Interior Point Methods for Nonlinear Optimization

Imre Pólik and Tamás Terlaky

1 Introduction

1.1 Historical Background

Interior-point methods (IPMs) are among the most efficient methods for solving linear, and also wide classes of other convex optimization problems. Since the path-breaking work of Karmarkar [48], much research was invested in IPMs. Many algorithmic variants were developed for Linear Optimization (LO). The new approach forced to reconsider all aspects of optimization problems. Not only the research on algorithms and complexity issues, but implementation strategies, duality theory and research on sensitivity analysis got also a new impulse. After more than a decade of turbulent research, the IPM community reached a good understanding of the basics of IPMs. Several books were published that summarize and explore different aspects of IPMs. The seminal work of Nesterov and Nemirovski [63] provides the most general framework for polynomial IPMs for convex optimization. Den Hertog [42] gives a thorough survey of primal and dual path-following IPMs for linear and structured convex optimization problems. Jansen [45] discusses primal-dual target following algorithms for linear optimization and complementarity problems. Wright [93] also concentrates on primal-dual IPMs, with special attention on infeasible IPMs, numerical issues and local, asymptotic convergence properties. The volume [80] contains 13 survey papers that cover almost all aspects of IPMs, their extensions and some applications. The book of Ye [96] is a rich source of polynomial IPMs not only for LO, but for convex optimization problems as well. It extends the IPM theory to derive bounds

I. Pólik and T. Terlaky (✉)
Department of Industrial and Systems Engineering, Lehigh University,
200 West Packer Avenue, 18015-1582, Bethlehem, PA, USA
e-mail: imre@polik.net; terlaky@lehigh.edu

G. Di Pillo et al. (eds.), *Nonlinear Optimization*, Lecture Notes in Mathematics 1989, DOI 10.1007/978-3-642-11339-0_4,

and approximations for classes of nonconvex optimization problems as well. Finally, Roos, Terlaky and Vial [72] present a thorough treatment of the IPM based theory – duality, complexity, sensitivity analysis – and wide classes of IPMs for LO.

Before going in a detailed discussion of our approach, some remarks are made on implementations of IPMs and on extensions and generalizations.

IPMs have also been implemented with great success for linear, conic and general nonlinear optimization. It is now a common sense that for large-scale, sparse, structured LO problems, IPMs are the method of choice and by today all leading commercial optimization software systems contain implementations of IPMs. The reader can find thorough discussions of implementation strategies in the following papers: [5, 53, 55, 94]. The books [72, 93, 96] also devote a chapter to that subject.

Some of the earlier mentioned books [42, 45, 63, 80, 96] discuss extensions of IPMs for classes of nonlinear problems. In recent years the majority of research is devoted to IPMs for nonlinear optimization, specifically for second order (SOCO) and semidefinite optimization (SDO). SDO has a wide range of interesting applications not only in such traditional areas as combinatorial optimization [1], but also in control, and different areas of engineering, more specifically structural [17] and electrical engineering [88]. For surveys on algorithmic and complexity issues the reader may consult [16, 18–20, 63, 64, 69, 75].

In the following sections we will build up the theory gradually, starting with linear optimization and generalizing through conic optimization to nonlinear optimization. We will demonstrate that the main idea behind the algorithms is similar but the details and most importantly the analysis of the algorithms are slightly different.

1.2 Notation and Preliminaries

After years of intensive research a deep understanding of IPMs is developed. There are easy to understand, simple variants of polynomial IPMs. The self-dual embedding strategy [47, 72, 97] provides an elegant solution for the initialization problem of IPMs. It is also possible to build up not only the complete duality theory of [72] of LO, but to perform sensitivity analysis [45, 46, 58, 72] on the basis of IPMs. We also demonstrate that IPMs not only converge to an optimal solution (if it exists), but after a finite number of iterations also allow a strongly polynomial rounding procedure [56, 72] to generate exact solutions. This all requires only the knowledge of elementary calculus and can be taught not only at a graduate, but at an advanced undergraduate level as well. Our aim is to present such an approach, based on the one presented in [72].

This chapter is structured as follows. First, in Section 2.1 we briefly review the general LO problem in canonical form and discuss how Goldman and Tucker's [32, 85] self-dual and homogeneous model is derived. In Section 2.2 the Goldman-Tucker theorem, i.e., the existence of a strictly complementary solution for the skew-symmetric self-dual model will be proved. Here such basic IPM objects, as the interior solution, the central path, the Newton step, the analytic center of polytopes will be introduced. We will show that the central path converges to a strictly complementary solution, and that an exact strictly complementary solution for LO, or a certificate for infeasibility can be obtained after a finite number of iterations. Our theoretical development is summarized in Section 2.3. Finally, in Section 2.4 a general scheme of IPM algorithms is presented. This is the scheme that we refer back to in later sections. In Section 3 we extend the theory to conic (second order and semidefinite) optimization, discuss some applications and present a variant of the algorithm. Convex nonlinear optimization is discussed in Section 4 and a suitable interior point method is presented. Available software implementations are discussed in Section 5. Some current research directions and open problems are discussed in Section 6.

1.2.1 Notation

$\mathbb{R}^n_+$ denotes the set of nonnegative vectors in $\mathbb{R}^n$. Throughout, we use $\|\cdot\|_p$ ($p \in \{1, 2, \infty\}$) to denote the p-norm on $\mathbb{R}^n$, with $\|\cdot\|$ denoting the Euclidean norm $\|\cdot\|_2$. I denotes the identity matrix, e is used to denote the vector which has all its components equal to one. Given an n-dimensional vector x, we denote by X the $n \times n$ diagonal matrix whose diagonal entries are the coordinates x_j of x. If $x,\ s \in \mathbb{R}^n$ then $x^T s$ denotes the dot product of the two vectors. Further, xs, x^α for $\alpha \in \mathbb{R}$ and $\max\{x, y\}$ denotes the vectors resulting from coordinatewise operations. For any matrix $A \in \mathbb{R}^{m\times n}$, A_j denotes the j^{th} column of A. Furthermore,

$$\pi(A) := \prod_{j=1}^{n} \|A_j\|. \tag{1.1}$$

For any index set $J \subseteq \{1, 2, \dots, n\}$, $|J|$ denotes the cardinality of J and $A_J \in \mathbb{R}^{m\times|J|}$ the submatrix of A whose columns are indexed by the elements in J. Moreover, if $K \subseteq \{1, 2, \dots, m\}$, $A_{KJ} \in \mathbb{R}^{|K|\times|J|}$ is the submatrix of A_J whose rows are indexed by the elements in K.

Vectors are assumed to be column vectors. The (vertical) concatenation of two vectors (or matrices of appropriate size) u and v is denoted by $(u; v)$, while the horizontal concatenation is (u, v).

2 Interior Point Methods for Linear Optimization

This section is based on [81]. Here we build the theory of interior point methods for linear optimization including almost all the proofs. In later sections we refer back to these results.

2.1 The Linear Optimization Problem

We consider the general LO problem (P) and its dual (D) in canonical form:

$$\min \left\{ c^T u : Au \geq b, \quad u \geq 0 \right\} \tag{P}$$

$$\max \left\{ b^T v : A^T v \leq c, \quad v \geq 0 \right\}, \tag{D}$$

where A is an $m \times k$ matrix, $b, v \in \mathbb{R}^m$ and $c, u \in \mathbb{R}^k$. It is well known that by using only elementary transformations, any given LO problem can easily be transformed into a "minimal" canonical form. These transformations can be summarized as follows:

- introduce slacks in order to get equations (if a variable has a lower and an upper bound, then one of these bounds is considered as an inequality constraint);
- shift the variables with lower or upper bound so that the respective bound becomes 0 and, if needed replace the variable by its negative;
- eliminate free variables;[1]
- use Gaussian elimination to transform the problem into a form where all equations have a singleton column (i.e., choose a basis and multiply the equations by the inverse basis) while dependent constraints are eliminated.

The weak duality theorem for the canonical LO problem is easily proved.

[1] Free variables can easily be eliminated one-by-one. If we assume that x_1 is a free variable and has a nonzero coefficient in a constraint, e.g., we have

$$\sum_{i=1}^{n} \alpha_i x_i = \beta$$

with $\alpha_1 \neq 0$, then we can express x_1 as

$$x_1 = \frac{\beta}{\alpha_1} - \sum_{i=1}^{n-1} \frac{\alpha_i}{\alpha_1} x_i. \tag{2.1}$$

Because x_1 has no lower or upper bounds, this expression for x_1 can be substituted into all the other constraints and in the objective function.

Theorem 2.1 (Weak duality for linear optimization). *Let us assume that $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^m$ are feasible solutions for the primal problem (P) and dual problem (D), respectively. Then one has*

$$c^T u \geq b^T v$$

where equality holds if and only if

(i) $u_i(c - A^T v)_i = 0$ *for all* $i = 1, \dots, k$ *and*
(ii) $v_j(Au - b)_j = 0$ *for all* $j = 1, \dots, m$.[2]

Proof. Using primal and dual feasibility of u and v we may write

$$(c - A^T v)^T u \geq 0 \qquad \text{and} \qquad v^T(Au - b) \geq 0$$

with equality if and only if (i), respectively (ii) holds. Summing up these two inequalities we have the desired inequality

$$0 \leq (c - A^T v)^T u + v^T(Au - b) = c^T u - b^T v.$$

The theorem is proved. □

One easily derives the following sufficient condition for optimality.

Corollary 2.2. *Let a primal and dual feasible solution $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^m$ with $c^T u = b^T v$ be given. Then u is an optimal solution of the primal problem (P) and v is an optimal solution of the dual problem (D).* □

The Weak Duality Theorem provides a sufficient condition to check optimality of a feasible solution pair. However, it does not guarantee that, in case of feasibility, an optimal pair with zero duality gap always exists. This is the content of the so-called Strong Duality Theorem that we are going to prove in the next sections by using only simple calculus and basic concepts of IPMs.

As we are looking for optimal solutions of the LO problem with zero duality gap, we need to find a solution of the system formed by the primal and the dual feasibility constraints and by requiring that the dual objective is at least as large as the primal one. By the Weak Duality Theorem (Thm. 2.1) we know that any solution of this system is both primal and dual feasible with equal objective values. Thus, by Corollary 2.2, they are optimal. By introducing appropriate slack variables the following inequality system is derived.

$$\begin{aligned} Au - z = b, &\quad u \geq 0, \quad z \geq 0 \\ A^T v + w = c, &\quad v \geq 0, \quad w \geq 0 \\ b^T v - c^T u - \rho = 0, &\quad \rho \geq 0. \end{aligned} \tag{2.2}$$

[2] These conditions are in general referred to as the *complementarity conditions*. Using the coordinatewise notation we may write $u(c - A^T v) = 0$ and $v(Au - b) = 0$. By the weak duality theorem complementarity and feasibility imply optimality.

By homogenizing, the *Goldman-Tucker model* [32, 85] is obtained.

$$
\begin{array}{rrrrrrll}
 & Au & -\tau b & -z & & & = 0, & u \geq 0, \quad z \geq 0 \\
-A^T v & & +\tau c & & -w & & = 0, & v \geq 0, \quad w \geq 0 \\
b^T v & -c^T u & & & & -\rho & = 0, & \tau \geq 0, \quad \rho \geq 0.
\end{array}
\tag{2.3}
$$

One easily verifies that if (v, u, τ, z, w, ρ) is a solution of the Goldman-Tucker system (2.3), then $\tau\rho > 0$ cannot hold. Indeed, if $\tau\rho$ were positive then the we would have

$$0 < \tau\rho = \tau b^T v - \tau c^T u = u^T A^T v - z^T v - v^T A u - w^T u = -z^T v - w^T u \leq 0$$

yielding a contradiction.

The homogeneous Goldman-Tucker system admits the trivial zero solution, but that has no value for our discussions. We are looking for some specific nontrivial solutions of this system. Clearly any solution with $\tau > 0$ gives a primal and dual optimal pair $(\frac{u}{\tau}, \frac{v}{\tau})$ with zero duality gap because ρ must be zero if $\tau > 0$. On the other hand, any optimal pair (u, v) with zero duality gap is a solution of the Goldman-Tucker system with $\tau = 1$ and $\rho = 0$.

Finally, if the Goldman-Tucker system admits a nontrivial feasible solution $(\overline{v}, \overline{u}, \overline{\tau}, \overline{z}, \overline{w}, \overline{\rho})$ with $\overline{\tau} = 0$ and $\overline{\rho} > 0$, then we may conclude that either (P), or (D), or both of them are infeasible. Indeed, $\overline{\tau} = 0$ implies that $A\overline{u} \geq 0$ and $A^T\overline{v} \leq 0$. Further, if $\overline{\rho} > 0$ then we have either $b^T\overline{v} > 0$, or $c^T\overline{u} < 0$, or both. If $b^T\overline{v} > 0$, then by assuming that there is a feasible solution $u \geq 0$ for (P) we have

$$0 < b^T\overline{v} \leq u^T A^T \overline{v} \leq 0$$

which is a contradiction, thus if $b^T\overline{v} > 0$, then (P) must be infeasible. Similarly, if $c^T\overline{u} < 0$, then by assuming that there is a dual feasible solution $v \geq 0$ for (D) we have

$$0 > c^T\overline{u} \geq v^T A \overline{u} \geq 0$$

which is a contradiction, thus if $c^T\overline{u} > 0$, then (D) must be infeasible.

Summarizing the results obtained so far, we have the following theorem.

Theorem 2.3. *Let a primal dual pair (P) and (D) of LO problems be given. The following statements hold for the solutions of the Goldman-Tucker system* (2.3).

1. *Any optimal pair (u, v) of (P) and (D) with zero duality gap is a solution of the corresponding Goldman-Tucker system with $\tau = 1$.*
2. *If (v, u, τ, z, w, ρ) is a solution of the Goldman-Tucker system then either $\tau = 0$ or $\rho = 0$, i.e., $\tau\rho > 0$ cannot happen.*
3. *Any solution (v, u, τ, z, w, ρ) of the Goldman-Tucker system, where $\tau > 0$ and $\rho = 0$, gives a primal and dual optimal pair $(\frac{u}{\tau}, \frac{v}{\tau})$ with zero duality gap.*

4. *If the Goldman-Tucker system admits a feasible solution* $(\overline{v}, \overline{u}, \overline{\tau}, \overline{z}, \overline{w}, \overline{\rho})$ *with* $\overline{\tau} = 0$ *and* $\overline{\rho} > 0$*, then we may conclude that either* (P)*, or* (D)*, or both of them are infeasible.*

Our interior-point approach will lead us to a solution of the Goldman-Tucker system, where either $\tau > 0$ or $\rho > 0$, avoiding the undesired situation when $\tau = \rho = 0$.

Before proceeding, we simplify our notations. Observe that the Goldman-Tucker system can be written in the following compact form

$$Mx \geq 0, \qquad x \geq 0, \qquad s(x) = Mx, \tag{2.4}$$

where

$$x = \begin{pmatrix} v \\ u \\ \tau \end{pmatrix}, \qquad s(x) = \begin{pmatrix} z \\ w \\ \rho \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix}$$

is a skew-symmetric matrix, i.e., $M^T = -M$. The Goldman-Tucker theorem [32, 72, 85] says that system (2.4) admits a strictly complementary solution. This theorem will be proved in the next section.

Theorem 2.4 (Goldman, Tucker). *System* (2.4) *has a strictly complementary feasible solution, i.e., a solution for which* $x + s(x) > 0$.

Observe that this theorem ensures that either case 3 or case 4 of Theorem 2.3 must occur when one solves the Goldman-Tucker system of LO. This is in fact the strong duality theorem of LO.

Theorem 2.5. *Let a primal and dual LO problem be given. Exactly one of the following statements hold:*

- (P) *and* (D) *are feasible and there are optimal solutions* u^* *and* v^* *such that* $c^T u^* = b^T v^*$.
- *Either problem* (P)*, or* (D)*, or both are infeasible.*

Proof. Theorem 2.4 implies that the Goldman-Tucker system of the LO problem admits a strictly complementary solution. Thus, in such a solution, either $\tau > 0$, and in that case item 3 of Theorem 2.3 implies the existence of an optimal pair with zero duality gap. On the other hand, when $\rho > 0$, item 4 of Theorem 2.3 proves that either (P) or (D) or both are infeasible. $\square$

Our next goal is to give an elementary constructive proof of Theorem 2.4. When this project is finished, we have the complete duality theory for LO.

2.2 The Skew-Symmetric Self-Dual Model

2.2.1 Basic Properties of the Skew-Symmetric Self-Dual Model

Following the approach in [72] we make our skew-symmetric model (2.4) a bit more general. Thus our prototype problem is

$$\min \left\{ q^T x : Mx \geq -q,\, x \geq 0 \right\}, \tag{SP}$$

where the matrix $M \in \mathbb{R}^{n \times n}$ is *skew-symmetric* and $q \in \mathbb{R}^n_+$. The set of feasible solutions of (SP) is denoted by

$$SP := \{ x \ : \ x \geq 0,\ Mx \geq -q \ \}.$$

By using the assumption that the coefficient matrix M is skew-symmetric and the right-hand-side vector $-q$ is the negative of the objective coefficient vector, one easily verifies that the dual of (SP) is equivalent to (SP) itself, i.e., problem (SP) is *self-dual*. Due to the self-dual property the following result is trivial.

Lemma 2.6. *The optimal value of* (SP) *is zero and (SP) admits the zero vector $x = 0$ as a feasible and optimal solution.*

Given $(x, s(x))$, where $s(x) = Mx + q$ we may write

$$q^T x = x^T (s(x) - Mx) = x^T s(x) = e^T (xs(x)),$$

i.e., for any optimal solution $e^T(xs(x)) = 0$ implying that the vectors x and $s(x)$ are complementary. For further use, the *optimal set* of (SP) is denoted by

$$SP^* := \{ x \ : \ x \geq 0,\ s(x) \geq 0,\ xs(x) = 0 \}.$$

A useful property of optimal solutions is given by the following lemma.

Lemma 2.7. *Let x and y be feasible for (SP). Then x and y are optimal if and only if*

$$xs(y) = ys(x) = xs(x) = ys(y) = 0. \tag{2.5}$$

Proof. Because M is skew-symmetric we have $(x - y)^T M (x - y) = 0$, which implies that $(x - y)^T (s(x) - s(y)) = 0$. Hence $x^T s(y) + y^T s(x) = x^T s(x) + y^T s(y)$ and this vanishes if and only if x and y are optimal. □

Thus, optimal solutions are complementary in the general sense, i.e., they are not only complementary w.r.t. their own slack vector, but complementary w.r.t. the slack vector for any other optimal solution as well.

All of the above results, including to find a trivial optimal solution were straightforward for (SP). The only nontrivial result that we need to prove is the existence of a strictly complementary solution.

First we prove the existence of a strictly complementary solution if the so-called interior-point condition holds.

Assumption 2.8 (Interior-Point Condition (IPC))
There exists a point $x^0 \in SP$ such that

$$(x^0, s(x^0)) > 0. \tag{2.6}$$

Before proceeding, we show that this condition can be assumed without loss of generality. If the reader is eager to know the proof of the existence of a strictly complementary solution for the self dual model (SP), he/she might temporarily skip the following subsection and return to it when all the results for the problem (SP) are derived under the IPC.

