Interior Point Methods Based on Kernel Functions*. PhD thesis, Faculty of Mathematics and Computer Science, TU Delft, NL–2628 CD Delft, The Netherlands, Oktober 2005. Ph. D. thesis.

[24] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin, and M. H. Wright. On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method. *Mathematical Programming*, 36:183–209, 1986.

[25] F. Glineur. *Topics in convex optimization: interior-point methods, conic duality and approximations*. PhD thesis, Faculté Polytechnique de Mons, 2001. Ph. D. thesis.

[26] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the Association for Computing Machinery*, 42(6):1115–1145, 1995.

[27] D. Goldfarb and M. J. Todd. Linear Programming. In G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors, *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, pages 141–170. North Holland, Amsterdam, The Netherlands, 1989.

[28] A. J. Goldman and A. W. Tucker. Theory of linear programming. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, Annals of Mathematical Studies, No. 38, pages 53–97. Princeton University Press, Princeton, New Jersey, 1956.

[29] C. C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34(2):167–227, 1992.

[30] O. Güler. Limiting behavior of the weighted central paths in linear programming. *Mathematical Programming*, 60:215–225, 1993.

[31] M. Halická, E. de Klerk, and C. Roos. On the convergence of the central path in semidefinite optimization. *SIAM Journal on Optimization*, 12(4):1090–1099, 2002.

[32] C. Helmberg. Semidefinite programming for combinatorial optimization. Technical Report 00-34, Konrad-Zuse-Zentrum für Informationstechink Berlin, Takustraße 7, D-14195 Berlin, Germany, Germany, Oktober 2000.

[33] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming*, volume 277 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[34] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[35] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1991.

[36] P. Huard. Resolution of mathematical programming with nonlinear constraints by the method of centres. In *Nonlinear Programming (NATO Summer School, Menton, 1964)*, pages 207–219. North-Holland, Amsterdam, 1967.

[37] P. Huard. A method of centers by upper-bounding functions with applications. In J.B. Rosen, O.L. Mangasarian, and K. Ritter, editors, *Nonlinear Programming : Proceedings of a Symposium held at the University of Wisconsin, Madison, Wisconsin, USA, May 1970*, pages 1–30. Academic Press, New York, USA, 1970.

[38] B. Jansen. *Interior Point Techniques in Optimization. Complexity, Sensitivity and Algorithms*. PhD thesis, Faculty of Mathematics and Computer Science, TU Delft, NL–2628 CD Delft, The Netherlands, December 1995. Ph. D. thesis.

[39] B. Jansen, C. Roos, T. Terlaky, and J.-Ph. Vial. Primal-dual algorithms for linear programming based on the logarithmic barrier method. *Journal of Optimization Theory and Applications*, 83:1–26, 1994.

[40] E. S. Karisch, F. Rendl, and J. Clausen. Solving graph bisection problems with semidefinite programming. *INFORMS Journal on Computing*, 12(3):177–191, 2000.

[41] N. K. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[42] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244:1093–1096, 1979. Translated into English in *Soviet Mathematics Doklady* 20, 191–194.

[43] V. Klee and G. J. Minty. How good is the simplex method? *Acadimic Press in inequalities-III, O Shisha, ed New York,*, pages 159–175, 1972.

[44] M. Kojima, N. Megiddo, and S. Mizuno. A primal-dual infeasible-interior-point algorithm for linear programming. *Mathematical Programming*, 61:263–280, 1993.

[45] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 29–47. Springer Verlag, New York, 1989.

[46] M. Kojima, T. Noma, and A. Yoshise. Global convergence in infeasible-interior-point algorithms. *Mathematical Programming*, 65:43–72, 1994.

[47] M. Kojima, M. Shida, and S. Shindoh. Local convergence of predictor-corrector infeasible interior-point algorithm for SDPs and SDLCPs. *Mathematical Programming, Series A*, 80:129–160, 1998.

[48] M. Kojima, M. Shindoh, and S. Hara. Interior point methods for the monotone linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7:86–125, 1997.

[49] J. B. Lasserre. Global optimization with polynomials and the problems of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[50] J. B. Lasserre and T. Prieto-Rumeau. SDP vs. LP Relaxations for the moment approach in some performance evaluation problems. *Stochastic Models*, 20:439–456, 2004.