2.2.2 IPC for the Goldman-Tucker Model

Recall that (SP) is just the abstract model of the Goldman-Tucker problem (2.4) and our goal is to prove Theorem 2.4. In order to apply the results of the coming sections we need to modify problem (2.4) so that the resulting equivalent problem satisfies the IPC.

Self-dual embedding of (2.4) with IPC

Due to the second statement of Theorem 2.3, problem (2.4) cannot satisfy the IPC. However, because problem (2.4) is just a homogeneous feasibility problem, it can be transformed into an equivalent problem (SP) which satisfies the IPC. This happens by enlarging, i.e., embedding the problem and defining an appropriate nonnegative vector q.

Let us take $x = s(x) = e$. These vectors are positive, but they do not satisfy (2.4). Let us further define the error vector r obtained this way by

$$r := e - Me, \quad \text{and let} \quad \lambda := n + 1.$$

Then we have

$$\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix} \begin{pmatrix} e \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda \end{pmatrix} = \begin{pmatrix} Me + r \\ -r^T e + \lambda \end{pmatrix} = \begin{pmatrix} e \\ 1 \end{pmatrix}. \tag{2.7}$$

Hence, the following problem

$$\min \left\{ \lambda\vartheta \;:\; -\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \vartheta \end{pmatrix} + \begin{pmatrix} s \\ \nu \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda \end{pmatrix};\; \begin{pmatrix} x \\ \vartheta \end{pmatrix}, \begin{pmatrix} s \\ \nu \end{pmatrix} \geq 0 \right\} \tag{$\overline{\text{SP}}$}$$

satisfies the IPC because for this problem the all-one vector is feasible. This problem is in the form of (SP), where

$$\overline{M} = \begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}, \qquad \overline{x} = \begin{pmatrix} x \\ \vartheta \end{pmatrix} \qquad \text{and} \qquad \bar{q} = \begin{pmatrix} 0 \\ \lambda \end{pmatrix}.$$

We claim that finding a strictly complementary solution to (2.4) is equivalent to finding a strictly complementary optimal solution to problem ($\overline{\text{SP}}$). This claim is valid, because ($\overline{\text{SP}}$) satisfies the IPC and thus, as we will see, it admits a strictly complementary optimal solution. Because the objective function is just a constant multiple of ϑ, this variable must be zero in any optimal solution, by Lemma 2.6. This observation implies the claimed result.

Conclusion

Every LO problem can be embedded in a self-dual problem ($\overline{\text{SP}}$) of the form (SP). This can be done in such a way that $\overline{x} = e$ is feasible for ($\overline{\text{SP}}$) and $\overline{s}(e) = e$. Having a strictly complementary solution of (SP) we either find an optimal solution of the embedded LO problem, or we can conclude that the LO problem does not have an optimal solution.

After this intermezzo, we return to the study of our prototype problem (SP) by assuming the IPC.

2.2.3 The Level Sets of (SP)

Let $x \in SP$ and $s = s(x)$ be a *feasible pair*. Due to self duality, the *duality gap* for this pair is twice the value

$$q^T x = x^T s,$$

however, for the sake of simplicity, the quantity $q^T x = x^T s$ itself will be referred to as *the duality gap*. First we show that the IPC implies the boundedness of the level sets.

Lemma 2.9. *Let the IPC be satisfied. Then, for each positive K, the set of all feasible pairs (x, s) such that $x^T s \le K$ is bounded.*

Proof. Let (x^0, s^0) be an interior-point. Because the matrix M is skew-symmetric, we may write

$$\begin{aligned} 0 &= (x - x^0)^T M (x - x^0) = (x - x^0)^T (s - s^0) \\ &= x^T s + (x^0)^T s^0 - x^T s^0 - s^T x^0. \end{aligned} \tag{2.8}$$

From here we get

$$x_j s_j^0 \le x^T s^0 + s^T x^0 = x^T s + (x^0)^T s^0 \le K + (x^0)^T s^0.$$

The proof is complete. □

In particular, this lemma implies that the set of optimal solutions SP^* is bounded as well.[3]

2.2.4 Central Path, Optimal Partition

First we define the central path [23, 27, 54, 74] of (SP).

Definition 2.11. Let the IPC be satisfied. The set of solutions

$$\{(x(\mu), s(x(\mu))) \,:\, Mx + q = s, \;\; xs = \mu e, \;\; x > 0 \quad \text{for some} \quad \mu > 0\} \tag{2.9}$$

is called the *central path* of (SP).

If no confusion is possible, instead of $s(x(\mu))$ the notation $s(\mu)$ will be used. Now we are ready to present our main theorem. This in fact establishes the existence of the central path. At this point our discussion deviates from the one presented in [72]. The proof presented here is more elementary because it does not make use of the logarithmic barrier function.

Theorem 2.12. *The next statements are equivalent.*

i. (SP) satisfies the interior-point condition;
ii. For each $0 < \mu \in \mathbb{R}$ there exists $(x(\mu), s(\mu)) > 0$ such that

$$\begin{aligned} Mx + q &= s \\ xs &= \mu e. \end{aligned} \tag{2.10}$$

iii. For each $0 < w \in \mathbb{R}^n$ there exists $(x, s) > 0$ such that

$$\begin{aligned} Mx + q &= s \\ xs &= w. \end{aligned} \tag{2.11}$$

---

[3] The following result shows that the IPC not only implies the boundedness of the level sets, but the converse is also true. We do not need this property in developing our main results, so this is presented without proof.

Corollary 2.10. *Let* (SP) *be feasible. Then the following statements are equivalent:*

i. the interior-point condition is satisfied;
ii. the level sets of $x^T s$ are bounded;
iii. the optimal set SP^* of (SP) *is bounded.*

Moreover, the solutions of these systems are unique.

Before proving this highly important result we introduce the notion of optimal partition and present our main result. The partition (B, N) of the index set $\{1, ..., n\}$ given by

$$B := \{i \ : \ x_i > 0, \ \ \text{for some} \ \ x \in SP^*\}, \tag{2.12a}$$

$$N := \{i \ : \ s(x)_i > 0, \ \ \text{for some} \ \ x \in SP^*\}, \tag{2.12b}$$

is called the *optimal partition*. By Lemma 2.7 the sets B and N are disjoint. Our main result says that the central path converges to a strictly complementary optimal solution, and this result proves that $B \cup N = \{1, ..., n\}$. When this result is established, the Goldman-Tucker theorem (Theorem 2.4) for the general LO problem is proved because we use the embedding method presented in Section 2.2.2.

Theorem 2.13. *If the IPC holds then there exists an optimal solution* x^* *and* $s^* = s(x^*)$ *of problem* (SP) *such that* $x^*_B > 0$, $s^*_N > 0$ *and* $x^* + s^* > 0$.

First we prove Theorem 2.12.

Proof. We start the proof by demonstrating that the systems in (*ii*) and (*iii*) may have at most one solution. Because (*ii*) is a special case of (*iii*), it is sufficient to prove uniqueness for (*iii*).

Let us assume to the contrary that for a certain $w > 0$ there are two vectors $(x, s) \neq (\overline{x}, \overline{s}) > 0$ solving (*iii*). Then using the fact that matrix M is skew-symmetric, we may write

$$0 = (x - \overline{x})^T M (x - \overline{x}) = (x - \overline{x})^T (s - \overline{s}) = \sum_{x_i \neq \overline{x}_i} (x - \overline{x})_i (s - \overline{s})_i. \tag{2.13}$$

Due to $xs = w = \overline{x}\,\overline{s}$ we have

$$x_i < \overline{x}_i \Longleftrightarrow s_i > \overline{s}_i \tag{2.14a}$$

$$x_i > \overline{x}_i \Longleftrightarrow s_i < \overline{s}_i. \tag{2.14b}$$

By considering these sign properties one easily verifies that the relation

$$0 = \sum_{x_i \neq \overline{x}_i} (x - \overline{x})_i (s - \overline{s})_i < 0 \tag{2.15}$$

should hold, but this is an obvious contradiction. As a result, we may conclude that if the systems in (*ii*) and (*iii*) admit a feasible solution, then such a solution is unique. □

The Newton step

In proving the existence of a solution for the systems in (*ii*) and (*iii*) our main tool is a careful analysis of the Newton step when applied to the nonlinear systems in (*iii*).[4]

Let a vector $(x, s) > 0$ with $s = Mx + q$ be given. For a particular $w > 0$ one wants to find the displacement $(\Delta x, \Delta s)$ that solves

$$\begin{aligned} M(x + \Delta x) + q &= s + \Delta s \\ (x + \Delta x)(s + \Delta s) &= w. \end{aligned} \tag{2.16}$$

This reduces to

$$\begin{aligned} M\Delta x &= \Delta s \\ x\Delta s + s\Delta x + \Delta x\Delta s &= w - xs. \end{aligned} \tag{2.17}$$

This equation system is still nonlinear. When we neglect the second order term $\Delta x\Delta s$ the *Newton equation*

$$\begin{aligned} M\Delta x &= \Delta s \\ x\Delta s + s\Delta x &= w - xs \end{aligned} \tag{2.18}$$

is obtained. This is a linear equation system and the reader easily verifies that the *Newton direction* Δx is the solution of the nonsingular system of equations[5]

$$(M + X^{-1}S)\Delta x = x^{-1}w - s. \tag{2.19}$$

When we perform a step in the Newton direction with step-length α, for the new solutions $(x^+, s^+) = (x + \alpha\Delta x, s + \alpha\Delta s)$ we have

$$\begin{aligned} x^+s^+ &= (x + \alpha\Delta x)(s + \alpha\Delta s) = xs + \alpha(x\Delta s + s\Delta x) + \alpha^2\Delta x\Delta s \\ &= xs + \alpha(w - xs) + \alpha^2\Delta x\Delta s. \end{aligned} \tag{2.20}$$

This relation clarifies that the local change of xs is determined by the vector $w - xs$. Luckily this vector is known in advance when we apply a Newton step, thus for sufficiently small α we know precisely which coordinates of xs

[4] Observe that no preliminary knowledge on any variants of Newton's method is assumed. We just define and analyze the Newton step for our particular situation.

[5] Nonsingularity follows from the fact that the sum of a skew-symmetric, thus positive semi-definite, and a positive definite matrix is positive definite. Although it is not advised to use for numerical computations, the Newton direction can be expressed as $\Delta x = (M + X^{-1}S)^{-1}\ (x^{-1}w - s)$.

decrease locally (precisely those for which the related coordinate of $w - xs$ is negative) and which coordinate of xs increase locally (precisely those for which the related coordinate of $w - xs$ is positive).

The equivalence of the three statements in Theorem 2.12.

Clearly (ii) is a special case of (iii) and the implication $(ii) \to (i)$ is trivial.

It only remains to be proved that (i), i.e., the IPC, ensures that for each $w > 0$ the nonlinear system in (iii) is solvable. To this end, let us assume that an $x^0 \in SP$ with $(x^0, s(x^0)) > 0$ is given. We use the notation $w^0 := x^0 s(x^0)$. The claim is proved in two steps.

Step 1. *For each* $0 < \underline{w} < \overline{w} \in \mathbb{R}^n$ *the following two sets are compact:*

$$L_{\overline{w}} := \{x \in SP \,:\, xs(x) \leq \overline{w}\} \text{ and}$$
$$U(\underline{w}, \overline{w}) := \{w \,:\, \underline{w} \leq w \leq \overline{w},\ w = xs(x) \text{ for some } x \in L_{\overline{w}}\}.$$

Let us first prove that $L_{\overline{w}}$ is compact. For each $\overline{w} > 0$, the set $L_{\overline{w}}$ is obviously closed. By definition $L_{\overline{w}}$ is included in the level set $x^T s \leq e^T \overline{w}$, which by Lemma 2.9 is bounded, thus $L_{\overline{w}}$ is compact.

By definition the set $U(\underline{w}, \overline{w})$ is bounded. We only need to prove that it is closed. Let a convergent sequence $w^i \to \hat{w}$, $w^i \in U(\underline{w}, \overline{w})$, $i = 1, 2, \ldots$ be given. Then clearly $\underline{w} \leq \hat{w} \leq \overline{w}$ holds. Further, for each i there exists $x^i \in L_{\overline{w}}$ such that $w^i = x^i s(x^i)$. Because the set $L_{\overline{w}}$ is compact, there is an $\hat{x} \in L_{\overline{w}}$ and a convergent subsequence $x^i \to \hat{x}$ (for ease of notation the subsequence is denoted again the same way). Then we have $\hat{x}s(\hat{x}) = \hat{w}$, proving that $U(\underline{w}, \overline{w})$ is closed, thus compact.

Observe that for each $w \in U(\underline{w}, \overline{w})$ by definition we have an $x \in SP$ with $w = xs(x)$. Due to $w > 0$ this relation implies that $x > 0$ and $s(x) > 0$.

Step 2. *For each* $\hat{w} > 0$, *the system* $Mx + q = s$, $xs = \hat{w}$, $x > 0$ *has a solution.*

If we have $\hat{w} = w^0 = x^0 s(x^0)$, then the claim is trivial. If $\hat{w} \neq w^0$ then we define $\overline{w} := \max\{\hat{w}, w^0\}$, $\overline{\eta} = \|\overline{w}\|_\infty + 1$, $\underline{w} := \min\{\hat{w}, w^0\}$ and $\underline{\eta} = \frac{1}{2} \min_i \underline{w}_i$. Then $\underline{\eta} e < \hat{w} < \overline{\eta} e$ and $\underline{\eta} e < w^0 < \overline{\eta} e$. Due to the last relation the set $\overline{U} := U(\underline{\eta} e, \overline{\eta} e)$ is nonempty and compact. We define the nonnegative function $d(w) : \overline{U} \to \mathbb{R}$ as

$$d(w) := \|w - \hat{w}\|_\infty.$$

The function $d(w)$ is continuous on the compact set $\overline{U}$, thus it attains its minimum

$$\tilde{w} := \arg\min_{w \in \overline{U}} \{d(w)\}.$$

If $d(\tilde{w}) = 0$, then $\tilde{w} = \hat{w} \Rightarrow \hat{w} \in \overline{U}$ and hence by the definition of $\overline{U}$ there is an $x \in SP$ satisfying $xs(x) = \hat{w}$ and the claim is proved.

If $d(\tilde{w}) > 0$ then we will show that a damped Newton step from $\tilde{w}$ towards $\hat{w}$ gives a point $w(\mathbf{a}) \in \overline{U}$ such that $d(w(\mathbf{a})) < d(\tilde{w})$, contradicting the fact that $\tilde{w}$ minimizes $d(w)$. This situation is illustrated in Figure 1.

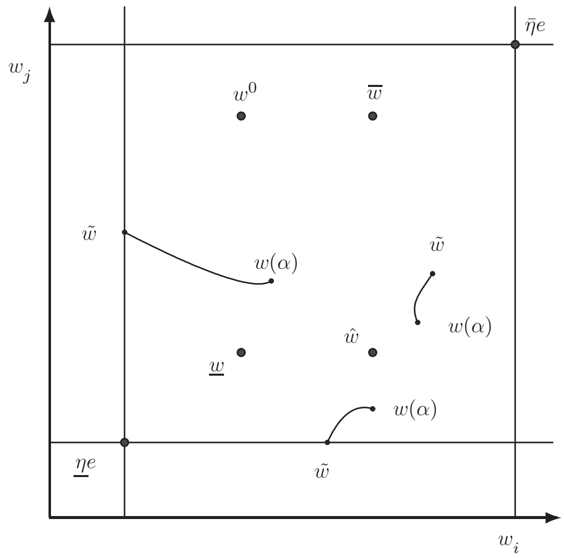

Fig. 1 The situation when $\hat{w} \neq \tilde{w}$. A damped Newton step from $\tilde{w}$ to $\hat{w}$ is getting closer to $\hat{w}$. For illustration three possible different $\tilde{w}$ values are chosen.

The Newton step is well defined, because for the vector $\tilde{x} \in SP$ defining $\tilde{w}$ the relations $\tilde{x} > 0$ and $\tilde{s} = s(\tilde{x}) > 0$ hold. A damped Newton step from $\tilde{w}$ to $\hat{w}$ with sufficiently small α results in a point closer (measured by $d(\cdot) = \|\cdot\|_\infty$) to $\hat{w}$, because

$$\begin{aligned} w(\alpha) = x(\alpha)s(\alpha) &:= (\tilde{x} + \alpha \Delta x)(\tilde{s} + \alpha \Delta s) = \tilde{x}\tilde{s} + \alpha(\hat{w} - \tilde{x}\tilde{s}) + \alpha^2 \Delta x \Delta s \\ &= \tilde{w} + \alpha(\hat{w} - \tilde{w}) + \alpha^2 \Delta x \Delta s. \end{aligned} \tag{2.21}$$

This relation implies that

$$w(\mathbf{a}) - \hat{w} = (1 - \alpha)(\tilde{w} - \hat{w}) + \alpha^2 \Delta x \Delta s, \tag{2.22}$$

i.e., for α small enough[6] all nonzero coordinates of $|w(\mathbf{a}) - \hat{w}|$ are smaller than the respective coordinates of $|\tilde{w} - \hat{w}|$. Hence, $w(\mathbf{a})$ is getting closer to $\hat{w}$, closer than $\tilde{w}$. Due to $\underline{\eta}e < \hat{w} < \overline{\eta}e$ this result also implies that for the chosen

[6] The reader easily verifies that any value of α satisfying

$$\alpha < \min\left\{ \frac{\tilde{w}_i - \hat{w}_i}{\Delta x_i \Delta s_i} \ : \ (\tilde{w}_i - \hat{w}_i)(\Delta x_i \Delta s_i) > 0 \right\}$$

satisfies the requirement.

small **a** value the vector $w(\mathbf{a})$ stays in $\overline{U}$. Thus $\tilde{w} \neq \hat{w}$ cannot be a minimizer of $d(w)$, which is a contradiction. This completes the proof. □

Now we are ready to prove our main theorem, the existence of a strictly complementary solution, when the IPC holds.

Proof of Theorem 2.13.