[51] A. Lisser and F. Rendl. Graph partitioning using linear and semidefinite programming. *Mathematical Programming, Series B*, 95(1):91–101, 2003.

[52] I. J. Lustig. Feasibility issues in aprimal-dual interior point method for linear programming. *Mathematical Programming*, 49:145–162, 1990/91.

[53] H. Lütkepohl. *Handbook of matrices*. John Wiley & Sons, Chichester, 1996.

[54] H. Mansouri and C. Roos. A new full-Newton step $O(n)$ infeasible interior-point algorithm for semidefinite optimization. *Submitted to Journal of Optimization Theory and Applications*.

[55] H. Mansouri and C. Roos. Simplified $O(n)$ infeasible interior-point algorithm for linear optimization using full-Newton step. *Optimization Methods and Software*, 22(3):519–530, 2007.

[56] L. McLinden. The analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88:101–161, 1980.

[57] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 131–158. Springer Verlag, New York, 1989. Identical version in : *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, pages 1–35, 1986.

[58] S. Mizuno. Polynomiality of infeasible interior point algorithms for linear programming. *Mathematical Programming*, 67:109–119, 1994.

[59] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 1993.

[60] S. Mizuno, M. J. Todd, and Y. Ye. A surface of analytic centers and infeasible-interior-point algorithms for linear programming. *Mathematics of Operations Research*, 20:135–162, 1995.

[61] R. D. C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7(3):663–678, 1997.

[62] R. D. C. Monteiro and I. Adler. Interior-path following primal-dual algorithms : Part I : Linear programming. *Mathematical Programming*, 44:27–41, 1989.

[63] Y. Nesterov. *Introductory Lectures on Convex Optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[64] Y. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1993.

[65] Y. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, 22(1):1–42, 1997.

[66] Y. Nesterov and M. J. Todd. Primal-dual interior point methods for self-scaled cones. *SIAM Journal on Optimization.*, 8(2):324–364, 1998.

[67] Y. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, 2(4):333–361, 2002.

[68] Y. Nesterov, M. J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Mathematical Programming*, 84(2):227–267, 1999.

[69] Y. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics. SIAM Publications, Philadelphia, USA, 1994.

[70] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96:293–320, 2003.

[71] J. Peña, J. Vera, and L. F. Zuluaga. Computing the stability number of a graph via linear and semidefinite programming. *SIAM Journal on Optimization*, 18(1):87–105, 2007.

[72] J. Peng, C. Roos, and T. Terlaky. A new and efficient large-update interior-point method for linear optimization. *Journal of Computational Technologies*, 6(4):61–80, 2001.

[73] J. Peng, C. Roos, and T. Terlaky. A new class of polynomial primal-dual methods for linear and semidefinite optimization. *European Journal of Operations Research*, 143(2):234–256, 2002.

[74] J. Peng, C. Roos, and T. Terlaky. Primal-dual interior-point methods for second-order conic optimization based on self-regular proximities. *SIAM Journal on Optimization*, 13(1):179–203, 2002.

[75] J. Peng, C. Roos, and T. Terlaky. Self-regular functions and new search directions for linear and semidefinite optimization. *Mathematical Programming*, 93:129–171, 2002.

[76] J. Peng, C. Roos, and T. Terlaky. *Self-Regularity. A New Paradigm for Primal-Dual Interior-Point Algorithms*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2002.

[77] F. A. Potra. A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. *Mathematical Programming*, 67(3):383–406, 1994.

[78] F. A. Potra. An infeasible-interior-point predictor-corrector algorithm for linear programming. *SIAM Journal on Optimization*, 6(1):19–32, 1996.

[79] F. A. Potra and R. Sheng. A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8:1007–1028, 1998.

[80] J. Renegar. A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40:59–93, 1988.

[81] J. Renegar. *A mathematical view of interior-point methods in convex optimization*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[82] C. Roos. A full-Newton step $O(n)$ infeasible interior-point algorithm for linear optimization. *SIAM Journal on Optimization*, 16(4):1110–1136, 2006.

[83] C. Roos and H. Mansouri. Full-Newton step polynomial-time methods for linear optimization based on locally self-concordant barrier function. *Submitted to Mathematical Programming*.

[84] C. Roos, T. Terlaky, and J. Ph. Vial. *Theory and Algorithms for Linear Optimization. An Interior-Point Approach*. John Wiley & Sons, Chichester, UK, 1997 (2nd Edition, Springer, 2006).