Let $\mu_t \to 0$ $(t = 1, 2, \cdots)$ be a monotone decreasing sequence, hence for all t we have $x(\mu_t) \in L_{\mu_1 e}$. Because $L_{\mu_1 e}$ is compact the sequence $x(\mu_t)$ has an accumulation point x^* and without loss of generality we may assume that $x^* = \lim\limits_{t\to\infty} x(\mu_t)$. Let $s^* := s(x^*)$. Clearly x^* is optimal because

$$x^* s^* = \lim_{t\to\infty} x(\mu_t) s(x(\mu_t)) = \lim_{t\to\infty} \mu_t e = 0. \tag{2.23}$$

We still have to prove that $(x^*, s(x^*))$ is strictly complementary, i.e., $x^* + s^* > 0$. Let $\mathbf{B} = \{i : x_i^* > 0\}$ and $\mathbf{N} = \{i : s_i^* > 0\}$. Using the fact that M is skew-symmetric, we have

$$0 = (x^* - x(\mu_t))^T (s^* - s(\mu_t)) = x(\mu_t)^T s(\mu_t) - {x^*}^T s(\mu_t) - x(\mu_t)^T s^*, \tag{2.24}$$

which, by using that $x(\mu_t)_i s(\mu_t)_i = \mu_t$, can be rewritten as

$$\sum_{i\in\mathbf{B}} x_i^* s(\mu_t)_i + \sum_{i\in\mathbf{N}} s_i^* x(\mu_t)_i = n\mu_t, \tag{2.25a}$$

$$\sum_{i\in\mathbf{B}} \frac{x_i^*}{x(\mu_t)_i} + \sum_{i\in\mathbf{N}} \frac{s_i^*}{s(\mu_t)_i} = n. \tag{2.25b}$$

By taking the limit as μ_t goes to zero we obtain that

$$|\mathbf{B}| + |\mathbf{N}| = n,$$

i.e., $(\mathbf{B}, \mathbf{N})$ is a partition of the index set, hence $(x^*, s(x^*))$ is a strictly complementary solution. The proof of Theorem 2.13 is complete. □

As we mentioned earlier, this result is powerful enough to prove the strong duality theorem of LO in the strong form, including strict complementarity, i.e., the Goldman-Tucker Theorem (Thm. 2.4) for SP and for (P) and (D).

Our next step is to prove that the accumulation point x^* is unique.

2.2.5 Convergence to the Analytic Center

In this subsection we prove that the central path has only one accumulation point, i.e., it converges to a unique point, the so-called analytic center [74] of the optimal set SP^*.

Definition 2.14. Let $\overline{x} \in SP^*$, $\overline{s} = s(\overline{x})$ maximize the product

$$\prod_{i\in\mathbf{B}} x_i \prod_{i\in\mathbf{N}} s_i \tag{2.26}$$

over $x \in SP^*$. Then $\overline{x}$ is called the *analytic center* of SP^*.

It is easily to verify that the analytic center is unique. Let us assume to the contrary that there are two different vectors $\overline{x} \neq \tilde{x}$ with $\overline{x}, \tilde{x} \in SP^*$ which satisfy the definition of analytic center, i.e.,

$$\vartheta^* = \prod_{i\in\mathbf{B}} \overline{x}_i \prod_{i\in\mathbf{N}} \overline{s}_i = \prod_{i\in\mathbf{B}} \tilde{x}_i \prod_{i\in\mathbf{N}} \tilde{s}_i = \max_{x\in SP^*} \prod_{i\in\mathbf{B}} x_i \prod_{i\in\mathbf{N}} s_i. \tag{2.27}$$

Let us define $x^* = \frac{\overline{x}+\tilde{x}}{2}$. Then we have

$$\begin{aligned}
\prod_{i\in\mathbf{B}} x_i^* \prod_{i\in\mathbf{N}} s_i^* &= \prod_{i\in\mathbf{B}} \frac{1}{2}(\overline{x}_i + \tilde{x}_i) \prod_{i\in\mathbf{N}} (\overline{s}_i + \tilde{s}_i) \\
&= \prod_{i\in\mathbf{B}} \frac{1}{2}\left(\sqrt{\frac{\overline{x}_i}{\tilde{x}_i}} + \sqrt{\frac{\tilde{x}_i}{\overline{x}_i}}\right) \prod_{i\in\mathbf{N}} \frac{1}{2}\left(\sqrt{\frac{\overline{s}_i}{\tilde{s}_i}} + \sqrt{\frac{\tilde{s}_i}{\overline{s}_i}}\right) \\
&\sqrt{\prod_{i\in\mathbf{B}} \overline{x}_i \prod_{i\in\mathbf{N}} \overline{s}_i \prod_{i\in\mathbf{B}} \tilde{x}_i \prod_{i\in\mathbf{N}} \tilde{s}_i} > \prod_{i\in\mathbf{B}} \overline{x}_i \prod_{i\in\mathbf{N}} \overline{s}_i = \vartheta^*,
\end{aligned} \tag{2.28}$$

which shows that $\overline{x}$ is not the analytic center. Here the last inequality follows from the classical inequality $\mathbf{a} + \frac{1}{\mathbf{a}} \geq 2$ if $\mathbf{a} \in \mathbb{R}_+$ and strict inequality holds when $\mathbf{a} \neq 1$.

Theorem 2.15. *The limit point x^* of the central path is the analytic center of SP^*.*

Proof. The same way as in the proof of Theorem 2.13 we derive

$$\sum_{i\in\mathbf{B}} \frac{\overline{x}_i}{x_i^*} + \sum_{i\in\mathbf{N}} \frac{\overline{s}_i}{s_i^*} = n. \tag{2.29}$$

Now we apply the arithmetic-geometric mean inequality to derive

$$\left(\prod_{i\in\mathbf{B}} \frac{\overline{x}_i}{x_i^*} \prod_{i\in\mathbf{N}} \frac{\overline{s}_i}{s_i^*}\right)^{\frac{1}{n}} \leq \frac{1}{n}\left(\sum_{i\in\mathbf{B}} \frac{\overline{x}_i}{x_i^*} + \sum_{i\in\mathbf{N}} \frac{\overline{s}_i}{s_i^*}\right) = 1. \tag{2.30}$$

Hence,

$$\prod_{i\in\mathbf{B}} \overline{x}_i \prod_{i\in\mathbf{N}} \overline{s}_i \leq \prod_{i\in\mathbf{B}} x_i^* \prod_{i\in\mathbf{N}} s_i^* \tag{2.31}$$

proving that x^* is the analytic center of SP^*. The proof is complete. □

2.2.6 Identifying the Optimal Partition

The condition number

In order to give bounds on the size of the variables along the central path we need to find a quantity that in some sense characterizes the set of optimal solutions. For an optimal solution $x \in SP^*$ we have

$$xs(x) = 0 \quad \text{and} \quad x + s(x) \geq 0.$$

Our next question is about the size of the nonzero coordinates of optimal solutions. Following the definitions in [72,96] we define a condition number of the problem (SP) which characterizes the magnitude of the nonzero variables on the optimal set SP^*.

Definition 2.16. Let us define

$$\sigma^x := \min_{i\in\mathbf{B}} \max_{x\in SP^*} \{x_i\} \tag{2.32a}$$

$$\sigma^s := \min_{i\in\mathbf{N}} \max_{x\in SP^*} \{s(x)_i\}. \tag{2.32b}$$

Then the *condition number* of (SP) is defined as

$$\sigma = \min\{\sigma^x, \sigma^s\} = \min_i \max_{x\in SP^*} \{x_i + s(x)_i\}. \tag{2.33}$$

To determine the condition number σ is in general more difficult than to solve the optimization problem itself. However, we can give an easily computable lower bound for σ. This bound depends only on the problem data.

Lemma 2.17 (Lower bound for $\boldsymbol{\sigma}$:). *If M and q are integral*[7] *and all the columns of M are nonzero, then*

$$\sigma \geq \frac{1}{\pi(M)}, \tag{2.34}$$

where $\pi(M) = \prod_{i=1}^{n} \|M_i\|$.

---

[7] If the problem data is rational, then by multiplying by the least common multiple of the denominators an equivalent LO problem with integer data is obtained.

Proof. The proof is based on Cramer's rule and on the estimation of determinants by using Hadamard's inequality.[8] Let $z = (x, s)$ be an optimal solution. Without loss of generality we may assume that the columns of the matrix $D = (-M, I)$ corresponding to the nonzero coordinates of $z = (x, s)$ are linearly independent. If they are not independent, then by using Gaussian elimination we can reduce the solution to get one with linearly independent columns. Let us denote this index set by J. Further, let the index set K be such that D_{KJ} is a nonsingular square submatrix of D. Such K exists, because the columns in D_J are linearly independent. Now we have $D_{KJ} z_J = q_K$, and hence, by Cramer's rule,

$$z_j = \frac{\det\left(D_{KJ}^{(j)}\right)}{\det\,(D_{KJ})}, \quad \forall j \in J, \tag{2.35}$$

where $D_{KJ}^{(j)}$ denotes the matrix obtained when the jth column in D_{KJ} is replaced by q_K. Assuming that $z_j > 0$ then, because the data is integral, the numerator in the quotient given above is at least one. Thus we obtain $z_j \geq \frac{1}{\det\,(D_{KJ})}$. By Hadamard's inequality the last determinant can be estimated by the product of the norm of its columns, which can further be bounded by the product of the norms of all the columns of the matrix M. □

The condition that none of the columns of the matrix M is a zero vector is not restrictive. For the general problem (SP) a zero column M_i would imply that $s_i = q_i$ for each feasible solution, thus the pair (x_i, s_i) could be removed. More important is that for our embedding problem $(\overline{\mathrm{SP}})$ none of the columns of the coefficient matrix

$$\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}$$

is zero. By definition we have $r = e - Me$ nonzero, because $e^T r = e^T e - e^T Me = n$. Moreover, if $M_i = 0$, then by using that matrix M is skew-symmetric we have $r_i = 1$, thus the ith column of the coefficient matrix is again nonzero.

---

[8] Let G be a nonsingular $n \times n$ matrix. Hadamard's inequality states that

$$\det\,(G) \leq \prod_{i=1}^{n} \|G_i\|$$

holds, see [37] for a reference.

The size of the variables along the central path

Now, by using the condition number σ, we are able to derive lower and upper bounds for the variables along the central path. Let (B, N) be the optimal partition of the problem (SP).

Lemma 2.18. *For each positive μ one has*

$$x_i(\mu) \geq \frac{\sigma}{n} \quad i \in \mathbf{B}, \quad x_i(\mu) \leq \frac{n\mu}{\sigma} \quad i \in \mathbf{N}, \tag{2.36a}$$

$$s_i(\mu) \leq \frac{n\mu}{\sigma} \quad i \in \mathbf{B}, \quad s_i(\mu) \geq \frac{\sigma}{n} \quad i \in \mathbf{N}. \tag{2.36b}$$

Proof. Let (x^*, s^*) be optimal, then by orthogonality we have

$$\begin{aligned}
(x(\mu) - x^*)^T(s(\mu) - s^*) &= 0, \\
x(\mu)^T s^* + s(\mu)^T x^* &= n\mu, \\
x_i(\mu)s_i^* \leq x(\mu)^T s^* &\leq n\mu, \ \ 1 \leq i \leq n.
\end{aligned}$$

Since we can choose (x^*, s^*) such that $s_i^* \geq \sigma$ and because $x_i(\mu)s_i(\mu) = \mu$, for $i \in \mathbf{N}$, we have

$$x_i(\mu) \leq \frac{n\mu}{s_i^*} \leq \frac{n\mu}{\sigma} \quad \text{and} \quad s_i(\mu) \geq \frac{\sigma}{n}, \ \ i \in \mathbf{N}.$$

The proofs of the other bounds are analogous. □

Identifying the optimal partition

The bounds presented in Lemma 2.18 make it possible to identify the optimal partition $(\mathbf{B}, \mathbf{N})$, when μ is sufficiently small. We just have to calculate the μ value that ensures that the coordinates going to zero are certainly smaller than the coordinates that converge to a positive number.

Corollary 2.19. *If we have a central solution $x(\mu) \in SP$ with*

$$\mu < \frac{\sigma^2}{n^2}, \tag{2.37}$$

then the optimal partition $(\mathbf{B}, \mathbf{N})$ can be identified.

The results of Lemma 2.18 and Corollary 2.19 can be generalized to the situation when a vector (x, s) is not on, but just in a certain neighbourhood of the central path. In order to keep our discussion short, we do not go into those details. The interested reader is referred to [72].

2.2.7 Rounding to an Exact Solution

Our next goal is to find a strictly complementary solution. This could be done by moving along the central path as $\mu \to 0$. Here we show that we do not have to do that, we can stop at a sufficiently small $\mu > 0$, and round off the current "almost optimal" solution to a strictly complementary optimal one. We need some new notation. Let the optimal partition be denoted by $(\mathbf{B}, \mathbf{N})$, let $\omega := \|M\|_\infty = \max_{1\le i\le n} \sum_{j=1}^n |M_{ij}|$ and $\pi := \pi(M) = \prod_{i=1}^n \|M_i\|$.

Lemma 2.20. *Let M and q be integral and all the columns of M be nonzero. If $(x, s) := (x(\mu), s(x(\mu)))$ is a central solution with*

$$x^T s = n\mu < \frac{\sigma^2}{n^{\frac{3}{2}}(1+\omega)^2\pi}, \quad \textit{which certainly holds if} \quad n\mu \le \frac{1}{n^{\frac{3}{2}}(1+\omega)^2\pi^3},$$

then by a simple rounding procedure a strictly complementary optimal solution can be found in $\mathcal{O}(n^3)$ arithmetic operations.

Proof. Proof. Let $x := x(\mu) > 0$ and $s := s(x) > 0$ be given. Because

$$\mu < \frac{\sigma^2}{n^{\frac{5}{2}}(1+\omega)^2\pi} < \frac{\sigma^2}{n^2} \tag{2.38}$$

the optimal partition $(\mathbf{B}, \mathbf{N})$ is known. Let us simply set the small variables $x_{\mathbf{N}}$ and $s_{\mathbf{B}}$ to zero. Then we correct the created error and estimate the size of the correction.

For (x, s) we have

$$M_{\mathbf{BB}} x_{\mathbf{B}} + M_{\mathbf{BN}} x_{\mathbf{N}} + q_{\mathbf{B}} = s_{\mathbf{B}}, \tag{2.39}$$

but by rounding $x_{\mathbf{N}}$ and $s_{\mathbf{B}}$ to zero the error $\hat{q}_{\mathbf{B}} = s_{\mathbf{B}} - M_{\mathbf{BN}} x_{\mathbf{N}}$ occurs. Similarly, we have

$$M_{\mathbf{NB}} x_{\mathbf{B}} + M_{\mathbf{NN}} x_{\mathbf{N}} + q_{\mathbf{N}} = s_{\mathbf{N}} \tag{2.40}$$

but by rounding $x_{\mathbf{N}}$ and $s_{\mathbf{B}}$ to zero the error $\hat{q}_{\mathbf{N}} = -M_{\mathbf{NN}} x_{\mathbf{N}}$ occurs.

Let us first estimate $\hat{q}_{\mathbf{B}}$ and $\hat{q}_{\mathbf{N}}$ by using the results of Lemma 2.18. For $\hat{q}_{\mathbf{B}}$ we have

$$\begin{aligned} \|\hat{q}_{\mathbf{B}}\| &\le \sqrt{n}\|\hat{q}_{\mathbf{B}}\|_\infty = \sqrt{n}\|s_{\mathbf{B}} - M_{\mathbf{BN}} x_{\mathbf{N}}\|_\infty \le \sqrt{n}\|(I, -M_{\mathbf{BN}})\|_\infty \left\| \begin{matrix} s_{\mathbf{B}} \\ x_{\mathbf{N}} \end{matrix} \right\|_\infty \\ &\le \sqrt{n}(1+\omega)\frac{n\mu}{\sigma} = \frac{n^{\frac{3}{2}}\mu(1+\omega)}{\sigma}. \end{aligned} \tag{2.41}$$

We give a bound for the infinity norm of $\hat{q}_N$ as well:

$$\|\hat{q}_{\mathbf{N}}\|_\infty = \|-M_{\mathbf{NN}}x_{\mathbf{N}}\|_\infty \le \|M_{\mathbf{NN}}\|_\infty \|x_{\mathbf{N}}\|_\infty \le \omega \frac{n\mu}{\sigma}. \tag{2.42}$$

Now we are going to correct these errors by adjusting $x_{\mathbf{B}}$ and $s_{\mathbf{N}}$. Let us denote the correction by ξ for $x_{\mathbf{B}}$ and by ζ for $s_{\mathbf{N}}$, further let $(\hat{x}, \hat{s})$ be given by $\hat{x}_{\mathbf{B}} := x_{\mathbf{B}} + \xi > 0$, $\hat{x}_{\mathbf{N}} = 0$, $\hat{s}_{\mathbf{B}} = 0$ and $\hat{s}_{\mathbf{N}} := s_{\mathbf{N}} + \zeta > 0$.

If we know the correction ξ of $x_{\mathbf{B}}$, then from equation (2.40) the necessary correction ζ of $s_{\mathbf{N}}$ can easily be calculated. Equation (2.39) does not contain s_N, thus by solving the equation

$$M_{\mathbf{BB}}\xi = -\hat{q}_{\mathbf{B}} \tag{2.43}$$

the corrected value $\hat{x}_{\mathbf{B}} = x_{\mathbf{B}} + \xi$ can be obtained.

First we observe that the equation $M_{\mathbf{BB}}\xi = -\hat{q}_{\mathbf{B}}$ is solvable, because any optimal solution x^* satisfies $M_{\mathbf{BB}}x^*_{\mathbf{B}} = -q_{\mathbf{B}}$, thus we may write $M_{\mathbf{BB}}(x_{\mathbf{B}}+\xi)$ $= M_{\mathbf{BB}}x^*_{\mathbf{B}} = -q_{\mathbf{B}}$, hence

$$\begin{aligned} M_{\mathbf{BB}}\xi = M_{\mathbf{BB}}(x^*_{\mathbf{B}} - x_{\mathbf{B}}) &= -q_{\mathbf{B}} - s_{\mathbf{B}} + M_{\mathbf{BN}}x_{\mathbf{N}} + q_{\mathbf{B}} \\ &= -s_{\mathbf{B}} + M_{\mathbf{BN}}x_{\mathbf{N}} = -\hat{q}_{\mathbf{B}}. \end{aligned} \tag{2.44}$$

This equation system can be solved by Gaussian elimination. The size of ξ obtained this way can be estimated by applying Cramer's rule and Hadamard's inequality, the same way as we have estimated σ in Lemma 2.17. If $M_{\mathbf{BB}}$ is zero, then we have $q_{\mathbf{B}} = 0$ and $M_{\mathbf{BN}}x_{\mathbf{N}} = s_{\mathbf{B}}$, thus rounding $x_{\mathbf{N}}$ and $s_{\mathbf{B}}$ to zero does not produce any error here, hence we can choose $\xi = 0$. If $M_{\mathbf{BB}}$ is not the zero matrix, then let $\overline{M}_{\mathbf{BB}}$ be a maximal nonsingular square submatrix of $M_{\mathbf{BB}}$ and let $\bar{q}_{\mathbf{B}}$ be the corresponding part of $\hat{q}_{\mathbf{B}}$. By using the upper bounds on $x_{\mathbf{N}}$ and $s_{\mathbf{B}}$ by Lemma 2.18 we have

$$\begin{aligned} |\xi_i| = \frac{|\det(\overline{M}^{(i)}_{\mathbf{BB}})|}{|\det(\overline{M}_{\mathbf{BB}})|} &\le |\det(\overline{M}^{(i)}_{\mathbf{BB}})| \\ &\le \|\bar{q}_{\mathbf{B}}\|\,|\det(\overline{M}_{\mathbf{BB}})| \le \frac{n^{\frac{3}{2}}\mu(1+\omega)}{\sigma}\pi, \end{aligned} \tag{2.45}$$

where (2.41) was used in the last estimation. This result, due to $\|x_{\mathbf{B}}\|_\infty \ge \frac{\sigma}{n}$, implies that $\hat{x}_{\mathbf{B}} = x_{\mathbf{B}} + \xi > 0$ certainly holds if $n\mu < \frac{\sigma^2}{n^{\frac{3}{2}}(1+\omega)\pi}$, and this is implied by the hypothesis of the theorem which was involving $(1+\omega)^2$ instead of $(1+\omega)$.