[85] C. Roos and J. Ph. Vial. A polynomial method of approximate centers for the linear programming problem. *Mathematical Programming*, 54:295–306, 1992.

[86] R. Sheng and F. A. Potra. A quadratically convergent infeasible-interior-point algorithm for LCP with polynomial complexity. *SIAM Journal on Optimization*, 7(2):304–317, 1997.

[87] N. Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and System Sciences*, 25:1–11, 1987.

[88] G. Sonnevend. An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In A. Prékopa, J. Szelezsán, and B. Strazicky, editors, *System Modelling and Optimization : Proceedings of the 12th IFIP-Conference held in Budapest, Hungary, September 1985*, volume 84 of *Lecture Notes in Control and Information Sciences*, pages 866–876. Springer Verlag, Berlin, West–Germany, 1986.

[89] G. Strang. *Linear Algebra and its Applications*. Harcourt Brace Jovanovich, Orlando, Florida, USA, 1988.

[90] J. F. Sturm and S. Zhang. Symmetric primal-dual path following algorithms for semidefinite programming. *Applied Numerical Mathematics*, 29(3):301–315, 1999.

[91] A. Suzuka, M. Ryuhei, Y. Akiko, and M. Tomomi. Semidefinite programming based approaches to home-away assignment problems in sports scheduling. *Lecture Notes in Computer Science*, 3521:95–103, 2005.

[92] K. Tanabe. Centered Newton method for mathematical programming. *In System Modelling and Optimization*, pages 197–206, 1988.

[93] K. Tanabe. Centered Newton method for linear programming: Interior and 'exterior' point method (in Janpanese). *In: New Methods for Linear Programming. K. Tone (Ed.) 3*, pages 98–100, 1990.

[94] M. J. Todd, K. C. Toh, and R. H. Tutuncu. On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization*, 8(3):769–796, 1998.

[95] 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.

[96] L. Vandenberghe and S. Boyd. A primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming, Series B*, 69:205–236, 1995.

[97] R. J. Vanderbei. *Linear Programming: Foundations and Extensions*. Kluwer Academic Publishers, Boston, USA, 1996.

[98] G. Q. Wang, Y. Q. Bai, and C. Roos. Primal-dual interior-point algorithms for semidefinite optimization based on a simple kernel function. *Journal of Mathematical Modelling and Algorithms*, 4:409–433, 2005.

[99] D. S. Watkins. *Fundamentals of Matrix Computations*. John Wiley & Sons, New York, 1991.

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

[101] S. J. Wright. A path-following infeasible-interior-point algorithm for linear complementarity problems. *Optimization Methods and Software*, 2:79–106, 1993.

[102] S. J. Wright. An infeasible-interior-point algorithm for linear complementarity problems. *Mathematical Programming*, 67(1):29–52, 1994.

[103] S. J. Wright. *Primal-dual Interior-Point Methods*. SIAM, Philadelphia, 1996.

[104] F. Wu, S. Wu, and Y. Ye. On quadratic convergence of the $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Annals of Operations Research*, 87:393–406, 1999.

[105] Y. Ye. *Interior Point Algorithms, Theory and Analysis*. John Wiley & Sons, Chichester, UK, 1997.

[106] Y. Ye, M. J. Todd, and S. Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19:53–67, 1994.

[107] Y. Zhang. On the convergence of a class of infeasible-interior-point methods for the horizantal linear complementarity problem. *SIAM Journal on Optimization*, 4:208–227, 1994.

[108] Y. Zhang. On extending primal-dual interior-point algorithms from linear programming to semidefinite programming. *SIAM Journal on Optimization*, 8(2):365–386, 1998.

Summary

Full-Newton Step Interior-Point Methods for Conic Optimization

In the theory of polynomial-time interior-point methods (IPMs) two important classes of methods are distinguished: small-update and large-update methods, respectively. Small-update IPMs have the best theoretical iteration bound and IPMs with full-Newton steps belong to this class of methods. Within each of these classes one has feasible and infeasible interior-point methods (IIPMs). In this thesis we first deal with full-Newton step IIPMs (Chapter 2 and Chapter 3), and in Chapter 4 we consider feasible full-Newton step IPMs.

In [82] Roos designed a primal-dual IIPM with full-Newton steps for linear optimization (LO) problems. The algorithm constructs strictly feasible iterates for a sequence of perturbations of the given problem and its dual problem, close to their central paths. Two types of full-Newton steps are used, feasibility steps and (ordinary) centering steps, respectively. In each iteration the algorithm starts with strictly feasible iterates of a perturbed primal-dual pair, on (or close to) its central path, and feasibility steps are used to find strictly feasible iterates for the next perturbed pair. By using centering steps for the new perturbed pair, strictly feasible iterates close to the central path of the new perturbed pair are obtained. During this both the duality gap and the infeasibility are reduced by the same factor. By repeating the same procedure the algorithm terminates in at most $O(n \log \frac{n}{\epsilon})$ steps either by finding an ϵ-solution for the given primal-dual pair, or detecting infeasibility or unboundedness of the given problem.

In Chapter 2, we present a slightly different IIPM for linear optimization. This algorithm is obtained by slightly changing the search direction for the feasibility step as used in [82]. Due to this the analysis of our algorithm is slightly easier than the analysis of the algorithm in [82] at some places, whereas the iteration bound is the same. This means that the iteration bound of our algorithm coin-

cides with the best known iteration bound for IIPMs for LO.

The results of Chapter 2 are extended to semidefinite optimization (SDO) in Chapter 3, where we also obtain an iteration bound that coincides with the best known iteration bound for IIPMs for SDO.

In Chapter 4 we consider (feasible) full-Newton step IPMs that are defined by barrier functions based on so-called kernel functions. Until very recently the theoretical iteration bound for all large-update methods was a factor $\sqrt{n}$ worse than the bound for small-update methods, where n denotes the number of inequalities in the problem. In practice the situation is opposite: in practice large-update methods are much more efficient than small-update methods. This is called the 'gap between theory and practice', also called the 'irony of IPMs' [80]. Some progress has recently been made in this respect by Peng et al. [74, 76] and Bai et al. [6–9]. They introduced a wide class of new barrier functions that are defined by univariate functions, called kernel functions. In these results the gap factor $\sqrt{n}$ was reduced to $\log n$. Remarkably enough, small-update versions of these methods all share the same iteration bound, namely $O(\sqrt{n}\log\frac{n}{\epsilon})$, which is the best known iteration bound for IPMs for LO. We show that the barrier functions underlying these methods are self-concordant in the region where the iterates of a full-Newton step algorithm occur. We therefore call these functions locally self-concordant. Using properties of (locally) self-concordant functions we find an explanation for the fact that these small-update methods have the iteration bound $O(\sqrt{n}\log\frac{n}{\epsilon})$.

Samenvatting

Inwendige Punt Methoden met volle Newton stappen voor Kegeloptimalisering

In de theorie van polynomiale inwendige punt methoden (IPMs) worden twee klassen van methoden onderscheiden: methoden met kleine en methoden met grote herzieningnen (van de barrière parameter). Methoden met kleine herzieningen hebben de beste iteratiegrens, en methoden met volle Newton stappen zijn van deze soort. Binnen beide klassen heeft men respectievelijk 'toelaatbare' en 'ontoelaatbare' methoden (IIPMs). In dit proefschrift kijken we eerst naar ontoelaatbare methoden met volle Newton stappen (in de hoofdstukken 2 en 3) en daarna, in Hoofdstuk 4, naar toelaatbare methoden met volle Newton stappen.

In [82] ontwierp Roos een primaal-duale ontoelaatbare methode met volle Newton stappen voor het oplossen van lineaire optimaliseringsproblemen. Het algoritme construeert strict toelaatbare oplossingen voor een rij van perturbaties van het gegeven probleem en het duale probleem, dichtbij hun centrale paden. Twee typen van volle Newton stappen worden gebruikt, achtereenvolgens toelaatbaar makende stappen en (gewone) centreerstappen. In elke iteratie begint het algoritme met een strict toelaatbare oplossing van een geperturbeerd primaal-duaal paar, op (of bijna op) het centrale pad van dit paar, en een toelaatbaar makende stap genereert een strict toelaatbare oplossing van het volgende geperturbeerde primaal-duale paar, voldoende dicht bij het centrale pad van dit nieuwe paar. Door middel van centreerstappen met betrekking tot het nieuwe geperturbeerde primaal-duale paar wordt een strict toelaatbare oplossing van het nieuwe geperturbeerde primaal-duale paar geconstrueerd op (of bijna op) het centrale pad van dit nieuwe paar. In elke iteratie worden zowel het dualiteitsgat als de primale en duale restvectoren gereduceerd met dezelfde factor. Door dezelfde procedure herhaald uit te voeren genereert het algoritme in ten hoogste $O(n \log \frac{n}{\epsilon})$ iteraties een ϵ-oplossing voor het gegeven primaal-duale paar, of kan worden vastgesteld

dat een van beide problemen onbegrensd dan wel ontoelaatbaar is.