Finally, we simply correct $s_{\mathbf{N}}$ by using (2.40), i.e., we define $\zeta := \hat{q}_{\mathbf{N}} + M_{\mathbf{NB}}\xi$. We still must ensure that

$$\hat{s}_{\mathbf{N}} := s_{\mathbf{N}} + \hat{q}_{\mathbf{N}} + M_{\mathbf{NB}}\xi > 0. \tag{2.46}$$

Using again the bounds given in Lemma 2.17, the bound (2.42) and the estimate on ξ, one easily verifies that

$$\begin{aligned}\|\hat{q}_{\mathbf{N}} + M_{\mathbf{NB}}\xi\|_\infty &\le \|(I, M_{\mathbf{NB}})\|_\infty \left\| \begin{matrix} \hat{q}_{\mathbf{N}} \\ \xi \end{matrix} \right\|_\infty \qquad (2.47)\\ &\le (1+\omega)\max\left\{\omega\frac{n\mu}{\sigma}, \frac{n^{\frac{3}{2}}\mu(1+\omega)\pi}{\sigma}\right\} = \frac{n^{\frac{3}{2}}\mu(1+\omega)^2\pi}{\sigma}.\end{aligned}$$

Thus, due to $\|s_{\mathbf{N}}\|_\infty \ge \frac{\sigma}{n}$, the vector $\hat{s}_{\mathbf{N}}$ is certainly positive if

$$\frac{\sigma}{n} > \frac{n^{\frac{3}{2}}\mu(1+\omega)^2\pi}{\sigma}. \qquad (2.48)$$

This is exactly the first inequality given in the lemma. The second inequality follows by observing that $\pi\sigma \ge 1$, by Lemma 2.17.

The proof is completed by noting that the solution of an equation system by using Gaussian elimination, some matrix-vector multiplications and vector-vector summations, all with a dimension not exceeding n, are needed to perform our rounding procedure. Thus the computational complexity of our rounding procedure is at most $\mathcal{O}(n^3)$. □

Note that this rounding result can also be generalized to the situation when a vector (x, s) is not on, but just in a certain neighbourhood of the central path. For details the reader is referred again to [72].[9]

2.3 Summary of the Theoretical Results

Let us return to our general LO problem in canonical form

$$\min\left\{c^T u \ : \ Au - z = b, \ u \ge 0, \ z \ge 0\right\} \qquad (P)$$

$$\max\left\{b^T v \ : \ A^T v + w = c, \ v \ge 0, \ w \ge 0\right\}, \qquad (D)$$

where the slack variables are already included in the problem formulation. In what follows we recapitulate the results obtained so far.

- In Section 2.1 we have seen that to solve the LO problem it is sufficient to find a strictly complementary solution to the Goldman-Tucker model

[9] This result makes clear that when one solves an LO problem by using an IPM, the iterative process can be stopped at a sufficiently small value of μ. At that point a strictly complementary optimal solution can be identified easily.

$$\begin{array}{rrrrrl} & Au & -\tau b & -z & & = 0 \\ -A^T v & & +\tau c & & -w & = 0 \\ b^T v & -c^T u & & & & -\rho = 0 \end{array}$$

$$v \geq 0,\ u \geq 0,\ \tau \geq 0,\ z \geq 0,\ w \geq 0,\ \rho \geq 0.$$

- This *homogeneous* system always admits the zero solution, but we need a solution for which $\tau + \rho > 0$ holds.
- If (u^*, z^*) is optimal for (P) and (v^*, w^*) for (D) then $(v^*, u^*, 1, z^*, w^*, 0)$ is a solution for the Goldman-Tucker model with the requested property $\tau + \rho > 0$. See Theorem 2.3.
- Any solution of the Goldman-Tucker model (v, u, τ, z, w, ρ) with $\tau > 0$ yields an optimal solution pair (scale the variables (u, z) and (v, w) by $\frac{1}{\tau}$) for LO. See Theorem 2.3.
- Any solution of the Goldman-Tucker model (u, z, v, w, τ, ρ) with $\rho > 0$ provides a certificate of primal or dual infeasibility. See Theorem 2.3.
- If $\tau = 0$ in every solution (v, u, τ, z, w, ρ) then (P) and (D) have no optimal solutions with zero duality gap.
- The Goldman-Tucker model can be transformed into a skew-symmetric self-dual problem (SP) satisfying the IPC. See Section 2.2.2.
- If problem (SP) satisfies the IPC then
  - the central path exists (see Theorem 2.12);
  - the central path converges to a strictly complementary solution (see Theorem 2.13);
  - the limit point of the central path is the analytic center of the optimal set (see Theorem 2.15);
  - if the problem data is integral and a solution on the central path with a sufficiently small μ is given, then the optimal partition (see Corollary 2.19) and an exact strictly complementary optimal solution (see Lemma 2.20) can be found.
- These results give a constructive proof of Theorem 2.4.
- This way, as we have seen in Section 2.1, the Strong Duality theorem of linear optimization (Theorem 2.5) is proved.

The above summary shows that we have completed our project. The duality theory of LO is built up by using only elementary calculus and fundamental concepts of IPMs. In the following sections we follow this recipe to derive interior point methods for conic and general nonlinear optimization.

In the rest of this section a generic IP algorithm is presented.

2.4 A General Scheme of IP Algorithms for Linear Optimization

In this section a glimpse of the main elements of IPMs is given. We keep on working with our model problem (SP). In Sections 2.1 and 2.2.2 we have shown that a general LO problem can be transformed into a problem of the form (SP), and that problem satisfies the IPC. Some notes are due to the linear algebra involved. We know that the size of the resulting embedding problem (SP) is more than doubled comparing to the size of the original LO problem. Despite the size increase, the linear algebra of an IPM can be organized so that the computational cost of an iteration stays essentially the same.

Let us consider the problem (cf. page 223)

$$\min\left\{\lambda\vartheta \ : \ \begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}\begin{pmatrix} x \\ \vartheta \end{pmatrix} + \begin{pmatrix} s \\ \nu \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda \end{pmatrix}; \ \begin{pmatrix} x \\ \vartheta \end{pmatrix}, \begin{pmatrix} s \\ \nu \end{pmatrix} \geq 0\right\}, \qquad (\overline{\text{SP}})$$

where $r = e - Me$, $\lambda = n+1$ and the matrix M is given by (2.4). This problem satisfies the IPC, because the all one vector $(x^0, \vartheta^0, s^0, \nu^0) = (e, 1, e, 1)$ is a feasible solution, moreover it is also on the central path by taking $\mu = 1$. In other words, it is a positive solution of the equation system

$$\begin{gathered} \begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}\begin{pmatrix} x \\ \vartheta \end{pmatrix} + \begin{pmatrix} s \\ \nu \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda \end{pmatrix}; \ \begin{pmatrix} x \\ \vartheta \end{pmatrix}, \begin{pmatrix} s \\ \nu \end{pmatrix} \geq 0 \\ \begin{pmatrix} x \\ \vartheta \end{pmatrix}\begin{pmatrix} s \\ \nu \end{pmatrix} = \begin{pmatrix} \mu e \\ \mu \end{pmatrix}, \end{gathered} \qquad (2.49)$$

which defines the central path of problem $(\overline{\text{SP}})$. As we have seen, for each $\mu > 0$, this system has a unique solution. However, in general this solution cannot be calculated exactly. Therefore we are making Newton steps to get approximate solutions.

Newton step

Let us assume that a feasible interior-point $(x, \vartheta, s, \nu) > 0$ is given.[10] We want to find the solution of (2.49) for a given $\mu \geq 0$, in other words we want to determine the displacements $(\Delta x, \Delta\vartheta, \Delta s, \Delta\nu)$ so that

[10] Here we assume that all the linear equality constraints are satisfied. The resulting IPM is a feasible IPM. In the literature one can find infeasible IPMs [93] that do not assume that the linear equality constraints are satisfied.

$$\begin{aligned}
\begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix}\begin{pmatrix} x+\Delta x \\ \vartheta+\Delta\vartheta \end{pmatrix} + \begin{pmatrix} s+\Delta s \\ \nu+\Delta\nu \end{pmatrix} &= \begin{pmatrix} 0 \\ \lambda \end{pmatrix}; \\
\begin{pmatrix} x+\Delta x \\ \vartheta+\Delta\vartheta \end{pmatrix}, \begin{pmatrix} s+\Delta s \\ \nu+\Delta\nu \end{pmatrix} &\geq 0; \\
\begin{pmatrix} x+\Delta x \\ \vartheta+\Delta\vartheta \end{pmatrix}\begin{pmatrix} s+\Delta s \\ \nu+\Delta\nu \end{pmatrix} &= \begin{pmatrix} \mu e \\ \mu \end{pmatrix}.
\end{aligned} \tag{2.50}$$

By neglecting the second order terms $\Delta x\Delta s$ and $\Delta\vartheta\Delta\nu$, and the nonnegativity constraints, the Newton equation system is obtained (cf. page 227)

$$\begin{array}{lllll}
-M\Delta x & -r\Delta\vartheta & +\Delta s & & = 0 \\
r^T\Delta x & & & +\Delta\nu & = 0 \\
s\Delta x & & +x\Delta s & & = \mu e - xs \\
 & \nu\Delta\vartheta & & +\vartheta\Delta\nu & = \mu - \vartheta\nu.
\end{array} \tag{2.51}$$

We start by making some observations. For any vector (x, ϑ, s, ν) that satisfies the equality constraints of $(\overline{\mathrm{SP}})$ we have

$$x^T s + \vartheta\nu = \vartheta\lambda. \tag{2.52}$$

Applying this to the solution obtained after making a Newton step we may write

$$(x+\Delta x)^T(s+\Delta s) + (\vartheta+\Delta\vartheta)^T(\nu+\Delta\nu) = (\vartheta+\Delta\vartheta)\lambda. \tag{2.53}$$

By rearranging the terms we have

$$(x^T s+\vartheta\nu)+(\Delta x^T\Delta s+\Delta\vartheta\Delta\nu)+(x^T\Delta s+s^T\Delta x+\vartheta\Delta\nu+\nu\Delta\vartheta) = \vartheta\lambda+\Delta\vartheta\lambda.$$

As we mentioned above, the first term in the left hand side sum equals to $\vartheta\lambda$, while from (2.51) we derive that the second sum is zero. From the last equations of (2.51) one easily derives that the third expression equals to $\mu(n+1) - x^T s - \vartheta\nu = \mu\lambda - \vartheta\lambda$. This way the equation $\mu\lambda - \vartheta\lambda = \Delta\vartheta\lambda$ is obtained, i.e., an explicit expression for $\Delta\vartheta$

$$\Delta\vartheta = \mu - \vartheta$$

is derived. This value can be substituted in the last equation of (2.51) to derive the solution

$$\Delta\nu = \frac{\mu}{\vartheta} - \nu - \frac{\nu(\mu-\vartheta)}{\vartheta},$$

i.e.,

$$\Delta\nu = \frac{\mu(1-\nu)}{\vartheta}.$$

On the other hand, Δs can be expressed from the third equation of (2.51) as

$$\Delta s = \mu X^{-1}e - s - X^{-1}S\Delta x.$$

Finally, substituting all these values in the first equation of (2.51) we have

$$M\Delta x + X^{-1}S\Delta x = \mu X^{-1}e - s - (\mu - \vartheta)r,$$

i.e., Δx is the unique solution of the positive definite system[11]

$$(M + X^{-1}S)\Delta x = \mu X^{-1}e - s - (\mu - \vartheta)r.$$

Having determined the displacements, we can make a (possibly damped) Newton step to update our current iterate:

$$\begin{aligned} x &:= x + \Delta x \\ \vartheta &:= \vartheta + \Delta\vartheta = \mu \\ s &:= s + \Delta s \\ \nu &:= \nu + \Delta\nu. \end{aligned}$$

Proximity measures

We have seen that the central path is our guide to a strictly complementary solution. However, due to the nonlinearity of the equation system determining the central path, we cannot stay on the central path with our iterates, even if our initial interior-point was perfectly centred. For this reason we need some centrality, or with other words proximity, measure that enables us to control and keep our iterates in an appropriate neighbourhood of the central path. In general this measure depends on the current primal-dual iterate x and s, and a value of μ on the central path. This measure quantifies how close the iterate is to the point corresponding to μ on the central path. We use $\delta(x, s, \mu)$ to denote this general proximity measure.

Let the vectors $\overline{x}$ and $\overline{s}$ be composed from x and ϑ, and from s and ν respectively. Note that on the central path all the coordinates of the vector $\overline{x}\,\overline{s}$ are equal. This observation indicates that the proximity measure

[11] Observe that although the dimensions of problem ($\overline{\text{SP}}$) are larger than problem (SP), to determine the Newton step for both systems requires essentially the same computational effort. Note also, that the special structure of the matrix M (see (2.4)) can be utilized when one solves this positive definite linear system. For details the reader is referred to $[5, 72, 93, 97]$.

$$\delta_c(\overline{x}\,\overline{s}) := \frac{\max(\overline{x}\,\overline{s})}{\min(\overline{x}\,\overline{s})}, \tag{2.54}$$

where $\max(\overline{x}\,\overline{s})$ and $\min(\overline{x}\,\overline{s})$ denote the largest and smallest coordinate of the vector $\overline{x}\,\overline{s}$, is an appropriate measure of centrality. In the literature of IPMs various centrality measures were developed (see the books [42, 45, 72, 93, 97]). Here we present just another one, extensively used in [72]:

$$\delta_0(\overline{x}\,\overline{s}, \mu) := \frac{1}{2}\left\| \left(\frac{\overline{x}\,\overline{s}}{\mu}\right)^{\frac{1}{2}} - \left(\frac{\mu}{\overline{x}\,\overline{s}}\right)^{\frac{1}{2}} \right\|. \tag{2.55}$$

Both of these proximity measures allow us to design polynomial IPMs.

A generic interior point algorithm

Algorithm 1 gives a general framework for an interior point method.

Algorithm 1 Generic Interior-Point Newton Algorithm

Input:
  A proximity parameter γ;
  an accuracy parameter $\mathbf{e} > 0$;
  a variable damping factor $\mathbf{a}$;
  update parameter θ, $0 < \theta < 1$;
  $(\overline{x}^0, \overline{s}^0)$, $\mu^0 \le 1$ s.t. $\mathbf{d}(\overline{x}^0\overline{s}^0, \mu^0) \le \gamma$.
begin
  $\overline{x} := \overline{x}^0$; $\overline{s} := \overline{s}^0$; $\mu := \mu^0$;
  while $(n+1)\mu \ge \mathbf{e}$ **do**
  begin
    $\mu := (1-\theta)\mu$;
    while $\mathbf{d}(\overline{x}\,\overline{s}, \mu) \ge \gamma$ **do**
    begin
      $\overline{x} := \overline{x} + \alpha\Delta\overline{x}$;
      $\overline{s} := \overline{s} + \alpha\Delta\overline{s}$;
    end
  end
end

The following crucial issues remain:

- choose the proximity parameter γ,
- choose a proximity measure $\delta(x, s, \mu)$,
- choose an update scheme for μ and
- specify how to damp the Newton step when needed.

Our goal with the selection of these parameters is to be able to prove polynomial iteration complexity of the resulting algorithm.

Three sets of parameters are presented, which ensure that the resulting IPMs are polynomial. The proofs of complexity can, e.g., be found in [72]. Recall that $(\overline{\mathrm{SP}})$ admits the all one vector as a perfectly centred initial solution with $\mu = 1$.

The first algorithm is a *primal-dual logarithmic barrier algorithm with full Newton steps*, studied e.g. in [72]. This IPM enjoys the best complexity known to date. Let us make the following choice:

- $\mathbf{d}(\overline{x}\,\overline{s}, \mu) := \delta_0(\overline{x}\,\overline{s}, \mu)$, this measure is zero on the central path;
- $\mu^0 := 1$;
- $\theta := \frac{1}{2\sqrt{n+1}}$;
- $\gamma = \frac{1}{\sqrt{2}}$;
- $(\Delta\overline{x}, \Delta\overline{s})$ is the solution of (2.51);
- $\alpha = 1$.

Theorem 2.21 (Theorem II.52 in [72]). *With the given parameter set the full Newton step algorithm requires not more than*

$$\left\lceil 2\sqrt{n+1}\log\frac{n+1}{\mathbf{e}} \right\rceil$$

iterations to produce a feasible solution $(\overline{x}, \overline{s})$ *for* $(\overline{\mathrm{SP}})$ *such that* $\delta_0(\overline{x}\,\overline{s}, \mu) \le \gamma$ *and* $(n+1)\mu \le \mathbf{e}$.

The second algorithm is a *large-update primal-dual logarithmic barrier algorithm*, studied also e.g. in [72]. Among our three algorithms, this is the most practical. Let us make the following choice:

- $\mathbf{d}(\overline{x}\,\overline{s}, \mu) := \delta_0(\overline{x}\,\overline{s}, \mu)$, this measure is zero on the central path;
- $\mu^0 := 1$;
- $0 < \theta < \frac{n+1}{n+1+\sqrt{n+1}}$;
- $\gamma = \frac{\sqrt{R}}{2\sqrt{1+\sqrt{R}}}$, where $R := \frac{\theta\sqrt{n+1}}{1-\theta}$;
- $(\Delta\overline{x}, \Delta\overline{s})$ is the solution of (2.51);
- $\mathbf{a}$ is the result of a line search, when along the search direction the primal-dual logarithmic barrier function

$$\frac{\overline{x}^T\overline{s}}{\mu} - (n+1) - \sum_{i=1}^{n+1} \log\frac{\overline{x}_i\overline{s}_i}{\mu}$$

is minimized.

Theorem 2.22 (Theorem II.74 in [72]). *With the given parameter set the large-update primal-dual logarithmic barrier algorithm requires not more than*

$$\left\lceil \frac{1}{\theta} \left\lceil 2\left(1+\sqrt{\frac{\theta\sqrt{n+1}}{1-\theta}}\right)^4 \right\rceil \log \frac{n+1}{\mathbf{e}} \right\rceil$$

iterations to produce a feasible solution $(\overline{x}, \overline{s})$ *for* $(\overline{\mathrm{SP}})$ *such that* $\delta_0(\overline{x}\,\overline{s}, \mu) \leq \tau$ *and* $(n+1)\mu \leq \mathbf{e}$.

When we choose $\theta = \frac{1}{2}$, then the total complexity becomes $\mathcal{O}\big((n+1)\log\frac{n+1}{\mathbf{e}}\big)$, while the choice $\theta = \frac{K}{\sqrt{n+1}}$, with any fixed positive value K gives a complexity of $\mathcal{O}\left(\sqrt{n+1}\log\frac{n+1}{\mathbf{e}}\right)$.

Other versions of this algorithm were studied in [66], where the analysis of large-update methods was based purely on the use of the proximity $\delta_0(\overline{x}\,\overline{s}, \mu)$.

The last algorithm is the *Dikin step algorithm* studied in [72]. This is one of the simplest IPMs, with an extremely elementary complexity analysis. The price for simplicity is that the polynomial complexity result is not the best possible. Let us make the following choices:

- $\mathbf{d}(\overline{x}\,\overline{s}, \mu) := \delta_c(\overline{x}\,\overline{s})$, this measure is always larger than or equal to 1;
- $\mu^0 := 0$, this implies that μ stays equal to zero, thus θ is irrelevant;
- $\gamma = 2$;
- $(\Delta\overline{x}, \Delta\overline{s})$ is the solution of (2.51) when the right-hand sides of the last two equations are replaced by $-\frac{x^2 s^2}{\|\overline{x}\,\overline{s}\|}$ and $-\frac{\vartheta\nu}{\|\overline{x}\,\overline{s}\|}$, respectively;
- $\alpha = \frac{1}{2\sqrt{n+1}}$.

Theorem 2.23 (Theorem I.27 in [72]). *With the given parameter set the Dikin step algorithm requires not more than*

$$\left\lceil 2(n+1)\log\frac{n+1}{\mathbf{e}} \right\rceil$$

iterations to produce a feasible solution $(\overline{x}, \overline{s})$ *for* $(\overline{\mathrm{SP}})$ *such that* $\delta_c(\overline{x}\,\overline{s}) \leq 2$ *and* $(n+1)\mu \leq \mathbf{e}$.

*2.5 *The Barrier Approach*

In our approach so far we perturbed the optimality conditions for the primal dual linear optimization problem to get the central path. In what follows we show an alternative, sometimes more intuitive, sometimes more technical route. Consider again the linear optimization problem in primal form:

$$\begin{aligned} \min\ & c^T u \\ Au &\geq b \\ u &\geq 0. \end{aligned} \tag{P}$$

A standard convex optimization trick to treat inequalities is to add them to the objective function with a barrier term:

$$\min\ c^T u - \mu \sum_{i=1}^{n} \ln u_i - \mu \sum_{j=1}^{m} \ln{(Au-b)_j}, \tag{PBar}$$

where $\mu > 0$. The function $-\ln t$ is a *barrier* function. In particular it goes to ∞ if t goes to 0, and for normalization, it is 0 at 1. If u_i is getting close to 0 then the modified objective function will converge to ∞. This way we received an unconstrained problem defined on the positive orthant, for which we can easily write the optimality conditions. The idea behind this method is to gradually reduce μ and at the same time try to solve the unconstrained problem approximately. If μ is decreased at the right rate then the algorithm will converge to the optimal solution of the original problem.

The first order necessary optimality conditions for system (PBar) are:

$$c_i - \mu \frac{1}{u_i} - \mu \sum_{j=1}^{m} \frac{A_{ji}}{(Au-b)_j} = 0,\ i = 1, \ldots, n. \tag{2.56}$$

This equation yields the same central path equations that we obtained in Definition 2.11. An identical result can be derived starting from the dual for of the linear optimization problem.

A natural extension of this idea is to replace the $-\ln t$ function with another barrier function. Sometimes we can achieve better complexity results by doing so, see [63] (universal barrier), [9,10,87] (volumetric barrier), [66,67] (self-regular barrier) for details.

3 Interior Point Methods for Conic Optimization

3.1 Problem Description

Conic optimization is a natural generalization of linear optimization. As we will see, most of the results in Section 2.3 carry over to the conic case with some minor modifications and the structure and analysis of the algorithm will be similar to the linear case.

A general conic optimization problem in primal form can be stated as

$$\begin{aligned} \min\ & c^T x \\ Ax &= b \\ x &\in \mathcal{K}, \end{aligned} \tag{PCon}$$

where $c, x \in \mathbb{R}^N$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times N}$ and $\mathcal{K} \subseteq \mathbb{R}^N$ is a cone. The standard Lagrange dual of this problem is

$$\begin{aligned} \max\, & b^T y \\ A^T y + s &= c \\ s &\in \mathcal{K}^*, \end{aligned} \tag{DCon}$$

where $y \in \mathbb{R}^m$, $s \in \mathbb{R}^N$ and $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$, namely $\mathcal{K}^* = \left\{ s \in \mathbb{R}^N : s^T x \geq 0, \, \forall x \in \mathcal{K} \right\}$. The weak duality theorem follows without any further assumption:

Theorem 3.1 (Weak duality for conic optimization). *If x, y and s are feasible solutions of the problems* (PCon) *and* (DCon) *then*

$$s^T x = c^T x - b^T y \geq 0. \tag{3.1}$$

This quantity is the duality gap. Consequently, if the duality gap is 0 for some solutions x, y and s, then they form an optimal solution.

Proof. Let x, y and s be feasible solutions, then

$$c^T x = (A^T y + s)^T x = x^T A^T y + x^T s = b^T y + x^T s \geq b^T y, \tag{3.2}$$

since $x \in \mathcal{K}$ and $s \in \mathcal{K}^*$ implies $x^T s \geq 0$. $\square$

In order for this problem to be tractable we have to make some assumptions.

Assumption 3.2 *Let us assume that $\mathcal{K}$ is a closed, convex, pointed (not containing a line) and solid (has nonempty interior) cone, and that it is self-dual, i.e., $\mathcal{K} = \mathcal{K}^*$.*

Cones in the focus of our study are called symmetric.

Theorem 3.3 (Real symmetric cones). *Any symmetric cone over the real numbers is a direct product of cones of the following type:*

nonnegative orthant: *the set of nonnegative vectors, $\mathbb{R}^n_+$,*
Lorentz or quadratic cone: *the set $\mathbb{L}_{n+1} = \{(u_0, u) \in \mathbb{R}_+ \times \mathbb{R}^n : u_0 \geq \|u\|\}$, and the*
positive semidefinite cone: *the cone $\mathbb{PS}^{n \times n}$ of $n \times n$ real symmetric positive semidefinite matrices.*

The dimensions of the cones forming the product can be arbitrary.

Let us assume further that the interior point condition is satisfied, i.e., there is a strictly feasible solution.[12] The strong duality theorem follows:

[12] This assumption is not needed if $\mathcal{K}$ is the linear cone, $\mathbb{R}^n_+$.

Theorem 3.4 (Strong duality for conic optimization). *If the primal problem* (PCon) *is* strictly feasible, *i.e., there exists an* x *for which* $Ax = b$ *and* $x \in \operatorname{int}(\mathcal{K})$*, then the dual problem is solvable (the maximum is attained) and the optimal values of the primal and dual problems are the same.*
If the dual problem (DCon) *is* strictly feasible, *i.e., there exists a* y *for which* $s = c - A^T y \in \operatorname{int}(\mathcal{K})$*, then the primal problem is solvable (the minimum is attained) and the optimal values of the primal and dual problems are the same.*
If both problems are strictly feasible then both are solvable and the optimal values are the same.

Remark 3.5. In conic optimization it can happen that one problem is infeasible but there is no certificate of infeasibility. Such problems are called weakly infeasible. Also, even if the duality gap is zero, the minimum or maximum might not be attained, meaning the problem is not solvable.

In what follows we treat the second order and the semidefinite cones separately. This simplification is necessary to keep the notation simple and to make the material more accessible. Interested readers can easily assemble the parts to get the whole picture.

First we introduce the following primal-dual second-order cone optimization problems:

$$\begin{array}{llr}
\min \displaystyle\sum_{i=1}^{k} {c^i}^T x^i & \max b^T y & \\
\displaystyle\sum_{i=1}^{k} A^i x^i = b & {A^i}^T y + s^i = c^i, i = 1, \dots, k & \text{(SOCO)} \\
x^i \in \mathbb{L}_{n_i}, i = 1, \dots, k & s^i \in \mathbb{L}_{n_i}, i = 1, \dots, k, &
\end{array}$$

where $x^i, s^i, c^i \in \mathbb{R}^{n_i}$, $b, y \in \mathbb{R}^m$ and $A^i \in \mathbb{R}^{m \times n^i}$, the number of cones is k and the i^{th} cone is of dimension n_i.

The semidefinite optimization problem requires slightly different notation:

$$\begin{array}{llr}
\min\ \operatorname{Tr}(CX) & \max b^T y & \\
\operatorname{Tr}\left(A^{(i)} X\right) = b_i, i = 1, \dots, m & \displaystyle\sum_{i=1}^{m} A^{(i)} y_i + S = C & \text{(SDO)} \\
X \in \mathbb{PS}^{n \times n} & S \in \mathbb{PS}^{n \times n}, &
\end{array}$$

where $X, S, C, A^{(i)} \in \mathbb{R}^{n \times n}$, $b, y \in \mathbb{R}^m$. For symmetric matrices U and V the quantity $\operatorname{Tr}(UV)$ is actually a scalar product defined on symmetric matrices, and is identical to the sum of the componentwise products of the matrix elements.

3.2 Applications of Conic Optimization

Let us briefly present three applications of conic optimization. For more details see [3, 11, 88, 91] and the references therein.

3.2.1 Robust Linear Optimization

Consider the standard linear optimization problem:

$$\begin{aligned} &\min c^T x \\ a_j^T x - b_j &\geq 0, \forall j = 1, \ldots, m, \end{aligned} \tag{3.3}$$

where the data (a_j, b_j) is uncertain. The uncertainty is usually due to some noise, or implementation or measurement error, and thus it is modelled by Gaussian distribution. The level sets of the distribution are ellipsoids, so we assume that the data vectors $(a_j; b_j)$ come from an ellipsoid. The inequalities then have to be satisfied for all possible values of the data. More precisely, the set off all possible data values is

$$\left\{ \begin{pmatrix} a_j \\ -b_j \end{pmatrix} = \begin{pmatrix} a_j^0 \\ -b_j^0 \end{pmatrix} + Pu : u \in \mathbb{R}^m, \, \|u\| \leq 1 \right\}, \tag{3.4}$$

and the new, robust constraint is represented as the following set of infinitely many constraints

$$\left(\begin{pmatrix} a_j^0 \\ -b_j^0 \end{pmatrix} + Pu \right)^T \begin{pmatrix} x \\ 1 \end{pmatrix} \geq 0, \, \forall u : \|u\| \leq 1. \tag{3.5}$$

This constraint is equivalent to

$$\begin{pmatrix} a_j^0 \\ -b_j^0 \end{pmatrix}^T \begin{pmatrix} x \\ 1 \end{pmatrix} \geq \max_{\|u\| \leq 1} \left\{ -u^T P^T \begin{pmatrix} x \\ 1 \end{pmatrix} \right\}. \tag{3.6}$$

The maximum on right hand side is the maximum of a linear function over a sphere, so it can be computed explicitly. This gives a finite form of the robust constraint:

$$\left(a_j^0\right)^T x - b_j^0 \geq \left\| P^T \begin{pmatrix} x \\ 1 \end{pmatrix} \right\|. \tag{3.7}$$

Introducing the linear equalities $z^j = \left(a_j^0\right)^T x - b_j^0$ and $z = P^T \begin{pmatrix} x \\ 1 \end{pmatrix}$ this constraint is a standard second order conic constraint. For more details on this approach see [11].

3.2.2 Eigenvalue Optimization

Given the $n \times n$ matrices $A^{(1)}, \ldots, A^{(m)}$ it is often required to find a nonnegative combination of them such that the smallest eigenvalue of the resulting matrix is maximal. The smallest eigenvalue function is not differentiable, thus we could not use it directly to solve the problem. Semidefinite optimization offers an efficient framework to solve these problems. The maximal smallest eigenvalue problem can be written as

$$\begin{aligned} &\max\ \lambda \\ &\sum_{i=1}^{m} A_i y_i - \lambda I \in \mathbb{PS}^{n \times n} \\ &y_i \geq 0,\ i = 1, \ldots, m. \end{aligned} \tag{3.8}$$

See [2, 63, 65] for more details.

3.2.3 Relaxing Binary Variables

A classical method to solve problems with binary variables is to apply a continuous relaxation. Given the binary variables $z_1, \ldots, z_n \in \{0, 1\}$ the most common solution is the linear relaxation $z_1, \ldots, z_n \in [0, 1]$. However, in many cases tighter relaxations can be obtained by introducing the new variables $x_i = (2z_i - 1)$ and relaxing the nonlinear nonconvex equalities $x_i^2 = 1$. Now consider the matrix $X = xx^T$. This matrix is symmetric, positive semidefinite, it has rank one and all the diagonal elements are 1. By relaxing the rank constraint we get a positive semidefinite relaxation of the original optimization problem. This technique was used extensively by Goemans and Williamson [29] to derive tight bounds for max-cut and satisfiability problems. For a survey of this area see [51] or the books [11, 40].

3.3 Initialization by Embedding

The key assumption for both the operation of an interior point method and the validity of the strong duality theorem is the existence of a strictly feasible solution of the primal-dual systems. Fortunately, the embedding technique we used for linear optimization generalizes to conic optimization [26]. Consider the following larger problem based on (PCon) and (DCon):

$$
\begin{array}{rrrrrrl}
 & \min(\bar{x}^T\bar{s}+1)\theta & & & & & \\
 & Ax & -b\tau & +\bar{b}\theta & & & = 0 \\
-A^Ty & & +c\tau & -\bar{c}\theta & -s & & = 0 \\
b^Ty & -c^Tx & & +\bar{z}\theta & & -\kappa & = 0 \\
-\bar{b}^Ty & +\bar{c}^Tx & -\bar{z}^T\tau & & & & = -\bar{x}^T\bar{s}-1 \\
 & x\in\mathcal{K}, & \tau\geq 0 & & s\in\mathcal{K} & \kappa\geq 0, &
\end{array}
\tag{HSD}
$$

where $\bar{x},\bar{s}\in\operatorname{int}(\mathcal{K})$, $\bar{y}\in\mathbb{R}^m$ are arbitrary starting points, τ,θ are scalars, $\bar{b}=b-A\bar{x}$, $\bar{c}=c-A^T\bar{y}-\bar{s}$ and $\bar{z}=c^T\bar{x}-b^T\bar{y}+1$. This model has the following properties [19, 52].

Theorem 3.6 (Properties of the HSD model). *System* (HSD) *is self-dual and it has a strictly feasible starting point, namely* $(x,s,y,\tau,\theta,\kappa)=(\bar{x},\bar{s},\bar{y},1,1,0)$. *The optimal value of these problems is* $\theta=0$, *and if* $\tau>0$ *at optimality then* $(x/\tau,y/\tau,s/\tau)$ *is an optimal solution for the original primal-dual problem with equal objective values, i.e., the duality gap is zero. If* $\tau=0$ *and* $\kappa>0$, *then the problem is either unbounded, infeasible, or the duality gap at optimality is nonzero. If* $\tau=\kappa=0$, *then either the problem is infeasible without a certificate (weakly infeasible) or the optimum is not attained.*

Remark 3.7. Due to strict complementarity, the $\tau=\kappa=0$ case cannot happen in linear optimization. The duality theory of conic optimization is weaker, this leads to all those ill-behaved problems.

The importance of this model is that the resulting system is strictly feasible with a known interior point, thus it can be solved directly with interior point methods.

3.4 Conic Optimization as a Complementarity Problem

3.4.1 Second Order Conic Case

In order to be able to present the second order conic case we need to define some elements of the theory of Jordan algebras for our particular case. All the proofs, along with the general theory can be found in [22]. Here we include as much of the theory (without proofs) as needed for the discussion. Our main source here is [3].

Given two vectors $u,v\in\mathbb{R}^n$ we can define a special product on them, namely:

$$u\circ v=(u^Tv;u_1v_{2:n}+v_1u_{2:n}). \tag{3.9}$$

The most important properties of this bilinear product are summarized in the following theorem:

Theorem 3.8 (Properties of $\circ$).

1. *Distributive law:* $u \circ (v + w) = u \circ v + u \circ w$.
2. *Commutative law:* $u \circ v = v \circ u$.
3. *The unit element is* $\iota = (1; 0)$, *i.e.,* $u \circ \iota = \iota \circ u = u$.
4. *Using the notation* $u^2 = u \circ u$ *we have* $u \circ (u^2 \circ v) = u^2 \circ (u \circ v)$.
5. *Power associativity:* $u^p = u \circ \cdots \circ u$ *is well-defined, regardless of the order of multiplication. In particular,* $u^p \circ u^q = u^{p+q}$.
6. *Associativity does not hold in general.*

The importance of this bilinear function lies in the fact that it can be used to generate the second order cone:

Theorem 3.9. *A vector x is in a second order cone (i.e., $x_1 \geq \|x_{2:n}\|_2$) if and only if it can be written as the square of a vector under the multiplication $\circ$, i.e., $x = u \circ u$.*

Moreover, analogously to the spectral decomposition theorem of symmetric matrices, every vector $u \in \mathbb{R}^n$ can be written as

$$u = \lambda_1 c^{(1)} + \lambda_2 c^{(2)}, \tag{3.10}$$

where $c^{(1)}$ and $c^{(2)}$ are on the boundary of the cone, and

$$c^{(1)^T} c^{(2)} = 0 \tag{3.11a}$$
$$c^{(1)} \circ c^{(2)} = 0 \tag{3.11b}$$
$$c^{(1)} \circ c^{(1)} = c^{(1)} \tag{3.11c}$$
$$c^{(2)} \circ c^{(2)} = c^{(2)} \tag{3.11d}$$
$$c^{(1)} + c^{(2)} = \iota \tag{3.11e}$$

The vectors $c^{(1)}$ and $c^{(2)}$ are called the Jordan frame and they play the role of rank one matrices. The numbers λ_1 and λ_2 are called eigenvalues of u. They behave much the same way as eigenvalues of symmetric matrices, except that in our case there is an easy formula to compute them:

$$\lambda_{1,2}(u) = u_1 \pm \|u_{2:n}\|_2 . \tag{3.12}$$

This also shows that a vector is in the second order cone if and only if both of its eigenvalues are nonnegative.

The spectral decomposition enables us to compute functions over the vectors:

$$\|u\|_F = \sqrt{\lambda_1^2 + \lambda_2^2} = \sqrt{2}\,\|u\|_2 , \tag{3.13a}$$

$$\|u\|_2 = \max\{|\lambda_1|, |\lambda_2|\} = |u_1| + \|u_{2:n}\|_2, \tag{3.13b}$$

$$u^{-1} = \lambda_1^{-1} c^{(1)} + \lambda_2^{-1} c^{(2)}, \tag{3.13c}$$

$$u^{\frac{1}{2}} = \lambda_1^{\frac{1}{2}} c^{(1)} + \lambda_2^{\frac{1}{2}} c^{(2)}, \tag{3.13d}$$

where $u \circ u^{-1} = u^{-1} \circ u = \iota$ and $u^{\frac{1}{2}} \circ u^{\frac{1}{2}} = u$.