In Hoofdstuk 2 presenteren we een variant van het hierboven beschreven algoritme voor lineaire optimalisering. Dit algoritme ontstaat door de zoekrichting in de toelaatbaar makende stap enigszins te wijzigen ten opzichte van [82]. Het gevolg is dat de analyse iets eenvoudiger wordt dan in [82], terwijl de iteratiegrens dezelfde is. Dit betekent dat de iteratiegrens van het algoritme samenvalt met de best bekende iteratiegrens voor ontoelaatbare inwendige punt methoden voor lineaire optimalisering.

De in Hoofdstuk 2 gepresenteerde resultaten worden in Hoofdstuk 3 gegeneraliseerd naar semidefiniete optimalisering; de verkregen iteratiegrens valt ook hier samen met de best bekende iteratiegrens voor ontoelaatbare inwendige punt methoden voor semidefiniete optimalisering.

In Hoofdstuk 4 beschouwen we (toelaatbare) inwendige punt methoden, met volle Newton stappen, die zijn gebaseerd op zogenaamde kernfuncties. Tot nog maar kort geleden was de theoretische iteratiegrens voor methoden met grote (of gulzige) herzieningen een factor $\sqrt{n}$ slechter dan de iteratiegrens voor methoden met kleine herzieningen, waarbij n het aantal ongelijkheden voorstelt in het gegeven probleem. In de praktijk treedt het omgekeerde op: methoden met gulzige herzieningen zijn veel efficiënter dan methoden met kleine herzieningen. Dit verschijnsel staat bekend als 'de kloof tussen theorie en praktijk', en is ook wel de 'ironie van inwendige punt methoden' genoemd [80]. Enige vooruitgang in dit opzicht werd onlangs gemaakt door Peng et al. [74, 76] and Bai et al. [6–9]. Zij introduceerden een grote klasse van nieuwe barriëre functies die worden gedefinieerd met behulp van functies met één variabele, kernfunctions genoemd. In deze resultaten werd de 'kloof factor' $\sqrt{n}$ vekleind tot $\log n$. Opvallend genoeg is de iteratiegrens voor methoden met kleine herzieningen altijd $O(\sqrt{n}\log\frac{n}{\epsilon})$; dit is de best bekende iteratiegrens voor inwendige punt methoden voor lineaire optimalisering. Wij tonen aan dat de nieuwe barriëre functies 'zelf-gelijkvormig' zijn in het gebied waar de iteranden van een volle Newton stap optreden. We noemen deze functies daarom 'lokaal zelf-gelijkvormig'. Door gebruik te maken van eigenschappen van lokaal zelf-gelijkvormige functies vinden we een verklaring voor het feit dat deze methoden allemaal de best mogelijke iteratiegrens, $O(\sqrt{n}\log\frac{n}{\epsilon})$, hebben.

Curriculum Vitae

Hossein Mansouri was born in Zarand, Kerman, Iran on March 22, 1972. He finished high school in 1990, then he started his studies at the Vali-asr University, Rafsanjan, Iran and received the Bachelor degree with high honor-first position in Applied Mathematics in 1996. He got his master degree in Applied Mathematics at Sharif University of Technology, Tehran, Iran in 1998. From September 1998 until April 2004 he was working as a lecturer at the Department of Mathematics faculty of Basic Sciences at the University of Shahrekord, Shahrekord, Iran.

In 2004, he got a PhD scholarship from the Iranian Ministry of Science, Research and Technology. He started his PhD in April 2004, at the Optimization Group, Department of Software Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, TUDelft, under the supervision of Prof. dr. ir. C. Roos. This thesis covers his PhD study. During this period he got a LNMB (Dutch Network for Mathematics of Operational Research) diploma.

Index