Since the mapping $v \mapsto u \circ v$ is linear, it can be represented with a matrix. Indeed, introducing the arrowhead matrix

$$\operatorname{Arr}(u) = \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ u_2 & u_1 & & \\ \vdots & & \ddots & \\ u_n & & & u_1 \end{pmatrix}, \tag{3.14}$$

we have $u \circ v = \operatorname{Arr}(u)\, v = \operatorname{Arr}(u) \operatorname{Arr}(v)\, \iota$. Another operator is the quadratic representation, which is defined as

$$Q_u = 2 \operatorname{Arr}(u)^2 - \operatorname{Arr}\left(u^2\right), \tag{3.15}$$

thus $Q_u(v) = 2u \circ (u \circ v) - u^2 \circ v$ is a quadratic function[13] in u. This operator will play a crucial role in the construction of the Newton system.

Remember that second order cone optimization problems usually include several cones, i.e., $\mathcal{K} = \mathbb{L}_{n_1} \times \cdots \times \mathbb{L}_{n_k}$. For simplicity let us introduce the notation

$$\begin{aligned} A &= \left(A^1, \dots, A^k\right), \\ x &= \left(x^1; \dots; x^k\right), \\ s &= \left(s^1; \dots; s^k\right), \\ c &= \left(c^1; \dots; c^k\right). \end{aligned} \tag{3.16}$$

With this notation we can write

$$\begin{aligned} Ax &= \sum_{i=1}^{k} A^i x^i, \\ A^T y &= \left({A^1}^T y; , \dots; {A^k}^T y\right). \end{aligned} \tag{3.17}$$

Moreover, for a partitioned vector $u = (u^1; \dots; u^k)$, $\operatorname{Arr}(u)$ and Q_u are block diagonal matrices built from the blocks $\operatorname{Arr}\left(u^i\right)$ and Q_{u^i}, respectively.

[13] In fact, this operation is analogous to the mapping $V \mapsto UVU$ for symmetric matrices.

The optimality conditions for second order conic optimization are

$$\begin{aligned} Ax &= b,\, x \in \mathcal{K} \\ A^T y + s &= c,\, s \in \mathcal{K} \\ x \circ s &= 0. \end{aligned} \tag{3.18}$$

The first four conditions represent the primal and dual feasibility, while the last condition is called the complementarity condition. An equivalent form of the complementarity condition is $x^T s = 0$.

Now we perturb[14] the complementarity condition to get the central path:

$$\begin{aligned} Ax &= b,\, x \in \mathcal{K} \\ A^T y + s &= c,\, s \in \mathcal{K} \\ x^i \circ s^i &= 2\mu\iota^i,\, i = 1, \ldots, k, \end{aligned} \tag{3.19}$$

where $\iota^i = (1; 0; \ldots; 0) \in \mathbb{R}^{n_i}$. Finally, we apply the Newton method to this system to get the Newton step:

$$\begin{aligned} A\Delta x &= 0 \\ A^T \Delta y + \Delta s &= 0, \\ x^i \circ \Delta s^i + \Delta x^i \circ s^i &= 2\mu\iota^i - x^i \circ s^i,\, i = 1, \ldots, k, \end{aligned} \tag{3.20}$$

where $\Delta x = (\Delta x^1; \ldots; \Delta x^k)$ and $\Delta s = (\Delta s^1; \ldots; \Delta s^k)$. To solve this system we first rewrite it using the operator $\operatorname{Arr}()$:

$$\begin{pmatrix} & A & \\ A^T & & I \\ & \operatorname{Arr}(s) & \operatorname{Arr}(x) \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta x \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2\mu\iota - x \circ s \end{pmatrix}, \tag{3.21}$$

where $\iota = (\iota^1; \ldots; \iota^k)$. Eliminating Δx and Δs we get the so-called normal equation:

$$\left(A \operatorname{Arr}(s)^{-1} \operatorname{Arr}(x) A^T\right) \Delta y = -A \operatorname{Arr}(s)^{-1} (2\mu\iota - x \circ s). \tag{3.22}$$

The coefficient matrix is a $m \times m$. Unfortunately, not only this system is not symmetric, which is a disadvantage in practice, but in general it can be

[14] Our choice of perturbation might seem arbitrary but in fact this is the exact analog of what we did for linear optimization, since the vector $(1; 0)$ on the right hand side is the unit element for the multiplication $\circ$. See Section 3.6 to understand where the multiplier 2 comes from.

singular, even if x and s are in the interior of the cone $\mathcal{K}$. As an example[15] take $A = \left(0, \sqrt{3.69} + 0.7, 1\right)$, $\mathcal{K} = \left\{x \in \mathbb{R}^3 : x_1 \geq \sqrt{x_2^2 + x_3^2}\right\}$. The points $x = (1; 0.8; 0.5)$ and $s = (1; 0.7; 0.7;)$ are strictly primal and dual feasible, but $A \operatorname{Arr}(s)^{-1} \operatorname{Arr}(x) A^T = 0$.

To prevent singularity and to get a symmetric system we rewrite the original optimization problem (SOCO) in an equivalent form. Let us fix a scaling vector $p \in \operatorname{int}(\mathcal{K})$ and consider the scaled problem[16]

$$\begin{array}{lcr} \min \ \left(Q_{p^{-1}} c\right)^T (Q_p x) & \max b^T y & \text{(SOCOscaled)} \\ \left(A Q_{p^{-1}}\right)(Q_p x) = b & \left(A Q_{p^{-1}}\right)^T y + Q_{p^{-1}} s = Q_{p^{-1}} c & \\ Q_p x \in \mathcal{K} & Q_{p^{-1}} s \in \mathcal{K} & \end{array}$$

where p^{-1} is defined by (3.13c), and Q_p is given by (3.15). The exact form of p will be specified later. This scaling has the following properties:

Lemma 3.10. *If* $p \in \operatorname{int}(\mathcal{K})$, *then*

1. Q_p *and* $Q_{p^{-1}}$ *are inverses of each other, i.e.,* $Q_p Q_{p^{-1}} = I$.
2. *The cone* $\mathcal{K}$ *is invariant, i.e.,* $Q_p(\mathcal{K}) = \mathcal{K}$.
3. *Problems* (SOCO) *and* (SOCOscaled) *are equivalent.*

We can write the optimality conditions (3.18) for the scaled problem and perturb them to arrive at the central path for the symmetrized system. This defines a new Newton system:

$$\begin{aligned} \left(A Q_{p^{-1}}\right)(Q_p \Delta x) &= 0 \\ \left(A Q_{p^{-1}}\right)^T \Delta y + Q_{p^{-1}} \Delta s &= 0, \\ (Q_p x) \circ \left(Q_{p^{-1}} \Delta s\right) + (Q_p \Delta x) \circ \left(Q_{p^{-1}} s\right) &= 2\mu\iota - (Q_p x) \circ \left(Q_{p^{-1}} s\right). \end{aligned} \tag{3.23}$$

Using Lemma 3.10 we can eliminate the scaling matrices from the first two equations, but not the third one. Although rather complicated, this system is still a linear system in the variables Δx, Δy and Δs.

Before we can turn our attention to other elements of the algorithm we need to specify p. The most natural choice, i.e., $p = \iota$ is not viable as it does not provide a nonsingular Newton system. Another popular choice is the pair of primal-dual HKM directions, i.e.,

$$p = s^{1/2} \text{ or } p = x^{1/2}, \tag{3.24}$$

---

[15] See [67, S6.3.1].

[16] This scaling technique was originally developed for semidefinite optimization by Monteiro [57] and Zhang [99], and later generalized for second order cone optimization by Schmieta and Alizadeh [73].

in which case

$$Q_{p^{-1}}s = \iota \text{ or } Q_p x = \iota. \tag{3.25}$$

These directions are implemented as the default choice in the SOCO solver package SDPT3. Finally, probably the most studied and applied direction is the NT direction, defined as:

$$p = \left(Q_{x^{1/2}}\left(Q_{x^{1/2}}s\right)^{-1/2}\right)^{-1/2} = \left(Q_{s^{-1/2}}\left(Q_{s^{1/2}}x\right)^{1/2}\right)^{-1/2}. \tag{3.26}$$

This very complicated formula actually simplifies the variables, since

$$Q_p x = Q_{p^{-1}}s. \tag{3.27}$$

The NT scaling is implemented in SeDuMi and MOSEK and is also available in SDPT3.

We will now customize the generic IPM algorithm (see Algorithm 1 on page 242) for second order conic optimization. Let $\mu = \mu(x, s)$ be defined as

$$\mu(x, s) = \sum_{i=1}^{k} \frac{{x^i}^T s^i}{n_i}. \tag{3.28}$$

First let us define some centrality measures (see [3]). These measures are defined in terms of the scaled variable $w = (w_1; \ldots; w_k)$, where $w_i = Q_{x_i^{1/2}} s_i$.

$$\delta_F(x, s) := \|Q_{x^{1/2}}s - \mu\iota\|_F := \sqrt{\sum_{i=1}^{k} \left(\lambda_1(w_i) - \mu\right)^2 + \left(\lambda_2(w_i) - \mu\right)^2} \tag{3.29a}$$

$$\delta_\infty(x, s) := \|Q_{x^{1/2}}s - \mu\iota\|_2 := \max_{i=1,\ldots,k} \left\{|\lambda_1(w_i) - \mu|, |\lambda_2(w_i) - \mu|\right\} \tag{3.29b}$$

$$\delta_\infty^-(x, s) := \left\|(Q_{x^{1/2}}s - \mu\iota)^-\right\|_\infty := \mu - \min_{i=1,\ldots,k} \left\{\lambda_1(w_i), \lambda_2(w_2)\right\}, \tag{3.29c}$$

where the norms are special norms defined in (3.13) for the Jordan algebra. We can establish the following relations for these measures:

$$\delta_\infty^-(x, s) \le \delta_\infty(x, s) \le \delta_F(x, s). \tag{3.30}$$

The neighbourhoods are now defined as

$$\mathcal{N}(\gamma) := \{(x, y, s) \text{ strictly feasible} : \delta(x, s) \le \gamma\mu(x, s)\}. \tag{3.31}$$

Choosing $\delta(x, s) = \delta_F(x, s)$ gives a narrow neighbourhood, while $\delta(x, s) = \delta_\infty^-(x, s)$ defines a wide one.

The results are summarized in the following theorem, taken from [3, 60].

Theorem 3.11 (Short-step IPM for SOCO). *Choose*[17] $\gamma = 0.088$ *and* $\zeta = 0.06$. *Assume that we have a starting point* $(x^0, y^0, s^0) \in \mathcal{N}_F(\gamma)$. *Compute the Newton step from the scaled Newton system* (3.23). *In every iteration,* μ *is decreased to* $\left(1 - \frac{\zeta}{\sqrt{k}}\right)\mu$, *i.e.,* $\theta = \frac{\zeta}{\sqrt{k}}$, *and the stepsize is* $\alpha = 1$. *This algorithm finds an* ε*-optimal solution for the second order conic optimization problem* (SOCO) *with* k *second order cones in at most*

$$\mathcal{O}\left(\sqrt{k}\log\frac{1}{\varepsilon}\right) \tag{3.33}$$

iterations. The cost of one iteration depends on the sparsity structure of the coefficient matrix A. *If all the data is dense then it is*

$$\mathcal{O}\left(m^3 + m^2 n + \sum_{i=1}^{k} n_i^2\right). \tag{3.34}$$

It might be surprising that the iteration complexity of the algorithm is independent of the dimensions of the cones. However, the cost of one iteration depends on the dimension of the cones.

Although this is essentially the best possible complexity result for second order cone optimization, this algorithm is not efficient enough in practice since θ is too small. Practical implementations use predictor-corrector schemes, see [67, 73, 77, 84] for more details.

Unlike the case of linear optimization, here we do not have a way to round an almost optimal interior solution to an optimal one, we have to live with approximate solutions.

3.4.2 Semidefinite Optimization

Interior point methods for semidefinite optimization have a very similar structure to the methods presented so far. We will apply the Newton method to the perturbed optimality conditions of semidefinite optimization.

---

[17] Any values $\gamma \in (0, 1/3)$ and $\zeta \in (0, 1)$ satisfying

$$\frac{4(\gamma^2 + \zeta^2)}{(1 - 3\gamma)^2}\left(1 - \frac{\zeta}{\sqrt{2n}}\right)^{-1} \leq \gamma \tag{3.32}$$

would work here.

The KKT optimality conditions for semidefinite optimization are:

$$\begin{aligned} \operatorname{Tr}\left(A^{(i)}X\right) &= b_i, i = 1, \ldots, m,\ X \in \mathbb{PS}^{n\times n} \\ \sum_{i=1}^{m} y_i A^{(i)} + S &= C,\ S \in \mathbb{PS}^{n\times n} \\ XS &= 0. \end{aligned} \tag{3.35}$$

Again, the first four conditions ensure feasibility, while the last equation is the complementarity condition. The last equation can be written equivalently as $\operatorname{Tr}(XS) = 0$. Now we perturb the complementarity condition, this way we arrive at the central path:

$$\begin{aligned} \operatorname{Tr}\left(A^{(i)}X\right) &= b_i, i = 1, \ldots, m,\ X \in \mathbb{PS}^{n\times n} \\ \sum_{i=1}^{m} y_i A^{(i)} + S &= C,\ S \in \mathbb{PS}^{n\times n} \\ XS &= \mu I, \end{aligned} \tag{3.36}$$

where I is the identity matrix. Now we try to apply the Newton method the same way we did for SOCO and LO, i.e., replace the variables with the updated ones and ignore the quadratic terms. This way we get:

$$\begin{aligned} \operatorname{Tr}\left(A^{(i)}\Delta X\right) &= 0, i = 1, \ldots, m \\ \sum_{i=1}^{m} \Delta y_i A^{(i)} + \Delta S &= 0 \\ X\Delta S + \Delta X S &= \mu I - XS. \end{aligned} \tag{3.37}$$

We want to keep the iterates X and S symmetric and positive definite, thus we need ΔX and ΔS to be symmetric as well. However, solving (3.37) the displacement ΔX is typically not symmetric, simply due to the fact that the product of two symmetric matrices is not symmetric. Moreover, forcing the symmetry of ΔX by adding $\Delta X = \Delta X^T$ as a new constraint will make the problem overdetermined. Our first attempt at formulating the Newton system fails spectacularly.

Scaling techniques for semidefinite optimization

The solution to the problem we encountered at the end of the previous section is again to rewrite the optimality conditions (3.35) in an equivalent form and use that system to derive the central path. This technique is called scaling or symmetrization and there are many ways to rewrite the optimality conditions, see [82] for a thorough review. This symmetrization replaces $XS = \mu I$ in (3.36) with $\frac{1}{2}(MXS + SXM) = \mu M$, where M might depend on X and S, and can thus change from iteration to iteration. This choice defines the Monteiro-Zhang family of search directions. The new symmetrized central path equations are

$$\begin{aligned} \operatorname{Tr}\left(A^{(i)}X\right) &= b_i, i = 1, \ldots, m,\ X \in \mathbb{PS}^{n\times n} \\ \sum_{i=1}^m y_i A^{(i)} + S &= C,\ S \in \mathbb{PS}^{n\times n} \\ MXS + SXM &= \mu M, \end{aligned} \tag{3.38}$$

and the Newton system is

$$\begin{aligned} \operatorname{Tr}\left(A^{(i)}\Delta X\right) &= 0, i = 1, \ldots, m \\ \sum_{i=1}^m \Delta y_i A^{(i)} + \Delta S &= 0 \\ MX\Delta S + M\Delta XS + S\Delta XM + \Delta SXM &= 2\mu I - MXS - SXM. \end{aligned} \tag{3.39}$$

The solution matrices ΔX and ΔS of this system are symmetric, thus we can update the current iterates maintaining the symmetry of the matrices. Details on how to solve this system can be found in [77].

Some standard choices of the scaling matrix M are (see [82] for more directions):

AHO scaling: The most natural choice, $M = I$. Unfortunately, the resulting system will have a solution only if X and S are in a small neighbourhood of the central path.

NT scaling: Probably the most popular choice,

$$M = S^{1/2}\left(S^{1/2}XS^{1/2}\right)^{-1/2}S^{1/2}. \tag{3.40}$$

This type of scaling has the strongest theoretical properties. Not surprisingly, most algorithmic variants use this scaling. It also facilitates the use of sparse linear algebra, see [77].

HKM scaling: In this case $M = S$ or $M = X^{-1}$. Typically, these scalings are somewhat faster to compute than the NT scaling, but certain large portions of the theory (such as [67]) are only developed for NT scaling.

Proximity measures

Let μ be defined as $\mu = \mu(X, S) := \frac{\mathrm{Tr}(XS)}{n}$ for the rest of this section. Now we need to define some centrality measures similar to (2.55) and (3.29). The most popular choices for semidefinite optimization include

$$\delta_F(X,S) := \left\| X^{1/2}SX^{1/2} - \mu I \right\|_F = \sqrt{\sum_{i=1}^{n} \left(\lambda_i(X^{1/2}SX^{1/2}) - \mu\right)^2} \tag{3.41a}$$

$$\delta_\infty(X,S) := \left\| X^{1/2}SX^{1/2} - \mu I \right\| = \max_i \left| \lambda_i(X^{1/2}SX^{1/2}) - \mu \right| \tag{3.41b}$$

$$\begin{aligned} \delta_\infty^-(X,S) &:= \left\| \left(X^{1/2}SX^{1/2} - \mu I \right)^- \right\|_\infty \\ &:= \max_i \left(\mu - \lambda_i(X^{1/2}SX^{1/2}) \right), \end{aligned} \tag{3.41c}$$

see [59] and the references therein for more details. For strictly feasible X and S, these measures are zero only on the central path. Due to the properties of norms we have the following relationships:

$$\delta_\infty^-(X,S) \le \delta_\infty(X,S) \le \delta_F(X,S). \tag{3.42}$$

The neighbourhoods are defined as

$$\mathcal{N}(\gamma) := \{(X,y,S) \text{ strictly feasible} : \delta(X,S) \le \gamma\mu(X,S)\}. \tag{3.43}$$

Choosing $\delta(X,S) = \delta_F(X,S)$ gives a narrow neighbourhood, while $\delta(X,S) = \delta_\infty^-(X,S)$ defines a wide one.

A short-step interior point method

The following theorem, taken from [59], summarizes the details and the complexity of a short-step interior point algorithm for semidefinite optimization. Refer to Algorithm 1 on page 242 for the generic interior point algorithm.

Theorem 3.12 (Short-step IPM for SDO). *Choose*[18] $\gamma = 0.15$ *and* $\zeta = 0.13$. *Assume that we have a starting point* $(X^0, y^0, S^0) \in \mathcal{N}_F(\gamma)$. *We get the Newton step from* (3.39). *In every iteration,* μ *is decreased to* $\left(1 - \frac{\zeta}{\sqrt{n}}\right)\mu$,

---

[18] Any values $\gamma \in (0, 1/\sqrt{2})$ and $\zeta \in (0,1)$ satisfying

$$\frac{2(\gamma^2 + \zeta^2)}{(1-\sqrt{2}\gamma)^2}\left(1 - \frac{\zeta}{\sqrt{n}}\right)^{-1} \le \gamma \tag{3.44}$$

would work here.

i.e., $\theta = \frac{\zeta}{\sqrt{n}}$*, and the stepsize is* $\alpha = 1$*. This algorithm finds and* ε*-optimal solution for the semidefinite optimization problem* (SDO) *with an* n *dimensional cone in at most*

$$\mathcal{O}\left(\sqrt{n}\log\frac{1}{\varepsilon}\right) \tag{3.45}$$

iterations. If all the data matrices are dense[19] *then the cost of one iteration is* $\mathcal{O}\left(mn^3 + m^2n^2 + m^3\right)$*.*

Remark 3.13. Depending on the magnitude of m compared to n any of the three terms of this expression can be dominant. The problem has $\mathcal{O}\left(n^2\right)$ variables, thus $m \leq n^2$. If m is close to n^2 then the complexity of one iteration is $\mathcal{O}\left(n^6\right)$, while with a much smaller m of order $\sqrt{n}$ the complexity is $\mathcal{O}\left(n^{3.5}\right)$.

Although this algorithmic variant is not very efficient in practice, this is still the best possible theoretical complexity result. Practical implementations usually use predictor-corrector schemes, see [77] for more details.

As we have already seen with second order conic optimization, it is not possible to obtain an exact solution to the problem. All we can get is an ε-optimal solution, see [68] for detailed complexity results.

3.5 Summary

To summarize the results about conic optimization let us go through our checklist from Section 2.3.

- We showed that the duality properties of conic optimization are slightly weaker than that of linear optimization, we need to assume strict feasibility (the interior point condition) for strong duality.
- We embedded the conic optimization problems (PCon) and (DCon) into a strictly feasible self-dual problem (HSD). From the optimal solutions of the self-dual model we can
  - derive optimal solutions for the original problem, or
  - decide primal or dual infeasibility, or
  - conclude that no optimal primal-dual solution pair exists with zero duality gap.
- If a strictly feasible solution exists (either in the original problem or in the self-dual model) then
  - the central path exists;
  - the central path converges to a maximally (not necessarily strictly) complementary solution;

[19] The complexity can be greatly reduced by exploiting the sparsity of the data, see [77] and the references therein.

  - the limit point of the central path is not necessarily the analytic center of the optimal set (only if the problem has a strictly complementary solution).

- Due to the lack of a rounding scheme we cannot get exact optimal solutions from our algorithm and thus cannot use the algorithm to get exact solutions.

*3.6 *Barrier Functions in Conic Optimization*

Interior point methods for conic optimization can also be introduced through barrier functions in a similar fashion as we did in Section 2.5 for linear optimization. However, the barrier functions for conic optimization are more complicated and the discussion is a lot more technical, much less intuitive.

A suitable logarithmic barrier function for a second order cone is

$$\phi(x) = -\ln\left(x_1^2 - \|x_{2:n}\|_2^2\right) = -\ln\lambda_1(x) - \ln\lambda_2(x), \tag{3.46}$$

assuming that x is in the interior of the second order cone. We can see that when the point x is getting close to the boundary, then at least one of its eigenvalues is getting close to 0 and $\phi(x)$ is diverging to infinity. For the optimality conditions of this problem we will need the derivatives of the barrier function $\phi(x)$:

$$\nabla\phi(x) = -2\frac{(x_1; -x_{2:n})^T}{x_1^2 - \|x_{2:n}\|_2^2} = -2\left(x^{-1}\right)^T, \tag{3.47}$$

where the inverse is taken in the Jordan algebra. The multiplier 2 appears due to the differentiation of a quadratic function, and it will also appear in the central path equations (3.19).

For the cone of positive semidefinite matrices we can use the barrier function

$$\phi(X) = -\ln\det\,(X) = -\sum_{i=1}^{n}\ln\lambda_i(X), \tag{3.48}$$

which has the derivative

$$\nabla\phi(X) = -\left(X^{-1}\right)^T. \tag{3.49}$$

Having these functions we can rewrite the conic optimization problem (PCon) as a linearly constrained problem

$$\begin{aligned} &\min\ c^T x + \mu\phi(x) \\ &\quad Ax = b, \end{aligned} \tag{PCon-Barrier}$$

where $\mu \geq 0$. The KKT optimality conditions for this problem are the same systems as (3.19) and (3.36) defining the central path, thus the barrier approach again provides an alternative description of the central path. For more details on the barrier approach for conic optimization see, e.g., [4].

4 Interior Point Methods for Nonlinear Optimization

First we will solve the nonlinear optimization problem by converting it into a nonlinear complementarity problem. We will present an interior point algorithm for this problem, analyze its properties and discuss conditions for polynomial complexity. Then we present a direct approach of handling nonlinear inequality constraints using barrier functions and introduce the concept of self-concordant barrier functions.

4.1 Nonlinear Optimization as a Complementarity Problem

Let us consider the nonlinear optimization problem in the form

$$\begin{aligned} \min\ & f(x) \\ g_j(x) &\leq 0,\ j = 1, \ldots, m \\ x &\geq 0, \end{aligned} \tag{NLO}$$

where $x \in \mathbb{R}^n$ and $f, g_j : \mathbb{R}^n \to \mathbb{R}$, are continuously differentiable convex functions. We will use the notation $g(x) = (g_1(x); \ldots; g_m(x))$. The KKT optimality conditions for this problem are

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^{m} \nabla g_j(x) y_j &\geq 0 \\ g_j(x) &\leq 0 \\ x, y &\geq 0 \\ \left(\nabla f(x) + \sum_{i=1}^{m} \nabla g_j(x) y_j \right)^T x &= 0 \\ g(x)^T y &= 0. \end{aligned} \tag{4.1}$$

Introducing

$$L(x, y) := f(x) + g(x)^T y \tag{4.2a}$$

$$F(\bar{x}) := \begin{pmatrix} \nabla_x L(x, y) \\ -g(x) \end{pmatrix} \tag{4.2b}$$

$$\bar{x} := \begin{pmatrix} x \\ y \end{pmatrix} \tag{4.2c}$$

we can write the nonlinear optimization problem as an equivalent nonlinear complementarity problem:

$$\begin{aligned} F(\bar{x}) - \bar{s} &= 0 \\ \bar{x}, \bar{s} &\geq 0 \\ \bar{x}\bar{s} &= 0. \end{aligned} \tag{4.3}$$

4.2 Interior Point Methods for Nonlinear Complementarity Problems

In this section we derive an algorithm for this problem based on [70].

Let us now simplify the notation and focus on the nonlinear complementarity problem in the following form:

$$\begin{aligned} F(x) - s &= 0 \\ x, s &\geq 0 \\ xs &= 0, \end{aligned} \tag{NCP}$$

where $x, s \in \mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{R}^n$. After perturbing the third equation (the complementarity condition) we receive the equations for the central path. Note that the existence of the central path requires stronger assumptions than in the linear or conic case, see [25] and the references therein for details.

$$\begin{aligned} F(x) - s &= 0 \\ x, s &\geq 0 \\ xs &= \mu e, \end{aligned} \tag{4.4}$$

where $\mu \geq 0$ and e is the all one vector. We use the Newton method to solve this system, the corresponding equation for the Newton step is:

$$\begin{aligned} F'(x)\Delta x - \Delta s &= 0 \\ s\Delta x + x\Delta s &= \mu e - xs, \end{aligned} \tag{4.5}$$

where $F'(x)$ is the Jacobian of $F(x)$. In general, the point $x + \Delta x$ is not feasible, i.e., $F(x + \Delta x) \geq 0$ and/or $x + \Delta x \geq 0$ is not satisfied, thus we will need to use a stepsize $\alpha > 0$ and consider a strictly feasible $x(\alpha) := x + \alpha\Delta x$ as the new (primal) iterate. The new dual iterate will be defined as $s(\alpha) = F(x + \alpha\Delta x)$. Note that unlike in linear and conic optimization, here $s(\alpha) \neq s + \alpha\Delta s$.

The algorithm is structured analogously to the generic structure of IPMs presented as Algorithm 1. All we need to do is to specify the details: the proximity measure $\delta(x, s)$, the choice of stepsize α and the update strategy of μ.

The proximity measure

There are several variants in existing implementations. The most important ones are

$$\delta_2(x,s) = \|xs - \mu e\|_2 \tag{4.6a}$$
$$\delta_\infty(x,s) = \|xs - \mu e\|_\infty \tag{4.6b}$$
$$\delta_\infty^-(x,s) = \left\|(xs - \mu e)^-\right\|_\infty := \max_i\left(\mu - x_i s_i\right), \tag{4.6c}$$

where $\mu = x^T s/n$. This enables us to define a neighbourhood of the central path:

$$\mathcal{N}(\gamma) = \left\{(x,s) \text{ strictly feasible} : \delta(x,s) \leq \gamma\mu\right\}, \tag{4.7}$$

where $\gamma \in (0, 1)$.

Choosing the stepsize α

For nonlinear optimization problems the stepsize is chosen using a line-search. We want to get a large step but stay away from the boundary of the feasible set. Let $\alpha_{\max}$ be the maximum feasible stepsize, i.e., the maximal value of α such that $x + \alpha\Delta x \geq 0$ and $F(x + \alpha\Delta x) \geq 0$.

We are looking for a stepsize $\alpha < \alpha_{\max}$ such that

- $(x(\alpha), s(\alpha))$ is inside the neighbourhood $\mathcal{N}(\gamma)$, and
- the complementarity gap $x(\alpha)^T F(x(\alpha))$ is minimized.

In some practical implementations $\alpha = 0.95\alpha_{\max}$ (or $\alpha = 0.99\alpha_{\max}$) is used as the stepsize, enhanced with a safeguarded backtracking strategy. The extra difficulty with general nonlinear optimization problems is that the line-search can get stuck in a local minimum, thus some globalization scheme is needed. Such ideas are implemented in the IPOPT solver [90].

Updating μ

Usually we try to decrease μ at a superlinear rate, if possible. In short-step methods, μ is changed to $\mu\left(1-\frac{\zeta}{\sqrt{n}}\right)$ after every iteration, i.e., $\theta=\frac{\zeta}{\sqrt{n}}$ in the general IPM framework on page 242, ζ is a constant depending on the neighbourhood parameter γ and the smoothness of the mapping F. The smoothness is quantified with a Lipschitz constant L in Assumption 4.1.

4.2.1 Complexity of IPM for NCP

Now assume that the Jacobian $F'(x)$ of $F(x)$ is a positive semidefinite matrix for all values of x. Then problem (NCP) is called a monotone nonlinear complementarity problem. If the original nonlinear optimization problem (NLO) is convex, then this always holds. To be able to prove polynomial convergence of IPMs for convex nonlinear problems we need to control the difference between $s(\alpha)=F(x(\alpha))$ and $s+\alpha\Delta s$. We assume a smoothness condition [8]:

Assumption 4.1 *Consider the nonlinear complementarity problem* (NCP). *Assume that $F(x)$ satisfies the scaled Lipschitz property, i.e., for any $x>0$, $h\in\mathbb{R}^n$, satisfying $|h_i/x_i|\le\beta<1$, there exists a constant $L(\beta)>1$ such that*

$$\|x\cdot(F(x+h)-F(x)-F'(x)h)\|_1\le L(\beta)h^TF'(x)h. \tag{4.8}$$

The complexity result is summarized in the following theorem:

Theorem 4.2 (Complexity of short-step IPM for monotone NCP). *Assume that $F(x)$ is a monotone mapping satisfying the scaled Lipschitz property. The proximity measure is based on the 2-norm and assume that a strictly feasible starting point in $\mathcal{N}_2(\gamma)$ with $x^Ts/n\le 1$ is available.*

The Newton step is computed from (4.5). *If γ and ζ are chosen properly, then $\alpha=1$ is a valid stepsize, i.e., no line-search is necessary.*

This algorithm yields an ε-complementary solution for (NCP) *in at most $\mathcal{O}\left(\sqrt{n}L\log(1/\varepsilon)\right)$ iterations.*

Explicit forms of the constants and detailed proofs can be found in [8]. The cost of one iteration depends on the actual form of $F(x)$. It includes computing the Jacobian of F at every iteration and solving an $n\times n$ linear system. When full Newton steps are not possible,[20] then finding $\alpha_{\max}$ and determining the stepsize α with a line-search are significant extra costs.

[20] This is the typical situation, as in practice we rarely have explicit information on the Lipschitz constant L.

4.3 Initialization by Embedding

Interior point methods require a strictly feasible starting point, but for nonlinear optimization problems even finding is feasible point is quite challenging. Moreover, if the original problem has nonlinear equality constraints which are modelled as two inequalities then the resulting system will not have an interior point solution. To remedy these problems we use a homogeneous embedding, similar to the ones presented in Section 2.2 and Section 3.3. Consider the following system [7, 8, 96]:

$$\begin{aligned} \nu F(x/\nu) - s &= 0 \\ x^T F(x/\nu) - \rho &= 0 \\ x, s, \nu, \rho &\geq 0 \\ xs &= 0 \\ \nu\rho &= 0. \end{aligned} \tag{NCP-H}$$

This is a nonlinear complementarity problem similar to (NCP). The properties of the homogenized system are summarized in the following theorem.

Theorem 4.3. *Consider the nonlinear complementarity problem* (NCP) *and its homogenized version* (NCP-H). *The following results hold:*

1. *The homogenized problem* (NCP-H) *is an* (NCP).
2. *If the original* (NCP) *is monotone then the homogenized* (NCP) *is monotone, too, thus we can use the algorithm presented in Section 4.2.*
3. *If the homogenized* (NCP) *has a solution* (x, s, ν, ρ) *with* $\nu > 0$ *then* $(x/\nu, s/\nu)$ *is a solution for the original system.*
4. *If* $\nu = 0$ *for all the solutions of* (NCP-H) *then the original system* (NCP) *does not have a solution.*

*4.4 *The Barrier Method*

An alternative way to introduce interior point methods for nonlinear optimization is to use the barrier technique already presented in Section 2.5 and Section 3.6. The basic idea is to place the nonlinear inequalities in the objective function inside a barrier function. Most barrier function are based on logarithmic functions.

The nonlinear optimization problem (NLO) can be rewritten as

$$\min\ f(x) - \mu \sum_{j=1}^{m} \ln(-g_j(x)) - \mu \sum_{i=1}^{n} \ln(x_i). \tag{4.9}$$

If x_i or $-g_j(x)$ gets close to 0, then the objective function grows to infinity. Our goal is to solve this barrier problem approximately for a given μ, then decrease μ and resolve the problem. If μ is decreased at the right rate and the approximate solutions are good enough, then this method will converge to an optimal solution of the nonlinear optimization problem. See [63] for details on the barrier approach for nonlinear optimization.

5 Existing Software Implementations

After their early discovery in the 1950s, by the end of the 1960s IPMs were sidelined because their efficient implementation was quite problematic. As IPMs are based on Newton steps, they require significantly more memory than first order methods. Computers at the time had very limited memory. Furthermore, the Newton system is inherently becoming ill-conditioned as the iterates approach the optimal solution set. Double precision floating point arithmetic and regularization techniques were in their very early stage at that time. Solving large scale linear systems would have required sparse linear algebra routines, which were also unavailable. Most of these difficulties have been solved by now and so IPMs have become a standard choice in many branches of optimization.

In the following we give an overview of existing implementations of interior point methods. See Table 1 for a quick comparison their features. The web site of the solvers and the bibliographic references are listed in Table 2.

Table 1 A comparison of existing implementations of interior point methods

Solver	License	LO	SOCO	SDO	NLO
CLP barrier	open source	✓	QO		
LIPSOL	open source	✓			
GLPK ipm	open source	✓			
HOPDM	commercial	✓	QO		✓
MOSEK barrier	commercial	✓	✓		✓
CPLEX barrier	commercial	✓	✓[21]		
XPRESS barrier	commercial	✓	QO		
CSDP	open source	✓		✓	
SDPA	open source	✓		✓	
SDPT3	open source	✓	✓	✓	
SeDuMi	open source	✓	✓	✓	
IPOPT	open source	✓[22]	✓[23]		✓
KNITRO	commercial	✓[22]	✓[23]		✓
LOQO	commercial	✓[22]	✓[23]		✓

[21] CPLEX solves second-order conic problems by treating them as special (nonconvex) quadratically constrained optimization problems.

[22] In theory all NLO solvers can solve linear optimization problems, but their efficiency and accuracy is worse than that of dedicated LO solvers.

[23] LOQO does solve second-order conic optimization problems but it uses a different approach. It handles the constraint $x_1 - \|x_{2:n}\|_2 \geq 0$ as a general nonlinear constraint, with

Table 2 Availability of implementations of IPMs

CLP	[24], `http://www.coin-or.org/Clp`
LIPSOL	[100], `http://www.caam.rice.edu/~zhang/lipsol`
GLPK	[28], `http://www.gnu.org/software/glpk`
HOPDM	[15], `http://www.maths.ed.ac.uk/~gondzio/software/hopdm.html`
MOSEK	[6], `http://www.mosek.com`
CPLEX	[12], `http://www.ilog.com`
XPRESS-MP	[41], `http://www.dashoptimization.com`
CSDP	[13], `http://projects.coin-or.org/Csdp`
SDPA	[95], `http://homepage.mac.com/klabtitech/sdpa-homepage`
SDPT3	[86], `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`
SeDuMi	[76], `http://sedumi.ie.lehigh.edu/`
IPOPT	[90], `http://projects.coin-or.org/Ipopt`
KNITRO	[14], `http://www.ziena.com/knitro.htm`
LOQO	[89], `http://www.princeton.edu/~rvdb/loqo`

5.1 Linear Optimization

Interior point algorithms are the method of choice for large scale, sparse, degenerate linear optimization problems. Solvers using the simplex method are usually not competitive on those problems due to the large number of pivots needed to get to an optimal solution. However, interior point methods still do not have an efficient warm start strategy, something simplex based methods can do naturally, so their use for branch-and-bound type algorithms is limited.

IPMs have also been implemented in leading commercial packages, usually together with a simplex based solver. Comprehensive surveys of implementation strategies of IPMs can be found in, e.g., [5, 36]. For a review on the strengths and weaknesses of interior point methods versus variants of the simplex method see [43].

Linear optimization problems with up to a million variables can be solved routinely on a modern PC. On larger parallel architectures, linear and quadratic problems with billions of variables have been solved [34].

5.2 Conic Optimization

Interior point methods are practically the only choice for semidefinite optimization, most of the existing general purpose solvers fall into this category, only PENSDP[24] being a notable exception. Also, PENSDP is the only solver

some extra care taken due to the nondifferentiability of this form. In a similar way, other IPM based NLO solvers can solve SOCO problems in principle.

[24] [49], `http://www.penopt.com/pensdp.html`

that can handle nonlinear semidefinite problems and it is also the only commercial SDO solver (at least at the time this chapter is written).

The implementation of IPMs for conic optimization is more complicated than that for linear optimization, see [13, 77, 83] for more details.

Unfortunately, commercial modelling languages do not support SDO, thus limit its use in the commercial sector. Second order conic optimization is in a slightly better situation, since it is easily formulated, but there are only very few specialized solvers available. Only very few solvers can solve problems including both second order and semidefinite constraints, currently only SeDuMi and SDPT3. Both of these packages run under Matlab.

There are two open source modelling languages that support conic optimization: Yalmip[25] and CVX[26]. Both of these packages are written in Matlab.

5.3 Nonlinear Optimization

There are literally hundreds of solvers available for nonlinear optimization and only a small fraction of those use interior point methods. On the other hand, arguably, the most powerful, robust solvers are actually based on interior point methods, IPOPT, KNITRO and LOQO being the most successful ones. These are all general use nonlinear optimization solvers, they can handle nonconvex problems as well (yielding a locally optimal solution). Some codes have been specialized for optimization problems with complementarity constraints. The best known variant is IPOPT-C [71], an extension of IPOPT.

The implementation of these methods poses further challenges, see [90] for details.

6 Some Open Questions

Interior point algorithms have proved to be very successful methods for linear and nonlinear optimization, especially for large-scale problems. The "interior-point revolution" [92] has completely changed the field of optimization. By today, the fundamental theoretical questions regarding complexity and convergence of interior point methods have been addressed, see also [62] for a recent survey. Most importantly, we know that results about the iteration complexity of these methods cannot be improved further, see [21] for details on the worst-case complexity of interior point methods.

---

[25] [50], `http://control.ee.ethz.ch/~joloef/yalmip.php`

[26] [38, 39], `http://www.stanford.edu/~boyd/cvx`

6.1 Numerical Behaviour

Current research is focusing on efficient implementations of the methods. Due to the ill-conditioned nature of the Newton system in the core of IP methods, people are looking for ways to improve the numerical behaviour of the implementations. Some notable results are included in [31, 78, 79]. Most of these ideas are implemented in leading interior point solvers.

6.2 Rounding Procedures

Iterates of interior point methods stay inside the set of feasible solutions, while with a linear objective, the optimal solution is on the boundary of the feasible set. Rounding procedures try to jump from the last iterate of the IPM to an optimal solution on the boundary. This theory has been well-developed for linear optimization and linear complementarity problems [56, 72]. For conic optimization, the mere existence of such a method is an open question. In general we cannot expect to be able to get an exact optimal solution, but under special circumstances we might be able to get one.

6.3 Special Structures

Exploiting sparsity has always been one of the easiest ways to improve the performance of an optimization algorithm. With the availability of efficient sparse linear algebra libraries and matrix factorization routines, general (unstructured) sparsity seems to have been taken care of. On the other hand, sparse problems containing some dense parts pose a different challenge [30]. Moreover, even very sparse semidefinite optimization problems lead to a fully dense Newton system, which puts a limit on the size of the problems that can be solved.

There are several other special types of structures that cannot be fully exploited by current implementations of interior point methods. This limits the size of the problems that can be solved with IPMs. At the same time it offers a wide open area of further research.

6.4 Warmstarting

A serious deficiency of interior point methods is the lack of an efficient warm-starting scheme. The purpose of a warm-start scheme is to significantly reduce the number of iterations needed to reoptimize the problem after changes to the data (constraints are added or deleted, numbers are changed). Despite numerous attempts (see [33, 35, 98]), none of the methods are particularly successful.

If the change in the problem data is small enough then simplex based methods can very quickly find a new optimal solution. If the change is large (hundreds or thousands of new constraints are added) then interior point methods have a slight edge over first order methods.

6.5 Parallelization

With the general availability of inexpensive multiple core workstations and distributed computing environments, parallelization of optimization algorithms is more important than ever. Most developers are working on a parallelized version of their codes. Some success stories are reported in [13, 34, 44, 61].

Acknowledgements The authors were supported by the NSERC Discovery Grant #5-48923, the Canada Research Chair Program and a MITACS project. Research was done while both authors working at the School of Computational Engineering and Science, McMaster University, Hamilton, Ontario, Canada.

References

1. F. Alizadeh. *Combinatorial optimization with interior point methods and semidefinite matrices*, Ph.D. thesis, University of Minnesota, Minneapolis, USA, 1991.
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 and D. Goldfarb. *Second-order cone programming*, Mathematical Programming, Series B, 95,3–51, 2002.
4. F. Alizadeh and S. H. Schmieta. *Symmetric cones, potential reduction methods*, In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 8, pages 195–233. Kluwer Academic Publishers, 2000.
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. E. D. Andersen, B. Jensen, R. Sandvik, and U. Worsøe. *The improvements in MOSEK version 5.*, Technical report 1-2007, MOSEK ApS, Fruebjergvej 3 Box 16, 2100 Copenhagen, Denmark, 2007.
7. E. D. Andersen and Y. Ye. *A computational study of the homogeneous algorithm for large-scale convex optimization*, Computational Optimization and Applications, 10(3), 243–280, 1998.
8. E. D. Andersen and Y. Ye. *On a homogeneous algorithm for the monotone complementarity problem*, Mathematical Programming, Series A, 84(2),375–399, 1999.

9. K. Anstreicher. *Large step volumetric potential reduction algorithms for linear programming*, Annals of Operations Research, 62, 521–538, 1996.
10. K. Anstreicher. *Volumetric path following algorithms for linear programming*, Mathematical Programming, 76, 245–263, 1997.
11. A. Ben-Tal and A. Nemirovski *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MPS-SIAM Series on Optimization. SIAM, Philadelphia, PA, 2001.
12. R. E. Bixby. *Solving real-world linear programs: A decade and more of progress*, Operations Research, 50(1), 3–15, 2002.
13. B. Borchers and J. G. Young. *Implementation of a primal-dual method for SDP on a shared memory parallel architecture*, Computational Optimization and Applications, 37(3), 355–369, 2007.
14. R. Byrd, J. Nocedal, and R. Waltz. *KNITRO: An integrated package for nonlinear optimization*, In G. Di Pillo and M. Roma, editors, *Large-Scale Nonlinear Optimization*, volume 83 of *Nonconvex Optimization and Its Applications*. Springer, 2006.
15. M. Colombo and J. Gondzio. *Further development of multiple centrality correctors for interior point methods*, Computational Optimization and Applications, 41(3), 277–305, 2008.
16. E. de Klerk, *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*, Kluwer Academic Publichers, Dordrecht, The Netherlands, 2002.
17. E. de Klerk, C. Roos, and T. Terlaky. *Semi-definite problems in truss topology optimization*, Technical Report 95-128, Faculty of Technical Mathematics and Informatics, T.U. Delft, The Netherlands, 1995.
18. E. de Klerk, C. Roos, and T. Terlaky. *Initialization in semidefinite programming via a self-dual, skew-symmetric embedding*, OR Letters, 20, 213–221, 1997.
19. E. de Klerk, C. Roos, and T. Terlaky. *Infeasible-start semidefinite programming algorithms via self-dual embeddings*. In P. M. Pardalos and H. Wolkowicz, editors, *Topics in Semidefinite and Interior Point Methods*,, volume 18 of *Fields Institute Communications*, pages 215–236. AMS, Providence, RI, 1998.
20. E. de Klerk, C. Roos, and T. Terlaky. *On primal-dual path-following algorithms for semidefinite programming*, In F. Gianessi, S. Komlósi, and T. Rapcsák, editors, *New Trends in Mathematical Programming*, pages 137–157. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
21. A. Deza, E. Nematollahi, and T. Terlaky. *How good are interior point methods? Klee-Minty cubes tighten iteration-complexity bounds*, Mathematical Programming, 113, 1–14, 2008.
22. J. Faraut and A. Korányi *Analysis on Symmetric Cones*, Oxford Mathematical Monographs. Oxford University Press, 1994.
23. A. V. Fiacco and G. P. McCormick *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, 1968. Reprint: Vol. 4. *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, USA, 1990.
24. J. Forrest, D. de le Nuez, and R. Lougee-Heimer *CLP User Guide*, IBM Corporation, 2004.
25. A. Forsgren, Ph. E. Gill, and M. H. Wright. *Interior methods for nonlinear optimization*, SIAM Review, 44(4), 525–597, 2002.
26. R. M. Freund. *On the behavior of the homogeneous self-dual model for conic convex optimization*, Mathematical Programming, 106(3), 527–545, 2006.
27. R. Frisch. *The logarithmic potential method for solving linear programming problems*, Memorandum, University Institute of Economics, Oslo, Norway, 1955.
28. *GNU Linear Programming Kit Reference Manual, Version 4.28*, 2008.
29. M. X. Goemans and D. P. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM, 42, 1115–1145, 1995.

30. D. Goldfarb and K. Scheinberg. *A product-form cholesky factorization method for handling dense columns in interior point methods for linear programming*, Mathematical Programming, 99(1), 1–34, 2004.
31. D. Goldfarb and K. Scheinberg. *Product-form cholesky factorization in interior point methods for second-order cone programming*, Mathematical Programming, 103(1), 153–179, 2005.
32. 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*, number 38 in Annals of Mathematical Studies, pages 53–97. Princeton University Press, Princeton, New Jersey, 1956.
33. J. Gondzio. *Warm start of the primal-dual method applied in the cutting plane scheme*, Mathematical Programming, 83(1), 125–143, 1998.
34. J. Gondzio and A. Grothey. *Direct solution of linear systems of size 10^9 arising in optimization with interior point methods*, In R. Wyrzykowski, J. Dongarra, N. Meyer, and J. Wasniewski, editors, *Parallel Processing and Applied Mathematics*, number 3911 in Lecture Notes in Computer Science, pages 513–525. Springer-Verlag, Berlin, 2006.
35. J. Gondzio and A. Grothey. *A new unblocking technique to warmstart interior point methods based on sensitivity analysis*, SIAM Journal on Optimization, 3, 1184–1210, 2008.
36. J. Gondzio and T. Terlaky. *A computational view of interior point methods for linear programming*, In J. E. Beasley, editor, *Advances in Linear and Integer Programming*, pages 103–185. Oxford University Press, Oxford, 1996.
37. I. S. Gradshteyn and I. M. Ryzhik *Tables of Integrals, Series, and Products*, Academic Press, San Diego, CA, 6th edition, 2000.
38. M. Grant and S. Boyd. *CVX: Matlab software for disciplined convex programming*, Web page and software, http://stanford.edu/∼boyd/cvx, 2008.
39. M. Grant and S. Boyd. *Graph implementations for nonsmooth convex programs*, In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*. Springer, 95–110, 2008.
40. M. Grötschel, L. Lovász, and A. Schrijver *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, 1988.
41. Ch. Guéret, Ch. Prins, and M. Sevaux *Applications of optimization with Xpress-MP*, Dash Optimization, 2002. Translated and revised by Susanne Heipcke.
42. 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.
43. T. Illés and T. Terlaky. *Pivot versus interior point methods: Pros and cons*, European Journal of Operational Research, 140(2), 170–190, 2002.
44. I. D. Ivanov and E. de Klerk. *Parallel implementation of a semidefinite programming solver based on CSDP in a distributed memory cluster*, Optimization Methods and Software, Appeared online first 28 September 2009.
45. B. Jansen, *Interior Point Techniques in Optimization. Complexity, Sensitivity and Algorithms*, volume 6 of *Applied Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
46. B. Jansen, J. J. de Jong, C. Roos, and T. Terlaky. *Sensitivity analysis in linear programming: just be careful!*, European Journal of Operations Research, 101(1), 15–28, 1997.
47. B. Jansen, C. Roos, and T. Terlaky. *The theory of linear programming: Skew symmetric self-dual problems and the central path*, Optimization, 29, 225–233, 1994.
48. N. K. Karmarkar. *A new polynomial-time algorithm for linear programming*, Combinatorica, 4, 373–395, 1984.
49. M. Kočvara and M. Stingl *PENSDP Users Guide (Version 2.2)*, PENOPT GbR, 2006.

50. J. Löfberg. *YALMIP: a toolbox for modeling and optimization in MATLAB*, In *Proceedings of the 2004 IEEE International Symposium on Computer Aided Control Systems Design*, pages 284–289, 2004.
51. L. Lovász and A. Schrijver. *Cones of matrices and setfunctions, and 0-1 optimization*, SIAM Journal on Optimization, 1(2), 166–190, 1991.
52. Z.-Q. Luo, J. F. Sturm, and S. Zhang. *Conic linear programming and self-dual embedding*, Optimization Methods and Software, 14, 169–218, 2000.
53. K. A. McShane, C. L. Monma, and D. F. Shanno. *An implementation of a primal-dual interior point method for linear programming*, ORSA Journal on Computing, 1, 70–83, 1989.
54. 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.
55. S. Mehrotra. *On the implementation of a (primal-dual) interior point method*, SIAM Journal on Optimization, 2(4), 575–601, 1992.
56. S. Mehrotra and Y. Ye. *On finding the optimal facet of linear programs*, Mathematical Programming, 62, 497–515, 1993.
57. R. D. C. Monteiro. *Primal-dual path-following algorithms for semidefinite programming*, SIAM Journal on Optimization, 7, 663–678, 1997.
58. R. D. C. Monteiro and S. Mehrotra. *A general parametric analysis approach and its implications to sensitivity analysis in interior point methods*, Mathematical Programming, 72, 65–82, 1996.
59. R. D. C. Monteiro and M. J. Todd. *Path-following methods.* In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 10, pages 267–306. Kluwer Academic Publishers, 2000.
60. R. D. C. Monteiro and T. Tsuchiya. *Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-family of directions*, Mathematical Programming, 88(1), 61–83, 2000.
61. K. Nakata, M. Yamashita, K. Fujisawa, and M. Kojima. *A parallel primal-dual interior-point method for semidefinite programs using positive definite matrix completion*, Parallel Computing, 32, 24–43, 2006.
62. A. S. Nemirovski and M. J. Todd. *Interior-point methods for optimization*, Acta Numerica, 17, 191–234, 2008.
63. Y. E. Nesterov and A. Nemirovski *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM Studies in Applied Mathematics.* SIAM Publications, Philadelphia, PA, 1994.
64. Y. E. Nesterov and M. J. Todd. *Self-scaled barriers and interior-point methods for convex programming*, Mathematics of Operations Research, 22(1), 1–42, 1997.
65. M. L. Overton. *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM Journal on Matrix Analysis and Applications, 9(2), 256–268, 1988.
66. J. Peng, C. Roos, and T. Terlaky. *New complexity analysis of the primal-dual Newton method for linear optimization*, Annals of Operations Research, 99, 23–39, 2000.
67. J. Peng, C. Roos, and T. Terlaky *Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms.* Princeton University Press, Princeton, NJ, 2002.
68. L. Porkoláb and L. Khachiyan. *On the complexity of semidefinite programs*, Journal of Global Optimization, 10(4), 351–365, 1997.
69. F. Potra and R. Sheng. *On homogeneous interior-point algorithms for semidefinite programming*, Optimization Methods and Software, 9(3), 161–184, 1998.
70. F. Potra and Y. Ye. *Interior-point methods for nonlinear complementarity problems*, Journal of Optimization Theory and Applications, 68, 617–642, 1996.

71. A. U. Raghunathan and L. T. Biegler. *An interior point method for mathematical programs with complementarity constraints (MPCCs)*, SIAM Journal on Optimization, 15(3), 720–750, 2005.
72. C. Roos, T. Terlaky, and J.-Ph. Vial *Theory and Algorithms for Linear Optimization. An Interior Approach*, Springer, New York, USA, 2nd edition, 2006.
73. S. H. Schmieta and F. Alizadeh. *Associative and Jordan algebras, and polynomial time interior-point algorithms*, Mathematics of Operations Research, 26(3), 543–564, 2001.
74. Gy. 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 Modeling 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.
75. J. F. Sturm. *Primal-dual interior point approach to semidefinite programming*, In J. B. G. Frenk, C. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
76. J. F. Sturm. *Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones*, Optimization Methods and Software, 11-12, 625–653, 1999.
77. J. F. Sturm. *Implementation of interior point methods for mixed semidefinite and second order cone optimization problems*, Optimization Methods and Software, 17(6), 1105–1154, 2002.
78. J. F. Sturm. *Avoiding numerical cancellation in the interior point method for solving semidefinite programs*, Mathematical Programming, 95(2), 219–247, 2003.
79. J. F. Sturm and S. Zhang. *An interior point method, based on rank-1 updates, for linear programming*, Mathematical Programming, 81, 77–87, 1998.
80. T. Terlaky, editor *Interior Point Methods of Mathematical Programming*, volume 5 of *Applied Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
81. T. Terlaky. *An easy way to teach interior-point methods*, European Journal of Operational Research, 130(1), 1–19, 2001.
82. M. J. Todd. *A study of search directions in primal-dual interior-point methods for semidefinite programming*, Optimization Methods and Software, 11, 1–46, 1999.
83. K. C. Toh, R. H. Tütüncü, and M. J. Todd. *On the implementation of SDPT3 (version 3.1) – a Matlab software package for semidefinite-quadratic-linear programming*, In *Proceedings of the IEEE Conference on Computer-Aided Control System Design*, 2004.
84. T. Tsuchiya. *A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming*, Optimization Methods and Software, 11, 141–182, 1999.
85. A. W. Tucker. *Dual systems of homogeneous linear relations*, In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, number 38 in Annals of Mathematical Studies, pages 3–18. Princeton University Press, Princeton, New Jersey, 1956.
86. R. H. Tütüncü, K. C. Toh, and M. J. Todd. *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming Series B, 95, 189–217, 2003.
87. P. M. Vaidya. *A new algorithm for minimizing convex functions over convex sets*, Mathematical Programming, 73(3), 291–341, 1996.
88. L. Vandenberghe and S. Boyd. *Semidefinite programming*, SIAM Review, 38, 49–95, 1996.
89. R. J. Vanderbei *LOQO User's Guide – Version 4.05*, Princeton University, School of Engineering and Applied Science, Department of Operations Research and Financial Engineering, Princeton, New Jersey, 2006.

90. A. Wächter and L. T. Biegler. *On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming*, Mathematical Programming, 106(1), 25–57, 2006.
91. H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Kluwer Academic Publishers, 2000.
92. M. H. Wright. *The interior-point revolution in optimization: History, recent developments, and lasting consequences*, Bulletin (New Series) of the American Mathematical Society, 42(1), 39–56, 2004.
93. S. J. Wright. *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, USA, 1997.
94. X. Xu, P. Hung, and Y. Ye. *A simplified homogeneous and self-dual linear programming algorithm and its implementation*, Annals of Operations Research, 62, 151–172, 1996.
95. M. Yamashita, K. Fujisawa, and M. Kojima. *Implementation and evaluation of SDPA 6.0 (SemiDefinite Programming Algorithm 6.0)*, Optimization Methods and Software, 18, 491–505, 2003.
96. Y. Ye *Interior-Point Algorithms: Theory and Analysis*, Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley and Sons, New York, 1997.
97. 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.
98. E. A. Yildirim and S. J. Wright. *Warm start strategies in interior-point methods for linear programming*, SIAM Journal on Optimization, 12(3), 782–810, 2002.
99. Y. Zhang. *On extending primal-dual interior-point algorithms from linear programming to semidefinite programming*, SIAM Journal of Optimization, 8, 356–386, 1998.
100. Y. Zhang. *Solving large-scale linear programs by interior-point methods under the MATLAB environment*, Optimization Methods Software, 10, 1–31, 1998.